UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering,
Mathematics & Computer Science

Finding Ways to Improve the Prediction Accuracy of a Model that Predicts the Outcome of a Football Match using Machine Learning

Jorn Boksem
June 2022

Supervisors: dr. F.A. Bukhsh
dr. A. Abhishta

Faculty of Electrical Engineering,
Mathematics and Computer Science
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Acknowledgements

In January 2022 I started working on this thesis at the DMB/DS group at the University of Twente. Now, six months later, the thesis has been completed. The route that we took has been very pleasant and educational, but also quite challenging. I developed an interest in machine learning and I hope that this interest will help me in my career as a data professional. Completing this thesis would not have been possible without the help and guidance of my supervisors. I would like to thank dr. F.A. Bukhsh and dr. A. Abhishta for all the effort they put in and the great feedback and advice they gave me. Additionally, I would like to thank them for their availability and flexibility while guiding me through this thesis. Next to that, I would like to thank all of the graduation committee members for attending my colloquium.

I would also like to thank my family and friends for supporting me during the work on my thesis. In particular, my sister and my girlfriend. I would like to thank my sister for supporting me by giving great advice during the course of this research. Next to that, I would like to thank Wendie who supported me on a daily basis and without whom I could not have done it.

Executive Summary

The goal of this research is to find new ways to improve the prediction accuracy of a model that predicts the outcome of a football match. In our literature review, we found out that there has been quite some research focusing on finding new and effective feature categories or algorithms. Most of these would make use of the known effective feature categories, like match attributes, match statistics and team performance, while introducing a new feature category. Also, some of these would experiment with a new algorithm to find out whether it would have potential. Studies like [1], [2] and [3] made use of these ways to improve the prediction accuracy and have had some success over the years. [1] made use of weather as a new feature category, while [2] focused on team/player ratings and team/player values. [3] did not focus on a new feature category but used the known effective ones, while experimenting with long short-term memory. As mentioned, these ways to improve the accuracy have had their success over the years but at a certain point the pile with new feature categories and algorithms will run out. In that case, we need to find other ways to improve the prediction accuracy of a model that predicts the outcome of a football match.

In this thesis, we propose two ways to improve the prediction accuracy other than finding new feature categories or algorithms. The first way is the use of feature category combinations. A feature category combination is a combination between two feature categories. The combination of these two categories could result into a new set of features which exists next to the feature category features themselves. This means that with the same data, we tend to create more value. We do this by taking a second look at the feature categories and reason which features could be created considering the data of both feature categories. This approach could lead to new features which could lead to an improved prediction accuracy. The second way we propose to improve the prediction accuracy is the use of ensembles. Specifically, ensembles that make use of the promising algorithms that we found in our literature review. Ensembles can be used like any individual algorithm but, in this case, make use of different algorithms to predict the outcome. Ensembles can make use of all promising algorithms or just a subset. Also, the algorithms that are a part of such an ensemble can be equally important or the ensemble could make use of a different

importance distribution. The use of different algorithms in an ensemble could lead to an improved prediction accuracy.

To evaluate whether these two ways can actually improve the prediction accuracy and are worthwhile we completed several of steps for each of the proposed ways. Starting with the feature category combinations. We looked at four different feature category combinations, namely team performance and team rating (TPTR), past match statistics and team rating (PMSTR), team performance and team value (TPTV), and past match statistics and team value (PMSTV). For each of these feature category combinations, we created three feature category combination feature sets. A feature set exists out of two features which are created by using both feature categories. We examined a feature category combination feature set by comparing a feature selection that did not make use of the feature category combination features and a feature selection that did make use of these features. The former consisted of four features. Out of each feature category, two features were created. In other words, these features were created out of the feature categories separately. The other feature selection also contained these features as well as the feature category combination features which are created by using both feature categories. These two feature selections were compared using five different algorithms (in their default state), namely random forest, XGBoost, logistic regression, support vector machine and an equally weighted ensemble. To compare these feature selections properly, we made use of a term called actual performance increase. The actual performance increase represents a percentage which describes how much bigger the increase in prediction accuracy due to using the feature selection that includes the feature category combination features is than the increase in prediction accuracy which is present due to making use of the feature selection that excludes these features. To calculate the increase in prediction accuracy due to making use of a certain feature selection, we need to take into account the ratio of the class that is present the most. This class represents the home team winning and the home win percentage is 45.9%. This means that when you constantly guess that the home team will win, the prediction accuracy will be 45.9%. To calculate the increase in prediction accuracy we subtract the home win percentage of the prediction accuracy achieved due to using a certain feature selection. When the actual performance increase is 10.0% or higher on average, the feature category combination or feature category combination feature set can be seen as worthwhile. This means that when the use of the feature selection that excludes the feature category combination features results in a prediction accuracy of 49.9%, the increase in prediction accuracy or delta is 4.0%. This also means that the prediction accuracy due to making use of the feature selection that includes the feature category combination features must be 4.4% or higher to be seen as worthwhile. We chose a minimum of 10.0% because the data is already available and familiar and it would only cost somewhat more resources.

To evaluate whether ensembles can actually improve the prediction accuracy and can be worthwhile, we compared the promising algorithms with the ensembles that contained that promising algorithm, individually. The possible ensembles we looked at during our research are all ensembles that consist of the promising algorithms or a subset of them while taking into account specific importance distributions. The importance distributions we took into account are all possible importance distribution where the algorithms could have a weight ranging from 1 to 4. For the broad comparison between the individual algorithm and the ensembles, we did not make use of the default state of the algorithms but applied hyperparameter optimisation to make sure the individual algorithm and the ensembles would perform at their best for the chosen feature selections. Every individual algorithm was compared over three feature selections, namely, the complete feature selection, a feature selection created using Pearson correlation coefficient, and a feature selection using the Fisher's Score ranking. We also made use of the actual performance increase in this part of the study to compare the individual algorithm and the ensembles properly. In this case, the actual performance increase describes how much bigger the increase in prediction accuracy due to using a specific ensemble is than the increase in prediction accuracy which is present when making use of the individual algorithm. We chose a minimum of 5.0% for an ensemble to be seen as worthwhile, due to only having to create the ensemble while not having to bother to collect and familiarize yourself with the data. Also, there won't be any features that have to be modified but the use of an ensemble would cost somewhat more resources due to using multiple algorithms.

In conclusion, we chose to explore four feature category combinations that we believed would have the best chance of improving the prediction models directly. Next to that, we reasoned which feature category combination features could be useful and came up with twelve sets of feature category combination features. The chosen four feature category combinations had a positive effect on the prediction accuracy. Even though not all feature category combination feature sets had an actual performance increase of 10.0% or higher, on average the feature selections that include the feature category combination features performed 16.2% better than the feature selections that exclude the feature category combination features. This leads to the conclusion that the use of these feature category combinations can indeed be seen as worthwhile and used to increase the accuracy of a model that predicts the outcome of a football match. The feature category combination team performance and team value is the best performing combination achieving an average actual performance increase of 18.0%.

Next to the feature category combinations, we took an interest in ensembles. More specifically, what subset of the promising algorithms could be most beneficial in an ensemble and what the importance distribution should be between these algorithms. We looked into these curiosities by comparing each promising algorithm with the relevant ensembles, separately. We found out that the most beneficial ensemble was using the composition containing random forest, XGBoost and support vector machine, the composition containing random forest, XGBoost and logistic regression or the composition containing all algorithms, each making use of an importance distribution where XGBoost was dominant. Next to that, we found out that for every individual algorithm, there is an ensemble that realises an increase in prediction accuracy. Also, we found out that only for some of the models that make use of an individual algorithm, it is worthwhile to make use of the found ensembles. For already really well performing individual algorithms, it seems that it is not worth the resources and effort to make use of these ensembles. In other words, the actual performance increase was too small to be seen as worthwhile. In cases where resources are not limited, the minimum actual performance increase could be lower which means that in such a scenario the use of ensembles could be seen as worthwhile.

Contents

List of Figures

List of Tables

List of Abbreviations

CFS Complete feature selection 101

F10FS Fisher-10 feature selection 107

F12FS Fisher-12 feature selection 108

F14FS Fisher-14 feature selection 105

F1FS Fisher-1 feature selection 106

FTR Feature that represents the outcome of the football match. It can only be a 0, 1 or 2 which represent a home win, a draw and an away in, respectively. 56

GCA Feature that represents the goals conceded this season by the away team. 68

GCH Feature that represents the goals conceded this season by the home team. 68

GCTRA Feature that represents the amount of goals conceded by the away team against the opponent's team rating class. 69

GCTRH Feature that represents the amount of goals conceded by the home team against the opponent's team rating class. 69

GCTVA Feature that represents the amount of goals conceded by the away team against the opponent's team value class. 89

GCTVH Feature that represents the amount of goals conceded by the home team against the opponent's team value class. 89

GSA Feature that represents the goals scored this season by the away team. 66

GSH Feature that represents the goals scored this season by the home team. 66

GSTRA Feature that represents the amount of goals scored by the away team against the opponent's team rating class. 66

GSTRH Feature that represents the amount of goals scored by the home team against the opponent's team rating class. 66

GSTVA Feature that represents the amount of goals scored by the away team against the opponent's team value class. 86

GSTVH Feature that represents the amount of goals scored by the home team against the opponent's team value class. 86

LPA Feature that represents the points collected by the away team during last season. 58

LPH Feature that represents the points collected by the home team during last season. 58

LPTRA Feature that represents the amount of points the away team collected against the opponent's team rating class during last season. 59

LPTRH Feature that represents the amount of points the home team collected against the opponent's team rating class during last season. 59

LPTVA Feature that represents the amount of points the away team collected against the opponent's team value class during last season. 79

LPTVH Feature that represents the amount of points the home team collected against the opponent's team value class during last season. 79

PA Feature that represents the points collected by the away team during this season. 56

PCCFS Pearson Correlation Coefficient feature selection 101

PH Feature that represents the points collected by the home team during this season. 56

PMSTR Feature category combination between past match statistics and team rating. 4, 95

PMSTV Feature category combination between past match statistics and team value. 4, 95

PTRA Feature that represents the amount of points the away team collected against the opponent's team rating class. 56

PTRH Feature that represents the amount of points the home team collected against the opponent's team rating class. 56

PTVA Feature that represents the amount of points the away team collected against the opponent's team value class. 76

PTVH Feature that represents the amount of points the home team collected against the opponent's team value class. 76

TPTR Feature category combination between team performance and team rating. 4, 95

TPTV Feature category combination between team performance and team value. 4, 95

TRA Feature that represents the class of the away team based on their team rating. 56

TRH Feature that represents the class of the home team based on their team rating. 56

TVA Feature that represents the class of the away team based on their team value. 76

TVH Feature that represents the class of the home team based on their team value. 76

Chapter 1

Introduction

In the last decade prediction in football while using machine learning has become quite popular. More and more research was published in this area exploring many sides of football prediction. Football prediction is used by football coaches, clubs, players and people who lay bets on football matches. Therefore, football prediction can be many things.

In football prediction there can be classification and regression problems. Classification is predicting in to what class something belongs. In this case, something can be a team, a player, a match etc. Regression is predicting us how much of something will be there. It is a continuous quantity. For example, goals, points, yellow cards, etc. There are many possible classification problems and regression problems in football prediction. A classification problem can be anything like predicting whether or not a player will get injured. Also, it can be predicting whether both teams will score or not, which will win, whether a goal will be score, if there will red cards in a game and much more. Regression problems can be just as comprehensive. It can predicting how many goals there will scored by each side or how many points a team will collect during a season.

When looking at the game of football, there is only thing one thing that is most important at the end of the day and that is winning. This means that being able to predict the outcome of a football match is quite valuable. Predicting the outcome of a football match is an interesting classification problem with three classes. The three classes are a win by the home team, a win by the away team, and a draw by both teams. There many things that could influence or cause one of the outcome of a football match.

When using machine learning to make a prediction for this classification problem, the two most important things a data scientist needs are features and algorithms. Because, the algorithm makes use of the features in a specific way to predict to which class the match belongs. Features can be any kind of data representing things that could influence or cause one of the outcome of a football match. Example of features are the team name, the side that wins the game, the team rating class, etc. There is one feature that is essential during training and testing which is called the target feature. This feature describes the actual class that the match belongs to. When the algorithm predicts the class that the match actually belongs to, we can speak of a correct prediction. When features can be grouped due to describing the same thing, they can be called a feature category.

Many combinations of feature categories and algorithms have already been tried. Some even with success. Most of the research consists of exploring the effectiveness of a new feature category with at least one well-known machine learning algorithm or the effectiveness of a new algorithm with at least one of the popular feature categories. Unfortunately, there is not much research out there trying to create more value out the already known and successful feature categories and algorithms.

1.1 Problem Definition

As mentioned before, there has been quite some research regarding new feature categories and algorithms which had quite some success over the years. And even though there are possibly still new feature categories and algorithms to be found, we can't just rely on those being an inexhaustible source for value. There are potentially other sources for value while using the resources already known.

Such a source of value could be the use of feature category combinations. Feature category combinations are a combination between two feature categories. The combining of these two categories will result into a new set of features which exists next to the feature categories themselves. Considering a different approach where combinations between feature categories are included could result in interesting new features. Interesting new features could on their turn improve the prediction accuracy of a model that predicts the outcome of a football match.

Another way to increase the prediction accuracy could be the use of ensembles that make use of different algorithms. They can be used in the same way as any individual algorithm but make use of multiple algorithms. The algorithms that are a part of the ensembles can be equally important or have different weights than each other. The usage of these ensembles could potentially improve the prediction accuracy of a model that predicts the outcome of a football match.

1.2 Research Questions

The research questions are stated below:

1. What is the state of the art in predicting the outcome of a football match using machine learning?
   (a) What algorithms have been used to predict the outcome of a football match?
   (b) What feature categories have been used to predict the outcome of a football match?
   (c) What algorithms and feature categories show promising results while being used to predict the outcome of a football match?
2. How can feature category combinations be used to improve the prediction accuracy when predicting the outcome of a football match using machine learning?
   (a) What feature categories can be combined into a new set of features?
   (b) What feature category combinations can be used to improve the prediction accuracy?
3. How can ensembles be used to achieve to improve the prediction accuracy when predicting the outcome of a football match using machine learning?
   (a) What of subset promising algorithms could be most beneficial to an ensemble?
   (b) What should the importance distribution be between these algorithms?

1.3 Research Goal

The goal of this research is to find new ways to improve the prediction accuracy of a model that predicts the outcome of a football match. We propose two ways to achieve that this. Namely, by using feature category combinations and ensembles. In this research these ways will be explored by evaluating several potentially beneficial feature category combinations and several potentially beneficial ensembles compositions that make use of a specific importance distribution.

1.4 Structure

This structure of this thesis is as follows: Chapter 2 consists of a systematic literature review regarding the state of the art in predicting the outcome of a football match using machine learning. Chapter 3 presents the methodology used and our machine learning pipeline. In Chapter 4 feature category combinations and feature category combination features are described, tested and evaluated. Next, different ensemble compositions using a specific importance distribution are described, tested and evaluated in Chapter 5. Finally, we will draw conclusions and reflect critically on our research in Chapter 6.

Chapter 2

Literature Review

In this chapter, we will provide insights regarding the state of the art in predicting the outcome of a football match using machine learning. We will describe the existing prediction models by looking at the algorithms and feature categories used to predict the outcome of a football match. Next to that, we will have a more detailed look at the most promising ones. This chapter is structured as follows: In the first section, we will describe our research method containing the research questions, search strategy, and execution of the data extraction process. Next, we will describe and discuss our results. Followed by a discussion regarding the threats to the validity of this review. And finally, we will conclude our findings and talk about potentially interesting research for future studies.

2.1 Research Method

In this section, we will provide an overview of the research method. First, we state our research questions. Next, we will describe our search strategy containing the inclusion and exclusion criteria. Finally, we will explain the execution of the data extraction process. Throughout this systematic literature review, the guidelines of [4] will be used.

2.1.1 Research Questions

The goal of this systematic literature review is to answer the following research question and sub-questions:

1. What is the state of the art in predicting the outcome of a football match using machine learning?
    (a) What algorithms have been used to predict the outcome of a football match?

(b) What feature categories have been used to predict the outcome of a football match?

(c) What algorithms and feature categories show promising results while being used to predict the outcome of a football match?

2.1.2 Search Strategy

For this systematic literature review, we used the search engine Scopus to find relevant and accessible literature. Scopus makes it possible for its users to search in a big pile of literature from various sources. It provides the ability to search for specific literature while making use of a customized search string. Furthermore, the user can decide to limit their search to specific sections of the literature (e.g. the abstract). For this systematic literature review, we limited the search to the title, abstract, and keywords. We made use of the following search string:

'(football OR soccer) AND (prediction OR predicting OR predict) AND (outcome OR result OR winner OR performance) AND "Machine Learning"'

The search was done on 30 January 2022 and resulted in 186 papers. The search string is a result of a learning process. This means that we explored other compositions and decided on a string that is complete but does not contain irrelevant additions.

Inclusion and Exclusion Criteria

We used inclusion and exclusion criteria to make sure only relevant and accessible literature will be used for this systematic literature review. The inclusion criteria are:

1. The paper directly relates to the topic of our review. Papers are only included if they talk about using machine learning to predict the outcome/winner of a football match.

2. The paper addresses the research questions directly.

3. The paper is published in a peer-reviewed journal, conference, or workshop.

4. The paper is in English.

The exclusion criteria are:

1. The paper does not talk about football prediction and machine learning as its main topics.

2. The paper is not peer-reviewed

3. The paper is not available for download.

2.1.3 Execution of the data extraction process

To make sure that the found papers are relevant, our research process contained three steps next to the search in Scopus. Namely, the removal of duplicate papers, the removal of irrelevant papers, and the removal of inaccessible papers. Our research process is displayed in Figure 2.1. As we can see in this figure, the search in Scopus provided us with 186 papers. After removing the duplicates, 177 papers remained. And finally, when the inclusion and exclusion criteria were applied 50 relevant papers lasted. 124 papers did not directly relate to the topic of our review and 3 papers were not available for download.

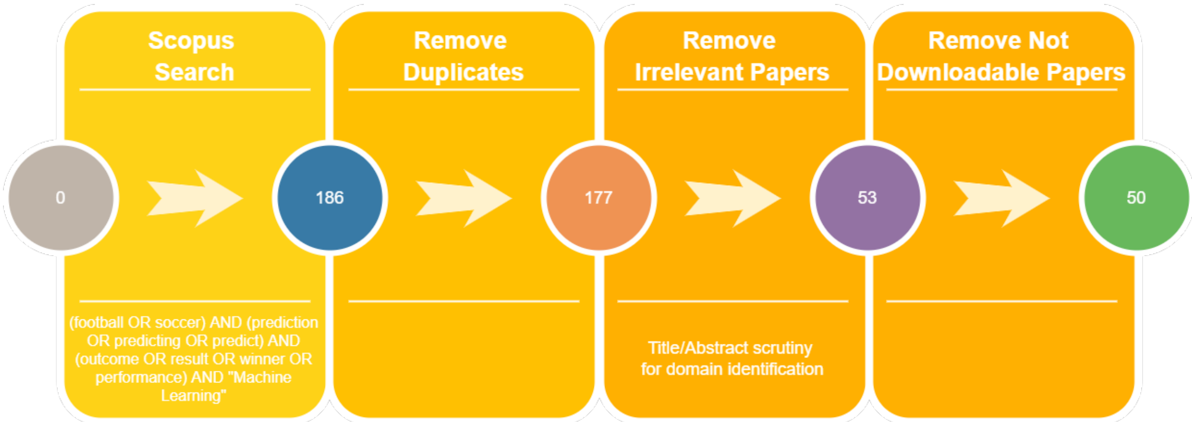

Figure 2.1: Study Selection Process

In Figure 2.2, the amount of relevant and accessible papers published per year is displayed. As we can see quite clearly, research regarding this subject has become more interesting in the last 5 years. This means that the selected literature can provide us with good insights regarding the use of today's technology in this area.

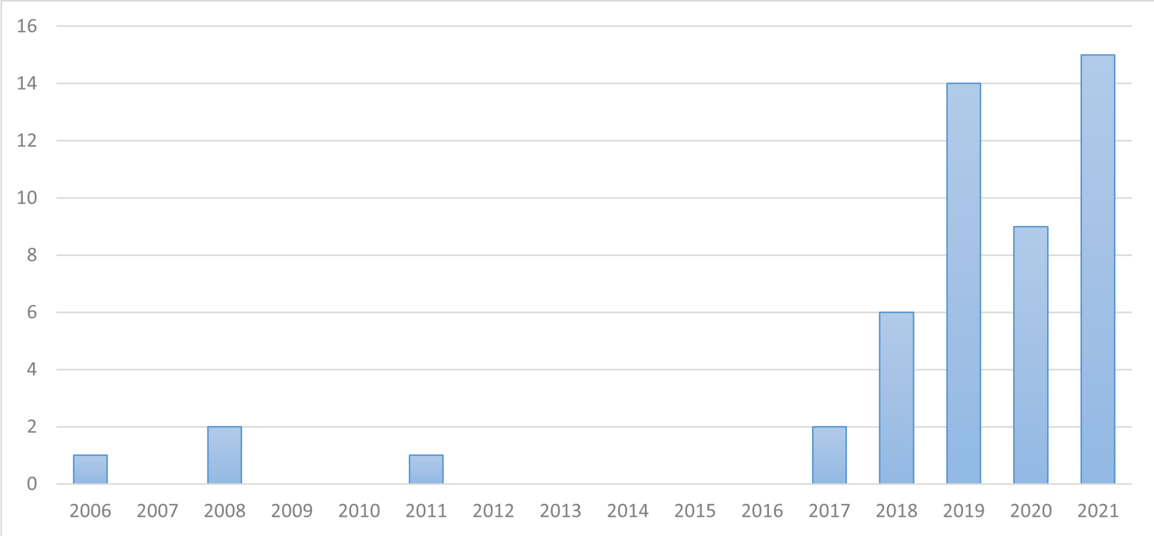

Figure 2.2: Selected literature with respect to year of publication

To answer the research questions stated in Section 2.1.1, we will look in the selected literature for prediction models that predict the outcome of a football match using machine learning. We will take a look at what algorithms and feature categories have been used. Furthermore, we will take a closer look at the prediction models that show promising results to find out which algorithms and feature categories have been successful in the past.

2.2 Results

In this section, the findings will be described. Each subsection contains the findings related to a single sub-question. First, we will describe the algorithms that have been used to predict the outcome of a football match. Next, we will describe the used feature categories. And finally, we will talk about the promising models and the related algorithms and feature categories.

2.2.1 RQ1: What algorithms have been used to predict the outcome of a football match?

To answer the first sub-question, we looked in the selected literature for algorithms that have been used to predict the outcome of a football match. The algorithms and the studies that used these algorithms can be found in Table 2.1. We used 11 different classes to classify the found algorithms. Nine of these are well-known machine learning algorithms. The other two consist of a class of ensemble methods and a class of other methods. The latter contains the methods Bayesian networks, linear and quadratic discriminant analyses, case-based reasoning, Poisson regression, extremely randomized trees, and customized probability models and algorithms. The class of ensemble methods consists of models that combine various well-known machine learning algorithms which means that the models inside this class can be very different. But even in the classes containing the well-known algorithms, there might be slight differences, due to having several algorithms that belong to the same class. For example, in the "Decision Tree" class, there could be a study regarding the C4.5 algorithm and a study regarding the CART algorithm. Both of these are decision tree algorithms.

Algorithms	**Studies**
Linear Regression	[5]
Artificial Neural Network	[6] [7] [8] [9] [10] [11] [2] [12] [13] [3] [14] [15]
Gradient Boosted Tress	[6] [8] [16] [17] [18] [10] [19] [5] [20] [21] [22] [23]
Decision Tree	[24] [7] [8] [25] [10] [11] [19] [26] [13] [21] [23] [27]
K-Nearest Neighbors	[24] [17] [10] [1] [26] [23]
Logistic Regression	[8] [28] [29] [10] [19] [26] [30] [14] [23]
Naive Bayes	[24] [6] [7] [31] [16] [32] [19] [2] [14]
Random Forest	[6] [16] [33] [32] [10] [1] [2] [26] [5] [14] [23] [27]
Support Vector Machine	[7] [8] [16] [18] [34] [32] [10] [11] [1] [35] [19] [2] [26] [5] [36] [23]
Ensemble of Multiple Algorithms	[32] [26] [5]
Other	[24] [6] [7] [37] [8] [38] [31] [29] [39] [1] [26] [21] [23]

Table 2.1: Algorithms used to predict the outcome of a football match

To gain more insights regarding the use of algorithms to predict the outcome of a football match, we will have a look at Figure 2.3. This Figure shows that support vector machines, artificial neural networks, gradient boosted trees, decision trees, logistic regression, naive Bayes classifiers, and random forests have been used a lot. This is not surprising due to their excellent ability to classify elements. In contrast to linear regression which is not a great classifier and is commonly used to handle regression problems. The expectation was that the usage of linear regression was quite limited in this area. Linear regression has been used only during a single study and thus confirming our expectations. Next to linear regression, the k-nearest neighbors algorithm, ensembles of different algorithms, and other methods have not been used that much as well. The Figure shows that there are quite some studies that made use of models that represent the "Other" class. But this does not necessarily mean that there is a lot of interest in models outside of the well-known machine learning algorithm group, since the class consists of a lot of different methods.

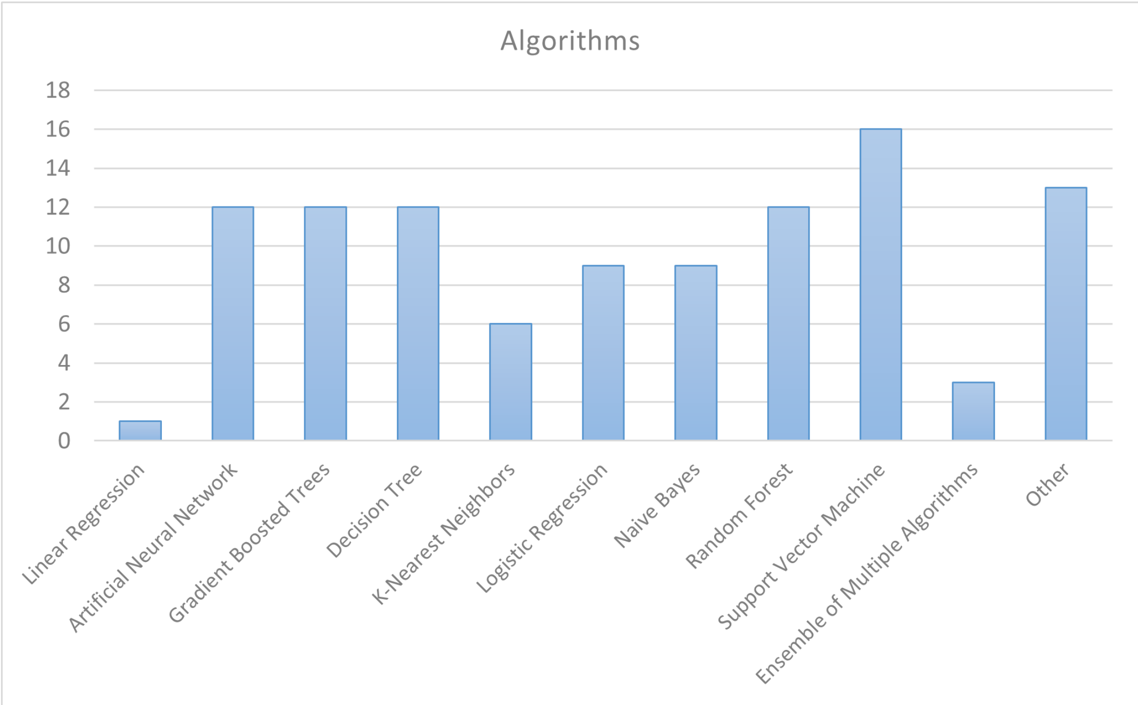

Figure 2.3: Algorithms

Another way to look at this data is to look at the number of studies per year that made use of the algorithms. In Figure 2.4, we can observe that interest in the well-known machine learning algorithms has increased over the years whereas interest in other methods has decreased. Next to that, there appears to be a peak of interest in the support vector machine algorithm in 2020 where it is used almost twice as much as any other algorithm. This could be the result of a high amount of studies focusing on binary classification in this area. Finally, we observe an increase in the use of logistic regression for predicting the outcome of a football match over the years.

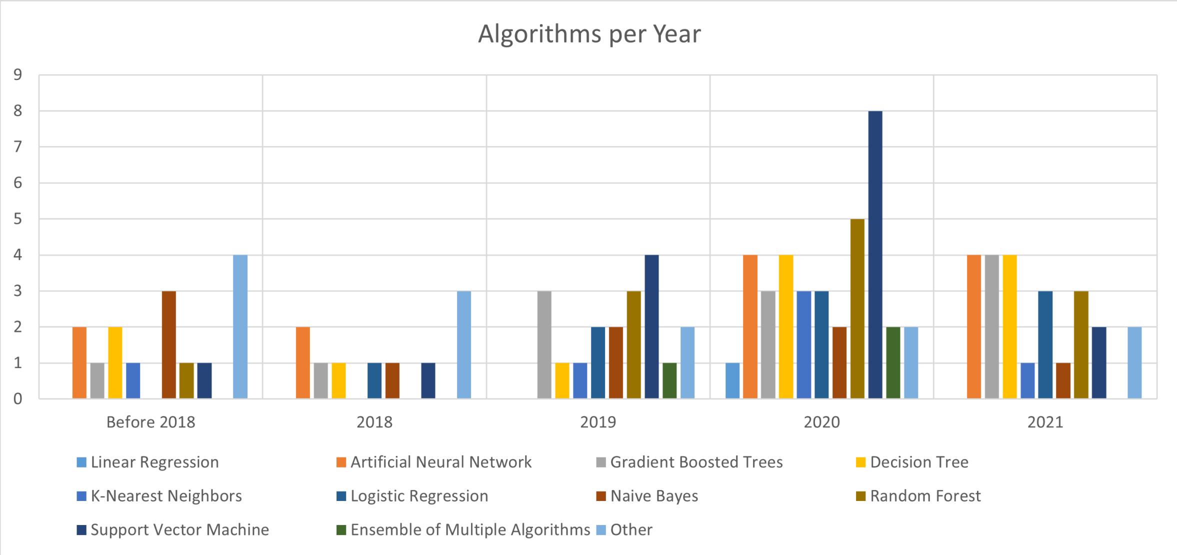

Figure 2.4: Algorithms per year

2.2.2 RQ2:What feature categories have been used to predict the outcome of a football match?

To answer the second sub-question, we looked in the selected literature for feature categories that have been used to predict the outcome of a football match. The feature categories and studies that used these feature categories can be found in Table 2.2. We made use of 21 different classes to distinguish all feature categories. Naturally, a model can make use of a feature selection that consists of multiple feature categories which means that it is possible for a study to be associated with several feature categories in the table below. Most of the classes are quite obvious, but in other contexts, some of these might be used differently. Therefore, we will discuss the most important ones, starting with match attributes. Match attributes are facts about the match that are known before it starts like match date, starting time, referee, and home and away teams. Match statistics are statistics produced during the match like yellow or red cards, shots on target, and the total amount of passes. Team performance describes features regarding the form of the team which can be long or short term. Head-to-head performance describes the historical success of the team when playing against its direct opponent. And finally, player attributes are facts regarding the player like age, height, and weight.

Feature Selection	Studies
Audience Data	[26]
Betting Odds	[9] [18] [28] [1] [12]
Coach Rating	[32]
Head-to-Head Performance	[6] [38] [33] [19] [23]
Match Attributes	[24] [37] [8] [9] [38] [31] [16] [17] [25] [28] [18] [34] [33] [39] [32] [10] [11] [1] [19] [2] [26] [5] [12] [13] [20] [3] [14] [22] [23] [27]
Match Statistics	[7] [37] [8] [9] [31] [25] [18] [34] [10] [11] [12] [13] [30] [20] [14] [15] [22] [27]
Passing Network Data	[14]
Past Match Statistics	[16]
Player Attributes	[5] [20]
Player Rating	[9] [2] [26] [5]
Player Statistics	[15]
Player Value	[2]
Position Tracking Data	[21]
Possession Chain Data	[35]
Team Performance	[24] [6] [38] [16] [17] [25] [18] [33] [29] [39] [32] [19] [2] [26] [12] [20] [3] [15] [23]
Team Rating	[8] [16] [29] [2] [26] [20] [14]
Team Value	[32] [20]
Twitter Data	[36]
Weather Data	[1]
Player Positions	[24] [8]
Injuries	[6]

Table 2.2: Feature categories used to predict the outcome of a football match

To analyze where the interests lay during these studies, we will take a look at Figure 2.5. Figure 2.5 displays the number of studies that used a specific feature category to predict the outcome of a football match. We can observe that the three most used feature categories are match attributes, match statistics, and team performance. This is probably the case due to the amount of available data. There are a lot of pre-processed data sets available on the internet containing data for these feature categories. Out of these three, the feature category named match attributes is used the most and is used almost twice as much as the others. This also seems logical because almost all studies took into account the home advantage. In contrast to feature categories that are used often, there is a big group of different feature categories that have been used only once or twice and a small group that has been used a couple of times. Part of this small group are the following feature categories: Betting odds, head-to-head performance, player rating, and team rating. This small group has probably also been used somewhat more due to the data availability. There are some sources that provide team and player ratings and quite a lot of sources that provide the betting odds.

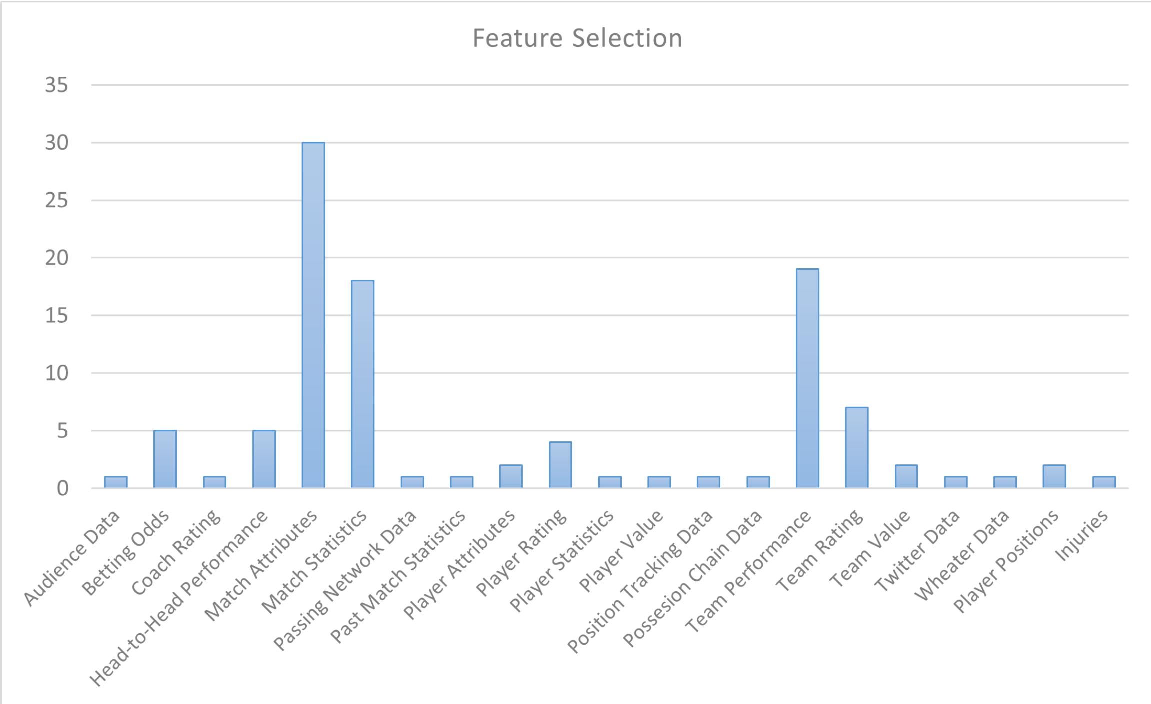

Figure 2.5: Feature selection

To discover whether the interests in feature categories shifted over time, we will take a look at Figure 2.6. In this Figure becomes clear that the use of match statistics has increased over the last years and the use of match attributes and team performance has decreased since 2019. As mentioned above, there is a quite big group of different feature categories that have been used only once or twice. The growth of this group over the years seems to be stable.

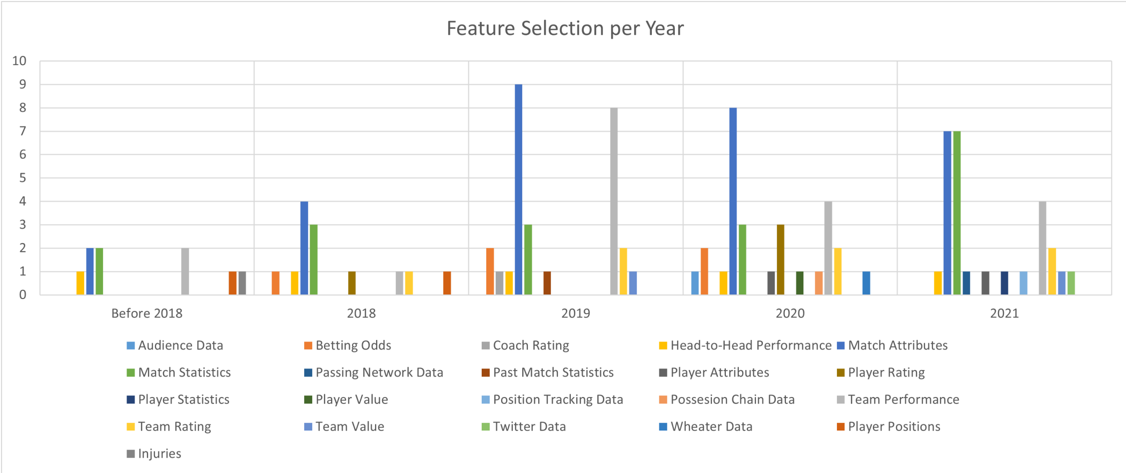

Figure 2.6: Feature categorie per year

2.2.3 RQ3: What algorithms and feature categories show promising results while used to predict the outcome of a football match?

To answer the third sub-question, we looked in the selected literature for models that have been used to predict the outcome of a football match. In particular, the used algorithms and feature selection and the accuracy of these models. We categorized these models into four categories, namely Binary and Pre-Game Data, Binary and In-Game Data, Multiclass and Pre-Game Data, and Multiclass and In-Game Data. A model applies binary classification when it classifies elements into exactly two classes. In this case, the classes would be a win for team A and no win for team A or a win for team A and a win for team B. A model applies multiclass classification when it classifies elements into 3 or more classes. In this case, the classes would be a win for team A, a draw, and a win for team B. A model is categorized as ”Pre-Game Data” when it takes into account only data that is known before the start of the match. In other words, it does not take data into account that is produced during the match. Unlike the models that are categorized as ”In-Game Data”. These make use of all data which means that they make use of the data known before the start of the match and the data produced during the match. For each of the four categories, we collected the number of models that are a part of the category and the average accuracy. Table 2.3 contains this data.

	Amount	Average Accuracy
Binary/Pre-Game Data	13	74.66%
Binary/In-Game Data	31	72.12%
Multiclass/Pre-Game Data	72	60.24%
Multiclass/In-Game Data	32	69.21%

Table 2.3: Prediction Model Categories

Logically, the accuracy of the binary prediction models should be higher than the accuracy of the multiclass prediction models due to making an easier prediction with fewer possible classes. Also, the models that make use of in-game data should be more accurate than the ones that make use of pre-game data due to having more up-to-date data and specifics of the match. As we can see in Table 2.3, the average accuracies of the binary models are higher than those of the multiclass models. This is as expected. This seems not to be the case for the expectation that pre-game data prediction models perform worse than in-game data prediction models. In particular, this expectation is wrong for the binary models. The binary models that make use of pre-game data perform better than the binary models that make use of in-game data. This could be the case due to the low amount of prediction models that applied binary classification and made use of pre-game data only. Having almost about a third of the number of data points the binary in-game data prediction models have, could result in a less valid comparison.

Binary and Pre-Game Data

As mentioned above, this category takes into account the prediction models that apply binary classification and which only make use of data that is known before the start of the match. Possible classes are a win for team A and no win for team A or a win for team A and a win for team B. This category is quite small compared to the others and has a high average accuracy of 74.66%. This is as expected due to only predicting two different classes. All models that are part of this category are part of a total of two studies. The models' accuracy, algorithm, and feature selection are displayed in Table 2.4.

When we take a look at the table, it becomes clear that both studies had success with the support vector machine algorithm. [1] and [19] achieved accuracies of 79.30% and 76.85%, respectively. [19] only performed better while using logistic regression, achieving an accuracy of 77.43%. Next to that, these studies achieved some decent accuracies using extremely randomized trees, AdaBoost, naive Bayes, decision trees, and random forests, ranging from 74.90% to 76.40%. The studies made use of different feature selections, [1] focusing on match attributes, betting odds, and weather data and [19] focusing on match attributes, team performance, and head-to-head performance. When looking at the bottom of the table, we see the support vector machine algorithm, the k-nearest neighbors algorithm with two different feature selections, and the random forest algorithm. All of these models are part of [1]. In that study, they created eight different prediction models using four algorithms and two feature selections. One feature selection that includes weather data and one feature selection that excludes weather data. Almost all algorithms in this study performed worse when excluding the weather data. Only the k-nearest neighbors algorithm performed better without the weather data. This could suggest weather data to be an interesting factor in predicting the outcome of a football match or it could mean that having more or various data is beneficial. Additionally, it would suggest that the combination of match attributes and betting odds is not a good feature selection. Next to that, the results in the bottom of the table suggest that there are better alternatives to predict the outcome of a football match than the k-nearest neighbors algorithm.

Study	Algorithm	Feature Selection	Highest Accuracy
[1]	Support Vector Machine	Match Attributes, Betting Odds, Weather Data	79.30%
[19]	Logistic Regression	Match Attributes, Team Performance, Head-to-Head Performance	77.43%
[19]	Support Vector Machine	Match Attributes, Team Performance, Head-to-Head Performance	76.85%
[1]	Extremely Randomized Trees Classifier	Match Attributes, Betting Odds, Weather Data	76.40%
[19]	AdaBoost	Match Attributes, Team Performance, Head-to-Head Performance	76.15%
[19]	Decision Tree	Match Attributes, Team Performance, Head-to-Head Performance	75.93%
[1]	Random Forest	Match Attributes, Betting Odds, Weather Data	75.60%
[19]	Naive Bayes	Match Attributes, Team Performance, Head-to-Head Performance	74.92%
[1]	Extremely Randomized Trees Classifier	Match Attributes, Betting Odds	74.90%
[1]	K-Nearest Neighbors	Match Attributes, Betting Odds	71.90%
[1]	Random Forest	Match Attributes, Betting Odds	71.80%
[1]	K-Nearest Neighbors	Match Attributes, Betting Odds, Weather Data	71.10%
[1]	Support Vector Machine	Match Attributes, Betting Odds	68.30%

Table 2.4: Binary and Pre-Game Data

Binary and In-Game Data

This category takes into account the prediction models that apply binary classification and which make use of data that is known before the start of the match and data that is produced during the match. Possible classes are a win for team A and no win for team A or a win for team A and a win for team B. This category is of average size compared to the others and has a high average accuracy of 72.12%. This is as expected due to only predicting two different classes and using in-game data. All models that are part of this category are part of a total of seven studies. The models' accuracy, algorithm, and feature selection are displayed in Table 2.5.

When looking at the table, we notice that the studies mostly used the same algorithms, but that their models perform very differently from each other. This means that the models are performing well or not due to their feature selection. Except for some of the models that did not use one of the well-known algorithms. For example, the best performing model, which is called a polynomial classifier and is a part of [7]. They described their customized algorithm as a parameterized nonlinear map which non-linearly expands a sequence of input vectors to a higher dimension and maps them to a desired output sequence. In combination with the feature category match statistics, they achieved an accuracy of 99.06%. As mentioned earlier, a lot of the performance is due to choosing the right feature selection. When looking at the top-performing algorithms of each study, we see that there is a small number of algorithms that perform well in multiple studies. Namely, gradient boosting and logistic regression, achieving an accuracy of 89.60% and 89.61%, respectively. Next to that, [8] shows that gradient boosting has the ability to perform about 20% better than the other algorithms used in their study. When looking at the average performance of the feature selections, we see that [7] achieved an average accuracy of 86.73% and [20] achieved an average accuracy of 89.60%. This difference could be due to [20] using a more various feature selection. Instead of only looking at match statistics, [20] also took into account match attributes, player attributes, team rating, team performance, and team value. To easily compare the top and bottom feature selections, we will take a look at [8] which, just like [20], used gradient boosting. [8] had a feature selection that consists of match attributes, match statistics, team rating, and player positions and is the worst-performing study in this category. The clear differences between [20] and [8] are that the former used a more various feature selection that contained player attributes, team performance, and team value and did not contain player positions. The other high-performing model did not have a feature selection that was very various but did also not contain player positions. This suggests player positions to be a bad feature selection. When looking at the worse-performing models for each study, we notice that artificial neural networks are always outperformed by other algorithms. This could suggest that there are better alternatives to predict the outcome of a football match.

Study	Algorithm	Feature Selection	Highest Accuracy
[7]	Polynomial Classifier	Match Statistics	99.06%
[7]	C4.5 Decision Tree	Match Statistics	89.61%
[20]	XGBoost	Match Attributes, Match Statisics, Player Attributes, Team Rating, Team Performance,Team Value	89.60%
[7]	Multi-Layer Perceptron (ANN)	Match Statistics	88.53%
[7]	Support Vector Machine	Match Statistics	86.28%
[7]	Radial Basis Function	Match Statistics	81.43%
[10]	Logistic Regression	Match Attributes, Match Statistics	80.12%
[10]	XGBoost	Match Attributes, Match Statistics	76.30%
[14]	Binomial Logistic Regression	Match Statistics, Team Rating, Passing Network Data	76.00%
[7]	Naive Bayes	Match Statistics	75.46%
[8]	Gradient Boosted Trees	Match Attributes, Match Statistics, Team Rating, Player Positions	75.38%
[14]	Naive Bayes	Match Attributes, Match Statistics, Team Rating, Passing Network Data	75.00%
[10]	Support Vector Machine with Polynomial Kernel	Match Attributes, Match Statistics	74.80%
[10]	Random Forest	Match Attributes, Match Statistics	74.50%
[21]	Linear Discriminant Analysis	Position Tracking Data	74.10%
[14]	Random Forest	Match Attributes, Match Statistics, Team Rating, Passing Network Data	74.00%
[10]	Multi-Layer Perceptron (ANN)	Match Attributes, Match Statistics	73.90%
[10]	Support Vector Machine with Linear Kernel	Match Attributes, Match Statistics	73.90%
[21]	Gradient Boosted Trees	Position Tracking Data	71.20%
[10]	Support Vector Machine with Radial Basis Function Kernel	Match Attributes, Match Statistics	70.30%
[14]	Artificial Neural Network	Match Attributes, Match Statistics, Team Rating, Passing Network Data	69.00%
[21]	Decision Tree	Position Tracking Data	67.10%
[35]	Support Vector Machine with Gaussian Radial Basis Function Kernel	Possesion Chain Data	66.60%
[21]	Quadratic Discriminant Analysis	Position Tracking Data	64.50%
[10]	K-Nearest Neighbors	Match Attributes, Match Statistics	63.20%
[10]	Decision Tree	Match Attributes, Match Statistics	59.60%
[8]	Support Vector Machine	Match Attributes, Match Statistics, Team Rating, Player Positions	56.92%
[8]	Logistic Regression	Match Attributes, Match Statistics, Team Rating, Player Positions	53.85%
[8]	Case-based Reasoning	Match Attributes, Match Statistics, Team Rating, Player Positions	52.31%
[8]	Decision Tree	Match Attributes, Match Statistics, Team Rating, Player Positions	52.31%
[8]	Artificial Neural Network	Match Attributes, Match Statistics, Team Rating, Player Positions	50.77%

Table 2.5: Binary and In-Game Data

Multiclass and Pre-Game Data

This category takes into account the prediction models that apply multiclass classification and which only make use of data that is known before the start of the match. Possible classes are a win for team A, a draw, and a win for team B. This category is quite big compared to the others and has a lower average accuracy of 60.24%. This is as expected due to predicting three different classes and only using pre-game data. The models' accuracy, algorithm, and feature selection are displayed in Table 2.6.

When looking at the top models which have an accuracy of higher than 75%, we notice several things. Ensembles of different algorithms are quite successful. But next to these, also random forests, gradient boosting and long short-term memory were quite successful. The feature categories that are a part of this group consist of match attributes, player attributes, player ratings, team performance, head-to-head performance, team values, and coach ratings. When looking at the bottom of the table, in particular, the models that have an accuracy of lower than 45%, we see that these made use of the k-nearest neighbors algorithm, naive Bayes, a decision tree algorithm, and a Hugin Bayesian learner. Two of these badly performing models made use of the k-nearest neighbors algorithm. As we saw in another category, the k-nearest neighbors algorithm does not seem a suitable algorithm to predict the outcome of a football match. Additionally, we observed that the feature category called player positions is not performing at all. This also seems to be the case in this category with the models of [24]. Another noticeable thing about the feature selections is that these are much richer than the feature selections in the last category. This is quite logical due to not being able to use match statistics which is obviously tells a great deal about the course of the match.

To find out which algorithms and feature categories are promising we will have to compare the studies properly. First, we will take a look at some studies that outperformed other studies which made use of roughly the same algorithms. From this group, we know that their feature selection was quite successful. [5] made use of match attributes, player attributes, and player ratings. [2] made use of match attributes, team ratings, team performance, player ratings, and player values. [19] and [33] made use of match attributes, team performance, and head-to-head performance. [32] made use of match attributes, team performance, team values, and coach ratings. And finally, [3] made use of match attributes and team performance. When looking at this group, we see several important parts of each feature selection. Match attributes are part of all of these feature selections. Followed by team performance which is a part of all but one. And finally, we observe that team/player/coach rating and values can be of use. When we take a look at these studies individually we can compare the algorithms because the algorithms are used in combination with the same feature selection. The best performing algorithms of the studies mentioned above are two ensembles of different algorithms, two artificial neural networks, a random forest, and logistic regression. Next to that, we see that the random forest algorithm performs second-best in three of the six studies.

Study	Algorithm	Feature Selection	Highest Accuracy
[5]	Ensemble (Random forest, Gradient Boosted Trees, Support Vector Machine, Linear Regression)	Match Attributes, Player Attributes, Player Rating	81.77%
[5]	Random Forest	Match Attributes, Player Attributes, Player Rating	81.26%
[3]	Long Short-Term Memory (RNN)	Match Attributes, Team Performance	80.75%
[33]	Random Forest	Match Attributes, Team Performance, Head-to-Head Performance	80.00%
[5]	Gradient Boosted Trees	Match Attributes, Player Attributes, Player Rating	79.12%
[32]	Ensemble (Support Vector Machine, Random Forest, Naïve Bayes) + Softmax	Match Attributes, Team Performance, Team Value, Coach Rating	76.50%
[2]	Multi-Layer Perceptron (ANN)	Match Attributes, Team Rating, Team Performance, Player Rating, Player Value	73.57%
[23]	Probabilistic Logic System	Match Attributes, Team Performance, Head-to-Head Performance	73.00%
[2]	Random Forest	Match Attributes, Team Rating, Team Performance, Player Rating, Player Value	72.92%
[5]	Linear Regression	Match Attributes, Player Attributes, Player Rating	72.92%
[32]	Random Forest	Match Attributes, Team Performance, Team Value, Coach Rating	71.40%
[19]	Logistic Regression	Match Attributes, Team Performance, Head-to-Head Performance	70.27%
[5]	Support Vector Machine	Match Attributes, Player Attributes, Player Rating	69.71%
[19]	Support Vector Machine	Match Attributes, Team Performance, Head-to-Head Performance	69.15%
[19]	AdaBoost	Match Attributes, Team Performance, Head-to-Head Performance	69.15%
[6]	LogitBoost	Team Performance, Head-to-Head Statistics, Injuries	68.80%
[6]	Artificial Neural Network	Team Performance, Head-to-Head Statistics, Injuries	68.80%
[32]	Naive Bayes	Match Attributes, Team Performance, Team Value, Coach Rating	67.90%
[19]	Decision Tree	Match Attributes, Team Performance, Head-to-Head Performance	67.76%
[19]	Naive Bayes	Match Attributes, Team Performance, Head-to-Head Performance	67.71%
[28]	Softmax Regression	Match Attributes, Betting Odds	67.60%
[1]	Extremely Randomized Trees Classifier	Match Attributes, Betting Odds, Weather Data	65.90%
[2]	Gaussian Naive Bayes	Match Attributes, Team Rating, Team Performance, Player Rating, Player Value	65.84%
[6]	Random Forest	Team Performance, Head-to-Head Statistics, Injuries	65.50%
[32]	Support Vector Machine with Radial Basis Function Kernel	Match Attributes, Team Performance, Team Value, Coach Rating	65.30%
[1]	Support Vector Machine	Match Attributes, Betting Odds, Weather Data	64.10%
[1]	Random Forest	Match Attributes, Betting Odds, Weather Data	63.30%
[6]	K-Nearest Neighbors	Team Performance, Head-to-Head Statistics, Injuries	62.50%
[1]	K-Nearest Neighbors	Match Attributes, Betting Odds, Weather Data	62.10%
[26]	Ensemble (Logistic Regression, Linear Discriminant Analysis, Quadratic Discriminant Analysis, K-Nearest Neighbors, Support Vector Machine, Decision Tree, Random Forest)	Match Attributes, Team Rating, Team Performance, Player Rating, Audience Data	62.03%
[26]	Linear Discriminant Analysis	Match Attributes, Team Rating, Team Performance, Player Rating, Audience Data	61.00%
[26]	Support Vector Machine	Match Attributes, Team Rating, Team Performance, Player Rating, Audience Data	59.40%
[29]	Ordered Logit Regression	Team Rating	59.38%
[24]	Bayesian Network	Match Attributes, Team Performance, Player Positions	59.21%
[29]	Bivariate Poisson	Team Rating	58.98%
[26]	Decision Tree	Match Attributes, Team Rating, Team Performance, Player Rating, Audience Data	58.80%
[2]	Support Vector Machine	Match Attributes, Team Rating, Team Performance, Player Rating, Player Value	58.77%
[29]	Ordered Logit Regression	Team Performance, Team Rating	58.20%
[26]	Logistic Regression	Match Attributes, Team Rating, Team Performance, Player Rating, Audience Data	57.80%
[26]	Quadratic Discriminant Analysis	Match Attributes, Team Rating, Team Performance, Player Rating, Audience Data	57.80%
[29]	Bivariate Poisson	Team Performance, Team Rating	57.03%
[23]	AdaBoost	Match Attributes, Team Performance, Head-to-Head Performance	57.00%
[16]	Gradient Boosted Trees	Match Attributes, Team Performance, Team Rating, Past Match Statistics	56.72%
[29]	Bivariate Poisson	Team Performance	56.64%
[16]	Random Forest	Match Attributes, Team Performance, Team Rating, Past Match Statistics	56.41%
[6]	Naive Bayes	Team Performance, Head-to-Head Statistics, Injuries	56.30%
[6]	Bayesian Network	Team Performance, Head-to-Head Statistics, Injuries	56.30%
[29]	Ordered Logit Regression	Team Performance	55.86%
[23]	Support Vector Machine	Match Attributes, Team Performance, Head-to-Head Performance	55.00%
[39]	Hierarchical Poisson Log-Linear	Match Attributes, Team Performance	54.85%
[16]	Support Vector Machine with Radial Basis Function Kernel	Match Attributes, Team Performance, Team Rating, Past Match Statistics	54.53%
[16]	Support Vector Machine with Linear Kernel	Match Attributes, Team Performance, Team Rating, Past Match Statistics	54.22%
[23]	K-Nearest Neighbors	Match Attributes, Team Performance, Head-to-Head Performance	54.00%
[23]	Random Forest	Match Attributes, Team Performance, Head-to-Head Performance	54.00%
[36]	Support Vector Machine	Twitter Data	54.00%
[39]	Bradley Terry	Match Attributes, Team Performance	53.88%
[1]	Random Forest	Match Attributes, Betting Odds	53.30%
[23]	Logistic Regression	Match Attributes, Team Performance, Head-to-Head Performance	53.00%
[1]	Extremely Randomized Trees Classifier	Match Attributes, Betting Odds	52.80%
[16]	Naive Bayes	Match Attributes, Team Performance, Team Rating, Past Match Statistics	52.66%
[17]	K-Nearest Neighbors	Match Attributes, Team Performance	51.94%
[1]	K-Nearest Neighbors	Match Attributes, Betting Odds	51.60%
[17]	XGBoost	Match Attributes, Team Performance	50.49%
[26]	Random Forest	Match Attributes, Team Rating, Team Performance, Player Rating, Audience Data	50.30%
[1]	Support Vector Machine	Match Attributes, Betting Odds	49.80%
[23]	Decision Tree	Match Attributes, Team Performance, Head-to-Head Performance	49.00%
[38]	Bayesian Function	Match Attributes, Team Performance, Head-to-Head Statistics	47.92%
[26]	K-Nearest Neighbors	Match Attributes, Team Rating, Team Performance, Player Rating, Audience Data	44.90%
[24]	Naive Bayes	Match Attributes, Team Performance, Player Positions	38.81%
[24]	MC4 Decision Tree	Match Attributes, Team Performance, Player Positions	38.65%
[24]	K-Nearest Neighbors	Match Attributes, Team Performance, Player Positions	37.06%
[24]	Hugin Bayesian Learner	Match Attributes, Team Performance, Player Positions	32.62%

Table 2.6: Multiclass and Pre-Game Data

Multiclass and In-Game Data

This category takes into account the prediction models that apply multiclass classification and which make use of data that is known before the start of the match and data that is produced during the match. Possible classes are a win for team A, a draw, and a win for team B. This category is of average size compared to the others and has an average accuracy of 69.21%. The models' accuracy, algorithm, and feature selection are displayed in Table 2.7.

When looking at the table, we notice several things. [34] achieved some impressive results with three models having an accuracy of over 90%. It is not completely fair to say that the accuracies would be as high if it concerned a full competition due to [34] using only a subset of the matches. The accuracy only applies to matches between the top six Premier League clubs. All of these models made use of a support vector machine and had a feature selection consisting of match attributes and match statistics. Other top models made use of decision tree algorithms, tree augmented naive Bayes, an artificial neural network, or gradient boosting. Most of these models made use of the same feature categories as [34], but two of these models also made use of features regarding team performance and betting odds. At the bottom of the table, we see quite some models of different studies that made use of artificial neural networks. Just like in the other categories, we see that the bottom models make use of a feature selection that contains player positions or betting odds.

Study	Algorithm	Feature Selection	Highest Accuracy
[34]	Support Vector Machine with Linear Kernel	Match Attributes, Match Statistics	100.00%
[34]	Support Vector Machine with Quadratic Kernel	Match Attributes, Match Statistics	98.75%
[34]	Support Vector Machine with Cubic Kernel	Match Attributes, Match Statistics	94.36%
[31]	Tree Augmented Naive Bayes	Match Attributes, Match Statistics	90.00%
[13]	Decision Forest	Match Attributes, Match Statistics	88.95%
[11]	Decision Forest	Match Attributes, Match Statistics	88.95%
[27]	C4.5 Decision Tree	Match Attributes, Match Statistics	85.00%
[34]	Support Vector Machine with Medium Radial Basis Function Kernel	Match Attributes, Match Statistics	83.70%
[18]	XGBoost	Match Attributes, Team Performance, Match Statistics, Betting Odds	82.40%
[12]	Long Short-Term Memory (RNN)	Match Attributes, Match Statistics, Team Performance, Betting Odds	80.75%
[31]	Bayesian Network	Match Attributes, Match Statistics	75.26%
[37]	Bayesian Networks	Match Attributes, Match Statistics	75.09%
[31]	Naive Bayes	Match Attributes, Match Statistics	74.03%
[25]	Decision Tree	Match Attributes, Match Statistics, Team Performance	73.40%
[13]	Artificial Neural Network	Match Attributes, Match Statistics	71.58%
[11]	Artificial Neural Network	Match Attributes, Match Statistics	71.58%
[27]	Random Forest	Match Attributes, Match Statistics	71.30%
[11]	Support Vector Machine	Match Attributes, Match Statistics	70.11%
[34]	Support Vector Machine with Coarse Radial Basis Function Kernel	Match Attributes, Match Statistics	69.91%
[22]	XGBoost	Match Attributes, Match Statistics	66.00%
[18]	Support Vector Machine	Match Attributes, Team Performance, Match Statistics, Betting Odds	66.00%
[30]	Logistic Regression	Match Statistics	62.67%
[34]	Support Vector Machine with Fine Radial Basis Function Kernel	Match Attributes, Match Statistics	62.28%
[8]	Gradient Boosted Trees	Match Attributes, Match Statistics, Team Rating, Player Positions	58.33%
[9]	LSTM Regression	Match Attributes, Match Statistics, Player Rating, Betting Odds	52.48%
[9]	Artificial Neural Network with Dense Layer	Match Attributes, Match Statistics, Player Rating, Betting Odds	52.41%
[9]	LSTM Classification	Match Attributes, Match Statistics, Player Rating, Betting Odds	52.06%
[8]	Artificial Neural Network	Match Attributes, Match Statistics, Team Rating, Player Positions	44.08%
[8]	Decision Tree	Match Attributes, Match Statistics, Team Rating, Player Positions	42.86%
[15]	Artificial Neural network with 3 Dense Layers	Match Statistics, Team Performance, Player Statistics	39.00%
[8]	Logistic Regression	Match Attributes, Match Statistics, Team Rating, Player Positions	36.90%
[8]	Case-based Reasoning	Match Attributes, Match Statistics, Team Rating, Player Positions	34.52%

Table 2.7: Multiclass and In-Game Data

2.3 Algorithms

Earlier, we discovered which algorithms have been used to predict the outcome of a football match. Also, we found out how many studies made use of specific algorithms. The algorithms that are used the most are support vector machine, artificial neural networks, gradient boosting, decision tree, logistic regression, naive Bayes, and random forest. The k-nearest neighbors algorithm and ensembles of different algorithms have been used less and linear regression has only been used once. The methods in the ”Other” group have not been used that much as well.

Additionally, we found out which algorithms have been successful and can be classified as ”promising”. The algorithms are divided into two groups. A group that contains the promising algorithms and a group that contains the less suitable algorithms. The latter contains algorithms that were outperformed on multiple occasions or which were very inconsistent. The promising algorithm group consists of the following algorithms: support vector machine, logistic regression, gradient boosting, random forest, and ensembles of different algorithms. The less suitable algorithm group consists of the k-nearest neighbor algorithm, decision tree algorithm, artificial neural networks, naive Bayes, and other methods. Even though some of these algorithms did show potential on a single occasion, they were not successful enough for different feature selections to classify them as ”promising”.

When taking into account the promising algorithms and the number of studies that used these, we see that ensembles of different algorithms achieved quite some success with only a small number of studies compared to the other algorithms. This means that research regarding ensembles in this area could be very beneficial.

2.4 Feature Categories

Next to the algorithms, we found out what feature categories have been used and how many different studies used them. Match attributes, match statistics, and team performance have been used the most. Followed by betting odds, head-to-head performance, player rating, and team rating. Furthermore, there is big group of different feature categories which have only been used once or twice.

Additionally, we found out which of these feature categories have been successful and can be classified as ”promising”. The following feature categories are classified as promising: match attributes, match statistics, team performance, head-to-head performance, coach ratings, player ratings, team ratings, team values, player attributes, and weather data. Next to that, we found that betting odds and player positions are not great feature categories to predict the outcome of a football match and can be classified as ”less suitable”. Furthermore, there were feature categories that were not classified as promising or less suitable due to not showing consistent success and not being bad either. When looking at the different categories, we concluded that variety in data is important, especially for the pre-game data categories.

2.5 Threats to Validity

There are some threats to the validity of this systematic literature review. Firstly, the incompleteness of the list containing relevant and accessible literature is a considerable threat. Several studies make use of different terms for football and the outcome of a football match. When using the most common way to describe the research questions, relevant literature could have been missed. To mitigate this threat, we defined a search string taking into account the other ways to describe football and the outcome of a football match. Secondly, not all studies that are a part of the relevant and accessible literature were included in the results. These studies did not present the accuracy or specific feature selection of their model. Not including these studies could result in having missed some relevant prediction models. Even though this could be the case, we believe the threat to validity to be limited due to the small number of studies that were not included.

2.6 Conclusion

The goal of this review is to find out what the state of the art in predicting the outcome of a football match using machine learning is. We used systematic literature review techniques to conduct the research and provided an overview of algorithms and feature categories that have been used to predict the outcome of a football match. Additionally, we found out which algorithms/feature categories are promising in this area.

Tables 2.1 and 2.2 contain the different algorithms and feature categories used to predict the outcome of a football match, respectively. As discussed in Section 2.3, the promising algorithms are support vector machine, logistic regression, gradient boosting, random forest, and ensembles of different algorithms. And as discussed in Section 2.4, the promising feature categories are match attributes, match statistics, team performance, head-to-head performance, coach ratings, player ratings, team ratings, team values, player attributes, and weather data.

In conclusion, in the area of football prediction has been quite some research regarding predicting the outcome of a football match. The research in this area experimented a lot with different combinations of algorithms and feature categories. A subset of these used algorithms/feature categories are promising but there is still potential in finding new algorithms/ensembles of different algorithms and feature selections.

2.6.1 Future Research

In the relevant and accessible literature, we found several interesting topics for future research. Some of the studies described their ideas for future research very specific while others described them quite abstract. Topics to do future research on are:

- Alternative Algorithms
- Alternative Features
- Improved Data
- Alternative Domains
- Alternative Leagues
- Include Draws/Scoring

• Alternative Process

The topics that we encountered most are ”Alternative Algorithms” and ”Alternative Features”. ”Alternative Algorithms” consists of the use of new or slightly different algorithms which potentially could improve the prediction accuracy. Several suggestions are light gradient boosting machine, ensembles, and deep learning methods such as recurrent neural networks and long short-term memory. ”Alternative Features” consists of the use of new or changed feature selections to improve the prediction accuracy. New feature suggestions are social media data, availability of key players, player transfers, new coach, level of injuries, attack and defense ratings, etcetera. Next to these two topics, we found five others. Namely, ”Improved Data”, ”Alternative Domains”, ”Alternative Leagues”, ”Include Draws/Scoring”, and ”Alternative Process”. ”Improved Data” describes making use of more detailed/up-to-date data. ”Alternative Domains” consists of the application of known prediction models in other areas, like tennis, golf, or basketball. ”Alternative Leagues” describes the use of known algorithms and features for different competitions where they might be more successful. ”Include Draws/Scoring” consists of including draws for the binary models and score predictions. And ”Alternative Process” describes the use of new feature engineering methods or looking at very specific and rare match events and predicting performance on an individual level.

For this thesis, we decided to conduct research in two of these future research topics. The topics will be ”Alternative Algorithms” and ”Alternative Features”. As mentioned before, the goal of this thesis is to find a way to improve the accuracy of predicting the outcome of a football match. We will attempt to achieve this goal by experimenting with different algorithms/ensembles of algorithms and feature selections.

In the literature review, it became clear that ensembles of different algorithms could potentially be used to increase the prediction accuracy of a model that predicts the outcome of a football match. Next to that, we found a group of algorithms which could be labeled as promising due to consistently achieving a higher accuracy than other algorithms. This suggests that ensembles consisting of promising algorithms have a great chance of increasing the prediction accuracy in this area. There are many ways to build such an ensemble based on the promising algorithms. We could, for instance, use all of them or just a subset. Also, we could make them equally important or make some more important than others.

We could also improve the feature selection to achieve a higher prediction accuracy. A lot of studies tried to do this by making use of a of subset of the promising feature categories while including a new feature category to find out whether it would have added value. What has not yet been experimented a lot with is finding combinations between the promising feature categories that we found. By not looking to create new features out of just one data source and combining different data sources into features, we could potentially create a whole new set of effective features. An example of this could be the combination of team performance and team rating. Combining these two could result in a feature selection that represents the team's performance against a specific class of team rating. This might result in a better prediction accuracy when comparing it with a feature selection which only uses features that are created out of a single feature category. As mentioned before, this thesis focuses on prediction models that only use data that is known before the match. This means we can not make use of all promising feature categories but only the ones that are not based on data which is produced during the match.

Chapter 3

Methodology

In this chapter, we will describe the methodology and machine learning pipeline used to answer the rest of the research questions. In the first section, we will describe the methodology CRISP-DM, how it's used in general and how we will use it to answer our research questions. Next, we will explain all steps of our machine learning pipeline.

3.1 CRISP-DM

Cross-industry standard process for data mining (CRISP-DM) is a methodology developed in 1996 to guide data mining projects. It is well-known and still widely used. CRISP-DM consists of six phases. Business understanding, data understanding, data preparation, modelling, evaluation and deployment. The flow of these phases is illustrated in Figure 3.1. The phases form an iterative process in which can be moved back and forth between several phases. In the next paragraph, we will describe what each phase consists of in general. What each phase consists of specifically for this thesis, can be found below the following paragraph.

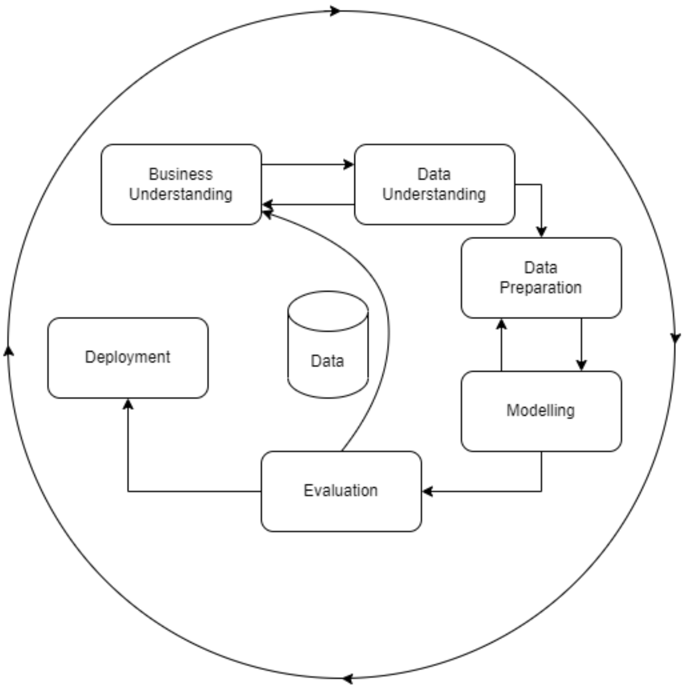

Figure 3.1: CRISP-DM

In the phase business understanding, we should focus on understanding the goals and requirements of a project. It is important to understand the main objectives of the project and the requirements of all involved actors. In the phase data understanding, the necessary data is identified and gathered. It is important to know with what data types and data sizes you are working. In the phase data preparation, the data is selected, cleaned, merged and manipulated. It is important for the modelling step that the data is complete and correct. In the modelling phase, we should decide on the necessary techniques, execute tests and build models. In the phase evaluation, the produced results will be looked at while taking into account the business understanding. Only then we can evaluate whether the objectives have been achieved and the requirements have been met. Furthermore, the business understanding will be updated. In the phase deployment, we should make sure that the customer can access the results. Also, in this phase, we should review on the project in its entirety.

For this research, two cycles within CRISP-DM are completed. Each cycle answers a research question. In the first cycle, our goal is to find out how feature category combinations can improve the accuracy in predicting the outcome of a football match. In the second cycle, the goal is to find out how ensembles can improve the accuracy in predicting the outcome of a football match. In Table 3.1, we can find which chapters and sections belong to the specific CRISP-DM cycles and phases. In the phase business understanding, we identify the research problem, define our research goal and create our research questions. It is important to understand the main topics of our research and the topics that are connected to our research. In the phase data understanding, we identify and gather the data necessary and find out what types of data we are going to deal with. In the data preparation phase, we clean, merge and modify the data in such a way that we can create the desired features. These features will be used in the modelling phase where we create the prediction models. In the modelling phase, we make use of several algorithms and feature selections to create prediction models. These will be compared with each other in the evaluation phase. In the evaluation phase, the results will be looked at and conclusions will be drawn while taking into account the business understanding. Only then we can evaluate whether the goal of our research have been met or not. In the phase deployment, all conclusions drawn from our research will be provided. Furthermore, the limitations and recommendations will be described.

	Business Understanding	**Data Understanding**	**Data Preparation**	**Modelling**	**Evaluation**	**Deployment**
1	Chapter 1	Section 4.4	Chapter 4	Chapter 4	Chapter 4	
2	Chapter 1	Section 4.4	Chapter 4	Chapter 5	Chapter 5	Chapter 6

Table 3.1: CRISP-DM Cycles

3.2 Machine Learning Pipeline

To be able to answer the research questions, we made use of three main steps in our machine learning pipeline. Namely, the preparation of the datasets, the preparation of the feature selection and the training and testing of the machine learning models. For each of these steps we made use of python to do several tasks like merging, deleting and transforming data. Next to that, python was used to train and test the machine learning models, while making use of 5-fold cross validation. In Figure 3.2, we can see the whole machine learning pipeline, the three main steps and the sub steps that are a part of them. Each step relates to at least one CRISP-DM phase and in some cases more. The step prepare datasets describes the actions that we took during the phase data understanding and some actions that we took during the phase data preparation. The step prepare feature selection describes the remaining actions that we took during the CRISP-DM phase data preparation. The step training and testing of the machine learning models describes the actions that we took during the CRISP-DM phases modelling and evaluation.

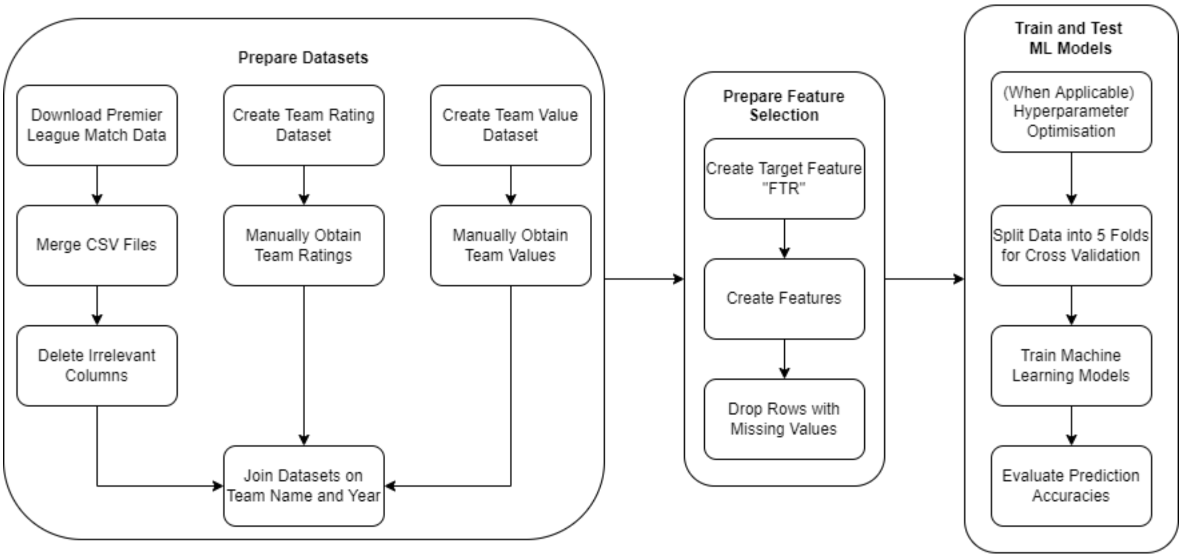

Figure 3.2: Machine Learning Pipeline

The first main step is preparing the datasets which contains several flows, each representing the preparation of a single dataset. The first flow starts with downloading twenty csv-files, each containing the Premier League match data of a single completed season. Next, the csv-files are merged into one file after which the irrelevant columns are deleted. Irrelevant columns are columns that will not be used throughout the whole machine learning pipeline. Finally, the dataset will be combined with the two other datasets. The second and third flow are somewhat different from the first due to the fact that these datasets could not be downloaded. We needed to create these datasets ourselves and manually obtain the data from the sources before we combined them with the other two datasets.

Now our data is present and does not contain any irrelevant data, we can start to prepare the feature selection. The first step in preparing the feature selection is the creation of a target feature. For the target feature, we used the full time result of a match which could be 'H', 'D', and 'A'. We changed these into numerical values 0, 1, and 2, respectively. Finally, the features should be created out of the prepared dataset and the rows with missing values should be dropped.

The last step in the machine learning pipeline is the training and testing of the machine learning models. The first sub step is splitting the data into 5 folds for cross validation, unless hyperparameter optimisation is applicable. Cross validation is a resampling method that makes use of different portions of the data when training and testing a model on different iterations and is necessary to reduce the bias and variance of the models. Next, the machine learning models will be trained and the prediction accuracies will be evaluated.

Chapter 4

Feature Category Combinations: Feature Creation, Modelling & Evaluation

This chapter describes what feature category combination features are and whether these have the potential to be new and interesting features. Possible combinations are discussed and some of these are tested to find out whether the use of these could result in a potential gain. During this chapter we will go through the data preparation, modelling and evaluation phase of the first CRISP-DM cycle, as illustrated in Table 3.1. Next to that, we will make use of each of the three steps of the machine learning pipeline which is displayed in Figure 3.2. There is only one sub step that will skipped due not being applicable in this chapter. That sub step is hyperparameter optimisation. In Section 4.1, we will explain the general idea behind feature category combinations. In Section 4.2, we will talk about possible feature category combinations based on the promising feature categories that we discovered in our systematic literature review. In Section 4.3, we will explain the evaluation process to decide whether the inclusion of a combination can increase performance. In Section 4.4, we will describe the necessary data sets. In Section 4.5, we will describe the feature category combination between team performance and team rating. In Section 4.6, we will describe the feature category combination between past match statistics and team rating. In Section 4.7, we will describe the feature category combination between team performance and team value. In Section 4.8, we will describe the feature category combination between past match statistics and team value. Each section describing a feature category combination will give a detailed view on the features that have created and tested. Also, they provide a summary to properly evaluate the whole combination. In Section 4.9, we will have a discussion regarding all four feature category combinations and whether feature category combination can be worthwhile in general. In Section 4.10, we will describe the key takeaways of this chapter.

4.1 Combining Feature Categories

During the systematic literature review, we found out that having a more various feature selection would result in a better prediction model. This means that adding more features from different feature categories would result in improved performance. A way to create new influential features is to find new influential feature categories to create these features from. There is already quite some research out there that focuses on this area. Another way to achieve this could be to consider combining existing feature categories into a new set of features. By default, features are created out of a single feature category, as visualised in Figure 4.1.

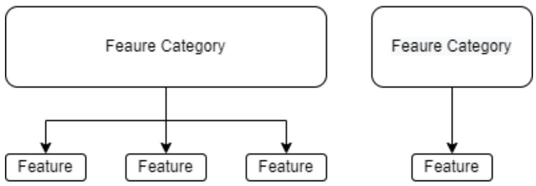

Figure 4.1: Standard Feature Creation

But by only considering such features, a potential gain might be missed. Considering a different approach where combinations between feature categories are included could result in interesting new features. This approach is visualised in Figure 4.2.

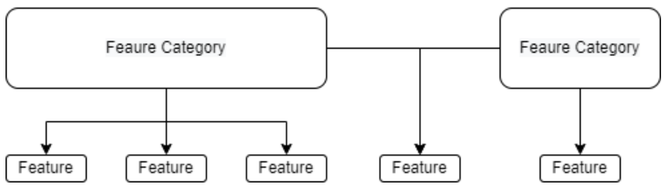

Figure 4.2: Combination Feature Creation

4.2 Possible Feature Category Combinations

In the systematic literature review became clear that some feature selections are better than others. As a result, we came up with a list of promising feature categories. This list consists of match attributes, match statistics, team performance, head-to-head performance, coach ratings, player ratings, team ratings, team values, player attributes, and weather data. For this thesis, we decided to only make use of data that is known before the start of a football match. This means that we can not make use of the match statistics, but we can make of past match statistics.

Obviously, there are countless possibilities when looking for feature category combinations but the idea is that a feature category combination should improve the accuracy of the prediction model without the addition of new data. In other words, it must provide valuable new information with the same data. Most feature categories do not fit together well, unfortunately due to not providing new information or information which directly relates to the outcome of a football match. Below is list containing feature category combinations based on the promising feature categories. Each combination can consist of multiple feature category combination features. These feature category combinations have the best chance of improving the prediction models directly.

- Team Performance and Team Rating
- Past Match Statistics and Team Rating
- Team Performance and Team Value
- Past Match Statistics and Team Value
- Team Performance and Coach Rating
- Past Match Statistics and Coach Rating
- Team Performance and Player Rating
- Past Match Statistics and Player Rating
- Team Performance and Player Attributes
- Past Match Statistics and Player Attributes
- Team Performance and Weather Data
- Past Match Statistics and Weather Data

4.3 Evaluation Process

To decide whether a feature category combination can be used to improve the prediction accuracy, we will compare a feature selection containing the feature category combination features and a feature selection without the feature category combination features. If the accuracy has increased after adding the feature category combination features, we can conclude that the feature category combination features can be used to improve the prediction accuracy. For both feature selections, we will create models with the promising algorithms that we found in our systematic literature review and an equally weighted ensemble of them. This means that we will compare the feature selections across five different algorithms. For this chapter, we only make use of the promising algorithms in their default state to make sure that the feature selection is the only difference between the models that are compared. During this procedure we took into account 5-fold cross validation to reduce the bias and variance of the models. To conclude whether or not an increase of an inclusion can be seen as worthwhile, we can not simply look at whether or not an inclusion leads to a higher prediction accuracy. We will need to take a look at whether the inclusion results in an actual performance increase that is 10.0% or higher. This does not mean an increase of 10.0% or more in terms of prediction accuracy but it means that the increase which is present due to the inclusion of the feature category combination features should be 10.0% or more than the increase which is present due to the exclusion of the feature category combination features. To make that more clear, if the increase in prediction accuracy due to exclusion is 4.0%, the increase in prediction accuracy due to inclusion must be 4.4% or above to be able to claim that the increase is worthwhile. We chose a minimum of 10.0% because the data is already available and familiar and it would only cost somewhat more resources. To find out the increase in prediction accuracy of these feature selections, we need to take into account the ratio of the class that is present the most. This class represents the home team winning and the home win percentage is 45.9%. This means that when you constantly guess that the home team will win, the prediction accuracy will be 45.9%. The prediction accuracy of the feature selections minus the home win percentage results in the increase in prediction accuracy realised by this feature selection. This means that if the use of the feature selection that excludes the feature category combination features results in a prediction accuracy of 49.9%, the increase in prediction accuracy or delta is 4.0%. This also means that the prediction accuracy due to including the feature category combination features must be 4.4% or higher to be able to claim that the increase is worthwhile.

4.4 Datasets

To be able to create the features data from three datasets was needed. We made use of a dataset containing match attributes, match statistics and betting odds, a dataset containing the team ratings obtained from the team stats database from the video game series FIFA and a dataset containing the team values obtained from transfermarkt.com.

4.4.1 Premier League Match Data

The main dataset is obtained from http://www.football-data.co.uk/. It contains match attributes, match statistics and betting odds of football matches played in the English Premier League. From all the data available here, we considered the last 20 seasons. Namely, data from the seasons 01/02 to 20/21. For each season they provided a csv-file containing data of all matches played in that season. There are 380 matches played each season which means that we obtained a total of 7600 rows except for the headers. Every row contains 48 fields. All fields can be found in Table 4.1. Most of these have a data type which is integer or float but there are also some that make use of a string or a date data type. An example of how this data looks can be found in appendix A.1,

League Division	Away Team Fouls Committed	Ladbrokes home win odds
Match Date (dd/mm/yy)	Home Team Corners	Ladbrokes draw odds
Time of match kick off	Away Team Corners	Ladbrokes away win odds
Home Team	Home Team Yellow Cards	Sporting Odds home win odds
Away Team	Away Team Yellow Cards	Sporting Odds draw odds
Full Time Home Team Goals	Home Team Red Cards	Sporting Odds away win odds
Full Time Away Team Goals	Away Team Red Cards	Sportingbet home win odds
Full Time Result	Bet365 home win odds	Sportingbet draw odds
Half Time Home Team Goals	Bet365 draw odds	Sportingbet away win odds
Half Time Away Team Goals	Bet365 away win odds	William Hill home win odds
Half Time Result	Gamebookers home win odds	William Hill draw odds
Home Team Shots	Gamebookers draw odds	William Hill away win odds
Away Team Shots	Gamebookers away win odds	Gamebookers over 2.5 goals
Home Team Shots on Target	Interwetten home win odds	Gamebookers over 2.5 goals
Away Team Shots on Target	Interwetten draw odds	Bet365 over 2.5 goals
Home Team Fouls Committed	Interwetten away win odds	Bet365 under 2.5 goals

Table 4.1: Fields Football-Data Dataset

4.4.2 Premier League FIFA Team Ratings

The team rating dataset was obtained from fifaindex.com. Instead of downloading a file containing everything we needed, we had to manually obtain this data. For the seasons 04/05 till 20/21, the team ratings are published online. For each season we created a csv-file containing twenty-one rows and five columns. For all twenty teams competing in the Premier League that season, it contains the team name and the team rating which can be divided into four different fields. The four different fields are the attack, midfield, defend and overall rating of a team. The team name column contains strings, unlike the others which are integers. An example of this data can be found in appendix A.2.

4.4.3 Premier League Transfermarkt Team Value

The team value dataset was obtained from transfermarkt.com. This dataset was also obtained manually. For the seasons 04/05 till 20/21, they have the team values published online. For each season we created a csv-file containing the team value for the twenty teams that competed that season. The value is noted in millions of euros. The csv-file contains twenty-one rows and two columns. A column which states the team name and a column which states the team value. The former which contains strings and the latter which contains floats. An example of this data can be found in appendix A.3.

4.5 Team Performance and Team Rating

This category combines the feature categories team performance and team rating. Team performance is often represented by several features. Features regarding points achieved in the current season, points achieved in the previous season, and form of the last few matches. Next to these, there are many others that could be used. Team rating can also be represented by different features, namely, attack, midfield, defence and overall rating. For this feature category, we calculated the total of these ratings and assigned a class to each team every season based on this rating. Classes can be 1 to 5 where 1 is the worst class and 5 is the best class. The main idea of this feature category combination is to look at a team's performance with respect to the team rating class of the opponent. Features that reflect a team's performance against a specific group of opponents can potentially be used to improve the prediction model.

4.5.1 Points Current Season and Team Rating Class

The first feature category combination feature set we are going to examine is the combination between points achieved in the current season and team rating class. When using standard feature creation we would look at four features regarding the total amount of points collected by the home and away team and their team rating classes. But when taking into account combination feature creation, we would also look at features regarding the amount of points collected by the home and away team against the team rating class of the opponent. These features are not created out of a single feature category, but out of both, as we can see in Figure 4.3. Before we will describe the results, we will explain what data is being used for training and testing.

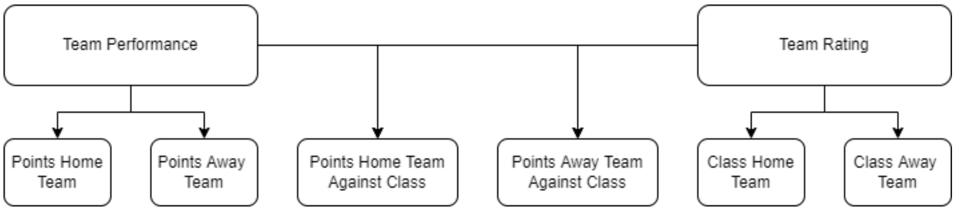

Figure 4.3: Points Current Season and Team Rating Class

Combination Excluded

When taking into account standard feature creation, the relevant data will consist of four features next to the target feature, ”FTR”. Namely, ”TRH”, ”TRA”, ”PH”, and ”PA”, as we can see in Table 4.2. ”FTR” represents the outcome of the football match. It can only be a 0, 1 or 2 which represent a home win, a draw and an away win, respectively. The features ”TRH” and ”TRA” represent the class of the home and away team based on their team rating. The features ”PH” and ”PA” represent the points collected by the home and away team during this season. As mentioned before, ”TRH” and ”TRA” are integers which range from 1 to 5 where 1 represents the worst class and 5 represents the best class, ”PH” and ”PA” are integers, as well, but can range from 0 to 114. In the worst case where a team would lose every match, it would still be zero till the end of the season. In the best case where a team would win every match, it would be 114 at the end of the season. As mentioned before, a team collects three points when winning a match, one point when the match ends in a draw, and zero points when losing the match.

FTR	**TRH**	**TRA**	**PH**	**PA**
(0,1,2)	*(1-5)*	*(1-5)*	*(0-114)*	*(0-114)*
0	3	2	25	20
1	1	2	13	13
0	2	4	33	24
2	2	2	20	19
1	3	2	20	21

Table 4.2: Points Current Season and Team Rating Class: Combination Excluded

Combination Included

When taking into account combination feature creation, the relevant data will contain two other features, namely ”PTRH” and ”PTRA”. An example of the data for all features can be found in Table 4.3. ”PTRH” and ”PTRA” represent the amount of points the home and away team collected against the opponent’s team rating class. These features are integers and can range from 0 to 114. The maximum of these integers will almost always be lower, due to it depending on the size of the opponent's class and it being very unlikely that all other teams would be part of the same class. Their minimum is zero which is only the case when every game against a certain class would be lost up to the relevant match.

FTR	**TRH**	**TRA**	**PH**	**PA**	**PTRH**	**PTRA**
(0,1,2)	*(1-5)*	*(1-5)*	*(0-114)*	*(0-114)*	*(0-114)*	*(0-114)*
0	3	2	25	20	14	0
1	1	2	13	13	8	2
0	2	4	33	24	0	12
2	2	2	20	19	12	9
1	3	2	20	21	12	1

Table 4.3: Points Current Season and Team Rating Class: Combination Included

Results

In Table 4.4, the accuracies of the ten different models can be found. The two different feature selections are used across five different algorithms. One feature selection that excludes the feature category combination features and one feature selection that includes the feature category combination features. As we can see in the table below, every algorithm except for XGBoost benefits from the inclusion of the feature category combination features. Especially, the random forest algorithm whose performance is increased by 2.6%. On average, performance increases by 0.5% when including the feature category combination features. This indicates that the inclusion can increase the accuracy of a model that predicts the outcome of a football match.

What it does not say is whether this increase is worthwhile. To find that out, we need the percentage of the class that is present the most. In other words, how often you will be correct when guessing that the home team will win. This percentage is 45.9%. The feature selection that excludes the feature category combination features has a prediction accuracy of 50.8% on average. This means that this feature selection increases the prediction accuracy by 4.9%. The feature selection that includes the feature category combination features increases the prediction accuracy by 51.3%. This means that the delta between the inclusion and exclusion of the feature category combination features is 0.5%. When dividing this delta with the delta between the feature selection that excludes the feature category combination features and the home win percentage, we can find the actual performance increase or in other words the effect of the inclusion of the feature category combination features. For the selected feature category combination features the actual performance increase is 10.2%. When using the same data in a more elaborate way, an increase of 10.2% can be seen as worthwhile.

Algorithm	Combination Excluded	Combination Included	Delta
Random Forest	0.461	0.487	0.026
XGBoost	0.489	0.482	-0.007
Logistic Regression	0.537	0.538	0.001
Support Vector Machine	0.528	0.529	0.001
Ensemble	0.527	0.529	0.002
Average	0.508	0.513	**0.005**

Table 4.4: Points Current Season and Team Rating Class: Results

4.5.2 Points Last Season and Team Rating Class

The next feature category combination feature set we are going to examine is the combination between points collected during last season and team rating class. When using standard feature creation we would look at four features regarding the total amount of points collected by the home and away team during last season and their team rating classes. But when taking into account combination feature creation, we would also look at features regarding the amount of points collected by the home and away team against the team rating class of the opponent during last season. As we can see in Figure 4.4, both of the feature categories are necessary to create these features. Before we will describe the results, we will explain what data is being used for training and testing.

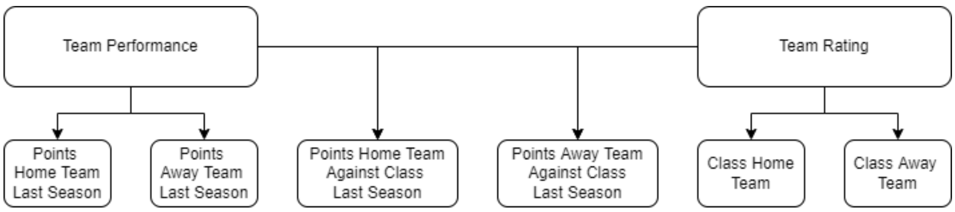

Figure 4.4: Points Last Season and Team Rating Class

Combination Excluded

When taking into account standard feature creation, the relevant data will consist of four features next to the target feature, ”FTR”. Namely, ”TRH”, ”TRA”, ”LPH”, and ”LPA”, as we can see in Table 4.5. ”FTR” represents the outcome of the football match. It can only be a 0, 1 or 2 which represent a home win, a draw and an away win, respectively. The features ”TRH” and ”TRA” represent the class of the home and away team based on their team rating. ”LPH” and ”LPA” represent the points collected by the home and away team during last season. ”TRH” and ”TRA” are integers and range from 1 to 5 where 1 represents the worst class and 5 represents the best class. ”LPH” and ”LPA” are also integers and can range from 0 to 114. In the worst case, they would be zero when every match has been lost last season. In the best case, they would be 114 when every match has been won last season.

FTR	**TRH**	**TRA**	**LPH**	**LPA**
(0,1,2)	*(1-5)*	*(1-5)*	*(0-114)*	*(0-114)*
0	4	2	52	44
0	2	2	42	34
2	2	4	46	52
2	2	5	44	77
1	2	4	39	44

Table 4.5: Points Last Season and Team Rating Class: Combination Excluded

Combination Included

When taking into account combination feature creation, the relevant data will contain another two features, namely ”LPTRH” and ”LPTRA”. An example of the relevant data containing all features can be found in Table 4.6. ”LPTRH” and ”LPTRA” represent the amount of points the home and away team collected against the opponent’s team rating class during last season. They can range from 0 to 114. The maximum of these integers will almost always be lower, due to it depending on the size of the opponent’s class and it being very unlikely that all other teams would be part of the same class. Their minimum is zero which is only the case when every game against a certain class would be lost during last season.

FTR	**TRH**	**TRA**	**LPH**	**LPA**	**LPTRH**	**LPTRA**
(0,1,2)	*(1-5)*	*(1-5)*	*(0-114)*	*(0-114)*	*(0-114)*	*(0-114)*
0	4	2	52	44	32	0
0	2	2	42	34	27	29
2	2	4	46	52	0	32
2	2	5	44	77	1	44
1	2	4	39	44	1	29

Table 4.6: Points Last Season and Team Rating Class: Combination Included

Results

In Table 4.7, the accuracies of the ten different models can be found. The two different feature selections are used across five different algorithms. One feature selection that excludes the feature category combination features and one feature selection that includes the feature category combination features. As we can see in the table below, every algorithm benefits from the inclusion of the feature category combination features. Especially, the random forest algorithm whose performance is increased by 4.4%. On average performance is, increases by 1.2% when including the feature category combination features. This indicates that the inclusion can increase the accuracy of a model that predicts the outcome of a football match.

What it does not say is whether this increase is worthwhile. To find that out, we need the percentage of the class that is present the most. In other words, how often you will be correct when guessing that the home team will win. This percentage is 45.9%. The feature selection that excludes the feature category combination features has a prediction accuracy of 49.6% on average. This means that this feature selection increases the prediction accuracy by 3.7%. The feature selection that includes the feature category combination features increases the prediction accuracy by 50.8%. This means that the delta between the inclusion and exclusion of the feature category combination features is 1.2%. When dividing this delta with the delta between the feature selection that excludes the feature category combination features and the home win percentage, we can find the actual performance increase or in other words the effect of the inclusion of the feature category combination features. For the selected feature category combination features the actual performance increase is 32.4%. When using the same data in a more elaborate way, an increase of 32.4% can be seen as worthwhile.

Algorithm	Combination Excluded	Combination Included	Delta
Random Forest	0.440	0.484	0.044
XGBoost	0.474	0.476	0.002
Logistic Regression	0.526	0.530	0.004
Support Vector Machine	0.522	0.523	0.001
Ensemble	0.520	0.528	0.008
Average	0.496	0.508	**0.012**

Table 4.7: Points Last Season and Team Rating Class: Results

4.5.3 Points Current and Last Season and Team Rating Class

The next feature category combination feature set we are going to examine is the combination between points collected in the current and last season and team rating class. When using standard feature creation we would look at six features regarding the points collected by the home and away team in the current and last season and their team rating classes. But when taking into account combination feature creation, we would also look at features regarding the amount of points collected by the home and away team against the team rating class of the opponent during the current and last season. As we can see in Figure 4.5, both of the feature categories are necessary to create these features. Before we will describe the results, we will explain what data is being used for training and testing.

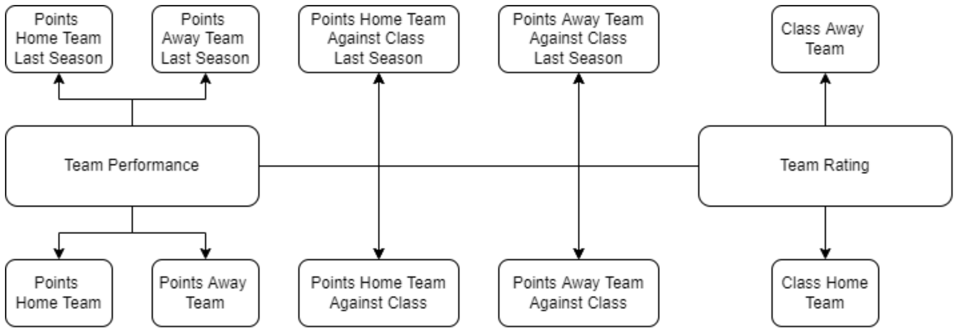

Figure 4.5: Points Current and Last Season and Team Rating Class

Combination Excluded

When taking into account standard feature creation, the relevant data will consist of six features next to the target feature, ”FTR”. Namely, ”TRH”, ”TRA”, ”PH”, ”PA”, ”LPH”, and ”LPA”, as we can see in Table 4.8. ”FTR” represents the outcome of the football match. It can only be a 0, 1 or 2 which represent a home win, a draw and an away win, respectively. The features ”TRH” and ”TRA” represent the class of the home and away team based on their team rating. The features ”PH” and ”PA” represent the points collected by the home and away team during this season. ”LPH” and ”LPA” represent the points collected by the home and away team during last season. As mentioned before, ”TRH” and ”TRA” are integers which range from 1 to 5 where 1 represents the worst class and 5 represents the best class, ”PH” and ”PA” are integers, as well, but can range from 0 to 114. In the worst case where a team would lose every match, it would still be zero till the end of the season. In the best case where a team would win every match, it would be 114 at the end of the season. ”LPH” and ”LPA” are also integers and can range from 0 to 114. In the the worst case, they would be zero when every match has been lost last season. In the best case, they would be 114 when every match has been won last season.

FTR	TRH	TRA	PH	PA	LPH	LPA
(0,1,2)	*(1-5)*	*(1-5)*	*(0-114)*	*(0-114)*	*(0-114)*	*(0-114)*
0	4	2	9	5	52	44
0	2	2	8	5	42	34
2	2	4	15	12	46	52
2	2	5	5	11	44	77
1	2	4	5	8	39	44

Table 4.8: Points Current and Last Season and Team Rating Class: Combination Excluded

Combination Included

When taking into account combination feature creation, the relevant data will contain another four features, namely ”PTRH”, ”PTRA”, ”LPTRH” and ”LPTRA”. An example of the relevant data containing all features can be found in Table 4.9. ”PTRH” and ”PTRA” represent the amount of points the home and away team collected against the opponent’s team rating class. These features are integers and can range from 0 to 114. The maximum of these integers will almost always be lower, due to it depending on the size of the opponent’s class and it being very unlikely that all other teams would be part of the same class. Their minimum is zero which is only the case when every game against a certain class would be lost up to the relevant match. ”LPTRH” and ”LPTRA” represent the amount of points the home and away team collected against the opponent’s team rating class during last season. They can also range from 0 to 114. The maximum of these integers will almost always be lower, due to it depending on the size of the opponent’s class and it being very unlikely that all other teams would be part of the same class. Their minimum is zero which is only the case when every game against a certain class would be lost during last season.

FTR	TRH	TRA	PH	PA	PTRH	PTRA	LPH	LPA	LPTRH	LPTRA
(0,1,2)	*(1-5)*	*(1-5)*	*(0-114)*	*(0-114)*	*(0-114)*	*(0-114)*	*(0-114)*	*(0-114)*	*(0-114)*	*(0-114)*
0	4	2	9	5	4	4	52	44	32	0
0	2	2	8	5	4	3	42	34	27	29
2	2	4	15	12	3	7	46	52	0	32
2	2	5	5	11	0	0	44	77	1	44
1	2	4	5	8	3	5	39	44	1	29

Table 4.9: Points Current and Last Season and Team Rating Class: Combination Included

Results

In Table 4.10, the accuracies of the ten different models can be found. The two different feature selections are used across five different algorithms. One feature selection that excludes the feature category combination features and one feature selection that includes the feature category combination features. As we can see in the table below, every algorithm except for logistic regression benefits from the inclusion of the feature category combination features. On average, performance increases by 0.3% when including the feature category combination features. This indicates that the inclusion can increase the accuracy of a model that predicts the outcome of a football match.

What it does not say is whether this increase is worthwhile. To find that out, we need the percentage of the class that is present the most. In other words, how often you will be correct when guessing that the home team will win. This percentage is 45.9%. The feature selection that excludes the feature category combination features has a prediction accuracy of 50.3% on average. This means that this feature selection increases the prediction accuracy by 4.4%. The feature selection that includes the feature category combination features increases the prediction accuracy by 50.6%. This means that the delta between the inclusion and exclusion of the feature category combination features is 0.3%. When dividing this delta with the delta between the feature selection that excludes the feature category combination features and the home win percentage, we can find the actual performance increase or in other words the effect of the inclusion of the feature category combination features. For the selected feature category combination features the actual performance increase is 6.8%. When using the same data in a more elaborate way, an increase of 6.8% can not be seen as worthwhile.

Algorithm	**Combination Excluded**	**Combination Included**	**Delta**
Random Forest	0.478	0.485	0.007
XGBoost	0.464	0.466	0.002
Logistic Regression	0.531	0.530	-0.001
Support Vector Machine	0.524	0.525	0.001
Ensemble	0.518	0.522	0.004
Average	0.503	0.506	**0.003**

Table 4.10: Points Current and Last Season and Team Rating Class: Results

4.5.4 Summary

In this subsection, we will look at the results of the whole feature category combination. We made use of three different feature category combination feature sets to evaluate whether the use of this feature category combination can improve the prediction accuracy. In Table 4.11, we can find the delta between the prediction accuracy when using the feature selection that excludes the feature category combination features and the home win percentage. The delta will be positive when the prediction accuracy while using the feature selection that excludes the feature category combination features is higher than the home win percentage. Also, we can find the delta between the prediction accuracy when using the feature selection that includes the feature category combination features and the home win percentage. The delta will be positive when the prediction accuracy while using the feature selection that includes the feature category combination features is higher than the home win percentage. When taking these two values into account, one can find out the delta between the prediction accuracies when including and excluding the feature category combination features. The delta will be positive when the prediction accuracy while using the feature selection that includes the feature category combination features is higher than the prediction accuracy while using the feature selection that excludes the feature category combination features. When dividing this delta with the delta between the prediction accuracy when using the feature selection that excludes the feature category combination features and the home win percentage, we can find the actual performance or in other words the effect of the inclusion of the feature category combination features compared to the effect of the feature selection that excludes them. This value is a percentage which can be positive and negative. As we can see in the table below, for each of the feature category combination sets the delta between the home win percentage and the prediction accuracy when using the feature selection that includes the feature category combination features is larger than the delta between the home win percentage and the prediction accuracy when using the feature selection that excludes the feature category combination features. We can also observe that two of the three feature category combination feature sets can be seen as worthwhile. The combination between points current and last season and team rating class only had an actual performance increase of 6.8% which means that it cannot be seen as worthwhile. The combination between points last season and team rating class performed the best achieving an actual performance increase of 32.4%. On average, the feature category combination achieved an actual performance increase of 16.5% which means that it can be seen as worthwhile.

Feature Category Combination Features	Delta Combination Excluded and Home Win	Delta Combination Included and Home Win	Delta Combination Included and Combination Excluded	Increase Actual Performance
Points Current Season and Team Rating Class	4.9%	5.4%	0.5%	10.2%
Points Last Season and Team Rating Class	3.7%	4.9%	1.2%	32.4%
Points Current and Last Season and Team Rating Class	4.4%	4.7%	0.3%	6.8%
Average	4.3%	5.0%	0.7%	16.5%

Table 4.11: Team Performance and Team Rating: Results

4.6 Past Match Statistics and Team Rating

This category combines the feature categories past match statistics and team rating. Past Match Statistics can be represented by several features. Examples of features are goals scored, goals conceded, shots on target and ball possession. Obviously, there are many more that can represent this feature category. In the previous feature category combination, we already described team rating in quite some detail. The way this feature category will be used will be unchanged. The main idea of this feature category combination is to look at a past match statistics with respect to the team rating class of the opponent. Features that reflect past match statistics against a specific group of opponents can potentially be used to improve the prediction model.

4.6.1 Goal Scored and Team Rating Class

The next feature category combination feature set we are going to examine is the combination between goals scored this season and team rating class. When using standard feature creation we would look at four features regarding the amount of goals scored by the home and away team during this season and their team rating classes. But when taking into account combination feature creation, we would also look at features regarding the amount of goals scored by the home and away team against the team rating class of the opponent during this season. As we can see in Figure 4.6, both of the feature categories are necessary to create these features. Before we will describe the results, we will explain what data is being used for training and testing.

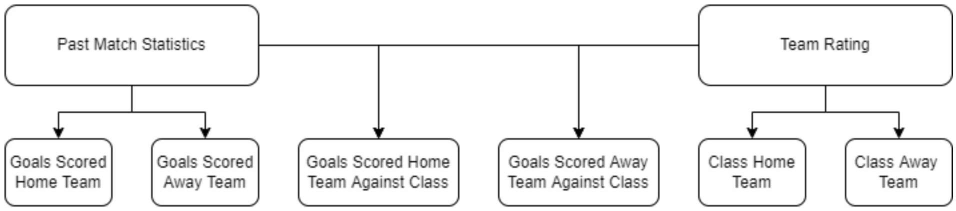

Figure 4.6: Goal Scored and Team Rating Class

Combination Excluded

When taking into account standard feature creation, the relevant data will consist of four features next to the target feature, ”FTR”. Namely, ”TRH”, ”TRA”, ”GSH”, and ”GSA”, as we can see in Table 4.12. ”FTR” represents the outcome of the football match. It can only be a 0, 1 or 2 which represent a home win, a draw and an away win, respectively. The features ”TRH” and ”TRA” represent the class of the home and away team based on their team rating. The features ”GSH” and ”GSA” represent the goals scored this season by the home and away team. As mentioned before, ”TRH” and ”TRA” can be integers from 1 to 5 where 1 represents the worst class and 5 represents the best class. ”GSH” and ”GSA” are integers as well but can range from 0 to infinity. In the worst case where a team would not score at all, it would still be zero till the end of the season. In the best case where a team scores a lot, it would be very high at the end of the season.

FTR	**TRH**	**TRA**	**GSH**	**GSA**
(0,1,2)	*(1-5)*	*(1-5)*	*(0-∞)*	*(0-∞)*
0	4	2	14	9
2	1	2	7	10
2	2	2	5	14
0	5	5	9	29
1	3	2	16	12

Table 4.12: Goal Scored and Team Rating Class: Combination Excluded

Combination Included

When taking into account combination feature creation, the relevant data will contain another two features, namely ”GSTRH” and ”GSTRA”. An example of the relevant data containing all features can be found in Table 4.13. ”GSTRH” and ”GSTRA” represent the amount of goals scored by the home and away team against the opponent’s team rating class. These features are integers. Their minimum is zero when a team does not score against a specific class in the season. And there is no maximum due to the team being allowed to score as much as they can against a specific class in the season.

FTR	TRH	TRA	GSH	GSA	GSTRH	GSTRA
(0,1,2)	*(1-5)*	*(1-5)*	*(0-∞)*	*(0-∞)*	*(0-∞)*	*(0-∞)*
0	4	2	14	9	7	0
2	1	2	7	10	2	5
2	2	2	5	14	2	7
0	5	5	9	29	0	0
1	3	2	16	12	8	0

Table 4.13: Goal Scored and Team Rating Class: Combination Included

Results

In Table 4.14, the accuracies of the ten different models can be found. The two different feature selections are used across five different algorithms. One feature selection that excludes the feature category combination features and one feature selection that includes the feature category combination features. As we can see in the table below, most algorithms benefit from the inclusion of the feature category combination features. Especially, the random forest algorithm whose performance is increased by 3.3%. Unfortunately, XGBoost and logistic regression perform worse due to the inclusion. On average, performance increases by 0.8% when including the feature category combination features. This indicates that the inclusion can increase the accuracy of a model that predicts the outcome of a football match. But this indication is not very convincing due to also decreasing the performance for two of the five algorithms. What it does not say is whether this increase is worthwhile. To find that out, we need the percentage of the class that is present the most. In other words, how often you will be correct when guessing that the home team will win. This percentage is 45.9%. The feature selection that excludes the feature category combination features has a prediction accuracy of 50.7% on average. This means that this feature selection increases the prediction accuracy by 4.8%. The feature selection that includes the feature category combination features increases the prediction accuracy by 51.5%. This means that the delta between the inclusion and exclusion of the feature category combination features is 0.8%. When dividing this delta with the delta between the feature selection that excludes the feature category combination features and the home win percentage, we can find the actual performance increase or in other words the effect of the inclusion of the feature category combination features. For the selected feature category combination features the actual performance increase is 16.7%. When using the same data in a more elaborate way, an increase of 16.7% can be seen as worthwhile.

Algorithm	Combination Excluded	Combination Included	Delta
Random Forest	0.450	0.483	0.033
XGBoost	0.496	0.493	-0.003
Logistic Regression	0.535	0.533	-0.002
Support Vector Machine	0.531	0.534	0.003
Ensemble	0.524	0.531	0.007
Average	0.507	0.515	**0.008**

Table 4.14: Goals Scored and Team Rating Class: Results

4.6.2 Goal Conceded and Team Rating Class

The next feature category combination feature set we are going to examine is the combination between goals conceded this season and team rating class. When using standard feature creation we would look at four features regarding the amount of goals conceded by the home and away team during this season and their team rating classes. But when taking into account combination feature creation, we would also look at features regarding the amount of goals conceded by the home and away team against the team rating class of the opponent during this season. As we can see in Figure 4.7, both of the feature categories are necessary to create these features. Before we will describe the results, we will explain what data is being used for training and testing.

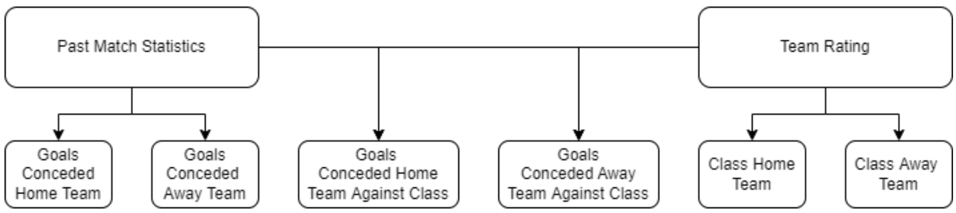

Figure 4.7: Goal Conceded and Team Rating Class

Combination Excluded

When taking into account standard feature creation, the relevant data will consist of four features next to the target feature, ”FTR”. Namely, ”TRH”, ”TRA”, ”GCH”, and ”GCA”, as we can see in Table 4.15. ”FTR” represents the outcome of the football match. It can only be a 0, 1 or 2 which represent a home win, a draw and an away win, respectively. The features ”TRH” and ”TRA” represent the class of the home and away team based on their team rating. The features ”GCH” and ”GCA” represent the goals conceded this season by the home and away team. As mentioned before, ”TRH” and ”TRA” can be integers from 1 to 5 where 1 represents the worst class and 5 represents the best class. ”GCH” and ”GCA” are integers as well but can range from 0 to infinity. In the worst case where a team would concede many goals, it would be very high at the end of the season. In the best case where a team concedes no goals, it would be zero at the end of the season.

FTR	**TRH**	**TRA**	**GCH**	**GCA**
(0,1,2)	*(1-5)*	*(1-5)*	*(0-∞)*	*(0-∞)*
1	2	1	11	17
0	2	2	13	12
0	5	2	3	10
1	1	5	15	12
2	4	2	10	10

Table 4.15: Goal Conceded and Team Rating Class: Combination Excluded

Combination Included

When taking into account combination feature creation, the relevant data will contain another two features, namely ”GCTRH” and ”GCTRA”. An example of the relevant data containing all features can be found in Table 4.16. ”GCTRH” and ”GCTRA” represent the amount of goals conceded by the home and away team against the opponent’s team rating class. These features are integers. Their minimum is zero when a team does not concede any goals against a specific class in the season. And there is no maximum due to the opponent being allowed to score as much as they can against a specific class in the season.

FTR	**TRH**	**TRA**	**GCH**	**GCA**	**GCTRH**	**GCTRA**
(0,1,2)	*(1-5)*	*(1-5)*	*(0-∞)*	*(0-∞)*	*(0-∞)*	*(0-∞)*
1	2	1	11	17	1	5
0	2	2	13	12	6	8
0	5	2	3	10	2	4
1	1	5	15	12	2	1
2	4	2	10	10	7	0

Table 4.16: Goal Conceded and Team Rating Class: Combination Included

Results

In Table 4.17, the accuracies of the ten different models can be found. The two different feature selections are used across five different algorithms. One feature

selection that excludes the feature category combination features and one feature selection that includes the feature category combination features. As we can see in the table below, only random forest benefits from the inclusion of the feature category combination features, as its performance is increased by 3.7%. The performance of logistic regression and support vector machine do not change when including the feature category combination features. And unfortunately, XGBoost and the ensemble perform worse due to the inclusion. On average, performance increases by 0.6% when including the feature category combination features. This indicates that the inclusion can increase the accuracy of a model that predicts the outcome of a football match. But this indication is not very convincing due to also decreasing the performance for two of the five algorithms.

What it does not say is whether this increase is worthwhile. To find that out, we need the percentage of the class that is present the most. In other words, how often you will be correct when guessing that the home team will win. This percentage is 45.9%. The feature selection that excludes the feature category combination features has a prediction accuracy of 50.6% on average. This means that this feature selection increases the prediction accuracy by 4.7%. The feature selection that includes the feature category combination features increases the prediction accuracy by 51.2%. This means that the delta between the inclusion and exclusion of the feature category combination features is 0.6%. When dividing this delta with the delta between the feature selection that excludes the feature category combination features and the home win percentage, we can find the actual performance increase or in other words the effect of the inclusion of the feature category combination features. For the selected feature category combination features the actual performance increase is 12.8%. When using the same data in a more elaborate way, an increase of 12.8% can be seen as worthwhile.

Algorithm	Combination Excluded	Combination Included	Delta
Random Forest	0.448	0.485	0.037
XGBoost	0.490	0.487	-0.003
Logistic Regression	0.533	0.533	0.000
Support Vector Machine	0.530	0.530	0.000
Ensemble	0.529	0.526	-0.003
Average	0.506	0.512	**0.006**

Table 4.17: Goals Conceded and Team Rating Class: Results

4.6.3 Goals Scored and Conceded and Team Rating Class

The next feature category combination feature set we are going to examine is the combination between goals scored and goals conceded this season and team rating class. When using standard feature creation we would look at six features regarding the amount of goals scored and goals conceded by the home and away team during this season and their team rating classes. But when taking into account combination feature creation, we would also look at features regarding the amount of goals scored and goals conceded by the home and away team against the team rating class of the opponent during this season. As we can see in Figure 4.8, both of the feature categories are necessary to create these features. Before we will describe the results, we will explain what data is being used for training and testing.

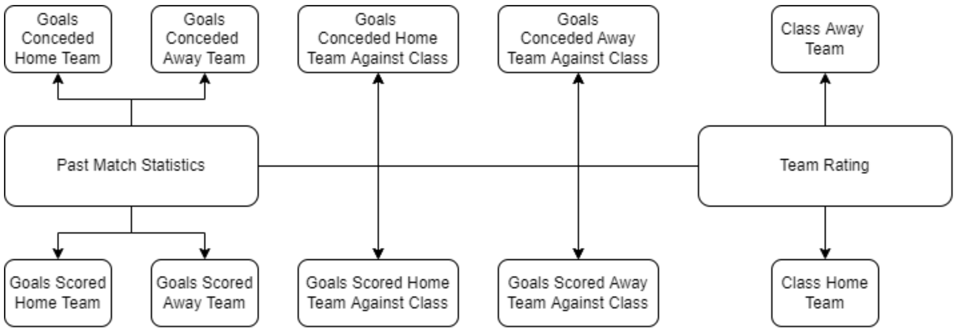

Figure 4.8: Goals Scored and Conceded and Team Rating Class

Combination Excluded

When taking into account standard feature creation, the relevant data will consist of six features next to the target feature, ”FTR”. Namely, ”TRH”, ”TRA”, ”GSH”, GSA””, ”GCH”, and ”GCA”, as we can see in Table 4.18. ”FTR” represents the outcome of the football match. It can only be a 0, 1 or 2 which represent a home win, a draw and an away win, respectively. The features ”TRH” and ”TRA” represent the class of the home and away team based on their team rating. The features ”GSH” and ”GSA” represent the goals scored this season by the home and away team. The features ”GCH” and ”GCA” represent the goals conceded this season by the home and away team. As mentioned before, ”TRH” and ”TRA” can be integers from 1 to 5 where 1 represents the worst class and 5 represents the best class. ”GSH” and ”GSA” are integers as well but can range from 0 to infinity. In the worst case where a team would not score at all, it would still be zero till the end of the season. In the best case where a team scores a lot, it would be very high at the end of the season. ”GCH” and ”GCA” are integers as well and can also range from 0 to infinity. In the worst case where a team would concede many goals, it would be very high at the end of the season. In the best case where a team concedes no goals, it would be zero at the end of the season.

FTR	TRH	TRA	GSH	GSA	GCH	GCA
(0,1,2)	*(1-5)*	*(1-5)*	*(0-∞)*	*(0-∞)*	*(0-∞)*	*(0-∞)*
0	4	2	14	9	8	15
2	1	2	7	10	14	7
2	2	2	5	14	4	11
0	5	5	9	29	7	8
1	3	2	16	12	12	11

Table 4.18: Goals Scored and Conceded and Team Rating Class: Combination Excluded

Combination Included

When taking into account combination feature creation, the relevant data will contain another four features, namely ”GSTRH”, ”GSTRA”, ”GCTRH”, and ”GCTRA”. An example of the relevant data containing all features can be found in Table 4.19. ”GSTRH” and ”GSTRA” represent the amount of goals scored by the home and away team against the opponent’s team rating class. These features are integers. Their minimum is zero when a team does not score against a specific class in the season. And there is no maximum due to the team being allowed to score as much as they can against a specific class in the season. ”GCTRH” and ”GCTRA” represent the amount of goals conceded by the home and away team against the opponent’s team rating class. These features are integers. Their minimum is zero when a team does not concede any goals against a specific class in the season. And there is no maximum due to the opponent being allowed to score as much as they can against a specific class in the season.

FTR	TRH	TRA	GSH	GSA	GSTRH	GSTRA	GCH	GCA	GCTRH	GCTRA
(0,1,2)	*(1-5)*	*(1-5)*	*(0-∞)*	*(0-∞)*	*(0-∞)*	*(0-∞)*	*(0-∞)*	*(0-∞)*	*(0-∞)*	*(0-∞)*
0	4	2	14	9	7	0	8	15	5	0
2	1	2	7	10	2	5	14	7	2	2
2	2	2	5	14	2	7	4	11	1	5
0	5	5	9	29	0	0	7	8	1	0
1	3	2	16	12	8	0	12	11	2	0

Table 4.19: Goals Scored and Conceded and Team Rating Class: Combination Included

Results

In Table 4.20, the accuracies of the ten different models can be found. The two different feature selections are used across five different algorithms. One feature selection that excludes the feature category combination features and one feature selection that includes the feature category combination features. As we can see in the table below, most algorithms benefit from the inclusion of the feature category combination features. Especially, the random forest algorithm and XGBoost whose performance are increased by 1.8% and 1.1%, respectively. Unfortunately, support vector machine and logistic regression perform worse due to the inclusion. On average, performance increases by 0.6% when including the feature category combination features. This indicates that the inclusion can increase the accuracy of a model that predicts the outcome of a football match. But this indication is not very convincing due to also decreasing the performance for two of the five algorithms.

What it does not say is whether this increase is worthwhile. To find that out, we need the percentage of the class that is present the most. In other words, how often you will be correct when guessing that the home team will win. This percentage is 45.9%. The feature selection that excludes the feature category combination features has a prediction accuracy of 51.2% on average. This means that this feature selection increases the prediction accuracy by 5.3%. The feature selection that includes the feature category combination features increases the prediction accuracy by 51.8%. This means that the delta between the inclusion and exclusion of the feature category combination features is 0.6%. When dividing this delta with the delta between the feature selection that excludes the feature category combination features and the home win percentage, we can find the actual performance increase or in other words the effect of the inclusion of the feature category combination features. For the selected feature category combination features the actual performance increase is 11.3%. When using the same data in a more elaborate way, an increase of 11.3% can be seen as worthwhile.

Algorithm	**Combination Excluded**	**Combination Included**	**Delta**
Random Forest	0.489	0.507	0.018
XGBoost	0.481	0.492	0.011
Logistic Regression	0.536	0.534	-0.002
Support Vector Machine	0.529	0.528	-0.001
Ensemble	0.527	0.529	0.002
Average	0.512	0.518	**0.006**

Table 4.20: Goals Scored and Conceded and Team Rating Class: Results

4.6.4 Summary

In this subsection, we will look at the results of the whole feature category combination. We made use of three different feature category combination feature sets to evaluate whether the use of this feature category combination can improve the prediction accuracy. In Table 4.21, we can find the delta between the prediction accuracy when using the feature selection that excludes the feature category combination features and the home win percentage. The delta will be positive when the prediction accuracy while using the feature selection that excludes the feature category combination features is higher than the home win percentage. Also, we can find the delta between the prediction accuracy when using the feature selection that includes the feature category combination features and the home win percentage. The delta will be positive when the prediction accuracy while using the feature selection that includes the feature category combination features is higher than the home win percentage. When taking these two values into account, one can find out the delta between the prediction accuracies when including and excluding the feature category combination features. The delta will be positive when the prediction accuracy while using the feature selection that includes the feature category combination features is higher than the prediction accuracy while using the feature selection that excludes the feature category combination features. When dividing this delta with the delta between the prediction accuracy when using the feature selection that excludes the feature category combination features and the home win percentage, we can find the actual performance or in other words the effect of the inclusion of the feature category combination features compared to the effect of the feature selection that excludes them. This value is a percentage which can be positive and negative. As we can see in the table below, for each of the feature category combination sets the delta between the home win percentage and the prediction accuracy when using the feature selection that includes the feature category combination features is larger than the delta between the home win percentage and the prediction accuracy when using the feature selection that excludes the feature category combination features. We can also observe that all three feature category combination feature sets can be seen as worthwhile. The combination between goals scored and team rating class performed the best achieving an actual performance increase of 16.7%. On average, the feature category combination achieved an actual performance increase of 13.6% which means that it can be seen as worthwhile.

Feature Category Combination Features	Delta Combination Excluded Home Win	Delta Combination Included Home Win	Delta Combination Included and Combination Excluded	Increase Actual Performance
Goal Scored and Team Rating Class	4.8%	5.6%	0.8%	16.7%
Goal Conceded and Team Rating Class	4.7%	5.3%	0.6%	12.8%
Goals Scored and Conceded and Team Rating Class	5.3%	5.9%	0.6%	11.3%
Average	4.9%	5.6%	0.7%	13.6%

Table 4.21: Past Match Statistics and Team Rating: Results

4.7 Team Performance and Team Value

This category combines the feature categories team performance and team value. Team performance is often represented by several features. Features regarding points achieved in the current season, points achieved in the previous season, and form of the last few matches. Next to these, there are many others that could be used. Team value is mostly represented by the total amount of transfer value a team has. Every player in a team has a specific worth and when calculating the total sum of all players in a team, one retrieves the team value. For this feature category, we calculated the maximum and minimum team value in a season of all teams. Based on the maximumum and minimum we assigned a class to each team every season. Classes can be 1 to 5 where 1 is the worst class and 5 is the best class. The main idea of this feature category combination is to look at a team's performance with respect to the team value class of the opponent. Features that reflect a team's performance against a specific group of opponents can potentially be used to improve the prediction model.

4.7.1 Points Current Season and Team Value Class

The next feature category combination feature set we are going to examine is the combination between points achieved in the current season and team value class. When using standard feature creation we would look at four features regarding the total amount of points collected by the home and away team and their team value classes. But when taking into account combination feature creation, we would also look at features regarding the amount of points collected by the home and away team against the team value class of the opponent. These features are not created out of a single feature category, but out of both, as we can see in Figure 4.9. Before we will describe the results, we will explain what data is being used for training and testing.

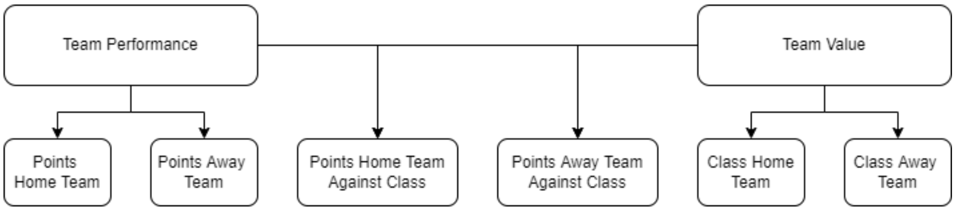

Figure 4.9: Points Current Season and Team Value Class

Combination Excluded

When taking into account standard feature creation, the relevant data will consist of four features next to the target feature, ”FTR”. Namely, ”TVH”, ”TVA”, ”PH”, and ”PA”, as we can see in Table 4.22. ”FTR” represents the outcome of the football match. It can only be a 0, 1 or 2 which represent a home win, a draw and an away win, respectively. The features ”TVH” and ”TVA” represent the class of the home and away team based on their team value. The features ”PH” and ”PA” represent the points collected by the home and away team during this season. As mentioned before, ”TVH” and ”TVA” are integers which range from 1 to 5 where 1 represents the worst class and 5 represents the best class, ”PH” and ”PA” are integers, as well, but can range from 0 to 114. In the worst case where a team would lose every match, it would still be zero till the end of the season. In the best case where a team would win every match, it would be 114 at the end of the season. As mentioned before, a team collects three points when winning a match, one point when the match ends in a draw, and zero points when losing the match.

FTR	**TVH**	**TVA**	**PH**	**PA**
(0,1,2)	*(1-5)*	*(1-5)*	*(0-114)*	*(0-114)*
0	4	1	29	14
1	1	4	12	40
2	1	3	17	34
1	1	1	19	20
2	1	5	27	49

Table 4.22: Points Current Season and Team Value Class: Combination Excluded

Combination Included

When taking into account combination feature creation, the relevant data will contain two other features, namely ”PTVH” and ”PTVA”. An example of the data for all features can be found in Table 4.23. ”PTVH” and ”PTVA” represent the amount of points the home and away team collected against the opponent’s team value class. These features are integers and can range from 0 to 114. The maximum of these integers will almost always be lower, due to it depending on the size of the opponent's class and it being very unlikely that all other teams would be part of the same class. Their minimum is zero which is only the case when every game against a certain class would be lost up to the relevant match.

FTR	**TVH**	**TVA**	**PH**	**PA**	**PTVH**	**PTVA**
(0,1,2)	*(1-5)*	*(1-5)*	*(0-114)*	*(0-114)*	*(0-114)*	*(0-114)*
0	4	1	29	14	25	0
1	1	4	12	40	0	32
2	1	3	17	34	0	29
1	1	1	19	20	15	18
2	1	5	27	49	0	34

Table 4.23: Points Current Season and Team Value Class: Combination Included

Results

In Table 4.24, the accuracies of the ten different models can be found. The two different feature selections are used across five different algorithms. One feature selection that excludes the feature category combination features and one feature selection that includes the feature category combination features. As we can see in the table below, every algorithm except for logistic regression benefits from the inclusion of the feature category combination features. Especially, the random forest algorithm whose performance is increased by 2.6%. On average, performance increases by 0.5% when including the feature category combination features. This indicates that the inclusion can increase the accuracy of a model that predicts the outcome of a football match.

What it does not say is whether this increase is worthwhile. To find that out, we need the percentage of the class that is present the most. In other words, how often you will be correct when guessing that the home team will win. This percentage is 45.9%. The feature selection that excludes the feature category combination features has a prediction accuracy of 50.9% on average. This means that this feature selection increases the prediction accuracy by 5.0%. The feature selection that includes the feature category combination features increases the prediction accuracy by 51.4%. This means that the delta between the inclusion and exclusion of the feature category combination features is 0.5%. When dividing this delta with the delta between the feature selection that excludes the feature category combination features and the home win percentage, we can find the actual performance increase or in other words the effect of the inclusion of the feature category combination features. For the selected feature category combination features the actual performance increase is 10.0%. When using the same data in a more elaborate way, an increase of 10.0% can be seen as worthwhile.

Algorithm	Combination Excluded	Combination Included	Delta
Random Forest	0.459	0.485	0.026
XGBoost	0.489	0.490	0.001
Logistic Regression	0.539	0.536	-0.003
Support Vector Machine	0.529	0.531	0.002
Ensemble	0.527	0.528	0.001
Average	0.509	0.514	**0.005**

Table 4.24: Points Current Season and Team Value Class: Results

4.7.2 Points Last Season and Team Value Class

The next feature category combination feature set we are going to examine is the combination between points collected during last season and team value class. When using standard feature creation we would look at four features regarding the total amount of points collected by the home and away team during last season and their team value classes. But when taking into account combination feature creation, we would also look at features regarding the amount of points collected by the home and away team against the team value class of the opponent during last season. As we can see in Figure 4.10, both of the feature categories are necessary to create these features. Before we will describe the results, we will explain what data is being used for training and testing.

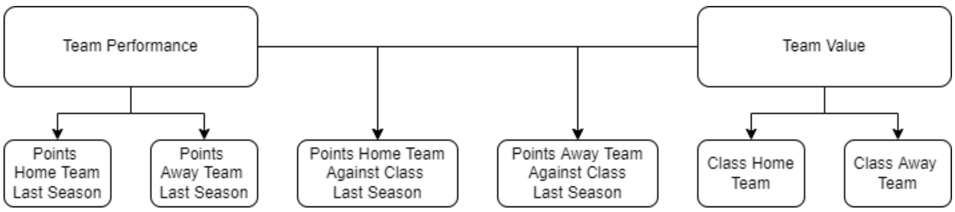

Figure 4.10: Points Last Season and Team Value Class

Combination Excluded

When taking into account standard feature creation, the relevant data will consist of four features next to the target feature, ”FTR”. Namely, ”TVH”, ”TVA”, ”LPH”, and ”LPA”, as we can see in Table 4.25. ”FTR” represents the outcome of the football match. It can only be a 0, 1 or 2 which represent a home win, a draw and an away win, respectively. The features ”TVH” and ”TVA” represent the class of the home and away team based on their team value. ”LPH” and ”LPA” represent the points collected by the home and away team during last season. ”TVH” and ”TVA” are integers and range from 1 to 5 where 1 represents the worst class and 5 represents the best class. ”LPH” and ”LPA” are also integers and can range from 0 to 114. In the worst case, they would be zero when every match has been lost last season. In the best case, they would be 114 when every match has been won last season.

FTR	**TVH**	**TVA**	**LPH**	**LPA**
(0,1,2)	*(1-5)*	*(1-5)*	*(0-114)*	*(0-114)*
0	4	1	83	39
1	1	4	45	77
2	1	3	61	58
1	1	1	44	47
2	1	5	52	95

Table 4.25: Points Last Season and Team Value Class: Combination Excluded

Combination Included

When taking into account combination feature creation, the relevant data will contain another two features, namely ”LPTVH” and ”LPTVA”. An example of the relevant data containing all features can be found in Table 4.26. ”LPTVH” and ”LPTVA” represent the amount of points the home and away team collected against the opponent’s team value class during last season. They can range from 0 to 114. The maximum of these integers will almost always be lower, due to it depending on the size of the opponent’s class and it being very unlikely that all other teams would be part of the same class. Their minimum is zero which is only the case when every game against a certain class would be lost during last season.

FTR	**TVH**	**TVA**	**LPH**	**LPA**	**LPTVH**	**LPTVA**
(0,1,2)	*(1-5)*	*(1-5)*	*(0-114)*	*(0-114)*	*(0-114)*	*(0-114)*
0	4	1	83	39	59	1
1	1	4	45	77	9	47
2	1	3	61	58	0	47
1	1	1	44	47	32	36
2	1	5	52	95	5	68

Table 4.26: Points Last Season and Team Value Class: Combination Included

Results

In Table 4.27, the accuracies of the ten different models can be found. The two different feature selections are used across five different algorithms. One feature selection that excludes the feature category combination features and one feature selection that includes the feature category combination features. As we can see in the table below, every algorithm benefits from the inclusion of the feature category combination features. Especially, the random forest algorithm and XGBoost whose performance are increased by 2.9% and 1.1%, respectively. On average, performance increases by 1.2% when including the feature category combination features. This indicates that the inclusion can increase the accuracy of a model that predicts the outcome of a football match.

What it does not say is whether this increase is worthwhile. To find that out, we need the percentage of the class that is present the most. In other words, how often you will be correct when guessing that the home team will win. This percentage is 45.9%. The feature selection that excludes the feature category combination features has a prediction accuracy of 49.5% on average. This means that this feature selection increases the prediction accuracy by 3.6%. The feature selection that includes the feature category combination features increases the prediction accuracy by 50.7%. This means that the delta between the inclusion and exclusion of the feature category combination features is 1.2%. When dividing this delta with the delta between the feature selection that excludes the feature category combination features and the home win percentage, we can find the actual performance increase or in other words the effect of the inclusion of the feature category combination features. For the selected feature category combination features the actual performance increase is 33.3%. When using the same data in a more elaborate way, an increase of 33.3% can be seen as worthwhile.

Algorithm	**Combination Excluded**	**Combination Included**	**Delta**
Random Forest	0.439	0.468	0.029
XGBoost	0.467	0.478	0.011
Logistic Regression	0.530	0.534	0.004
Support Vector Machine	0.524	0.531	0.007
Ensemble	0.513	0.522	0.009
Average	0.495	0.507	**0.012**

Table 4.27: Points Last Season and Team Value Class: Results

4.7.3 Points Current and Last Season and Team Value Class

The next feature category combination feature set we are going to examine is the combination between points collected in the current and last season and team value class. When using standard feature creation we would look at six features regarding the points collected by the home and away team in the current and last season and their team value classes. But when taking into account combination feature creation, we would also look at features regarding the amount of points collected by the home and away team against the team value class of the opponent during the current and last season. As we can see in Figure 4.11, both of the feature categories are necessary to create these features. Before we will describe the results, we will explain what data is being used for training and testing.

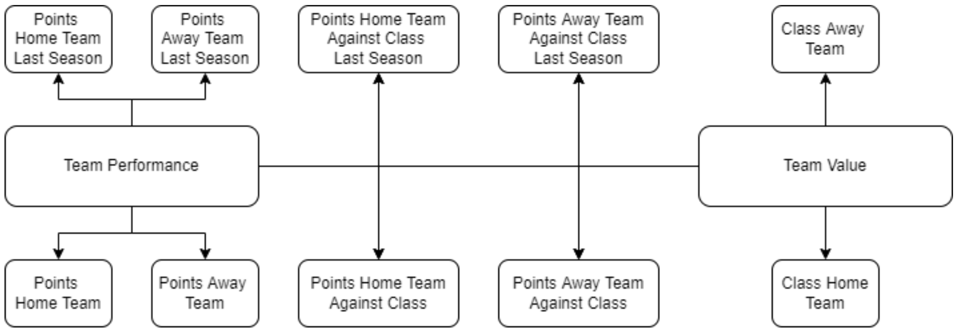

Figure 4.11: Points Current and Last Season and Team Value Class

Combination Excluded

When taking into account standard feature creation, the relevant data will consist of six features next to the target feature, ”FTR”. Namely, ”TVH”, ”TVA”, ”PH”, ”PA”, ”LPH”, and ”LPA”, as we can see in Table 4.28. ”FTR” represents the outcome of the football match. It can only be a 0, 1 or 2 which represent a home win, a draw and an away win, respectively. The features ”TVH” and ”TVA” represent the class of the home and away team based on their team value. The features ”PH” and ”PA” represent the points collected by the home and away team during this season.. ”LPH” and ”LPA” represent the points collected by the home and away team during last season. As mentioned before, ”TVH” and ”TVA” are integers which range from 1 to 5 where 1 represents the worst class and 5 represents the best class, ”PH” and ”PA” are integers, as well, but can range from 0 to 114. In the worst case where a team would lose every match, it would still be zero till the end of the season. In the best case where a team would win every match, it would be 114 at the end of the season. ”LPH” and ”LPA” are also integers and can range from 0 to 114. In the worst case, they would be zero when every match has been lost last season. In the best case, they would be 114 when every match has been won last season.

FTR	TVH	TVA	PH	PA	LPH	LPA
(0,1,2)	*(1-5)*	*(1-5)*	*(0-114)*	*(0-114)*	*(0-114)*	*(0-114)*
0	4	1	29	14	83	39
1	1	4	12	40	45	77
2	1	3	17	34	61	58
1	1	1	19	20	44	47
2	1	5	27	49	52	95

Table 4.28: Points Current and Last Season and Team Value Class: Combination Excluded

Combination Included

When taking into account combination feature creation, the relevant data will contain another four features, namely "PTVH", "PTVA", "LPTVH" and "LPTVA". An example of the relevant data containing all features can be found in Table 4.29. "PTVH" and "PTVA" represent the amount of points the home and away team collected against the opponent's team value class. These features are integers and can range from 0 to 114. The maximum of these integers will almost always be lower, due to it depending on the size of the opponent's class and it being very unlikely that all other teams would be part of the same class. Their minimum is zero which is only the case when every game against a certain class would be lost up to the relevant match. "LPTVH" and "LPTVA" represent the amount of points the home and away team collected against the opponent's team value class during last season. They can also range from 0 to 114. The maximum of these integers will almost always be lower, due to it depending on the size of the opponent's class and it being very unlikely that all other teams would be part of the same class. Their minimum is zero which is only the case when every game against a certain class would be lost during last season.

FTR	TVH	TVA	PH	PA	PTVH	PTVA	LPH	LPA	LPTVH	LPTVA
(0,1,2)	*(1-5)*	*(1-5)*	*(0-114)*	*(0-114)*	*(0-114)*	*(0-114)*	*(0-114)*	*(0-114)*	*(0-114)*	*(0-114)*
0	4	1	29	14	25	0	83	39	59	1
1	1	4	12	40	0	32	45	77	9	47
2	1	3	17	34	0	29	61	58	0	47
1	1	1	19	20	15	18	44	47	32	36
2	1	5	27	49	0	34	52	95	5	68

Table 4.29: Points Current and Last Season and Team Value Class: Combination Included

Results

In Table 4.30, the accuracies of the ten different models can be found. The two different feature selections are used across five different algorithms. One feature selection that excludes the feature category combination features and one feature selection that includes the feature category combination features. As we can see in the table below, every algorithm except for the ensemble benefits from the inclusion of the feature category combination features. On average, performance increases by 0.5% when including the feature category combination features. This indicates that the inclusion can increase the accuracy of a model that predicts the outcome of a football match.

What it does not say is whether this increase is worthwhile. To find that out, we need the percentage of the class that is present the most. In other words, how often you will be correct when guessing that the home team will win. This percentage is 45.9%. The feature selection that excludes the feature category combination features has a prediction accuracy of 50.6% on average. This means that this feature selection increases the prediction accuracy by 4.7%. The feature selection that includes the feature category combination features increases the prediction accuracy by 51.1%. This means that the delta between the inclusion and exclusion of the feature category combination features is 0.5%. When dividing this delta with the delta between the feature selection that excludes the feature category combination features and the home win percentage, we can find the actual performance increase or in other words the effect of the inclusion of the feature category combination features. For the selected feature category combination features the actual performance increase is 10.6%. When using the same data in a more elaborate way, an increase of 10.6% can be seen as worthwhile.

Algorithm	**Combination Excluded**	**Combination Included**	**Delta**
Random Forest	0.479	0.487	0.008
XGBoost	0.469	0.478	0.009
Logistic Regression	0.532	0.535	0.003
Support Vector Machine	0.525	0.533	0.008
Ensemble	0.523	0.521	-0.002
Average	0.506	0.511	**0.005**

Table 4.30: Points Current and Last Season and Team Value Class: Results

4.7.4 Summary

In this subsection, we will look at the results of the whole feature category combination. We made use of three different feature category combination feature sets to evaluate whether the use of this feature category combination can improve the prediction accuracy. In Table 4.31, we can find the delta between the prediction accuracy when using the feature selection that excludes the feature category combination features and the home win percentage. The delta will be positive when the prediction accuracy while using the feature selection that excludes the feature category combination features is higher than the home win percentage. Also, we can find the delta between the prediction accuracy when using the feature selection that includes the feature category combination features and the home win percentage. The delta will be positive when the prediction accuracy while using the feature selection that includes the feature category combination features is higher than the home win percentage. When taking these two values into account, one can find out the delta between the prediction accuracies when including and excluding the feature category combination features. The delta will be positive when the prediction accuracy while using the feature selection that includes the feature category combination features is higher than the prediction accuracy while using the feature selection that excludes the feature category combination features. When dividing this delta with the delta between the prediction accuracy when using the feature selection that excludes the feature category combination features and the home win percentage, we can find the actual performance or in other words the effect of the inclusion of the feature category combination features compared to the effect of the feature selection that excludes them. This value is a percentage which can be positive and negative. As we can see in the table below, for each of the feature category combination sets the delta between the home win percentage and the prediction accuracy when using the feature selection that includes the feature category combination features is larger than the delta between the home win percentage and the prediction accuracy when using the feature selection that excludes the feature category combination features. We can also observe that all three feature category combination feature sets can be seen as worthwhile. The combination between points last season and team value class performed the best achieving an actual performance increase of 33.3%. The combination between points current season and team value class can barely be seen as worthwhile due to achieving the minimum actual performance increase of 10.0%. On average, the feature category combination achieved an actual performance increase of 18.0% which means that it can be seen as worthwhile.

Feature Category Combination Features	Delta Combination Excluded Home Win	Delta Combination Included Home Win	Delta Combination Included and Combination Excluded	Increase Actual Performance
Points Current Season and Team Value Class	5.0%	5.5%	0.5%	10.0%
Points Last Season and Team Value Class	3.6%	4.8%	1.2%	33.3%
Points Current and Last Season and Team Value Class	4.7%	5.2%	0.5%	10.6%
Average	4.4%	5.2%	0.7%	18.0%

Table 4.31: Team Performance and Team Value: Results

4.8 Past Match Statistics and Team Value

This category combines the feature categories past match statistics and team value. Past Match Statistics can be represented by several features. Examples of features are goals scored, goals conceded, shots on target and ball possession. Obviously, there are many more that can represent this feature category. In the previous feature category combination, we already described team value in quite some detail. The way this feature category will be used will be unchanged. The main idea of this feature category combination is to look at a past match statistics with respect to the team value class of the opponent. Features that reflect past match statistics against a specific group of opponents can potentially be used to improve the prediction model.

4.8.1 Goals Scored and Team Value Class

The next feature category combination feature set we are going to examine is the combination between goals scored this season and team value class. When using standard feature creation we would look at four features regarding the amount of goals scored by the home and away team during this season and their team value classes. But when taking into account combination feature creation, we would also look at features regarding the amount of goals scored by the home and away team against the team value class of the opponent during this season. As we can see in Figure 4.12, both of the feature categories are necessary to create these features. Before we will describe the results, we will explain what data is being used for training and testing.

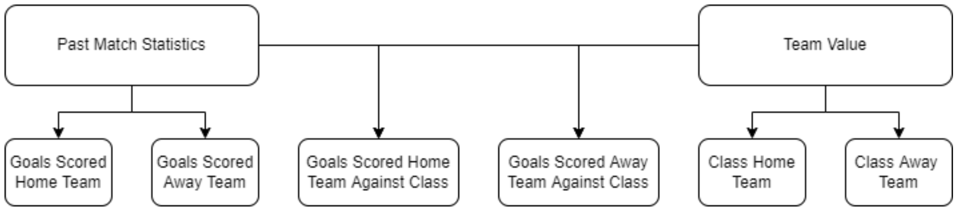

Figure 4.12: Goals Scored and Team Value Class

Combination Excluded

When taking into account standard feature creation, the relevant data will consist of four features next to the target feature, "FTR". Namely, "TVH", "TVA", "GSH", and "GSA", as we can see in Table 4.32. "FTR" represents the outcome of the football match. It can only be a 0, 1 or 2 which represent a home win, a draw and an away win, respectively. The features "TVH" and "TVA" represent the class of the home and away team based on their team value. The features "GSH" and "GSA" represent the goals scored this season by the home and away team. As mentioned before, "TVH" and "TVA" can be integers from 1 to 5 where 1 represents the worst class and 5 represents the best class. "GSH" and "GSA" are integers as well but can range from 0 to infinity. In the worst case where a team would not score at all, it would still be zero till the end of the season. In the best case where a team scores a lot, it would be very high at the end of the season.

FTR	**TVH**	**TVA**	**GSH**	**GSA**
(0,1,2)	*(1-5)*	*(1-5)*	*(0-∞)*	*(0-∞)*
0	4	1	14	9
2	1	1	7	10
2	2	1	5	14
0	5	4	9	29
1	1	1	16	12

Table 4.32: Goals Scored and Team Value Class: Combination Excluded

Combination Included

When taking into account combination feature creation, the relevant data will contain another two features, namely "GSTVH" and "GSTVA". An example of the relevant data containing all features can be found in Table 4.33. "GSTVH" and "GSTVA" represent the amount of goals scored by the home and away team against the opponent's team value class. These features are integers. Their minimum is zero when a team does not score against a specific class in the season. And there is no maximum due to the team being allowed to score as much as they can against a specific class in the season.

FTR	TVH	TVA	GSH	GSA	GSTVH	GSTVA
(0,1,2)	*(1-5)*	*(1-5)*	*(0-∞)*	*(0-∞)*	*(0-∞)*	*(0-∞)*
0	4	1	14	9	10	0
2	1	1	7	10	3	8
2	2	1	5	14	2	1
0	5	4	9	29	2	0
1	1	1	16	12	8	10

Table 4.33: Goals Scored and Team Value Class: Combination Included

Results

In Table 4.34, the accuracies of the ten different models can be found. The two different feature selections are used across five different algorithms. One feature selection that excludes the feature category combination features and one feature selection that includes the feature category combination features. As we can see in the table below, most algorithms benefit from the inclusion of the feature category combination features. Especially, the random forest algorithm whose performance is increased by 3.0%. The performance of logistic regression and support vector machine do not change when including the feature category combination features. On average, performance increases by 0.8% when including the feature category combination features. This indicates that the inclusion can increase the accuracy of a model that predicts the outcome of a football match.

What it does not say is whether this increase is worthwhile. To find that out, we need the percentage of the class that is present the most. In other words, how often you will be correct when guessing that the home team will win. This percentage is 45.9%. The feature selection that excludes the feature category combination features has a prediction accuracy of 50.5% on average. This means that this feature selection increases the prediction accuracy by 4.6%. The feature selection that includes the feature category combination features increases the prediction accuracy by 51.3%. This means that the delta between the inclusion and exclusion of the feature category combination features is 0.8%. When dividing this delta with the delta between the feature selection that excludes the feature category combination features and the home win percentage, we can find the actual performance increase or in other words the effect of the inclusion of the feature category combination features. For the selected feature category combination features the actual performance increase is 17.4%. When using the same data in a more elaborate way, an increase of 17.4% can be seen as worthwhile.

Algorithm	Combination Excluded	Combination Included	Delta
Random Forest	0.447	0.477	0.030
XGBoost	0.483	0.488	0.005
Logistic Regression	0.538	0.538	0.000
Support Vector Machine	0.531	0.531	0.000
Ensemble	0.526	0.531	0.005
Average	0.505	0.513	**0.008**

Table 4.34: Goals Scored and Team Value Class: Results

4.8.2 Goals Conceded and Team Value Class

The next feature category combination feature set we are going to examine is the combination between goals conceded this season and team value class. When using standard feature creation we would look at four features regarding the amount of goals conceded by the home and away team during this season and their team value classes. But when taking into account combination feature creation, we would also look at features regarding the amount of goals conceded by the home and away team against the team value class of the opponent during this season. As we can see in Figure 4.13, both of the feature categories are necessary to create these features. Before we will describe the results, we will explain what data is being used for training and testing.

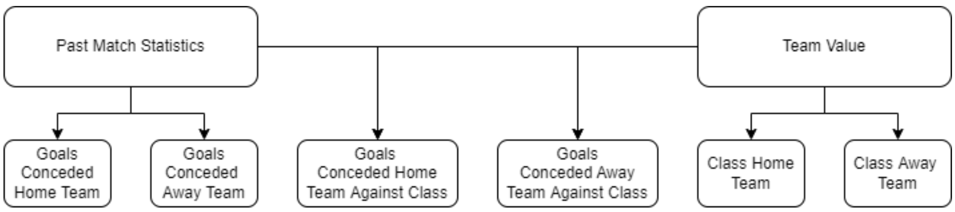

Figure 4.13: Goals Conceded and Team Value Class

Combination Excluded

When taking into account standard feature creation, the relevant data will consist of four features next to the target feature, ”FTR”. Namely, ”TVH”, ”TVA”, ”GCH”, and ”GCA”, as we can see in Table 4.35. ”FTR” represents the outcome of the football match. It can only be a 0, 1 or 2 which represent a home win, a draw and an away win, respectively. The features ”TVH” and ”TVA” represent the class of the home and away team based on their team value. The features ”GCH” and ”GCA” represent the goals conceded this season by the home and away team. As mentioned before,

”TVH” and ”TVA” can be integers from 1 to 5 where 1 represents the worst class and 5 represents the best class. ”GCH” and ”GCA” are integers as well but can range from 0 to infinity. In the worst case where a team would concede many goals, it would be very high at the end of the season. In the best case where a team concedes no goals, it would be zero at the end of the season.

FTR	TVH	TVA	GCH	GCA
(0,1,2)	*(1-5)*	*(1-5)*	*(0-∞)*	*(0-∞)*
0	4	1	8	15
2	1	1	14	7
2	2	1	4	11
0	5	4	7	8
1	1	1	12	11

Table 4.35: Goals Conceded and Team Value Class: Combination Excluded

Combination Included

When taking into account combination feature creation, the relevant data will contain another two features, namely ”GCTVH” and ”GCTVA”. An example of the relevant data containing all features can be found in Table 4.36. ”GCTVH” and ”GCTVA” represent the amount of goals conceded by the home and away team against the opponent’s team value class. These features are integers. Their minimum is zero when a team does not concede any goals against a specific class in the season. And there is no maximum due to the opponent being allowed to score as much as they can against a specific class in the season.

FTR	TVH	TVA	GCH	GCA	GCTVH	GCTVA
(0,1,2)	*(1-5)*	*(1-5)*	*(0-∞)*	*(0-∞)*	*(0-∞)*	*(0-∞)*
0	4	1	8	15	3	4
2	1	1	14	7	3	2
2	2	1	4	11	2	1
0	5	4	7	8	1	0
1	1	1	12	11	2	10

Table 4.36: Goals Conceded and Team Value Class: Combination Included

Results

In Table 4.37, the accuracies of the ten different models can be found. The two different feature selections are used across five different algorithms. One feature

selection that excludes the feature category combination features and one feature selection that includes the feature category combination features. As we can see in the table below, most algorithms benefit from the inclusion of the feature category combination features. Especially, the random forest algorithm whose performance is increased by 3.8%. The performance of XGBoost does not change when including the feature category combination features. On average, performance increases by 1.0% when including the feature category combination features. This indicates that the inclusion can increase the accuracy of a model that predicts the outcome of a football match.

What it does not say is whether this increase is worthwhile. To find that out, we need the percentage of the class that is present the most. In other words, how often you will be correct when guessing that the home team will win. This percentage is 45.9%. The feature selection that excludes the feature category combination features has a prediction accuracy of 50.3% on average. This means that this feature selection increases the prediction accuracy by 4.4%. The feature selection that includes the feature category combination features increases the prediction accuracy by 51.3%. This means that the delta between the inclusion and exclusion of the feature category combination features is 1.0%. When dividing this delta with the delta between the feature selection that excludes the feature category combination features and the home win percentage, we can find the actual performance increase or in other words the effect of the inclusion of the feature category combination features. For the selected feature category combination features the actual performance increase is 22.7%. When using the same data in a more elaborate way, an increase of 22.7% can be seen as worthwhile.

Algorithm	Combination Excluded	Combination Included	Delta
Random Forest	0.437	0.475	0.038
XGBoost	0.486	0.486	0.000
Logistic Regression	0.537	0.539	0.002
Support Vector Machine	0.527	0.532	0.005
Ensemble	0.526	0.531	0.005
Average	0.503	0.513	**0.010**

Table 4.37: Goals Conceded and Team Value Class: Results

4.8.3 Goals Scored and Conceded and Team Value Class

The next feature category combination feature set we are going to examine is the combination between goals scored and goals conceded this season and team value class. When using standard feature creation we would look at six features regarding the amount of goals scored and goals conceded by the home and away team during this season and their team value classes. But when taking into account combination feature creation, we would also look at features regarding the amount of goals scored and goals conceded by the home and away team against the team value class of the opponent during this season. As we can see in Figure 4.14, both of the feature categories are necessary to create these features. Before we will describe the results, we will explain what data is being used for training and testing.

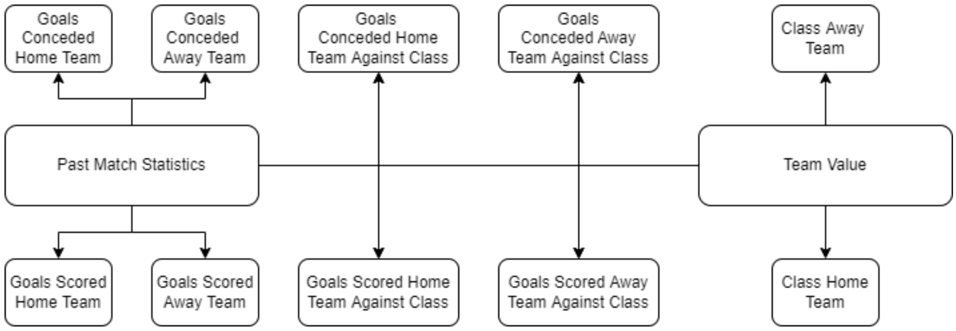

Figure 4.14: Goals Scored and Conceded and Team Value Class

Combination Excluded

When taking into account standard feature creation, the relevant data will consist of six features next to the target feature, ”FTR”. Namely, ”TVH”, ”TVA”, ”GSH”, GSA””, ”GCH”, and ”GCA”, as we can see in Table 4.38. ”FTR” represents the outcome of the football match. It can only be a 0, 1 or 2 which represent a home win, a draw and an away win, respectively. The features ”TVH” and ”TVA” represent the class of the home and away team based on their team value. The features ”GSH” and ”GSA” represent the goals scored this season by the home and away team. The features ”GCH” and ”GCA” represent the goals conceded this season by the home and away team. As mentioned before, ”TVH” and ”TVA” can be integers from 1 to 5 where 1 represents the worst class and 5 represents the best class. ”GSH” and ”GSA” are integers as well but can range from 0 to infinity. In the worst case where a team would not score at all, it would still be zero till the end of the season. In the best case where a team scores a lot, it would be very high at the end of the season. ”GCH” and ”GCA” are integers as well and can also range from 0 to infinity. In the worst case where a team would concede many goals, it would be very high at the end of the season. In the best case where a team concedes no goals, it would be zero at the end of the season.

FTR	TVH	TVA	GSH	GSA	GCH	GCA
(0,1,2)	*(1-5)*	*(1-5)*	*(0-∞)*	*(0-∞)*	*(0-∞)*	*(0-∞)*
0	4	1	14	9	8	15
2	1	1	7	10	14	7
2	2	1	5	14	4	11
0	5	4	9	29	7	8
1	1	1	16	12	12	11

Table 4.38: Goals Scored and Conceded and Team Value Class: Combination Excluded

Combination Included

When taking into account combination feature creation, the relevant data will contain another four features, namely ”GSTVH”, ”GSTVA”, ”GCTVH”, and ”GCTVA”. An example of the relevant data containing all features can be found in Table 4.39. ”GSTVH” and ”GSTVA” represent the amount of goals scored by the home and away team against the opponent’s team value class. These features are integers. Their minimum is zero when a team does not score against a specific class in the season. And there is no maximum due to the team being allowed to score as much as they can against a specific class in the season. ”GCTVH” and ”GCTVA” represent the amount of goals conceded by the home and away team against the opponent’s team value class. These features are integers. Their minimum is zero when a team does not concede any goals against a specific class in the season. And there is no maximum due to the opponent being allowed to score as much as they can against a specific class in the season.

FTR	TVH	TVA	GSH	GSA	GSTVH	GSTVA	GCH	GCA	GCTVH	GCTVA
(0,1,2)	*(1-5)*	*(1-5)*	*(0-∞)*	*(0-∞)*	*(0-∞)*	*(0-∞)*	*(0-∞)*	*(0-∞)*	*(0-∞)*	*(0-∞)*
0	4	1	14	9	10	0	8	15	3	4
2	1	1	7	10	3	8	14	7	3	2
2	2	1	5	14	2	1	4	11	2	1
0	5	4	9	29	2	0	7	8	1	0
1	1	1	16	12	8	10	12	11	2	10

Table 4.39: Goals Scored and Conceded and Team Value Class: Combination Included

Results

In Table 4.40, the accuracies of the ten different models can be found. The two different feature selections are used across five different algorithms. One feature

selection that excludes the feature category combination features and one feature selection that includes the feature category combination features. As we can see in the table below, every algorithm except for XGBoost benefits from the inclusion of the feature category combination features. On average, performance increases by 0.5% when including the feature category combination features. This indicates that the inclusion can increase the accuracy of a model that predicts the outcome of a football match.

What it does not say is whether this increase is worthwhile. To find that out, we need the percentage of the class that is present the most. In other words, how often you will be correct when guessing that the home team will win. This percentage is 45.9%. The feature selection that excludes the feature category combination features has a prediction accuracy of 51.3% on average. This means that this feature selection increases the prediction accuracy by 5.4%. The feature selection that includes the feature category combination features increases the prediction accuracy by 51.8%. The delta between the inclusion and exclusion of the feature category combination features is 0.5%. When dividing this delta with the delta between the feature selection that excludes the feature category combination features and the home win percentage, we can find the actual performance increase or in other words the effect of the inclusion of the feature category combination features. For the selected feature category combination features the actual performance increase is 9.3%. When using the same data in a more elaborate way, an increase of 9.3% can not be seen as worthwhile.

Algorithm	**Combination Excluded**	**Combination Included**	**Delta**
Random Forest	0.481	0.501	0.020
XGBoost	0.486	0.485	-0.001
Logistic Regression	0.537	0.538	0.001
Support Vector Machine	0.529	0.533	0.004
Ensemble	0.531	0.535	0.004
Average	0.513	0.518	**0.005**

Table 4.40: Goals Scored and Conceded and Team Value Class: Results

4.8.4 Summary

In this subsection, we will look at the results of the whole feature category combination. We made use of three different feature category combination feature sets to evaluate whether the use of this feature category combination can improve the prediction accuracy. In Table 4.41, we can find the delta between the prediction accuracy when using the feature selection that excludes the feature category combination features and the home win percentage. The delta will be positive when the prediction accuracy while using the feature selection that excludes the feature category combination features is higher than the home win percentage. Also, we can find the delta between the prediction accuracy when using the feature selection that includes the feature category combination features and the home win percentage. The delta will be positive when the prediction accuracy while using the feature selection that includes the feature category combination features is higher than the home win percentage. When taking these two values into account, one can find out the delta between the prediction accuracies when including and excluding the feature category combination features. The delta will be positive when the prediction accuracy while using the feature selection that includes the feature category combination features is higher than the prediction accuracy while using the feature selection that excludes the feature category combination features. When dividing this delta with the delta between the prediction accuracy when using the feature selection that excludes the feature category combination features and the home win percentage, we can find the actual performance or in other words the effect of the inclusion of the feature category combination features compared to the effect of the feature selection that excludes them. This value is a percentage which can be positive and negative. As we can see in the table below, for each of the feature category combination sets the delta between the home win percentage and the prediction accuracy when using the feature selection that includes the feature category combination features is larger than the delta between the home win percentage and the prediction accuracy when using the feature selection that excludes the feature category combination features. We can also observe that two of the three feature category combination feature sets can be seen as worthwhile. The combination between goals scored and conceded and team value class only had an actual performance increase of 9.3% which means it cannot be seen as worthwhile. The combination between goals conceded and team value class performed the best achieving an actual performance increase of 22.7%. On average, the feature category combination achieved an actual performance increase of 16.5% which means that it can be seen as worthwhile.

Feature Category Combination Features	Delta Combination Excluded Home Win	Delta Combination Included Home Win	Delta Combination Included and Combination Excluded	Increase Actual Performance
Goals Scored and Team Value Class	4.6%	5.4%	0.8%	17.4%
Goals Conceded and Team Value Class	4.4%	5.4%	1.0%	22.7%
Goals Scored and Conceded and Team Value Class	5.4%	5.8%	0.5%	9.3%
Average	4.8%	5.5%	0.8%	16.5%

Table 4.41: Past Match Statistics and Team Value: Results

4.9 Discussion

In this section, we will take a look at all feature category combinations that we have explored in this chapter. We looked at four different feature category combinations, namely team performance and team rating (TPTR), past match statistics and team rating (PMSTR), team performance and team value (TPTV), and past match statistics and team value (PMSTV). In Table 4.42, we can find the average delta between the prediction accuracy when using the feature selection that excludes the feature category combination features and the home win percentage. The delta will be positive when the prediction accuracy while using the feature selection that excludes the feature category combination features is higher than the home win percentage. Also, we can find the average delta between the prediction accuracy when using the feature selection that includes the feature category combination features and the home win percentage. The delta will be positive when the prediction accuracy while using the feature selection that includes the feature category combination features is higher than the home win percentage. When taking these two values into account, one can find out the average delta between the prediction accuracies when including and excluding the feature category combination features. The delta will be positive when the prediction accuracy while using the feature selection that includes the feature category combination features is higher than the prediction accuracy while using the feature selection that excludes the feature category combination features. When dividing this delta with the delta between the prediction accuracy when using the feature selection that excludes the feature category combination features and the home win percentage, we can find the average actual performance or in other words the effect of the inclusion of the feature category combination features compared to the effect of the feature selection that excludes them. This value is a percentage which can be positive and negative. As we can see in the table below, for each of the feature category combinations the delta between the home win percentage and the prediction accuracy when using the feature selection that includes the feature category combination features is larger than the delta between the home win percentage and the prediction accuracy when using the feature selection that excludes the feature category combination features. We can also observe that all feature category combinations can be seen as worthwhile. On average, the feature selections that include the feature category combination features perform 16.2% better than the feature selections that exclude the feature category combination features. The feature category combination team performance and team value is the best performing combination achieving an average actual performance increase of 18.0%.

Feature Category Combination	Delta Combination Excluded Home Win	Delta Combination Included Home Win	Delta Combination Included and Combination Excluded	Increase Actual Performance
TPTR	4.3%	5.0%	0.7%	16.5%
PMSTR	4.9%	5.6%	0.7%	13.6%
TPTV	4.4%	5.2%	0.7%	18.0%
PMSTV	4.8%	5.5%	0.8%	16.5%
Average	4.6%	5,3%	0.7%	16.2%

Table 4.42: Feature Category Combinations: Results

4.10 Key Takeaways

In the beginning of this chapter, we explained that there are many possible feature category combinations but that most feature categories do not fit well together due to not providing new information or information which directly relates to the outcome of a football match. This resulted in us choosing to explore four feature category combinations that we believed would have the best chance of improving the prediction models directly. Next to that, we reasoned which feature category combination features could be useful and came up with twelve sets of feature category combination features. As became clear in the last section, these four feature category combinations can be seen as worthwhile. On average, the feature selections that include the feature category combination features performed 16.2% better than the feature selections that exclude the feature category combination features. This leads to the conclusion that the use of these feature category combinations and the chosen feature category combination features can indeed be seen as worthwhile and used to increase the accuracy of a model that predicts the outcome of a football match. The feature category combination team performance and team value was the best performing combination achieving an average actual performance increase of 18.0%.

Chapter 5

Ensembles: Modelling & Evaluating

This chapter describes what ensembles can be created out of the promising algorithms that we found in our literature review and whether these can be used to increase the prediction accuracy of a model that predicts the outcome of a football match. During this chapter we will go through the modelling and evaluation phase of the second CRISP-DM cycle, as illustrated in Table 3.1. Next to that, we will make use of the last step of the machine learning pipeline which is displayed in Figure 3.2. In Section 5.1, we will explain the general idea behind ensembles consisting of the promising algorithms and what possible ensembles we are going to evaluate. In Section 5.2, we will describe how we are going to model and evaluate these ensembles while taking into account several feature selections and the best parameters for each algorithm to properly assess whether an ensemble can increase the prediction accuracy in this area. In Section 5.3, we will discuss the used feature selection methods and explain how these work. In Section 5.4, the hyperparameter optimisation will be described. In Section 5.5, the performance of the random forest algorithm and the possible ensembles that make use of this algorithm will be compared. In Section 5.6, the performance of the XGBoost algorithm and the possible ensembles that make use of this algorithm will be compared. In Section 5.7, the performance of logistic regression and the possible ensembles that make use of this algorithm will be compared. In Section 5.8, the performance of the support vector machine and the possible ensembles that make use of this algorithm will be compared. In Section 5.9, we will discuss our findings. In Section 5.10, we will describe the key takeaways of this chapter.

5.1 Ensembles consisting of Promising Algorithms

In the literature review, it became clear that ensembles that make use of different algorithms could potentially be used to increase the prediction accuracy of a model that predicts the outcome of a football match. Ensembles that make use of different algorithms can used in the same way as any individual algorithm but make use of multiple algorithms. Next to that, we found a group of algorithms which could be labeled as promising due to consistently achieving a higher accuracy than other algorithms. This suggests that ensembles consisting of promising algorithms have a great chance of increasing the prediction accuracy in this area. There are many ways to build an ensemble that consists of the promising algorithms. Namely, you could use all of them of just a subset. Also, you could make the algorithms in an ensemble equally weighted or make some more important. The total amount of combinations one could make with the promising algorithms is eleven. There are four promising algorithms, namely, random forest, XGBoost, logistic regression and support vector machine. With these four, one could make six combinations containing two algorithms, four combinations containing three algorithms and one combination containing all algorithms. Next to the combinations, the weight distribution or importance of each algorithm could be used to create a better ensemble. These can be any integer. They can be equally weighted or totally different. To get a good idea about good weight distributions, we will take into account a range from 1 to 4. This means that when testing the combination containing all algorithms, we will have to test 256 different ensembles. As predicting the outcome of a football match is a classification problem, we chose to work with a voting classifier that makes use of hard voting. In other words, the class with the most votes will be chosen. Each algorithm that is a part of the ensemble that is being tested casts a vote for a prediction class. This vote can be worth more than the other vote due to the weight distribution. All votes are summed up and the class with most votes will be the prediction class of the ensemble.

5.2 Modelling and Evaluation Process

To be able to assess whether ensembles can be used to increase the prediction accuracy of a model that predicts the outcome of a football match, we will need to compare the ensembles with the individual algorithms while taking into account multiple scenarios in which these individual algorithms perform optimally. This means that we will have to create several feature selections and apply hyperparameter optimisation before we create the ensembles. In Figure 5.1, the needed feature selection criteria, hyperparameter optimisation and comparison of the models are properly displayed. As we can see in the figure, we make use of three different feature selection criteria. Two of these are independent and one is dependent on the algorithms. This means that the feature selection criteria in total will result in six feature selections. A complete feature selection, a feature selection created using Pearson correlation coefficient, and four feature selections created using the Fisher's Score ranking. The next step to make sure these algorithms perform optimally in these scenario's is hyperparameter optimisation. For every feature selection, we will look for the best parameters for each algorithm in terms of accuracy. When these are found, the ensembles can be created. The evaluation of the individual algorithms and the ensembles will be divided into four parts where we compare each individual algorithm with the ensembles that contain the individual algorithm. Such a comparison will be divided into three parts where we first compare the individual algorithm and the ensembles using two different algorithms. Followed by the comparison of the individual algorithm and the ensembles using three different algorithms and, finally, the comparison of the individual algorithm and the ensembles using four different algorithms. The accuracies of the individual algorithms and the ensembles will be compared over the relevant feature selections. Namely, the complete feature selection, the Pearson correlation coefficient feature selection and the relevant Fisher's Score feature selection.

When the comparisons have been made, we can find out whether the use of the most beneficial ensembles is worthwhile or that the investment is not worth it. To find that out we need to take a look at the actual performance increase. The actual performance increase is the increase in prediction accuracy of the ensemble minus the increase in prediction accuracy of the individual algorithm divided by the increase in prediction accuracy of the individual algorithm. When the actual performance increase is higher than 5.0%, we believe the use of the ensemble to be worthwhile due to the only having to create the ensemble while not having to bother to collect and familiarize yourself with the data. There won't be any features that have to be modified but the use of an ensemble would cost somewhat more resources due to using multiple algorithms. As mentioned above, to calculate the actual performance increase, we must find out what the increase in prediction accuracy due to the individual algorithm and the ensembles. To do that we need to take into account the ratio of the class that is present the most. This class represents the home team winning and the home win percentage is 45.9%. In case the prediction accuracy of the individual algorithm is 49.9% the prediction accuracy increase is 4.0%. This means that for ensemble to be labeled as worthwhile the increase in prediction accuracy must be 4.2% or higher. In other words, the prediction accuracy must be 51.1% or higher.

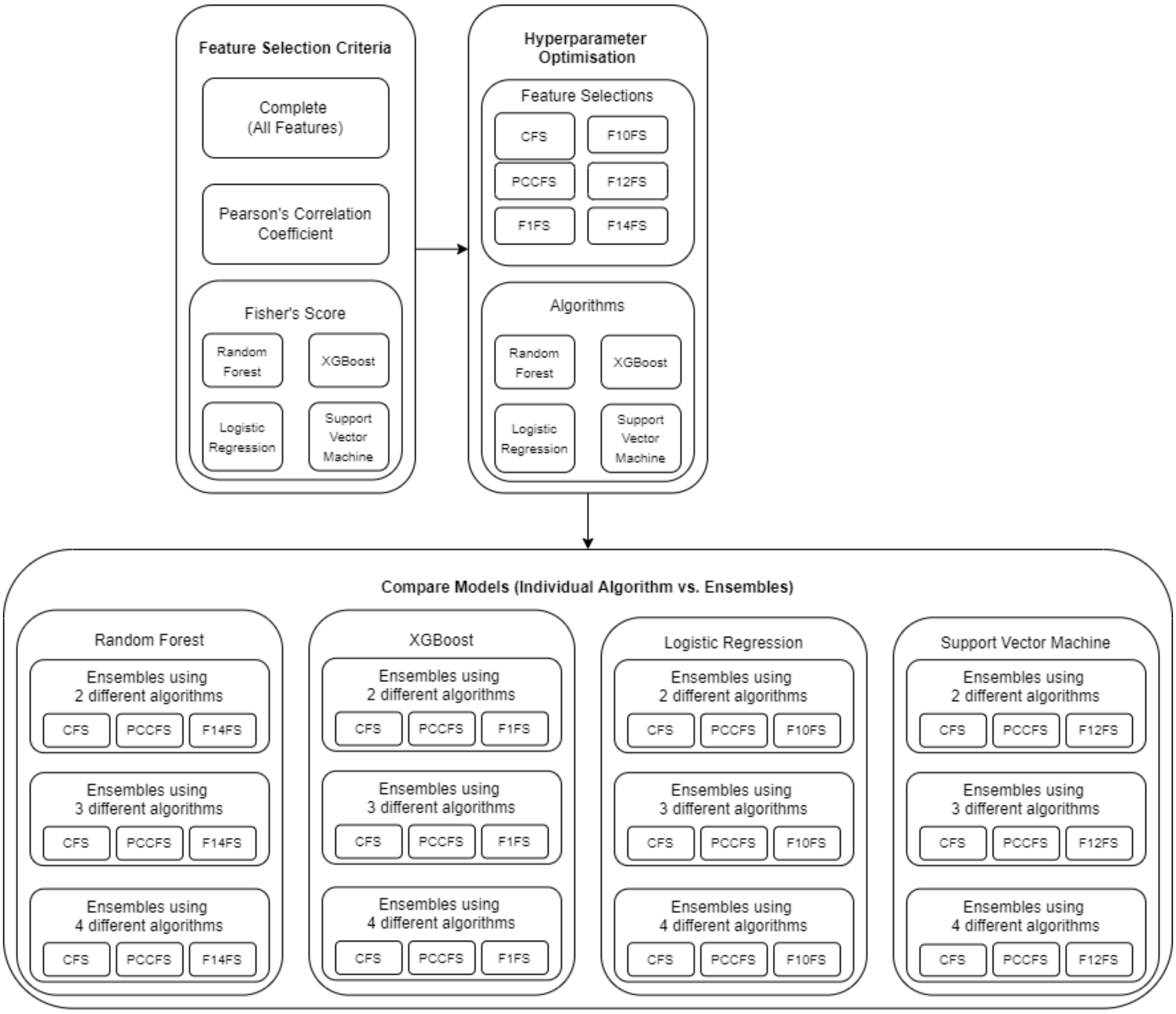

Figure 5.1: Modelling and Evaluation Process

5.3 Feature Selections

To fully assess whether ensembles could provide a better accuracy in this area than an individual algorithm, we made use of different feature selections. One feature selection which contains all features used in Chapter4, a feature selection created using Pearson correlation coefficient, and four feature selections created using the Fisher's Score ranking. To create all feature selections, except the complete feature selection, we made use of two filter feature selection methods. One dependent on the algorithm of the model and one independent of the algorithm of the model. We chose for these two filter feature selection methods due to the big amount of features we could use and their dependence/independence. Having many features makes it very difficult to run every possible feature combination to find the best one due to the big amount of possible combinations. Also, to get a good idea of whether ensembles could provide a better the accuracy in this area than an individual algorithm, we decided that we needed a feature selection method which is dependent on the algorithm of the model and a feature selection method which is independent of the algorithm of the model.

5.3.1 Complete Feature Selection

The list of features we used in Chapter4 can be found in Table 5.1. The complete feature selection (CFS) contains every of these 28 features.

TRH	PTRA	LPTVH	GSTVA
TRA	PTVH	LPTVA	GCH
TVH	PTVA	GSH	GCA
TVA	LPH	GSA	GCTRH
PH	LPA	GSTRH	GCTRA
PA	LPTRH	GSTRA	GCTVH
PTRH	LPTRA	GSTVH	GCTVA

Table 5.1: Complete Feature Selection

5.3.2 Pearson Correlation Coefficient Feature Selection

The idea of Pearson correlation coefficient is that good features are highly correlated with the target feature and uncorrelated amongst themselves. The Pearson correlation coefficient feature selection (PCCFS) is created in two steps and independent of the algorithm of the model. The first step is selecting features that are correlated with the target feature. We call this selection the relevant features. The second step is selecting the relevant features that are not correlated with any other features from this set or are correlated the most with the target feature. In case they are correlated with another feature in this set, we select the one that is correlated the most with the target feature. We do this because when two features are correlated, we can predict one from the other. Therefore, the model only really needs one of them.

As mentioned above, the first step is to find the features that are correlated with the target feature. The easiest way to do this is create a heat map containing the correlations between all features and thus also the target feature. The heat map is displayed in Figure 5.2. When looking at this heat map, we can find out which features are relevant features and are correlated with the target feature the most.

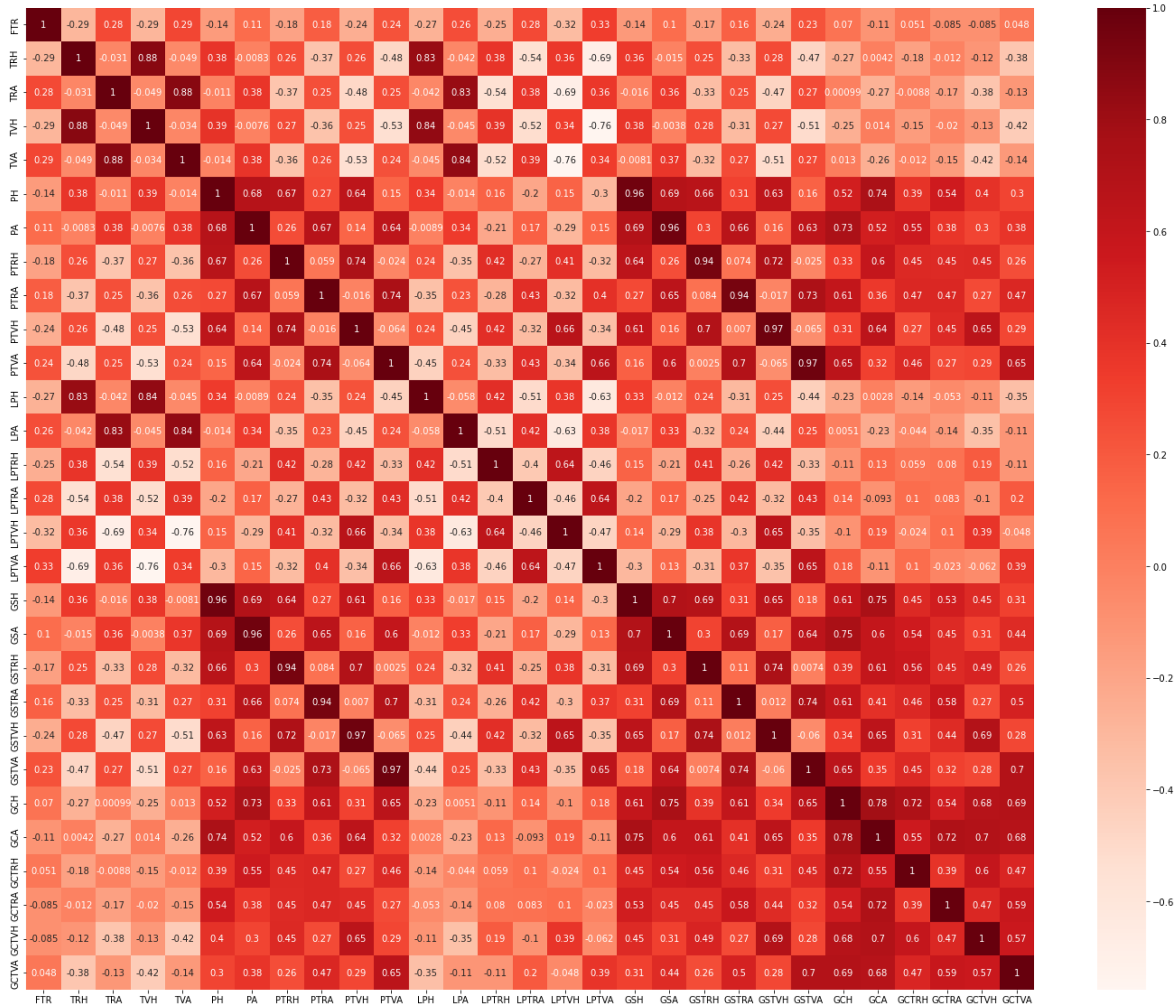

Figure 5.2: Pearson Correlation Coefficient Heat Map

When selecting the relevant features we made use use of a threshold of 0.25. The relevant features and their correlation with the target feature can be found in Table 5.2.

Feature	Correlation FTR
TRH	-0.291377
TRA	0.278276
TVH	-0.286619
TVA	0.292344
LPH	-0.265221
LPA	0.255368
LPTRA	0.275324
LPTVH	-0.323730
LPTVA	0.333190

Table 5.2: Pearson Correlation Coefficient Relevant Features

The next step is to find out which of the relevant features are not correlated with any other features from this set or are correlated the most with the target feature. To do this, a heat map containing the correlations between the relevant features should be created. The heat map is displayed in Figure 5.3.

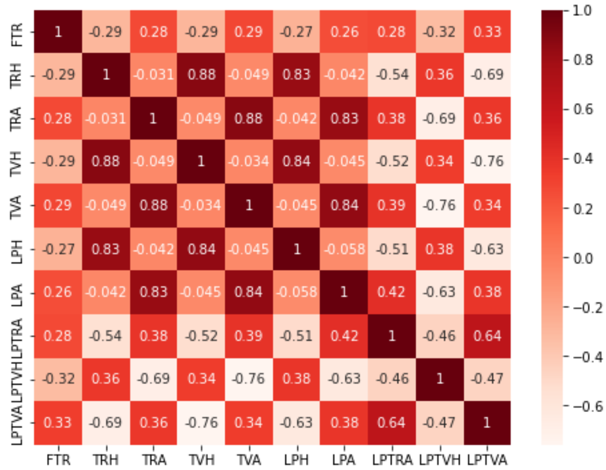

Figure 5.3: Pearson Correlation Coefficient Heat Map Relevant Features

When selecting which relevant features correlated amongst themselves, we made use of a threshold of 0.75. In Table 5.3, we can see what features are dropped due to being correlated amongst the other relevant features. The procedure of dropping features can be done in multiple ways but when starting the procedure with the feature that is correlating the most with the target feature, we make sure we do not drop any features too soon. According to the heat map, there are more correlated relevant features than these four cases. But when looking at these cases, we see that it concerns already dropped features which means that we did not forget to drop any features.

Relevant Feature 1	Relevant Feature 2	Dropped Feature
TVH	LPTVA	TVH
TVA	LPTVH	TVA
TRH	LPH	LPH
TRA	LPA	LPA

Table 5.3: Pearson Correlation Coefficient Drop Correlating Relevant Features

After dropping the correlated and weak relevant features, there are five features left. These features together form the Pearson correlation coefficient feature selection and can be found in Table 5.4.

TRH
TRA
LPTRA
LPTVH
LPTVA

Table 5.4: Pearson Correlation Coefficient Feature Selection

5.3.3 Fisher's Score Feature Selections

Fisher's score is one of the most widely used supervised feature selection methods. It returns the ranks of the features based on the fisher's score in descending order. To find the most suitable feature selection for an algorithm, we start by adding features to the potential feature selection based on their rank. The first potential feature selection contains only the first feature based on the ranking. The second potential feature selection contains the first and the second feature based on the ranking. This means that we will have 28 potential feature selections. Finally, you choose the potential feature selection with the highest accuracy for that particular algorithm. This feature selection method is dependent on the algorithm of the model due to the fact that the selection is highly influenced by the performance of an algorithm with a subset of the features. The ranking of every feature we used in Chapter4 can be found in Table 5.5.

n	Feature	n	Feature
1	LPTRH	15	GSTVH
2	GCTVA	16	GSTVA
3	GCTRA	17	PTVH
4	LPTRA	18	PTVA
5	GCTRH	19	GCA
6	GSA	20	LPA
7	LPTVH	21	PH
8	GCTVH	22	GCH
9	GSH	23	LPH
10	LPTVA	24	TRA
11	GSTRA	25	PA
12	PTRA	26	TVH
13	GSTRH	27	TRH
14	PTRH	28	TVA

Table 5.5: Fisher's Score Ranking

Random Forest

As we can see in Table 5.6, the random forest algorithm performs best with the feature selection containing the first 14 features of the fisher's score ranking. There are multiple feature selections which have an accuracy of 52.0% but these are larger and therefore contain features that do not add any new information. From now on this feature selection will be referred to as fisher-14 feature selection (F14FS).

n	Feature	Random Forest	n	Feature	Random Forest
1	LPTRH	0.501	15	GSTVH	0.514
2	GCTVA	0.434	16	GSTVA	0.516
3	GCTRA	0.442	17	PTVH	0.52
4	LPTRA	0.463	18	PTVA	0.516
5	GCTRH	0.473	19	GCA	0.519
6	GSA	0.478	20	LPA	0.507
7	LPTVH	0.496	21	PH	0.508
8	GCTVH	0.498	22	GCH	0.506
9	GSH	0.507	23	LPH	0.507
10	LPTVA	0.506	24	TRA	0.508
11	GSTRA	0.51	25	PA	0.505
12	PTRA	0.513	26	TVH	0.506
13	GSTRH	0.516	27	TRH	0.511
14	**PTRH**	**0.52**	28	TVA	0.507

Table 5.6: Fisher's Score Random Forest

XGBoost

As we can see in Table 5.7, the XGBoost algorithm performs best with the feature selection containing only the first feature of the fisher's score ranking. From now on this feature selection will be referred to as fisher-1 feature selection (F1FS). This result is a quite remarkable and suggests that the algorithm in its current form is overfitting. This means that it is possible that other feature selections could be more suitable to fully assess whether ensembles could provide a better accuracy in this area than the XGBoost algorithm. But do the fact that we do not apply hyperparameter optimisation before creating the feature selections and create them based on the default algorithm, the feature selection that will used is the fisher-1 feature selection.

n	Feature	XGBoost	n	Feature	XGBoost
1	**LPTRH**	**0.503**	15	GSTVH	0.49
2	GCTVA	0.477	16	GSTVA	0.487
3	GCTRA	0.474	17	PTVH	0.485
4	LPTRA	0.474	18	PTVA	0.495
5	GCTRH	0.469	19	GCA	0.494
6	GSA	0.467	20	LPA	0.473
7	LPTVH	0.472	21	PH	0.471
8	GCTVH	0.471	22	GCH	0.479
9	GSH	0.479	23	LPH	0.485
10	LPTVA	0.48	24	TRA	0.488
11	GSTRA	0.49	25	PA	0.493
12	PTRA	0.487	26	TVH	0.475
13	GSTRH	0.487	27	TRH	0.482
14	PTRH	0.489	28	TVA	0.483

Table 5.7: Fisher's Score XGBoost

Logistic Regression

As we can see in Table 5.8, the logistic regression algorithm performs best with the feature selection containing the first 10 features of the fisher's score ranking. There are multiple feature selections which have an accuracy of 53.5% but these are larger and therefore contain features that do not add any new information. From now on this feature selection will be referred to as fisher-10 feature selection F10FS.

n	Feature	Logistic Regression	n	Feature	Logistic Regression
1	LPTRH	0.504	15	GSTVH	0.535
2	GCTVA	0.501	16	GSTVA	0.535
3	GCTRA	0.506	17	PTVH	0.534
4	LPTRA	0.519	18	PTVA	0.534
5	GCTRH	0.518	19	GCA	0.535
6	GSA	0.52	20	LPA	0.522
7	LPTVH	0.526	21	PH	0.523
8	GCTVH	0.527	22	GCH	0.524
9	GSH	0.534	23	LPH	0.524
10	**LPTVA**	**0.535**	24	TRA	0.526
11	GSTRA	0.534	25	PA	0.526
12	PTRA	0.534	26	TVH	0.529
13	GSTRH	0.534	27	TRH	0.528
14	PTRH	0.534	28	TVA	0.529

Table 5.8: Fisher's Score Logistic Regression

Support Vector Machine

As we can see in Table 5.9, the support vector machine algorithm performs best with the feature selection containing the first 12 features of the fisher's score ranking.From now on this feature selection will be referred to as fisher-12 feature selection F12FS.

n	Feature	Support Vector Machine	n	Feature	Support Vector Machine
1	LPTRH	0.504	15	GSTVH	0.535
2	GCTVA	0.505	16	GSTVA	0.535
3	GCTRA	0.509	17	PTVH	0.534
4	LPTRA	0.517	18	PTVA	0.534
5	GCTRH	0.518	19	GCA	0.535
6	GSA	0.528	20	LPA	0.522
7	LPTVH	0.533	21	PH	0.523
8	GCTVH	0.532	22	GCH	0.524
9	GSH	0.537	23	LPH	0.524
10	LPTVA	0.537	24	TRA	0.526
11	GSTRA	0.539	25	PA	0.526
12	**PTRA**	**0.54**	26	TVH	0.529
13	GSTRH	0.539	27	TRH	0.528
14	PTRH	0.539	28	TVA	0.529

Table 5.9: Fisher's Score Support Vector Machine

5.4 Hyperparameter Optimisation

To be able to evaluate the individual algorithms and ensembles properly, we need to make sure they perform optimally in for all feature selections. This can be done by optimising the hyperparameters used by the algorithms. Our hyperparameter optimisation process is displayed in Figure 5.4. It contains two steps and a best parameter combination as a result. The first step is the randomised search which is done for the random forest algorithm and the XGBoost algorithm for every feature selection. The reason that a randomised search is necessary for these algorithm is that they have many important parameters that can have quite a broad range of values. Randomised search does not take a look at every parameter combination but a random set. As a result of the randomised search, the parameter area which is beneficial to the algorithm will have been discovered. The next step is the grid search. The grid search is much more precise and does take a look at every parameter combination that exists in the range selected by us. It will look for the best parameter combination for every possible combination between the algorithms and feature selections. As we can see in the figure, there are six feature selections and four algorithms which result in 24 best parameter combinations.

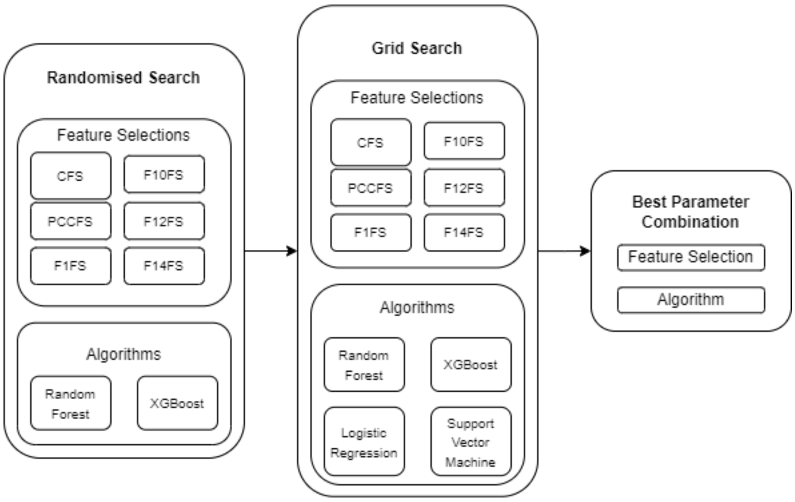

Figure 5.4: Hyperparameter Optimisation Process

We took into account typical values for the parameters. The ranges and possible values for each parameter are displayed in Table 5.10.

P	**Random Forest**	**P**	**XGBoost**	**P**	**Logistic Regression**	**P**	**Support Vector Machine**
max_depth	10-100, None	colsample_bytree	0.5-1.0	C	1e-3-1e5	C	1e-2-1e3
max_features	log2/sqrt	learning_rate	0.01-0.2	penalty	none, l1, l2, elasticnet	gamma	1e-4-1e1, scale
min_samples_leaf	1-30	max_depth	1-10	solver	newton-cg, lbfgs, liblinear		
min_samples_split	1-30	min_child_weight	1-30				
n_estimators	100-2000	n_estimators	100-2000				
		subsample	0.5-1.0				

Table 5.10: Hyperparameter Optimisation Range

5.4.1 Complete Feature Selection

In Table 5.11, the best parameters for the algorithms while using the complete feature selection are displayed.

P	**Random Forest**	**P**	**XGBoost**	**P**	**Logistic Regression**	**P**	**Support Vector Machine**
max_depth	65	colsample_bytree	0.9	C	0.001	C	1.0
max_features	sqrt	learning_rate	0.1	penalty	l2	gamma	0.0001
min_samples_leaf	20	max_depth	1	solver	newton-cg		
min_samples_split	15	min_child_weight	10				
n_estimators	1200	n_estimators	100				
		subsample	0.7				

Table 5.11: Hyperparameter Optimisation Complete

5.4.2 Pearson Correlation Coefficient Feature Selection

In Table 5.12, the best parameters for the algorithms while using the Pearson correlation coefficient feature selection are displayed.

P	**Random Forest**	**P**	**XGBoost**	**P**	**Logistic Regression**	**P**	**Support Vector Machine**
max_depth	10	colsample_bytree	1	C	0.01	C	100.0
max_features	log2	learning_rate	0.1	penalty	l1	gamma	0.001
min_samples_leaf	32	max_depth	1	solver	liblinear		
min_samples_split	18	min_child_weight	12				
n_estimators	1900	n_estimators	400				
		subsample	0.8				

Table 5.12: Hyperparameter Optimisation Pearson Correlation

5.4.3 Fisher-1 Feature Selection

In Table 5.13, the best parameters for the algorithms while using the fisher-1 feature selection are displayed.

P	**Random Forest**	**P**	**XGBoost**	**P**	**Logistic Regression**	**P**	**Support Vector Machine**
max_depth	10	colsample_bytree	0.6	C	0.001	C	10.0
max_features	log2	learning_rate	0.1	penalty	none	gamma	0.1
min_samples_leaf	2	max_depth	4	solver	newton-cg		
min_samples_split	10	min_child_weight	25				
n_estimators	500	n_estimators	1000				
		subsample	0.5				

Table 5.13: Hyperparameter Optimisation Fisher-1 Feature Selection

5.4.4 Fisher-10 Feature Selection

In Table 5.14, the best parameters for the algorithms while using the fisher-10 feature selection are displayed.

P	**Random Forest**	**P**	**XGBoost**	**P**	**Logistic Regression**	**P**	**Support Vector Machine**
max_depth	25	colsample_bytree	0.9	C	0.001	C	1.0
max_features	log2	learning_rate	0.05	penalty	none	gamma	0.0001
min_samples_leaf	22	max_depth	4	solver	newton-cg		
min_samples_split	32	min_child_weight	20				
n_estimators	1700	n_estimators	100				
		subsample	0.7				

Table 5.14: Hyperparameter Optimisation Fisher-10 Feature Selection

5.4.5 Fisher-12 Feature Selection

In Table 5.15, the best parameters for the algorithms while using the fisher-12 feature selection are displayed.

P	**Random Forest**	**P**	**XGBoost**	**P**	**Logistic Regression**	**P**	**Support Vector Machine**
max_depth	None	colsample_bytree	0.75	C	0.1	C	1.0
max_features	sqrt	learning_rate	0.05	penalty	l1	gamma	scale
min_samples_leaf	23	max_depth	4	solver	liblinear		
min_samples_split	3	min_child_weight	2				
n_estimators	200	n_estimators	50				
		subsample	0.5				

Table 5.15: Hyperparameter Optimisation Fisher-12 Feature Selection

5.4.6 Fisher-14 Feature Selection

In Table 5.16, the best parameters for the algorithms while using the fisher-14 feature selection are displayed.

P	**Random Forest**	**P**	**XGBoost**	**P**	**Logistic Regression**	**P**	**Support Vector Machine**
max_depth	40	colsample_bytree	0.6	C	0.01	C	1.0
max_features	log2	learning_rate	0.05	penalty	l2	gamma	scale
min_samples_leaf	18	max_depth	2	solver	newton-cg		
min_samples_split	13	min_child_weight	23				
n_estimators	1800	n_estimators	200				
		subsample	0.5				

Table 5.16: Hyperparameter Optimisation Fisher-14 Feature Selection

5.5 Random Forest and Ensembles

In this section, we will compare the individual algorithm called random forest with ensembles that make use of this algorithm including at least one other algorithm. We will start with a comparison between the individual algorithm and the ensembles using two algorithms. We will take a look at the best and worst performing ensembles to get an idea which weight distribution and composition of algorithms have potential to increase the prediction accuracy of a model that predicts the outcome of a football match. We will evaluate the ensembles by taking into account the complete feature selection, the Pearson correlation coefficient feature selection and the Fisher-14 feature selection. After looking at the ensembles using two algorithms, we will take a look at the ensembles using three algorithms and the ensembles using four algorithms. Finally, we will take a look at the best compositions of these three categories to find out whether a certain ensemble composition using a certain amount of algorithms performs better than the individual algorithm and the other ensemble compositions.

5.5.1 Ensembles using Two Algorithms

In this subsection, we will take a look at the ensembles that make use of two algorithms of which one is the random forest algorithm. The ensemble compositions that we will take a look at are:

- Random Forest and XGBoost
- Random Forest and Logistic Regression
- Random Forest and Support Vector Machine

Each ensemble composition has two algorithms which can differ in importance. Both algorithms can have a weight ranging from 1 to 4. This means that there are 16 possible ensembles for each ensemble composition. In appendix B.1, the accuracies of the models using the random forest algorithm and the accuracies of the models using the ensembles can be found.

Best Ensembles

The 20 best performing ensembles and the accuracies of the models can be found in Table 5.17. As we can see in this table, there are a lot of ensembles that perform better than the individual algorithm. The ensemble composition containing random forest and XGBoost performs best, especially when the XGBoost algorithm is more important than the random forest algorithm. When they are equally weighted, the ensemble perform slightly worse on average but still better than all other compositions. The second best composition contains random forest and support vector machine. This ensemble composition performs best when they are equally weighted, followed by the ensembles where support vector machine is dominant. Next to that, the best ensemble in this category achieves accuracies of 53.3%, 54.1% and 53.6% for the complete, Pearson correlation coefficient and Fisher-14 feature selections, respectively.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS	Total
RF	1	0	0	0	0.527	0.538	0.536	**1.601**
E-2	1	2	0	0	0.533	0.541	0.536	**1.610**
E-3	1	3	0	0	0.533	0.541	0.536	**1.610**
E-4	1	4	0	0	0.533	0.541	0.536	**1.610**
E-7	2	3	0	0	0.533	0.541	0.536	**1.610**
E-8	2	4	0	0	0.533	0.541	0.536	**1.610**
E-12	3	4	0	0	0.533	0.541	0.536	**1.610**
E-1	1	1	0	0	0.533	0.540	0.537	**1.610**
E-16	4	4	0	0	0.533	0.540	0.537	**1.610**
E-11	3	3	0	0	0.532	0.540	0.536	**1.608**
E-38	2	0	0	2	0.531	0.538	0.539	**1.608**
E-6	2	2	0	0	0.531	0.540	0.536	**1.607**
E-33	1	0	0	1	0.531	0.537	0.539	**1.607**
E-43	3	0	0	3	0.531	0.537	0.539	**1.607**
E-48	4	0	0	4	0.531	0.536	0.539	**1.606**
E-34	1	0	0	2	0.531	0.535	0.539	**1.605**
E-35	1	0	0	3	0.531	0.535	0.539	**1.605**
E-36	1	0	0	4	0.531	0.535	0.539	**1.605**
E-39	2	0	0	3	0.531	0.535	0.539	**1.605**
E-40	2	0	0	4	0.531	0.535	0.539	**1.605**
E-44	3	0	0	4	0.531	0.535	0.539	**1.605**

Table 5.17: Random Forest and Ensembles using Two Algorithms (Best 20)

Worst Ensembles

The 20 worst performing ensembles and the accuracies of the models can be found in Table 5.18. As we can see in this table, even in the 20 worst performing ensembles there are ensembles that perform better than the individual algorithm. The ensemble composition containing random forest and logistic regression performs worst, especially when logistic regression is dominant. When it is the other way around it performs slightly better. Also, the ensemble composition containing random forest and XGBoost performs slightly better when random forest is dominant.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS	Total
RF	1	0	0	0	0.527	0.538	0.536	**1.601**
E-18	1	0	2	0	0.531	0.531	0.534	**1.596**
E-19	1	0	3	0	0.531	0.531	0.534	**1.596**
E-20	1	0	4	0	0.531	0.531	0.534	**1.596**
E-23	2	0	3	0	0.531	0.531	0.534	**1.596**
E-24	2	0	4	0	0.531	0.531	0.534	**1.596**
E-28	3	0	4	0	0.531	0.531	0.534	**1.596**
E-31	4	0	3	0	0.526	0.538	0.534	**1.598**
E-26	3	0	2	0	0.526	0.537	0.536	**1.599**
E-42	3	0	0	2	0.526	0.537	0.536	**1.599**
E-13	4	1	0	0	0.527	0.538	0.535	**1.600**
E-14	4	2	0	0	0.528	0.537	0.535	**1.600**
E-15	4	3	0	0	0.527	0.538	0.535	**1.600**
E-10	3	2	0	0	0.527	0.538	0.536	**1.601**
E-21	2	0	1	0	0.528	0.538	0.535	**1.601**
E-25	3	0	1	0	0.528	0.538	0.535	**1.601**
E-29	4	0	1	0	0.527	0.539	0.535	**1.601**
E-37	2	0	0	1	0.527	0.538	0.536	**1.601**
E-45	4	0	0	1	0.527	0.538	0.536	**1.601**
E-27	3	0	3	0	0.529	0.536	0.537	**1.602**
E-9	3	1	0	0	0.528	0.538	0.536	**1.602**

Table 5.18: Random Forest and Ensembles using Two Algorithms (Worst 20)

5.5.2 Ensembles using Three Algorithms

In this subsection, we will take a look at the ensembles that make use of three algorithms of which one is the random forest algorithm. The ensemble compositions that we will take a look at are:

- Random Forest, XGBoost and Logistic Regression
- Random Forest, XGBoost and Support Vector Machine
- Random Forest, Logistic Regression and Support Vector Machine

Each ensemble composition has three algorithms which can differ in importance. All algorithms can have a weight ranging from 1 to 4. This means that there are 64 possible ensembles for each ensemble composition. In appendix B.2, the accuracies of the models using the random forest algorithm and the accuracies of the models using the ensembles can be found.

Best Ensembles

The 20 best performing ensembles and the accuracies of the models can be found in Table 5.19. As we can see in this table, there are a lot of ensembles that perform better than the individual algorithm. The ensemble composition containing random forest, XGBoost and logistic regression and the ensemble composition containing random forest, XGBoost and support vector machine perform great and best in this category, especially when XGBoost is most important. Furthermore, the best ensemble in this category achieves accuracies of 53.5%, 54.1% and 53.6% for the complete, Pearson correlation coefficient and Fisher-14 feature selections, respectively.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS	Total
RF	1	0	0	0	0.527	0.538	0.536	**1.601**
E-53	1	2	1	0	0.535	0.541	0.536	**1.612**
E-78	2	4	2	0	0.535	0.541	0.536	**1.612**
E-122	1	3	0	2	0.535	0.541	0.536	**1.612**
E-63	1	4	3	0	0.535	0.540	0.536	**1.611**
E-73	2	3	1	0	0.534	0.541	0.536	**1.611**
E-109	4	4	1	0	0.533	0.541	0.537	**1.611**
E-117	1	2	0	1	0.534	0.540	0.537	**1.611**
E-127	1	4	0	3	0.535	0.541	0.535	**1.611**
E-137	2	3	0	1	0.535	0.540	0.536	**1.611**
E-140	2	3	0	4	0.534	0.540	0.537	**1.611**
E-142	2	4	0	2	0.535	0.540	0.536	**1.611**
E-155	3	3	0	3	0.534	0.541	0.536	**1.611**
E-57	1	3	1	0	0.533	0.541	0.536	**1.610**
E-58	1	3	2	0	0.534	0.540	0.536	**1.610**
E-61	1	4	1	0	0.533	0.541	0.536	**1.610**
E-62	1	4	2	0	0.533	0.541	0.536	**1.610**
E-77	2	4	1	0	0.533	0.541	0.536	**1.610**
E-93	3	4	1	0	0.534	0.540	0.536	**1.610**
E-94	3	4	2	0	0.533	0.541	0.536	**1.610**
E-96	3	4	4	0	0.534	0.540	0.536	**1.610**

Table 5.19: Random Forest and Ensembles using Three Algorithms (Best 20)

Worst Ensembles

The 20 worst performing ensembles and the accuracies of the models can be found in Table 5.20. As we can see in this table, even in the 20 worst performing ensembles there are ensembles that perform better than the individual algorithm. The ensemble composition containing random forest, XGBoost and logistic regression and the ensemble composition containing random forest, logistic regression and support vector machine perform worst when logistic regression is dominant.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS	Total
RF	1	0	0	0	0.527	0.538	0.536	**1.601**
E-51	1	1	3	0	0.531	0.531	0.534	**1.596**
E-52	1	1	4	0	0.531	0.531	0.534	**1.596**
E-56	1	2	4	0	0.531	0.531	0.534	**1.596**
E-68	2	1	4	0	0.531	0.531	0.534	**1.596**
E-185	1	0	3	1	0.531	0.531	0.534	**1.596**
E-189	1	0	4	1	0.531	0.531	0.534	**1.596**
E-190	1	0	4	2	0.531	0.531	0.534	**1.596**
E-205	2	0	4	1	0.531	0.531	0.534	**1.596**
E-145	3	1	0	1	0.528	0.537	0.535	**1.600**
E-161	4	1	0	1	0.527	0.538	0.535	**1.600**
E-165	4	2	0	1	0.528	0.538	0.534	**1.600**
E-81	3	1	1	0	0.527	0.538	0.536	**1.601**
E-162	4	1	0	2	0.527	0.539	0.535	**1.601**
E-98	4	1	2	0	0.527	0.539	0.536	**1.602**
E-209	3	0	1	1	0.528	0.538	0.536	**1.602**
E-225	4	0	1	1	0.528	0.538	0.536	**1.602**
E-101	4	2	1	0	0.528	0.539	0.535	**1.602**
E-191	1	0	4	3	0.53	0.535	0.538	**1.603**
E-213	3	0	2	1	0.53	0.536	0.537	**1.603**
E-65	2	1	1	0	0.529	0.540	0.535	**1.604**

Table 5.20: Random Forest and Ensembles using Three Algorithms (Worst 20)

5.5.3 Ensembles using Four Algorithms

In this subsection, we will take a look at the ensembles that make use of four algorithms of which one is the random forest algorithm. The ensemble composition that we will take a look at is:

- Random Forest, XGBoost, Logistic Regression and Support Vector Machine

This ensemble composition has four algorithms which can differ in importance. All algorithms can have a weight ranging from 1 to 4. This means that there are 256 possible ensembles for this ensemble composition. In appendix B.3, the accuracies of the models using the random forest algorithm and the accuracies of the models using the ensembles can be found.

Best Ensembles

The 20 best performing ensembles and the accuracies of the models can be found in Table 5.21. As we can see in this table, there are a lot of ensembles that perform better than the individual algorithm. The ensembles where random forest and XGBoost are most important perform the best. Furthermore, the best ensemble in this category achieves accuracies of 53.4%, 54.1% and 53.7% for the complete, Pearson correlation coefficient and Fisher-14 feature selections, respectively.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS	Total
RF	1	0	0	0	0.527	0.538	0.536	**1.601**
E-481	4	4	1	1	0.534	0.541	0.537	**1.612**
E-357	2	4	2	1	0.534	0.541	0.536	**1.611**
E-401	3	3	1	1	0.534	0.541	0.536	**1.611**
E-407	3	3	2	3	0.534	0.541	0.536	**1.611**
E-421	3	4	2	1	0.534	0.540	0.537	**1.611**
E-465	4	3	1	1	0.533	0.541	0.537	**1.611**
E-466	4	3	1	2	0.534	0.540	0.537	**1.611**
E-491	4	4	3	3	0.534	0.541	0.536	**1.611**
E-321	2	2	1	1	0.534	0.541	0.535	**1.610**
E-439	4	1	2	3	0.534	0.537	0.539	**1.610**
E-273	1	3	1	1	0.534	0.540	0.536	**1.610**
E-289	1	4	1	1	0.533	0.541	0.536	**1.610**
E-293	1	4	2	1	0.534	0.540	0.536	**1.610**
E-306	2	1	1	2	0.535	0.537	0.538	**1.610**
E-309	2	1	2	1	0.535	0.537	0.538	**1.610**
E-353	2	4	1	1	0.534	0.540	0.536	**1.610**
E-361	2	4	3	1	0.533	0.541	0.536	**1.610**
E-370	3	1	1	2	0.534	0.540	0.536	**1.610**
E-371	3	1	1	3	0.535	0.537	0.538	**1.610**
E-386	3	2	1	2	0.534	0.540	0.536	**1.610**

Table 5.21: Random Forest and Ensembles using Four Algorithms (Best 20)

Worst Ensembles

The 20 worst performing ensembles and the accuracies of the models can be found in Table 5.22. As we can see in this table, even in the 20 worst performing ensembles almost every ensemble performs better than the individual algorithm. The ensembles where logistic regression is most important perform the worst. There is one ensemble that performs bad where random forest is dominant. But this does not suggest that ensembles where random forest is dominant perform badly due to the fact that there is only one ensemble like this in the 20 worst performing ensembles. This deviation is present due to the fact that we did not set the random state of the random forest algorithm which means that every time the classifier is called other and possibly unfavorable trees are created .

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS	Total
RF	1	0	0	0	0.527	0.538	0.536	**1.601**
E-253	1	1	4	1	0.531	0.531	0.534	**1.596**
E-433	4	1	1	1	0.525	0.538	0.535	**1.598**
E-249	1	1	3	1	0.53	0.534	0.538	**1.602**
E-317	2	1	4	1	0.53	0.535	0.538	**1.603**
E-335	2	2	4	3	0.53	0.535	0.538	**1.603**
E-250	1	1	3	2	0.53	0.535	0.539	**1.604**
E-254	1	1	4	2	0.531	0.534	0.539	**1.604**
E-255	1	1	4	3	0.53	0.535	0.539	**1.604**
E-265	1	2	3	1	0.53	0.535	0.539	**1.604**
E-267	1	2	3	3	0.531	0.535	0.538	**1.604**
E-269	1	2	4	1	0.531	0.535	0.538	**1.604**
E-272	1	2	4	4	0.531	0.535	0.538	**1.604**
E-284	1	3	3	4	0.531	0.535	0.538	**1.604**
E-285	1	3	4	1	0.53	0.535	0.539	**1.604**
E-288	1	3	4	4	0.531	0.535	0.538	**1.604**
E-318	2	1	4	2	0.53	0.535	0.539	**1.604**
E-330	2	2	3	2	0.531	0.535	0.538	**1.604**
E-334	2	2	4	2	0.53	0.535	0.539	**1.604**
E-432	3	4	4	4	0.531	0.535	0.538	**1.604**
E-437	4	1	2	1	0.529	0.549	0.535	**1.604**

Table 5.22: Random Forest and Ensembles using Four Algorithms (Worst 20)

5.5.4 Best Ensemble Composition

Combining the best results of the ensemble compositions into one overview will give us the opportunity to find out which subset of algorithms performs best. As we can see in Table 5.23, for every ensemble composition there is an ensemble with a specific importance distribution that performs better than the individual algorithm for each feature selection. The best performing ensemble composition is the one containing all algorithms. The best performing ensemble for this composition outperforms the individual algorithm by 0.8% while using the complete feature selection. Also, there is an ensemble with this composition that outperforms the individual algorithm by 0.3% when using the Pearson correlation coefficient feature selection. Next to that, there is at least one ensemble with that composition that outperforms the individual algorithm by 0.4% when using the Fisher-14 feature selection.

RF	XGB	LR	SVM	CFS	PCCFS	F14FS	Total
x				0.527	0.538	0.536	**1.601**
x	x			0.533	0.541	0.537	**1.611**
x		x		0.531	0.539	0.538	**1.608**
x			x	0.531	0.538	0.539	**1.608**
x	x	x		0.535	0.541	0.537	**1.613**
x	x		x	0.535	0.541	0.539	**1.615**
x		x	x	0.532	0.540	0.540	**1.612**
x	x	x	x	0.535	0.541	0.540	**1.616**

Table 5.23: Random Forest and Ensembles Best Composition

5.6 XGBoost and Ensembles

In this section, we will compare the individual algorithm called XGBoost with ensembles that make use of this algorithm including at least one other algorithm. We will start with a comparison between the individual algorithm and the ensembles using two algorithms. We will take a look at the best and worst performing ensembles to get an idea which weight distribution and composition of algorithms have potential to increase the prediction accuracy of a model that predicts the outcome of a football match. We will evaluate the ensembles by taking into account the complete feature selection, the Pearson correlation coefficient feature selection and the Fisher-1 feature selection. After looking at the ensembles using two algorithms, we will take a look at the ensembles using three algorithms and the ensembles using four algorithms. Finally, we will take a look at the best compositions of these three categories to find out whether a certain ensemble composition using a certain amount of algorithms performs better than the individual algorithm and the other ensemble compositions.

5.6.1 Ensembles using Two Algorithms

In this subsection, we will take a look at the ensembles that make use of two algorithms of which one is the XGBoost algorithm. The ensemble compositions that we will take a look at are:

- XGBoost and Random Forest
- XGBoost and Logistic Regression
- XGBoost and Support Vector Machine

Each ensemble composition has two algorithms which can differ in importance. Both algorithms can have a weight ranging from 1 to 4. This means that there are 16 possible ensembles for each ensemble composition. In appendix C.1, the accuracies of the models using the XGBoost algorithm and the accuracies of the models using the ensembles can be found.

Best Ensembles

The 20 best performing ensembles and the accuracies of the models can be found in Table 5.24. As we can see in this table, there are a lot of ensembles that perform as good as the individual algorithm but there is not any ensemble in this category that performs better than the individual algorithm. The ensemble compositions where XGBoost is dominant performs best. Furthermore, the best ensemble in this category achieves accuracies of 53.3%, 54.1% and 50.4% for the complete, Pearson correlation coefficient and Fisher-1 feature selections, respectively.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS	Total
XGB	0	1	0	0	0.533	0.541	0.504	**1.578**
E-5	1	2	0	0	0.533	0.541	0.504	**1.578**
E-6	2	2	0	0	0.533	0.541	0.504	**1.578**
E-9	1	3	0	0	0.533	0.541	0.504	**1.578**
E-10	2	3	0	0	0.533	0.541	0.504	**1.578**
E-13	1	4	0	0	0.533	0.541	0.504	**1.578**
E-14	2	4	0	0	0.533	0.541	0.504	**1.578**
E-15	3	4	0	0	0.533	0.541	0.504	**1.578**
E-21	0	2	1	0	0.533	0.541	0.504	**1.578**
E-25	0	3	1	0	0.533	0.541	0.504	**1.578**
E-26	0	3	2	0	0.533	0.541	0.504	**1.578**
E-29	0	4	1	0	0.533	0.541	0.504	**1.578**
E-30	0	4	2	0	0.533	0.541	0.504	**1.578**
E-31	0	4	3	0	0.533	0.541	0.504	**1.578**
E-37	0	2	0	1	0.533	0.541	0.504	**1.578**
E-41	0	3	0	1	0.533	0.541	0.504	**1.578**
E-42	0	3	0	2	0.533	0.541	0.504	**1.578**
E-45	0	4	0	1	0.533	0.541	0.504	**1.578**
E-46	0	4	0	2	0.533	0.541	0.504	**1.578**
E-47	0	4	0	3	0.533	0.541	0.504	**1.578**
E-16	4	4	0	0	0.533	0.540	0.504	**1.577**

Table 5.24: XGBoost and Ensembles using Two Algorithms (Best 20)

Worst Ensembles

The 20 worst performing ensembles and the accuracies of the models can be found in Table 5.25. As we can see in this table, the ensemble composition containing

XGBoost and logistic regression perform worst when logistic regression is dominant. Followed by other ensemble compositions where XGBoost is not as important as the other algorithm.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS	Total
XGB	0	1	0	0	0.533	0.541	0.504	**1.578**
E-18	0	1	2	0	0.531	0.531	0.504	**1.566**
E-19	0	1	3	0	0.531	0.531	0.504	**1.566**
E-20	0	1	4	0	0.531	0.531	0.504	**1.566**
E-23	0	2	3	0	0.531	0.531	0.504	**1.566**
E-24	0	2	4	0	0.531	0.531	0.504	**1.566**
E-28	0	3	4	0	0.531	0.531	0.504	**1.566**
E-8	4	2	0	0	0.527	0.538	0.503	**1.568**
E-3	3	1	0	0	0.527	0.538	0.504	**1.569**
E-7	3	2	0	0	0.528	0.538	0.504	**1.570**
E-34	0	1	0	2	0.531	0.535	0.504	**1.570**
E-35	0	1	0	3	0.531	0.535	0.504	**1.570**
E-36	0	1	0	4	0.531	0.535	0.504	**1.570**
E-39	0	2	0	3	0.531	0.535	0.504	**1.570**
E-40	0	2	0	4	0.531	0.535	0.504	**1.570**
E-44	0	3	0	4	0.531	0.535	0.504	**1.570**
E-4	4	1	0	0	0.529	0.538	0.503	**1.570**
E-2	2	1	0	0	0.529	0.538	0.504	**1.571**
E-12	4	3	0	0	0.528	0.539	0.504	**1.571**
E-33	0	1	0	1	0.533	0.537	0.504	**1.574**
E-38	0	2	0	2	0.533	0.537	0.504	**1.574**

Table 5.25: XGBoost and Ensembles using Two Algorithms (Worst 20)

5.6.2 Ensembles using Three Algorithms

In this subsection, we will take a look at the ensembles that make use of three algorithms of which one is the XGBoost algorithm. The ensemble compositions that we will take a look at are:

- XGBoost, Random Forest and Logistic Regression
- XGBoost, Random Forest and Support Vector Machine
- XGBoost, Logistic Regression and Support Vector Machine

Each ensemble composition has three algorithms which can differ in importance. All algorithms can have a weight ranging from 1 to 4. This means that there are 64 possible ensembles for each ensemble composition. In appendix C.2, the accuracies of the models using the XGBoost algorithm and the accuracies of the models using the ensembles can be found.

Best Ensembles

The 20 best performing ensembles and the accuracies of the models can be found in Table 5.26. As we can see in this table, there are some ensembles that perform better than the individual algorithm and a lot of ensembles that perform equally well. The ensemble composition containing XGBoost, random forest and logistic regression and the ensemble composition containing XGBoost, random forest and support vector machine perform best when XGBoost is dominant.The ensemble composition containing XGBoost, random forest and support vector machine also performed well with an ensemble where random forest and XGBoost where dominant. Next to that, the best ensemble in this category achieves accuracies of 53.6%, 54.0% and 50.4% for the complete, Pearson correlation coefficient and Fisher-1 feature selections, respectively.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS	Total
XGB	0	1	0	0	0.533	0.541	0.504	**1.578**
E-99	1	4	3	0	0.536	0.540	0.504	**1.580**
E-171	3	4	0	3	0.535	0.541	0.504	**1.580**
E-65	1	2	1	0	0.535	0.540	0.504	**1.579**
E-154	3	3	0	2	0.534	0.541	0.504	**1.579**
E-70	2	2	2	0	0.534	0.540	0.504	**1.578**
E-81	1	3	1	0	0.533	0.541	0.504	**1.578**
E-82	1	3	2	0	0.534	0.540	0.504	**1.578**
E-85	2	3	1	0	0.534	0.540	0.504	**1.578**
E-86	2	3	2	0	0.534	0.540	0.504	**1.578**
E-97	1	4	1	0	0.533	0.541	0.504	**1.578**
E-98	1	4	2	0	0.533	0.541	0.504	**1.578**
E-101	2	4	1	0	0.533	0.541	0.504	**1.578**
E-102	2	4	2	0	0.534	0.540	0.504	**1.578**
E-103	2	4	3	0	0.533	0.541	0.504	**1.578**
E-105	3	4	1	0	0.534	0.54	0.504	**1.578**
E-129	1	2	0	1	0.534	0.540	0.504	**1.578**
E-140	3	2	0	4	0.534	0.540	0.504	**1.578**
E-145	1	3	0	1	0.533	0.541	0.504	**1.578**
E-146	1	3	0	2	0.534	0.540	0.504	**1.578**
E-149	2	3	0	1	0.534	0.540	0.504	**1.578**

Table 5.26: XGBoost and Ensembles using Three Algorithms (Best 20)

Worst Ensembles

The 20 worst performing ensembles and the accuracies of the models can be found in Table 5.27. As we can see in this table, The ensemble composition containing XGBoost, random forest and logistic regression and the ensemble composition containing XGBoost, random forest and support vector machine perform worst when logistic regression is dominant. Followed by the ensemble composition containing

XGBoost, random forest and logistic regression where random forest is dominant.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS	Total
XGB	0	1	0	0	0.533	0.541	0.504	**1.578**
E-51	1	1	3	0	0.531	0.531	0.504	**1.566**
E-52	1	1	4	0	0.531	0.531	0.504	**1.566**
E-56	2	1	4	0	0.531	0.531	0.504	**1.566**
E-68	1	2	4	0	0.531	0.531	0.504	**1.566**
E-185	0	1	3	1	0.531	0.531	0.504	**1.566**
E-189	0	1	4	1	0.531	0.531	0.504	**1.566**
E-190	0	1	4	2	0.531	0.531	0.504	**1.566**
E-205	0	2	4	1	0.531	0.531	0.504	**1.566**
E-141	4	2	0	1	0.527	0.537	0.504	**1.568**
E-57	3	1	1	0	0.527	0.538	0.504	**1.569**
E-62	4	1	2	0	0.528	0.537	0.504	**1.569**
E-77	4	2	1	0	0.528	0.537	0.504	**1.569**
E-115	1	1	0	3	0.531	0.535	0.504	**1.570**
E-116	1	1	0	4	0.531	0.535	0.504	**1.570**
E-120	2	1	0	4	0.531	0.535	0.504	**1.570**
E-132	1	2	0	4	0.531	0.535	0.504	**1.570**
E-177	0	1	1	1	0.531	0.535	0.504	**1.570**
E-178	0	1	1	2	0.531	0.535	0.504	**1.570**
E-179	0	1	1	3	0.531	0.535	0.504	**1.570**
E-180	0	1	1	4	0.531	0.535	0.504	**1.570**

Table 5.27: XGBoost and Ensembles using Three Algorithms (Worst 20)

5.6.3 Ensembles using Four Algorithms

In this subsection, we will take a look at the ensembles that make use of four algorithms of which one is the XGBoost algorithm. The ensemble composition that we will take a look at is:

- XGBoost, Random Forest, Logistic Regression and Support Vector Machine

This ensemble composition has four algorithms which can differ in importance. All algorithms can have a weight ranging from 1 to 4. This means that there are 256 possible ensembles for this ensemble composition. In appendix C.3, the accuracies of the models using the XGBoost algorithm and the accuracies of the models using the ensembles can be found.

Best Ensembles

The 20 best performing ensembles and the accuracies of the models can be found in Table 5.28. As we can see in this table, there are some ensembles that perform better than the individual algorithm and a lot of ensembles that perform equally well.

The ensembles where XGBoost and random forest are dominant perform best. Furthermore, the best ensemble in this category achieves accuracies of 53.3%, 54.2% and 50.4% for the complete, Pearson correlation coefficient and Fisher-1 feature selections, respectively.

ALG	**RF**	**XGB**	**LR**	**SVM**	**CFS**	**PCCFS**	**F1FS**	**Total**
XGB	0	1	0	0	0.533	0.541	0.504	**1.578**
E-482	4	4	1	2	0.533	0.542	0.504	**1.579**
E-486	4	4	2	2	0.534	0.541	0.504	**1.579**
E-277	3	1	2	1	0.534	0.540	0.504	**1.578**
E-321	2	2	1	1	0.533	0.541	0.504	**1.578**
E-322	2	2	1	2	0.534	0.540	0.504	**1.578**
E-338	3	2	1	2	0.534	0.540	0.504	**1.578**
E-355	4	2	1	3	0.534	0.540	0.504	**1.578**
E-369	1	3	1	1	0.534	0.540	0.504	**1.578**
E-387	2	3	1	3	0.533	0.541	0.504	**1.578**
E-389	2	3	2	1	0.534	0.540	0.504	**1.578**
E-402	3	3	1	2	0.534	0.540	0.504	**1.578**
E-407	3	3	2	3	0.534	0.540	0.504	**1.578**
E-418	4	3	1	2	0.534	0.540	0.504	**1.578**
E-419	4	3	1	3	0.534	0.540	0.504	**1.578**
E-423	4	3	2	3	0.534	0.540	0.504	**1.578**
E-433	1	4	1	1	0.533	0.541	0.504	**1.578**
E-434	1	4	1	2	0.534	0.540	0.504	**1.578**
E-437	1	4	2	1	0.534	0.540	0.504	**1.578**
E-449	2	4	1	1	0.534	0.540	0.504	**1.578**
E-450	2	4	1	2	0.534	0.540	0.504	**1.578**

Table 5.28: XGBoost and Ensembles using Four Algorithms (Best 20)

Worst Ensembles

The 20 worst performing ensembles and the accuracies of the models can be found in Table 5.29. As we can see in this table, The ensembles where logistic regression is very dominant perform worst of all.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS	Total
XGB	0	1	0	0	0.533	0.541	0.504	**1.578**
E-253	1	1	4	1	0.531	0.531	0.504	**1.566**
E-254	1	1	4	2	0.530	0.534	0.504	**1.568**
E-317	1	2	4	1	0.530	0.534	0.504	**1.568**
E-249	1	1	3	1	0.531	0.534	0.504	**1.569**
E-250	1	1	3	2	0.530	0.535	0.504	**1.569**
E-265	2	1	3	1	0.530	0.535	0.504	**1.569**
E-285	3	1	4	1	0.530	0.535	0.504	**1.569**
E-334	2	2	4	2	0.530	0.535	0.504	**1.569**
E-335	2	2	4	3	0.530	0.535	0.504	**1.569**
E-243	1	1	1	3	0.531	0.535	0.504	**1.570**
E-244	1	1	1	4	0.531	0.535	0.504	**1.570**
E-245	1	1	2	1	0.531	0.535	0.504	**1.570**
E-248	1	1	2	4	0.531	0.535	0.504	**1.570**
E-255	1	1	4	3	0.531	0.535	0.504	**1.570**
E-260	2	1	1	4	0.531	0.535	0.504	**1.570**
E-266	2	1	3	2	0.531	0.535	0.504	**1.570**
E-267	2	1	3	3	0.531	0.535	0.504	**1.570**
E-269	2	1	4	1	0.531	0.535	0.504	**1.570**
E-270	2	1	4	2	0.531	0.535	0.504	**1.570**
E-271	2	1	4	3	0.531	0.535	0.504	**1.570**

Table 5.29: XGBoost and Ensembles using Four Algorithms (Worst 20)

5.6.4 Best Ensemble Composition

Combining the best results of the ensemble compositions into one overview will give us the opportunity to find out which subset of algorithms performs best. As we can see in Table 5.30, for most ensemble compositions there is an ensemble with a specific importance distribution that performs better than the individual algorithm for the complete feature selection. For the Pearson correlation coefficient feature selection this is not the case, there is only one ensemble composition that achieved a higher performance than the individual algorithm. For the Fisher-1 feature selections the best accuracies are all equal to the accuracy of the individual algorithm. The best performing ensemble compositions are the ensemble composition containing all algorithms and the ensemble composition containing random forest, XGBoost and logisitic regression. The best performing ensembles for both compositions outperform the individual algorithm while using the complete feature selection. The former by 0.2% and the latter by 0.3%. Also, there is an ensemble with the former composition that outperforms the individual algorithm by 0.1% when using the Pearson correlation coefficient feature selection. The latter performs equal to the individual algorithm. Next to that, both compositions perform equal to the individual algorithm when using the Fisher-1 feature selection.

RF	XGB	LR	SVM	CFS	PCCFS	F1FS	Total
	x			0.533	0.541	0.504	**1.578**
x	x			0.533	0.541	0.504	**1.578**
	x	x		0.534	0.541	0.504	**1.579**
	x		x	0.533	0.541	0.504	**1.578**
x	x	x		0.536	0.541	0.504	**1.581**
x	x		x	0.535	0.541	0.504	**1.580**
	x	x	x	0.534	0.541	0.504	**1.579**
x	x	x	x	0.535	0.542	0.504	**1.581**

Table 5.30: XGBoost and Ensembles Best Composition

5.7 Logistic Regression and Ensembles

In this section, we will compare the individual algorithm called logistic regression with ensembles that make use of this algorithm including at least one other algorithm. We will start with a comparison between the individual algorithm and the ensembles using two algorithms. We will take a look at the best and worst performing ensembles to get an idea which weight distribution and composition of algorithms have potential to increase the prediction accuracy of a model that predicts the outcome of a football match. We will evaluate the ensembles by taking into account the complete feature selection, the Pearson correlation coefficient feature selection and the Fisher-10 feature selection. After looking at the ensembles using two algorithms, we will take a look at the ensembles using three algorithms and the ensembles using four algorithms. Finally, we will take a look at the best compositions of these three categories to find out whether a certain ensemble composition using a certain amount of algorithms performs better than the individual algorithm and the other ensemble compositions.

5.7.1 Ensembles using Two Algorithms

In this subsection, we will take a look at the ensembles that make use of two algorithms of which one is the logistic regression algorithm. The ensemble compositions that we will take a look at are:

- Logistic Regression and Random Forest
- Logistic Regression and XGBoost
- Logistic Regression and Support Vector Machine

Each ensemble composition has two algorithms which can differ in importance. Both algorithms can have a weight ranging from 1 to 4. This means that there are 16 possible ensembles for each ensemble composition. In appendix D.1, the accuracies of the models using the logistic regression algorithm and the accuracies of the models using the ensembles can be found.

Best Ensembles

The 20 best performing ensembles and the accuracies of the models can be found in Table 5.31. As we can see in this table, there are a lot of ensembles that perform better than the individual algorithm. The ensemble composition containing logistic regression and XGBoost performs best when XGBoost is dominant. Followed by the same ensemble composition having an equal weight distribution. Next to that, the best ensemble in this category achieves accuracies of 53.3%, 54.1% and 53.6% for the complete, Pearson correlation coefficient and Fisher-10 feature selections, respectively.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS	Total
LR	0	0	1	0	0.531	0.531	0.535	**1.597**
E-2	0	2	1	0	0.533	0.541	0.536	**1.610**
E-3	0	3	1	0	0.533	0.541	0.536	**1.610**
E-4	0	4	1	0	0.533	0.541	0.536	**1.610**
E-7	0	3	2	0	0.533	0.541	0.536	**1.610**
E-8	0	4	2	0	0.533	0.541	0.536	**1.610**
E-12	0	4	3	0	0.533	0.541	0.536	**1.610**
E-1	0	1	1	0	0.534	0.537	0.536	**1.607**
E-6	0	2	2	0	0.534	0.537	0.536	**1.607**
E-11	0	3	3	0	0.534	0.537	0.536	**1.607**
E-16	0	4	4	0	0.534	0.537	0.536	**1.607**
E-33	0	0	1	1	0.531	0.535	0.538	**1.604**
E-38	0	0	2	2	0.531	0.535	0.538	**1.604**
E-17	1	0	1	0	0.530	0.536	0.537	**1.603**
E-27	3	0	3	0	0.530	0.536	0.537	**1.603**
E-34	0	0	1	2	0.531	0.535	0.537	**1.603**
E-35	0	0	1	3	0.531	0.535	0.537	**1.603**
E-36	0	0	1	4	0.531	0.535	0.537	**1.603**
E-39	0	0	2	3	0.531	0.535	0.537	**1.603**
E-40	0	0	2	4	0.531	0.535	0.537	**1.603**
E-44	0	0	3	4	0.531	0.535	0.537	**1.603**

Table 5.31: Logistic Regression and Ensembles using Two Algorithms (Best 20)

Worst Ensembles

The 20 worst performing ensembles and the accuracies of the models can be found in Table 5.32. As we can see in this table, even in the 20 worst performing ensembles there are ensembles that perform better than the individual algorithm. None of the ensembles perform worse than the individual algorithm. The ensemble compositions where logistic regression is very dominant perform worst.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS	Total
LR	0	0	1	0	0.531	0.531	0.535	**1.597**
E-5	0	1	2	0	0.531	0.531	0.535	**1.597**
E-9	0	1	3	0	0.531	0.531	0.535	**1.597**
E-10	0	2	3	0	0.531	0.531	0.535	**1.597**
E-13	0	1	4	0	0.531	0.531	0.535	**1.597**
E-14	0	2	4	0	0.531	0.531	0.535	**1.597**
E-15	0	3	4	0	0.531	0.531	0.535	**1.597**
E-21	1	0	2	0	0.531	0.531	0.535	**1.597**
E-25	1	0	3	0	0.531	0.531	0.535	**1.597**
E-26	2	0	3	0	0.531	0.531	0.535	**1.597**
E-29	1	0	4	0	0.531	0.531	0.535	**1.597**
E-30	2	0	4	0	0.531	0.531	0.535	**1.597**
E-31	3	0	4	0	0.531	0.531	0.535	**1.597**
E-37	0	0	2	1	0.531	0.531	0.535	**1.597**
E-41	0	0	3	1	0.531	0.531	0.535	**1.597**
E-42	0	0	3	2	0.531	0.531	0.535	**1.597**
E-45	0	0	4	1	0.531	0.531	0.535	**1.597**
E-46	0	0	4	2	0.531	0.531	0.535	**1.597**
E-47	0	0	4	3	0.531	0.531	0.535	**1.597**
E-19	3	0	1	0	0.527	0.537	0.535	**1.599**
E-23	3	0	2	0	0.528	0.538	0.534	**1.600**

Table 5.32: Logistic Regression and Ensembles using Two Algorithms (Worst 20)

5.7.2 Ensembles using Three Algorithms

In this subsection, we will take a look at the ensembles that make use of three algorithms of which one is the logistic regression algorithm. The ensemble compositions that we will take a look at are:

- Logistic Regression, Random Forest and XGBoost
- Logistic Regression, Random Forest and Support Vector Machine
- Logistic Regression, XGBoost and Support Vector Machine

Each ensemble composition has three algorithms which can differ in importance. All algorithms can have a weight ranging from 1 to 4. This means that there are 64 possible ensembles for each ensemble composition. In appendix D.2, the accuracies of the models using the logistic regression algorithm and the accuracies of the models using the ensembles can be found.

Best Ensembles

The 20 best performing ensembles and the accuracies of the models can be found in Table 5.33. As we can see in this table, there are a lot of ensembles that perform better than the individual algorithm. The ensemble composition containing random forest, XGBoost and logistic regression and the ensemble composition XGBoost, logistic regression and support vector machine perform best when XGBoost is dominant. Followed by the ensemble composition containing random forest, XGBoost and logistic regression where logistic regression is somewhat dominant. Next to that, the best ensemble in this category achieves accuracies of 53.4%, 54.1% and 53.5% for the complete, Pearson correlation coefficient and Fisher-10 feature selections, respectively.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS	Total
LR	0	0	1	0	0.531	0.531	0.535	**1.597**
E-93	1	4	3	0	0.534	0.541	0.535	**1.610**
E-53	1	2	1	0	0.535	0.541	0.534	**1.610**
E-57	1	3	1	0	0.533	0.541	0.536	**1.610**
E-61	1	4	1	0	0.533	0.541	0.536	**1.610**
E-62	2	4	1	0	0.533	0.541	0.536	**1.610**
E-77	1	4	2	0	0.533	0.541	0.536	**1.610**
E-78	2	4	2	0	0.535	0.541	0.534	**1.610**
E-121	0	3	1	1	0.533	0.541	0.536	**1.610**
E-125	0	4	1	1	0.533	0.541	0.536	**1.610**
E-126	0	4	1	2	0.533	0.541	0.536	**1.610**
E-141	0	4	2	1	0.533	0.541	0.536	**1.610**
E-58	2	3	1	0	0.534	0.541	0.534	**1.609**
E-73	1	3	2	0	0.535	0.540	0.534	**1.609**
E-63	3	4	1	0	0.534	0.540	0.535	**1.609**
E-86	2	2	3	0	0.532	0.541	0.536	**1.609**
E-103	3	2	4	0	0.533	0.540	0.536	**1.609**
E-107	3	3	4	0	0.533	0.540	0.536	**1.609**
E-49	1	1	1	0	0.533	0.540	0.535	**1.608**
E-69	1	2	2	0	0.533	0.540	0.535	**1.608**
E-75	3	3	2	0	0.532	0.541	0.535	**1.608**

Table 5.33: Logistic Regression and Ensembles using Three Algorithms (Best 20)

Worst Ensembles

The 20 worst performing ensembles and the accuracies of the models can be found in Table 5.34. As we can see in this table, even in the 20 worst performing ensembles there are ensembles that perform better than the individual algorithm. None of the ensembles perform worse than the individual algorithm. The ensemble compositions where logistic regression is very dominant performs worst. Followed by the ensemble composition containing random forest, XGBoost and logistic regression and the ensemble composition containing random forest, logistic regression and support vector machine where random forest is dominant.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS	Total
LR	0	0	1	0	0.531	0.531	0.535	**1.597**
E-81	1	1	3	0	0.531	0.531	0.535	**1.597**
E-97	1	1	4	0	0.531	0.531	0.535	**1.597**
E-98	2	1	4	0	0.531	0.531	0.535	**1.597**
E-101	1	2	4	0	0.531	0.531	0.535	**1.597**
E-145	0	1	3	1	0.531	0.531	0.535	**1.597**
E-161	0	1	4	1	0.531	0.531	0.535	**1.597**
E-162	0	1	4	2	0.531	0.531	0.535	**1.597**
E-165	0	2	4	1	0.531	0.531	0.535	**1.597**
E-209	1	0	3	1	0.531	0.531	0.535	**1.597**
E-225	1	0	4	1	0.531	0.531	0.535	**1.597**
E-226	1	0	4	2	0.531	0.531	0.535	**1.597**
E-229	2	0	4	1	0.531	0.531	0.535	**1.597**
E-52	4	1	1	0	0.527	0.538	0.535	**1.600**
E-68	4	1	2	0	0.527	0.538	0.535	**1.600**
E-190	4	0	1	2	0.528	0.538	0.534	**1.600**
E-237	4	0	4	1	0.530	0.535	0.535	**1.600**
E-56	4	2	1	0	0.528	0.538	0.535	**1.601**
E-113	0	1	1	1	0.531	0.535	0.535	**1.601**
E-118	0	2	1	2	0.531	0.535	0.535	**1.601**
E-123	0	3	1	3	0.531	0.535	0.535	**1.601**

Table 5.34: Logistic Regression and Ensembles using Three Algorithms (Worst 20)

5.7.3 Ensembles using Four Algorithms

In this subsection, we will take a look at the ensembles that make use of four algorithms of which one is the logistic regression algorithm. The ensemble composition that we will take a look at is:

- Logistic Regression, XGBoost, Random Forest and Support Vector Machine

This ensemble composition has four algorithms which can differ in importance. All algorithms can have a weight ranging from 1 to 4. This means that there are 256 possible ensembles for this ensemble composition. In appendix D.3, the accuracies of the models using the logistic regression algorithm and the accuracies of the models using the ensembles can be found.

Best Ensembles

The 20 best performing ensembles and the accuracies of the models can be found in Table 5.35. As we can see in this table, there are a lot of ensembles that perform better than the individual algorithm. The ensembles where random forest and

XGBoost are somewhat dominant performs best. Followed by the ensembles where random forest and XGBoost are very dominant. There are also some good performing ensembles where XGBoost is the only dominant algorithm and ensembles where XGBoost and support vector machine are dominant. Next to that, the best ensemble in this category achieves accuracies of 53.4%, 54.1% and 53.7% for the complete, Pearson correlation coefficient and Fisher-10 feature selections, respectively.

ALG	**RF**	**XGB**	**LR**	**SVM**	**CFS**	**PCCFS**	**F10FS**	**Total**
LR	0	0	1	0	0.531	0.531	0.535	**1.597**
E-431	4	4	3	3	0.534	0.541	0.537	**1.612**
E-271	4	2	1	3	0.535	0.540	0.536	**1.611**
E-301	4	4	1	1	0.534	0.541	0.536	**1.611**
E-281	3	3	1	1	0.534	0.541	0.535	**1.610**
E-361	3	4	2	1	0.534	0.541	0.535	**1.610**
E-261	2	2	1	1	0.534	0.540	0.536	**1.610**
E-286	4	3	1	2	0.534	0.540	0.536	**1.610**
E-289	1	4	1	1	0.533	0.541	0.536	**1.610**
E-300	3	4	1	4	0.534	0.540	0.536	**1.610**
E-346	3	3	2	2	0.533	0.541	0.536	**1.610**
E-354	1	4	2	2	0.533	0.541	0.536	**1.610**
E-366	4	4	2	2	0.534	0.540	0.536	**1.610**
E-368	4	4	2	4	0.534	0.540	0.536	**1.610**
E-278	2	3	1	2	0.533	0.540	0.537	**1.610**
E-288	4	3	1	4	0.533	0.540	0.537	**1.610**
E-294	2	4	1	2	0.534	0.541	0.534	**1.609**
E-267	3	2	1	3	0.533	0.540	0.536	**1.609**
E-283	3	3	1	3	0.534	0.540	0.535	**1.609**
E-290	1	4	1	2	0.534	0.540	0.535	**1.609**
E-298	3	4	1	2	0.534	0.540	0.535	**1.609**

Table 5.35: Logistic Regression and Ensembles using Four Algorithms (Best 20)

Worst Ensembles

The 20 worst performing ensembles and the accuracies of the models can be found in Table 5.36. As we can see in this table, even in the 20 worst performing ensembles almost every ensemble performs better than the individual algorithm. None of the ensembles perform worse than the individual algorithm. The ensembles where logistic regression is quite dominant perform worst. There is one ensemble that performs bad where random forest is dominant. But this does not suggest that ensembles where random forest is dominant perform badly due to the fact that there is not another ensemble like this in the 20 worst performing ensembles. This deviation is present due to the fact that we did not set the random state of the random forest algorithm which means that every time the classifier is called other and possibly unfavorable trees are created .

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS	Total
LR	0	0	1	0	0.531	0.531	0.535	**1.597**
E-433	1	1	4	1	0.531	0.531	0.535	**1.597**
E-253	4	1	1	1	0.528	0.538	0.534	**1.600**
E-455	2	2	4	3	0.529	0.535	0.536	**1.600**
E-370	1	1	3	2	0.531	0.535	0.535	**1.601**
E-373	2	1	3	1	0.530	0.535	0.536	**1.601**
E-378	3	1	3	2	0.531	0.535	0.535	**1.601**
E-387	1	2	3	3	0.531	0.535	0.535	**1.601**
E-404	1	3	3	4	0.531	0.535	0.535	**1.601**
E-435	1	1	4	3	0.530	0.535	0.536	**1.601**
E-438	2	1	4	2	0.530	0.535	0.536	**1.601**
E-440	2	1	4	4	0.531	0.535	0.535	**1.601**
E-441	3	1	4	1	0.530	0.535	0.536	**1.601**
E-443	3	1	4	3	0.531	0.535	0.535	**1.601**
E-452	1	2	4	4	0.531	0.535	0.535	**1.601**
E-454	2	2	4	2	0.530	0.535	0.536	**1.601**
E-468	1	3	4	4	0.531	0.535	0.535	**1.601**
E-470	2	3	4	2	0.530	0.535	0.536	**1.601**
E-434	1	1	4	2	0.530	0.535	0.537	**1.602**
E-305	1	1	2	1	0.531	0.535	0.536	**1.602**
E-322	1	2	2	2	0.531	0.535	0.536	**1.602**

Table 5.36: Logistic Regression and Ensembles using Four Algorithms (Worst 20)

5.7.4 Best Ensemble Composition

Combining the best results of the ensemble compositions into one overview will give us the opportunity to find out which subset of algorithms performs best. As we can see in Table 5.37, for every ensemble composition there is an ensemble with a specific importance distribution that performs equal to or better than the individual algorithm for each feature selection. The best performing ensemble compositions are the ensemble composition containing all algorithms and the ensemble composition using random forest, XGBoost and logisitic regression. The best performing ensembles for both compositions outperform the individual algorithm by 0.4% while using the complete feature selection. Also, there is an ensemble for both compositions outperforms the individual algorithm by 1.0% when using the Pearson correlation coefficient feature selection. Next to that, there is at least one ensemble for both compositions that outperforms the individual algorithm by 0.2% when using the Fisher-10 feature selection.

RF	XGB	LR	SVM	CFS	PCCFS	F10FS	Total
		x		0.531	0.531	0.535	**1.597**
	x	x		0.534	0.541	0.536	**1.611**
x		x		0.531	0.539	0.537	**1.607**
		x	x	0.531	0.535	0.538	**1.604**
x	x	x		0.535	0.541	0.537	**1.613**
	x	x	x	0.534	0.541	0.537	**1.612**
x		x	x	0.533	0.539	0.538	**1.610**
x	x	x	x	0.535	0.541	0.537	**1.613**

Table 5.37: Logistic Regression and Ensembles Best Composition

5.8 Support Vector Machine and Ensembles

In this section, we will compare the individual algorithm called support vector machine with ensembles that make use of this algorithm including at least one other algorithm. We will start with a comparison between the individual algorithm and the ensembles using two algorithms. We will take a look at the best and worst performing ensembles to get an idea which weight distribution and composition of algorithms have potential to increase the prediction accuracy of a model that predicts the outcome of a football match. We will evaluate the ensembles by taking into account the complete feature selection, the Pearson correlation coefficient feature selection and the Fisher-12 feature selection. After looking at the ensembles using two algorithms, we will take a look at the ensembles using three algorithms and the ensembles using four algorithms. Finally, we will take a look at the best compositions of these three categories to find out whether a certain ensemble composition using a certain amount of algorithms performs better than the individual algorithm and the other ensemble compositions.

5.8.1 Ensembles using Two Algorithms

In this subsection, we will take a look at the ensembles that make use of two algorithms of which one is the support vector machine algorithm. The ensemble compositions that we will take a look at are:

- Support Vector Machine and Random Forest
- Support Vector Machine and XGBoost
- Support Vector Machine and Logistic Regression

Each ensemble composition has two algorithms which can differ in importance. Both algorithms can have a weight ranging from 1 to 4. This means that there are 16 possible ensembles for each ensemble composition. In appendix E.1, the accuracies of the models using the support vector machine algorithm and the accuracies of the models using the ensembles can be found.

Best Ensembles

The 20 best performing ensembles and the accuracies of the models can be found in Table 5.38. As we can see in this table, there are a lot of ensembles that perform better than the individual algorithm. The ensemble composition containing support vector machine and XGBoost perform best when XGBoost is dominant. Followed by the equally weighted ensemble of support vector machine and XGBoost. Next to that, the best ensemble in this category achieves accuracies of 53.3%, 54.1% and 53.7% for the complete, Pearson correlation coefficient and Fisher-12 feature selections, respectively.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS	Total
SVM	0	0	0	1	0.531	0.535	0.540	**1.606**
E-2	0	2	0	1	0.533	0.541	0.537	**1.611**
E-3	0	3	0	1	0.533	0.541	0.537	**1.611**
E-4	0	4	0	1	0.533	0.541	0.537	**1.611**
E-7	0	3	0	2	0.533	0.541	0.537	**1.611**
E-8	0	4	0	2	0.533	0.541	0.537	**1.611**
E-12	0	4	0	3	0.533	0.541	0.537	**1.611**
E-1	0	1	0	1	0.533	0.537	0.540	**1.610**
E-6	0	2	0	2	0.533	0.537	0.540	**1.610**
E-11	0	3	0	3	0.533	0.537	0.540	**1.610**
E-16	0	4	0	4	0.533	0.537	0.540	**1.610**
E-33	1	0	0	1	0.531	0.537	0.540	**1.608**
E-38	2	0	0	2	0.531	0.537	0.539	**1.607**
E-48	4	0	0	4	0.531	0.536	0.540	**1.607**
E-5	0	1	0	2	0.531	0.535	0.540	**1.606**
E-9	0	1	0	3	0.531	0.535	0.540	**1.606**
E-10	0	2	0	3	0.531	0.535	0.540	**1.606**
E-13	0	1	0	4	0.531	0.535	0.540	**1.606**
E-14	0	2	0	4	0.531	0.535	0.540	**1.606**
E-15	0	3	0	4	0.531	0.535	0.540	**1.606**
E-17	0	0	1	1	0.531	0.535	0.540	**1.606**

Table 5.38: Support Vector Machine and Ensembles using Two Algorithms (Best 20)

Worst Ensembles

The 20 worst performing ensembles and the accuracies of the models can be found in Table 5.39. As we can see in this table, even in the 20 worst performing ensembles there are ensembles that perform equal to the individual algorithm. The ensemble composition containing logistic regression and support vector machine perform worst when logistic regression is dominant. Followed by the ensemble composition containing random forest and support vector machine when random forest is dominant.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS	Total
SVM	0	0	0	1	0.531	0.535	0.540	**1.606**
E-18	0	0	2	1	0.531	0.531	0.535	**1.597**
E-19	0	0	3	1	0.531	0.531	0.535	**1.597**
E-20	0	0	4	1	0.531	0.531	0.535	**1.597**
E-23	0	0	3	2	0.531	0.531	0.535	**1.597**
E-24	0	0	4	2	0.531	0.531	0.535	**1.597**
E-28	0	0	4	3	0.531	0.531	0.535	**1.597**
E-35	3	0	0	1	0.528	0.538	0.534	**1.600**
E-39	3	0	0	2	0.528	0.538	0.534	**1.600**
E-44	4	0	0	3	0.529	0.537	0.534	**1.600**
E-36	4	0	0	1	0.527	0.539	0.535	**1.601**
E-40	4	0	0	2	0.528	0.538	0.535	**1.601**
E-34	2	0	0	1	0.528	0.538	0.536	**1.602**
E-27	0	0	3	3	0.531	0.534	0.540	**1.605**
E-43	3	0	0	3	0.530	0.537	0.538	**1.605**
E-5	0	1	0	2	0.531	0.535	0.540	**1.606**
E-9	0	1	0	3	0.531	0.535	0.540	**1.606**
E-10	0	2	0	3	0.531	0.535	0.540	**1.606**
E-13	0	1	0	4	0.531	0.535	0.540	**1.606**
E-14	0	2	0	4	0.531	0.535	0.540	**1.606**
E-15	0	3	0	4	0.531	0.535	0.540	**1.606**

Table 5.39: Support Vector Machine and Ensembles using Two Algorithms (Worst 20)

5.8.2 Ensembles using Three Algorithms

In this subsection, we will take a look at the ensembles that make use of three algorithms of which one is the support vector machine algorithm. The ensemble compositions that we will take a look at are:

- Support Vector Machine, Random Forest and XGBoost
- Support Vector Machine, Random Forest and Logistic Regression
- Support Vector Machine, XGBoost and Logistic Regression

Each ensemble composition has three algorithms which can differ in importance. All algorithms can have a weight ranging from 1 to 4. This means that there are 64 possible ensembles for each ensemble composition. In appendix E.2, the accuracies of the models using the support vector machine algorithm and the accuracies of the models using the ensembles can be found.

Best Ensembles

The 20 best performing ensembles and the accuracies of the models can be found in Table 5.40. As we can see in this table, there are a lot of ensembles that perform better than the individual algorithm. The ensemble composition containing random forest, XGBoost and support vector machine when XGBoost is dominant or when XGBoost and random forest are dominant. The ensemble composition containing logistic regression, XGBoost and support vector machine perform slightly worse when XGBoost is dominant. Next to that, the best ensemble in this category achieves accuracies of 53.6%, 54.0% and 53.7% for the complete, Pearson correlation coefficient and Fisher-12 feature selections, respectively.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS	Total
SVM	0	0	0	1	0.531	0.535	0.540	**1.606**
E-157	1	4	0	3	0.536	0.540	0.537	**1.613**
E-117	1	2	0	1	0.534	0.540	0.538	**1.612**
E-122	2	3	0	1	0.535	0.540	0.537	**1.612**
E-123	3	3	0	1	0.533	0.540	0.539	**1.612**
E-128	4	4	0	1	0.534	0.540	0.538	**1.612**
E-134	2	2	0	2	0.534	0.540	0.538	**1.612**
E-142	2	4	0	2	0.534	0.541	0.537	**1.612**
E-143	3	4	0	2	0.535	0.540	0.537	**1.612**
E-158	2	4	0	3	0.534	0.540	0.538	**1.612**
E-159	3	4	0	3	0.534	0.541	0.537	**1.612**
E-57	0	3	1	1	0.533	0.541	0.537	**1.611**
E-61	0	4	1	1	0.533	0.541	0.537	**1.611**
E-62	0	4	2	1	0.533	0.541	0.537	**1.611**
E-77	0	4	1	2	0.533	0.541	0.537	**1.611**
E-118	2	2	0	1	0.534	0.540	0.537	**1.611**
E-121	1	3	0	1	0.533	0.541	0.537	**1.611**
E-125	1	4	0	1	0.533	0.541	0.537	**1.611**
E-126	2	4	0	1	0.533	0.541	0.537	**1.611**
E-127	3	4	0	1	0.534	0.540	0.537	**1.611**
E-133	1	2	0	2	0.534	0.540	0.537	**1.611**

Table 5.40: Support Vector Machine and Ensembles using Three Algorithms (Best 20)

Worst Ensembles

The 20 worst performing ensembles and the accuracies of the models can be found in Table 5.41. As we can see in this table, even in the 20 worst performing ensembles there are ensembles that perform better than the individual algorithm. The ensemble composition containing logistic regression, XGBoost and support vector machine and the ensemble composition containing logistic regression, random forest and support vector machine perform worst when logistic regression is dominant. Followed by several ensemble compositions where random forest is very dominant.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS	Total
SVM	0	0	0	1	0.531	0.535	0.540	**1.606**
E-51	0	1	3	1	0.531	0.531	0.535	**1.597**
E-52	0	1	4	1	0.531	0.531	0.535	**1.597**
E-56	0	2	4	1	0.531	0.531	0.535	**1.597**
E-68	0	1	4	2	0.531	0.531	0.535	**1.597**
E-185	1	0	3	1	0.531	0.531	0.535	**1.597**
E-189	1	0	4	1	0.531	0.531	0.535	**1.597**
E-190	2	0	4	1	0.531	0.531	0.535	**1.597**
E-205	1	0	4	2	0.531	0.531	0.535	**1.597**
E-180	4	0	1	1	0.529	0.538	0.533	**1.600**
E-115	3	1	0	1	0.528	0.538	0.535	**1.601**
E-120	4	2	0	1	0.529	0.539	0.533	**1.601**
E-179	3	0	1	1	0.530	0.537	0.534	**1.601**
E-184	4	0	2	1	0.528	0.539	0.534	**1.601**
E-196	4	0	1	2	0.528	0.538	0.536	**1.602**
E-116	4	1	0	1	0.528	0.539	0.535	**1.602**
E-132	4	1	0	2	0.528	0.539	0.535	**1.602**
E-178	2	0	1	1	0.531	0.536	0.536	**1.603**
E-188	4	0	3	1	0.531	0.536	0.537	**1.604**
E-191	3	0	4	1	0.531	0.535	0.538	**1.604**
E-192	4	0	4	1	0.531	0.535	0.538	**1.604**

Table 5.41: Support Vector Machine and Ensembles using Three Algorithms (Worst 20)

5.8.3 Ensembles using Four Algorithms

In this subsection, we will take a look at the ensembles that make use of four algorithms of which one is the support vector machine algorithm. The ensemble composition that we will take a look at is:

- Support Vector Machine, Logistic Regression, XGBoost and Random Forest

This ensemble composition has four algorithms which can differ in importance. All algorithms can have a weight ranging from 1 to 4. This means that there are 256 possible ensembles for this ensemble composition. In appendix E.3, the accuracies of the models using the support vector machine algorithm and the accuracies of the models using the ensembles can be found.

Best Ensembles

The 20 best performing ensembles and the accuracies of the models can be found in Table 5.42. As we can see in this table, there are a lot of ensembles that perform better than the individual algorithm. The ensembles where XGBoost and random forest are dominant. Followed by the ensembles where XGBoost is solely dominant. Next to that, the best ensemble in this category achieves accuracies of 53.4%, 54.1% and 53.8% for the complete, Pearson correlation coefficient and Fisher-12 feature selections, respectively.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS	Total
SVM	0	0	0	1	0.531	0.535	0.540	**1.606**
E-292	4	4	1	1	0.534	0.541	0.538	**1.613**
E-343	3	3	2	2	0.534	0.540	0.539	**1.613**
E-476	4	3	3	4	0.535	0.537	0.541	**1.613**
E-273	1	3	1	1	0.535	0.540	0.537	**1.612**
E-274	2	3	1	1	0.533	0.541	0.538	**1.612**
E-278	2	3	2	1	0.534	0.540	0.538	**1.612**
E-290	2	4	1	1	0.534	0.540	0.538	**1.612**
E-293	1	4	2	1	0.534	0.541	0.537	**1.612**
E-295	3	4	2	1	0.534	0.541	0.537	**1.612**
E-298	2	4	3	1	0.534	0.540	0.538	**1.612**
E-339	3	3	1	2	0.535	0.540	0.537	**1.612**
E-372	4	1	1	3	0.534	0.540	0.538	**1.612**
E-386	2	2	1	3	0.534	0.537	0.541	**1.612**
E-436	4	1	1	4	0.535	0.537	0.540	**1.612**
E-241	1	1	1	1	0.535	0.537	0.539	**1.611**
E-261	1	2	2	1	0.534	0.537	0.540	**1.611**
E-280	4	3	2	1	0.531	0.541	0.539	**1.611**
E-281	1	3	3	1	0.534	0.537	0.540	**1.611**
E-289	1	4	1	1	0.533	0.541	0.537	**1.611**
E-294	2	4	2	1	0.534	0.540	0.537	**1.611**

Table 5.42: Support Vector Machine and Ensembles using Four Algorithms (Best 20)

Worst Ensembles

The 20 worst performing ensembles and the accuracies of the models can be found in Table 5.43. As we can see in this table, even in the 20 worst performing ensembles almost every ensemble performs better than the individual algorithm. The ensemble where logistic regression is very dominant performs worst. Followed by the ensemble where random forest is very dominant.

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS	Total
SVM	0	0	0	1	0.531	0.535	0.540	**1.606**
E-253	1	1	4	1	0.531	0.531	0.535	**1.597**
E-244	4	1	1	1	0.528	0.539	0.534	**1.601**
E-243	3	1	1	1	0.530	0.539	0.535	**1.604**
E-248	4	1	2	1	0.529	0.541	0.534	**1.604**
E-254	2	1	4	1	0.530	0.535	0.539	**1.604**
E-255	3	1	4	1	0.530	0.535	0.539	**1.604**
E-269	1	2	4	1	0.531	0.535	0.538	**1.604**
E-317	1	1	4	2	0.531	0.534	0.539	**1.604**
E-245	1	1	2	1	0.530	0.535	0.540	**1.605**
E-249	1	1	3	1	0.531	0.535	0.539	**1.605**
E-250	2	1	3	1	0.531	0.535	0.539	**1.605**
E-270	2	2	4	1	0.530	0.535	0.540	**1.605**
E-313	1	1	3	2	0.530	0.535	0.540	**1.605**
E-334	2	2	4	2	0.530	0.535	0.540	**1.605**
E-350	2	3	4	2	0.530	0.535	0.540	**1.605**
E-375	3	1	2	3	0.531	0.535	0.539	**1.605**
E-398	2	2	4	3	0.530	0.535	0.540	**1.605**
E-447	3	1	4	4	0.531	0.535	0.539	**1.605**
E-457	1	2	3	4	0.531	0.535	0.539	**1.605**
E-463	3	2	4	4	0.532	0.535	0.538	**1.605**

Table 5.43: Support Vector Machine and Ensembles using Four Algorithms (Worst 20)

5.8.4 Best Ensemble Composition

Combining the best results of the ensemble compositions into one overview will give us the opportunity to find out which subset of algorithms performs best. As we can see in Table 5.30, for most ensemble compositions there is an ensemble with a specific importance distribution that performs better than the individual algorithm for the complete feature selection and the Pearson correlation coefficient feature selection. For the Fisher-12 feature selections this is not the case, there are only two ensemble compositions that achieved a higher performance than the individual algorithm. The best performing ensemble compositions are the the ensemble composition containing all algorithms and the ensemble composition containing random forest, XGBoost and support vector machine. The best performing ensembles for both compositions outperform the individual algorithm while using the complete feature selection. The former by 0.4% and the latter by 0.5%. Also, there is an ensemble for both compositions that outperforms the individual algorithm by 0.6% when using the Pearson correlation coefficient feature selection. Next to that, there is at least one ensemble with the ensemble composition containing all algorithms that outperforms the individual algorithm by 0.1% when using the Fisher-14 feature selection. The other composition performs equal to the individual algorithm.

RF	XGB	LR	SVM	CFS	PCCFS	F12FS	Total
			x	0.531	0.535	0.540	**1.606**
	x		x	0.533	0.541	0.540	**1.614**
		x	x	0.531	0.535	0.540	**1.606**
x			x	0.531	0.539	0.540	**1.610**
	x	x	x	0.534	0.541	0.540	**1.615**
x	x		x	0.536	0.541	0.540	**1.617**
x		x	x	0.533	0.539	0.541	**1.613**
x	x	x	x	0.535	0.541	0.541	**1.617**

Table 5.44: Support Vector Machine and Ensembles Best Composition

5.9 Discussion

In this section, we will discuss the findings and answer the relevant research questions for each algorithm separately. We will describe how ensembles of the promising algorithms or a subset of the promising algorithms can be used to improve the prediction accuracy of a model that predicts the outcome of a football match. We will discuss which subset of the promising algorithms could be beneficial in an ensemble and what the importance distribution should be of such an ensemble. We looked at these things in three different categories representing smaller and larger ensembles to find out how to improve the prediction accuracy when resources are limited and when resources are unlimited. As mentioned before, we will discuss each individual algorithm and the relevant ensembles separately. This means that we will only consider ensembles of which the algorithm is a part of.

5.9.1 Random Forest

The ensemble composition containing all algorithms achieved the highest prediction accuracy over the three feature selections. This does not mean that any ensemble making use of this composition will always outperform the others and the individual algorithm. It simply means that we could create an ensemble with this composition while taking into account a specific importance distribution that would achieve the highest prediction accuracy for at least one of these feature selections. To find out more about the ensembles that performed best for all feature selections, we will have to take a closer look at the three categories. We will start with the bigger ensembles containing all algorithms which can be very beneficial when resources are not an issue. The best performing ensembles in this category outperformed the individual algorithm for all three feature selections. They made use of an importance distribution where random forest and XGBoost were very dominant. Next to that, the best ensemble achieved accuracies of 53.4%, 54.1% and 53.7% for the complete, Pearson correlation coefficient and Fisher-14 feature selections, respectively. When resources are somewhat more limited, a smaller ensemble could also be beneficial. In the category containing only ensembles using three algorithms, the best ensembles were the ones using the composition containing random forest, XGBoost and support vector machine and the composition containing random forest, XGBoost and logistic regression. The most important about these ensembles is that the XGBoost algorithm is very dominant in all of them. The best ensemble in this category achieved accuracies of 53.5%, 54.1% and 53.6% for the complete, Pearson correlation coefficient and Fisher-14 feature selections, respectively. In the category containing only ensembles using two algorithms, the best ensembles were the ones using the composition containing random forest and XGBoost. They made use of an importance distribution where XGBoost was quite dominant. The best ensemble in this category achieved accuracies of 53.3%, 54.1% and 53.6% for the complete, Pearson correlation coefficient and Fisher-14 feature selections, respectively.

Our goal was to find out how to use ensembles to improve the prediction accuracy. For us to achieve this goal, we needed to find ensembles that outperformed the individual algorithm which achieved accuracies of 52.7%, 53.8% and 53.6% for the complete, Pearson correlation coefficient and Fisher-14 feature selections, respectively. The best ensembles of each category outperformed the individual algorithm for the complete and Pearson correlation coefficient feature selections. They performed equally well for the Fisher-14 feature selection. This means that we found a way to use ensembles to increase the prediction accuracy. The next step is to take a look at what subset of the promising algorithms and which importance distribution can be most beneficial to an ensemble. When comparing the best ensembles of each category with each other, we see that the ensemble using two algorithms performed slightly worse than the other two taking into account all feature selections. The others performed equally well but when also taking into account the amount of resources used, the most beneficial ensemble is the one using the composition containing random forest, XGBoost and support vector machine or the composition containing random forest, XGBoost and logistic regression while making use of an importance distribution where XGBoost was very dominant.

Now we have established that ensembles can be used to increase accuracy of a model that predicts the outcome of a football match and what the most beneficial ensembles should look like, we can find out whether the use of these particular ensembles is worthwhile or that the investment is not worth it at all. To find that out we need to take a look at the actual performance increase. The actual performance increase is the increase in prediction accuracy of the ensemble minus the increase in prediction accuracy of the individual algorithm divided by the increase in prediction accuracy of the individual algorithm. As mentioned before, to find out the increase in prediction accuracy, we need to take into account the ratio of the class that is present the most. This class represents the home team winning and the home win percentage is 45.9%. The individual algorithm is responsible for an increase of 6.8%, 7.9%, and 7.7% for the complete, Pearson correlation coefficient and Fisher-14 feature selections, respectively. The best ensemble is responsible for an increase of 7.6%, 8.2% and 7.7% for the complete, Pearson correlation coefficient and Fisher-14 feature selections, respectively. This means that the actual performance increase is 11.8%, 3.8%, 0% for the complete, Pearson correlation coefficient and Fisher-14 feature selections, respectively. The average actual performance increase is 5.2%. This means that making use of the most beneficial ensemble will be worthwhile in this case.

5.9.2 XGBoost

The ensemble composition containing all algorithms and the ensemble composition containing random forest, XGBoost and logisitic regression achieved the highest prediction accuracy over the three feature selections. This does not mean that any ensemble making use of these compositions will always outperform the others and the individual algorithm. It simply means that we could create an ensemble with one of these compositions while taking into account a specific importance distribution that would achieve the highest prediction accuracy for at least one of these feature selections. To find out more about the ensembles that performed best for all feature selections, we will have to take a closer look at the three categories. We will start with the bigger ensembles containing all algorithms which can be very beneficial when resources are not an issue. The best performing ensembles in this category outperformed the individual algorithm for one of the three feature selection. They performed equally well for the other feature selections compared to the individual algorithm. The ensembles made use of an importance distribution where random forest and XGBoost were very dominant. Next to that, the best ensemble achieved accuracies of 53.3%, 54.2% and 50.4% for the complete, Pearson correlation coefficient and Fisher-1 feature selections, respectively. When resources are somewhat more limited, a smaller ensemble could also be beneficial. In the category containing only ensembles using three algorithms, the best ensembles were the ones using the composition containing random forest, XGBoost and support vector machine and the composition containing random forest, XGBoost and logistic regression. The most important about these ensembles is that the XGBoost algorithm is dominant in all of them. The best ensemble in this category achieved accuracies of 53.6%, 54.0% and 50.4% for the complete, Pearson correlation coefficient and Fisher-1 feature selections, respectively. In the category containing only ensembles using two algorithms, the best ensembles were the ones that made use of an importance distribution where XGBoost was quite dominant. All composition performed equally well. The best ensemble in this category achieved accuracies of 53.3%, 54.1% and 50.4% for the complete, Pearson correlation coefficient and Fisher-1 feature selections, respectively.

Our goal was to find out how to use ensembles to improve the prediction accuracy. For us to achieve this goal, we needed to find ensembles that outperformed the individual algorithm which achieved accuracies of 53.3%, 54.1% and 50.4% for the complete, Pearson correlation coefficient and Fisher-1 feature selections, respectively. The best ensemble for the smallest category did not outperform the individual algorithm for any feature selection. The best ensembles of other categories did. Even though they only increased the accuracy for one feature selection, they did increase the prediction accuracy. The ensemble containing all algorithm outperformed the individual algorithm by 0.1% for the Pearson correlation coefficient feature selection and the ensemble containing random forest, XGBoost and logistic regression outperformed the individual algorithm by 0.3% for the complete feature selection. This means that we found a way to use ensembles to increase the prediction accuracy. The next step is to take a look at what subset of the promising algorithms and which importance distribution can be most beneficial to an ensemble. When comparing the best ensembles of each category with each other, we see that the ensembles using two algorithms performed slightly worse than the other two. Also, we see that the ensemble using three algorithms performed slightly better than the ensemble using all algorithms. This means that the most beneficial ensemble is the one using the composition containing random forest, XGBoost and support vector machine or the composition containing random forest, XGBoost and logistic regression while making use of an importance distribution where XGBoost was dominant.

Now we have established that ensembles can be used to increase accuracy of a model that predicts the outcome of a football match and what the most beneficial ensembles should look like, we can find out whether the use of these particular ensembles is worthwhile or that the investment is not worth it at all. To find that out we need to take a look at the actual performance increase. The actual performance increase is the increase in prediction accuracy of the ensemble minus the increase in prediction accuracy of the individual algorithm divided by the increase in prediction accuracy of the individual algorithm. As mentioned before, to find out the increase in prediction accuracy, we need to take into account the ratio of the class that is present the most. This class represents the home team winning and the home win percentage is 45.9%. The individual algorithm is responsible for an increase of

7.4%, 8.2%, and 4.5% for the complete, Pearson correlation coefficient and Fisher-1 feature selections, respectively. The best ensemble is responsible for an increase of 7.7%, 8.1% and 4.5% for the complete, Pearson correlation coefficient and Fisher-1 feature selections, respectively. This means that the actual performance increase is 4.1%, -1.2%, 0% for the complete, Pearson correlation coefficient and Fisher-1 feature selections, respectively. The average actual performance increase is 1.0%. This means that making use of the most beneficial ensemble will not be worthwhile in this case.

5.9.3 Logistic Regression

The ensemble composition containing all algorithms and the ensemble composition containing random forest, XGBoost and logisitic regression achieved the highest prediction accuracy over the three feature selections. This does not mean that any ensemble making use of these compositions will always outperform the others and the individual algorithm. It simply means that we could create an ensemble with one of these compositions while taking into account a specific importance distribution that would achieve the highest prediction accuracy for at least one of these feature selections. To find out more about the ensembles that performed best for all feature selections, we will have to take a closer look at the three categories. We will start with the bigger ensembles containing all algorithms which can be very beneficial when resources are not an issue. The best performing ensembles in this category outperformed the individual algorithm all feature selection. The ensembles made use of an importance distribution where random forest and XGBoost were somewhat dominant. Next to that, the best ensemble achieved accuracies of 53.4%, 54.1% and 53.7% for the complete, Pearson correlation coefficient and Fisher-10 feature selections, respectively. When resources are somewhat more limited, a smaller ensemble could also be beneficial. In the category containing only ensembles using three algorithms, the best ensembles were the ones using the composition containing random forest, XGBoost and logistic regression and the composition containing XGBoost, logistic regression and support vector machine. The most important about these ensembles is that the XGBoost algorithm is dominant in all of them. The best ensemble in this category achieved accuracies of 53.4%, 54.1% and 53.5% for the complete, Pearson correlation coefficient and Fisher-10 feature selections, respectively. In the category containing only ensembles using two algorithms, the best ensembles were the ones using the composition containing logistic regression and XGBoost. They made use of an importance distribution where XGBoost was quite dominant. The best ensemble in this category achieved accuracies of 53.3%, 54.1% and 53.6% for the complete, Pearson correlation coefficient and Fisher-10 feature selections, respectively.

Our goal was to find out how to use ensembles to improve the prediction accuracy. For us to achieve this goal, we needed to find ensembles that outperformed the individual algorithm which achieved accuracies of 53.1%, 53.1% and 53.5% for the complete, Pearson correlation coefficient and Fisher-10 feature selections, respectively. The best ensembles of each category outperformed the individual algorithm for the complete and Pearson correlation coefficient feature selections. They mostly outperformed the individual algorithm for the Fisher-10 feature selection. The best ensemble that made use of three algorithms performed equally well on that feature selection. This means that we found a way to use ensembles to increase the prediction accuracy. The next step is to take a look at what subset of the promising algorithms and which importance distribution can be most beneficial to an ensemble. When comparing the best ensembles of each category with each other, we see that the ensemble using four algorithms performed better than the others. This means that the most beneficial ensemble is the one using all algorithms while making use of an importance distribution where random forest and XGBoost were somewhat dominant.

Now we have established that ensembles can be used to increase accuracy of a model that predicts the outcome of a football match and what the most beneficial ensembles should look like, we can find out whether the use of these particular ensembles is worthwhile or that the investment is not worth it at all. To find that out we need to take a look at the actual performance increase. The actual performance increase is the increase in prediction accuracy of the ensemble minus the increase in prediction accuracy of the individual algorithm divided by the increase in prediction accuracy of the individual algorithm. As mentioned before, to find out the increase in prediction accuracy, we need to take into account the ratio of the class that is present the most. This class represents the home team winning and the home win percentage is 45.9%. The individual algorithm is responsible for an increase of 7.2%, 7.2%, and 7.6% for the complete, Pearson correlation coefficient and Fisher-10 feature selections, respectively. The best ensemble is responsible for an increase of 7.5%, 8.2% and 7.8% for the complete, Pearson correlation coefficient and Fisher-10 feature selections, respectively. This means that the actual performance increase is 4.2%, 13.9%, 2.6% for the complete, Pearson correlation coefficient and Fisher-10 feature selections, respectively. The average actual performance increase is 6.9%. This means that making use of the most beneficial ensemble will be worthwhile in this case.

5.9.4 Support Vector Machine

The ensemble composition containing all algorithms and the ensemble composition containing random forest, XGBoost and support vector machine achieved the highest prediction accuracy over the three feature selections. This does not mean that any ensemble making use of these compositions will always outperform the others and the individual algorithm. It simply means that we could create an ensemble with one of these compositions while taking into account a specific importance distribution that would achieve the highest prediction accuracy for at least one of these feature selections. To find out more about the ensembles that performed best for all feature selections, we will have to take a closer look at the three categories. We will start with the bigger ensembles containing all algorithms which can be very beneficial when resources are not an issue. The best performing ensembles in this category outperformed the individual algorithm two of the three feature selection. The ensembles made use of several importance distributions. Namely, one where random forest and XGBoost were very dominant, one where random forest and XGBoost were somewhat dominant and one where random forest and support vector machine were somewhat dominant. Next to that, the best ensemble achieved accuracies of 53.4%, 54.1% and 53.8% for the complete, Pearson correlation coefficient and Fisher-12 feature selections, respectively. When resources are somewhat more limited, a smaller ensemble could also be beneficial. In the category containing only ensembles using three algorithms, the best ensembles were the ones using the composition containing random forest, XGBoost and support vector machine. The most important about these ensembles is that the XGBoost algorithm is dominant in all of them. The best ensemble in this category achieved accuracies of 53.6%, 54.0% and 53.7% for the complete, Pearson correlation coefficient and Fisher-12 feature selections, respectively. In the category containing only ensembles using two algorithms, the best ensembles were the ones using the composition containing support vector machine and XGBoost. They made use of an importance distribution where XGBoost was quite dominant. The best ensemble in this category achieved accuracies of 53.3%, 54.1% and 53.7% for the complete, Pearson correlation coefficient and Fisher-12 feature selections, respectively.

Our goal was to find out how to use ensembles to improve the prediction accuracy. For us to achieve this goal, we needed to find ensembles that outperformed the individual algorithm which achieved accuracies of 53.1%, 53.5% and 54.0% for the complete, Pearson correlation coefficient and Fisher-12 feature selections, respectively. The best ensembles of each category strongly outperformed the individual algorithm for the complete and Pearson correlation coefficient feature selections. But they got outperformed of the individual algorithm for the Fisher-12 feature selection. Even though they did get outperformed for one feature selection, they strongly outperformed the individual algorithm for the other two feature selections. This means that we found a way to use ensembles to increase the prediction accuracy. The next step is to take a look at what subset of the promising algorithms and which importance distribution can be most beneficial to an ensemble.When comparing the best ensembles of each category with each other, we see that the ensemble using two algorithms performed slightly worse than the other two when taking into account all feature selections. The other two performed equally well but when also taking into account the amount of resources used, the most beneficial ensemble is the one using the composition containing random forest, XGBoost and support vector machine while making use of an importance distribution where XGBoost was dominant.

Now we have established that ensembles can be used to increase accuracy of a model that predicts the outcome of a football match and what the most beneficial ensembles should look like, we can find out whether the use of these particular ensembles is worthwhile or that the investment is not worth it. To find that out we need to take a look at the actual performance increase. The actual performance increase is the increase in prediction accuracy of the ensemble minus the increase in prediction accuracy of the individual algorithm divided by the increase in prediction accuracy of the individual algorithm. As mentioned before, to find out the increase in prediction accuracy, we need to take into account the ratio of the class that is present the most. This class represents the home team winning and the home win percentage is 45.9%. The individual algorithm is responsible for an increase of 7.2%, 7.6%, and 8.1% for the complete, Pearson correlation coefficient and Fisher-12 feature selections, respectively. The best ensemble is responsible for an increase of 7.7%, 8.1% and 7.8% for the complete, Pearson correlation coefficient and Fisher-12 feature selections, respectively. This means that the actual performance increase is 6.9%, 6.6%, -3.7% for the complete, Pearson correlation coefficient and Fisher-12 feature selections, respectively. The average actual performance increase is 3.3%. This means that making use of the most beneficial ensemble will not be worthwhile in this case.

5.10 Key Takeaways

The goal of this chapter was to find out how to make use of ensembles to improve the prediction accuracy of a model that predicts the outcome of a football match. We were interested in what subset of the promising algorithms could be most beneficial and what the importance distribution should be between these algorithms. We looked into these curiosities by comparing each promising algorithm with the relevant ensembles separately. The detailed discussion regarding this can be found in Section 5.9. We found out that the most beneficial ensemble is using the composition containing random forest, XGBoost and support vector machine, the composition containing random forest, XGBoost and logistic regression or the composition containing all algorithms, each making use of an importance distribution where XGBoost was dominant. Next to that, we found out that for every individual algorithm, there is an ensemble that realises an increase in prediction accuracy. Also, we found out that only for some of the models that make use of an individual algorithm, it is worthwhile to make use of the found ensembles. For already really well performing individual algorithms, it seems that it is not worth the resources and effort to make use of these ensembles. In other words, the actual performance increase was too small to be seen as worthwhile. In cases where resources are not limited, the minimum actual performance increase could be lower which means that in such a scenario the use of ensembles could be seen as worthwhile.

Chapter 6

Conclusions and Critical Reflection

The goal of this research was to find new ways to improve the prediction accuracy of a model that predicts the outcome of a football match. In our literature review, we found out that there has been quite some research focusing on finding new and effective feature categories or algorithms. Most of these would make use of the known effective feature categories, like match attributes, match statistics and team performance, while introducing a new feature category. Also, some of these would experiment with a new algorithm to find out whether it would have potential. Studies like [1], [2] and [3] made use of these ways to improve the prediction accuracy and have had some success over the years. [1] made use of weather as a new feature category, while [2] focused on team/player ratings and team/player values. [3] did not focus on a new feature category but used the known effective ones, while experimenting with long short-term memory. As mentioned, these ways to improve the accuracy have had their success over the years but at a certain point the pile with new feature categories and algorithms will run out. In that case, we need to find other ways to improve the prediction accuracy of a model that predicts the outcome of a football match.

In this thesis, we propose two ways to improve the prediction accuracy other than finding new feature categories or algorithms. The first way is the use of feature category combinations. A feature category combination is a combination between two feature categories. The combination of these two categories could result into a new set of features which exists next to the feature category features themselves. This means that with the same data, we tend to create more value. We do this by taking a second look at the feature categories and reason which features could be created considering the data of both feature categories. This approach could lead to new features which could lead to an improved prediction accuracy. The second way we propose to improve the prediction accuracy is the use of ensembles. Specifically, ensembles that make use of the promising algorithms that we found in our literature review. Ensembles can be used like any individual algorithm but, in this case, make use of multiple algorithms to predict the outcome. Ensembles can make use of all promising algorithms or just a subset. Also, the algorithms that are a part of such an ensemble can be equally important or the ensemble could make use of a different importance distribution. The use of different algorithms in an ensemble could lead to an improved prediction accuracy.

We chose to explore four feature category combinations that we believed would have the best chance of improving the prediction models directly. Next to that, we reasoned which feature category combination features could be useful and came up with twelve sets of feature category combination features. The chosen four feature category combinations had a positive effect on the prediction accuracy. Even though not all feature category combination feature sets had an actual performance increase of 10.0% or higher, on average the feature selections that include the feature category combination features performed 16.2% better than the feature selections that exclude the feature category combination features. This leads to the conclusion that the use of these feature category combinations can indeed be seen as worthwhile and used to increase the accuracy of a model that predicts the outcome of a football match. The feature category combination team performance and team value is the best performing combination achieving an average actual performance increase of 18.0%.

Next to the feature category combinations, we took an interest in ensembles. More specifically, what subset of the promising algorithms could be most beneficial in an ensemble and what the importance distribution should be between these algorithms. We looked into these curiosities by comparing each promising algorithm with the relevant ensembles, separately. We found out that the most beneficial ensemble was using the composition containing random forest, XGBoost and support vector machine, the composition containing random forest, XGBoost and logistic regression or the composition containing all algorithms, each making use of an importance distribution where XGBoost was dominant. Next to that, we found out that for every individual algorithm, there is an ensemble that realises an increase in prediction accuracy. Also, we found out that only for some of the models that make use of an individual algorithm, it is worthwhile to make use of the found ensembles. For already really well performing individual algorithms, it seems that it is not worth the resources and effort to make use of these ensembles. In other words, the actual performance increase was too small to be seen as worthwhile. In cases where resources are not limited, the minimum actual performance increase could be lower which means that in such a scenario the use of ensembles could be seen as worthwhile.

All in all, we can conclude that the two ways we proposed can be quite successful. These ways are not only successful in a scenario where there are unlimited resources, but can also be worthwhile when resources are more limited. Especially, the feature category combinations which performed great while simply using the same data in a more elaborate way.

The remainder of this chapter will be structured as follows: In Section 6.1, we will answer the research questions. In Section 6.2, we will describe our contributions to science. In Section 6.3 we will explain what the contributions to practice are. In Section 6.4, we will describe our limitations. Finally, in Section 6.5, we will talk about several possibilities for future research.

6.1 Research Questions

In this section, we will answer the research questions separately. We will start with answering each main research question, followed by their sub research questions.

6.1.1 RQ1: What is the state of the art in predicting the outcome of a football match using machine learning?

In the area of football prediction has been quite some research regarding the prediction of the outcome of a football match using machine learning. The research in this area experimented a lot with different combinations of algorithms and feature categories. A subset of these used algorithms/feature categories are promising but there is still a lot of potential in finding new algorithms/ensembles of different algorithms and feature selections.

What algorithms have been used to predict the outcome of a football match?

Table 2.1 contains the different algorithms used to predict the outcome of a football match.

What feature categories have been used to predict the outcome of a football match?

Table 2.2 contains the different feature categories used to predict the outcome of a football match.

What algorithms and feature categories show promising results while being used to predict the outcome of a football match?

As discussed in section 2.3, the promising algorithms are support vector machine, logistic regression, gradient boosting, random forest, and ensembles of different algorithms. And as discussed in section 2.4, the promising feature categories are match attributes, match statistics, team performance, head-to-head performance, coach ratings, player ratings, team ratings, team values, player attributes, and weather data.

6.1.2 RQ2: How can feature category combinations be used to improve the prediction accuracy when predicting the outcome of a football match using machine learning?

A feature category combination that provides enough worth can be used to increase the prediction accuracy. The chosen four feature category combinations had a positive effect on the prediction accuracy. Even though not all feature category combination feature sets had an actual performance increase of 10.0% or higher, on average the feature selections that include the feature category combination features performed 16.2% better than the feature selections that exclude the feature category combination features. This leads to the conclusion that the use of these feature category combinations can indeed be seen as worthwhile and used to increase the accuracy of a model that predicts the outcome of a football match. The feature category combination team performance and team value is the best performing combination achieving an average actual performance increase of 18.0%.

What feature categories can be combined into a new set of features?

In section 4.2, we describe the possible feature category combinations. We explained that there are many possible feature category combinations but that most feature categories do not fit well together due to not providing new information or information which directly relates to the outcome of a football match. The feature category combinations that are listed in section 4.2 have the best chance of improving the prediction models directly.

What feature category combinations can be used to improve the prediction accuracy?

We looked at four different feature category combinations, namely team performance and team rating (TPTR), past match statistics and team rating (PMSTR), team performance and team value (TPTV), and past match statistics and team value (PMSTV). In Section 4.9, the summarised performance of the feature category combinations can be found. All feature category combinations had a positive effect on the prediction accuracy.

6.1.3 RQ3: How can ensembles be used to achieve to improve the prediction accuracy when predicting the outcome of a football match using machine learning?

Ensembles that make use of the right composition and importance distribution can be used to improve the prediction accuracy. We found out that for every individual algorithm, there is an ensemble that realises an increase in prediction accuracy. Also, we found out that only for some of the models that make use of an individual algorithm, it is worthwhile to make use of the found ensembles. For already really well performing individual algorithms, it seems that it is not worth the resources and effort to make use of these ensembles. In other words, the actual performance increase was too small to be seen as worthwhile. In cases where resources are not limited, the minimum actual performance increase could be lower which means that in such a scenario the use of ensembles could be seen as worthwhile.

What of subset promising algorithms could be most beneficial to an ensemble?

The most beneficial ensemble is using the composition containing random forest, XGBoost and support vector machine, the composition containing random forest, XGBoost and logistic regression or the composition containing all algorithms.

What should the importance distribution be between these algorithms?

The importance distributions of the beneficial ensemble compositions were very much a like. Each ensemble composition made use of an importance distribution where XGBoost was dominant. In some of these importance distributions XGBoost appeared to be very dominant.

6.2 Contribution to Science

This thesis contributes to science by exploring multiple ways to improve the prediction accuracy of a model that predicts the outcome of a football match. As mentioned before, most studies tried to improve the prediction accuracy by looking for algorithms and feature categories. This approach has had some success over the years but at a certain point the pile with new feature categories and algorithms will run out. This thesis provides other ways to improve the prediction accuracy. It provides a broad comparison between the inclusion and exclusion of several feature category combinations and between individual algorithms and various ensemble compositions that make use of different importance distributions. Furthermore, it presents several worthwhile feature category combinations and ensemble compositions that make use of a successful importance distribution. It also suggests that there could be many feature category combination features in this area that have not been found yet. New and existing models could make use of these ways to increase their prediction accuracy.

6.3 Contribution to Practice

This thesis contributes to practice by potentially making it easier to predict the outcome of a football match. Football teams could make use of the ways found in this thesis to produce better predictions which they could use to their advantage. They could anticipate better when to give players rest and change their training schedule so the team is in optimal condition when they play a game that they predicted to draw or to lose. When the players are rested and in optimal condition, the team might be able to change a draw into a win or even a loss into win. There is a bigger chance of this happening when the opponent did not take the necessary precautions due to making worse predictions or none at all. Next to that, this thesis could be of use to people who lay bets on football matches. Due to only taking into account data that has been known before the start of the match, the ways found could also improve the predictions made by this group of people.

6.4 Limitations

This thesis has are several limitations that influenced the results and the conclusions drawn from them. The first one being that the feature category combinations and the feature category combination features were specifically chosen for what we believed was their great chance to improve the prediction accuracy. This means that others might not be worthwhile or even increase the prediction accuracy. In other words, these specific feature category combinations and the feature category combination features increase the prediction accuracy and can be seen as worthwhile but there is no guarantee others will do that as well. Next to that, we only looked at ensembles compositions that took into account all possible combinations between the promising algorithms. This group of ensembles might not be a proper representation of ensembles in this area. These ensembles were somewhat successful but there is a reasonable chance there are other ensembles out there which include other algorithms that are even more successful. Also, the evaluated ensemble compositions could have had more success with other importance distributions. We evaluated each ensemble composition with the importance distributions where an algorithm's weight could range from 1 to 4. There might be other importance distributions with different ranges of weight that are more successful than the ones we found. Finally, due to the nature of the random forest algorithm the prediction accuracies produced during these experiments can differ each time. This is due to the random forest algorithm using randomisation, it builds each tree in the forest using a randomly selected subset of features and a randomly selected subsample of the data. This means that the chances that all trees in one run are the same as in the other run are very slim. As a result, the produced accuracies can be somewhat different which could have resulted in more or less favorable accuracies for the feature category combinations and the ensembles. The feature category combinations could have been more or less successful due to this randomness which could result in more or less feature category combination feature sets to be worthwhile. This could also have influenced whether all feature category combinations can be seen as worthwhile. The ensembles could have also been affected by the randomness of the random forest algorithm. This could result in higher or lower prediction accuracies for the individual algorithm and the ensembles. This in turn could have influenced whether specific ensembles with respect to the individual algorithm can be seen as worthwhile.

6.5 Future Research

There are several interesting directions for future research. There could potentially be a lot more beneficial feature category combinations in this area which have not been found yet. Also, there could be other beneficial feature category combination features for these new beneficial feature category combinations and the one that we already found. Next to that, there probably are other beneficial ensembles of other sizes containing other algorithms that we haven't considered. Also, one could do more research on the importance distributions for the beneficial ensemble compositions that we found. We only took into account importance distributions where an algorithm's weight could range from 1 to 4 but the ensembles could potentially be even more beneficial using an importance distribution with a broader range of algorithm weights.

References

[1] D. Iskandaryan, F. Ramos, D. Palinggi, and S. Trilles, "The effect of weather in soccer results: An approach using machine learning techniques," *Applied Sciences (Switzerland)*, vol. 10, no. 19, pp. 1–14, 2020, cited By 2. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85092779559&doi=10.3390%2fapp10196750&partnerID=40&md5=9f261ae4f0633f81ea60b744bf1ec4e2

[2] D. Rudrapal, S. Boro, J. Srivastava, and S. Singh, "A deep learning approach to predict football match result," *Advances in Intelligent Systems and Computing*, vol. 990, pp. 93–99, 2020, cited By 1. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85077129600&doi=10.1007%2f978-981-13-8676-3_9&partnerID=40&md5=695a4945c4dfd07de39132b970eed220

[3] S. Jain, E. Tiwari, and P. Sardar, "Soccer result prediction using deep learning and neural networks," *Lecture Notes on Data Engineering and Communications Technologies*, vol. 57, pp. 697–707, 2021, cited By 1. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85101988991&doi=10.1007%2f978-981-15-9509-7_57&partnerID=40&md5=82f44d4aff5f6dcb1e594e8ccc1ae703

[4] B. Kitchenham, D. Budgen, and P. Brereton, *Evidence-Based Software Engineering and Systematic Reviews*, ser. Chapman & Hall/CRC Innovations in Software Engineering and Software Development Series. CRC Press, 2015. [Online]. Available: https://books.google.nl/books?id=bGfdCgAAQBAJ

[5] J. Stübinger, B. Mangold, and J. Knoll, "Machine learning in football betting: Prediction of match results based on player characteristics," *Applied Sciences (Switzerland)*, vol. 10, no. 1, 2020, cited By 10. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85079219903&doi=10.3390%2fapp10010046&partnerID=40&md5=dca665cffcb1eeff14bd18f0b67f0781

[6] J. Hucaljuk and A. Rakipović, "Predicting football scores using machine learning techniques," 2011, pp. 1623–1627, cited By 29. [Online]. Avail-

able: https://www.scopus.com/inward/record.uri?eid=2-s2.0-80052304370&partnerID=40&md5=166375e46b1d30a6ca2eeb51ace50a93

[7] R. Martins, A. Martins, L. Neves, L. Lima, E. Flores, and M. do Nascimento, "Exploring polynomial classifier to predict match results in football championships," *Expert Systems with Applications*, vol. 83, pp. 79–93, 2017, cited By 13. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85018636158&doi=10.1016%2fj.eswa.2017.04.040&partnerID=40&md5=8effbfd84512954f9412e19d4a06f085

[8] Y. Cho, J. Yoon, and S. Lee, "Using social network analysis and gradient boosting to develop a soccer win–lose prediction model," *Engineering Applications of Artificial Intelligence*, vol. 72, pp. 228–240, 2018, cited By 10. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85046805969&doi=10.1016%2fj.engappai.2018.04.010&partnerID=40&md5=c559cf225d04819bc97ef4047af72e5d

[9] N. Danisik, P. Lacko, and M. Farkas, "Football match prediction using players attributes," 2018, pp. 201–206, cited By 3. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85056778547&doi=10.1109%2fDISA.2018.8490613&partnerID=40&md5=f4c20a04e05e54e50533401b78b849a9

[10] J. Almulla and T. Alam, "Machine learning models reveal key performance metrics of football players to win matches in qatar stars league," *IEEE Access*, vol. 8, pp. 213 695–213 705, 2020, cited By 1. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85097186634&doi=10.1109%2fACCESS.2020.3038601&partnerID=40&md5=28f29e59610ec432de7db2ac27f8bc99

[11] A. Azeman, A. Mustapha, S. Mostafa, S. AbuSalim, M. Jubair, and M. Hassan, "Football match outcome prediction by applying three machine learning algorithms," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 1 Special Issue 1, pp. 73–77, 2020, cited By 1. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85091842473&doi=10.30534%2fijeter%2f2020%2f1181.12020&partnerID=40&md5=769c6d07239c6ebb120d668a72640099

[12] E. Tiwari, P. Sardar, and S. Jain, "Football match result prediction using neural networks and deep learning," 2020, pp. 229–231, cited By 2. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.

0-85093070911&doi=10.1109%2fICRITO48877.2020.9197811&partnerID=40&md5=fd2a231a4553500547ed0d9fa1696f69

[13] A. Azeman, A. Mustapha, N. Razali, A. Nanthaamomphong, and M. Abd Wahab, "Prediction of football matches results: Decision forest against neural networks," 2021, pp. 1032–1035, cited By 0. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85112821727&doi=10.1109%2fECTI-CON51831.2021.9454789&partnerID=40&md5=f68a9b4eade56359f8acbc9eea575b76

[14] R. Ievoli, L. Palazzo, and G. Ragozini, "On the use of passing network indicators to predict football outcomes[formula presented]," *Knowledge-Based Systems*, vol. 222, 2021, cited By 3. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85103694170&doi=10.1016%2fj.knosys.2021.106997&partnerID=40&md5=423b5de00bed5f4fd96050c1f5ca547a

[15] R. Nestoruk and G. Słowiński, "Prediction of football games results," vol. 2951, 2021, pp. 156–165, cited By 0. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85115879548&partnerID=40&md5=7edc79bd3c8a54eb59ac4901e3ac0fe0

[16] R. Baboota and H. Kaur, "Predictive analysis and modelling football results using machine learning approach for english premier league," *International Journal of Forecasting*, vol. 35, no. 2, pp. 741–755, 2019, cited By 45. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85062885097&doi=10.1016%2fj.ijforecast.2018.01.003&partnerID=40&md5=68a8f988e3266ccdc3b6b68ca971146e

[17] D. Berrar, P. Lopes, and W. Dubitzky, "Incorporating domain knowledge in machine learning for soccer outcome prediction," *Machine Learning*, vol. 108, no. 1, pp. 97–126, 2019, cited By 12. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85052115326&doi=10.1007%2fs10994-018-5747-8&partnerID=40&md5=69737557b619265f9e65b4c1e387c97e

[18] S. Khan and V. Kirubanand, "Comparing machine learning and ensemble learning in the field of football," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 5, pp. 4321–4325, 2019, cited By 4. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85067626731&doi=10.11591%2fijece.v9i5.pp4321.4325&partnerID=40&md5=534e827473864fadb972f19708e4682e

[19] M. Raju, M. Mia, M. Sayed, and M. Riaz Uddin, "Predicting the outcome of english premier league matches using machine learning," 2020, cited By 0. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85101661863&doi=10.1109%2fSTI50764.2020.9350327&partnerID=40&md5=57123be71d30c59b73b19c0a26d01823

[20] Y. Geurkink, J. Boone, S. Verstockt, and J. G. Bourgois, "Machine learning-based identification of the strongest predictive variables of winning and losing in belgian professional soccer," *Applied Sciences*, vol. 11, no. 5, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/5/2378

[21] F. Goes, M. Kempe, J. Van Norel, and K. Lemmink, "Modelling team performance in soccer using tactical features derived from position tracking data," *IMA Journal of Management Mathematics*, vol. 32, no. 4, pp. 519–533, 2021, cited By 3. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85111919144&doi=10.1093%2fimaman%2fdpab006&partnerID=40&md5=773ddfcd532c837e97ac534aebaea2ab

[22] A. Ranjan, V. Kumar, D. Malhotra, R. Jain, and P. Nagrath, "Predicting the result of english premier league matches," *Lecture Notes in Networks and Systems*, vol. 203 LNNS, pp. 435–446, 2021, cited By 0. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85111338927&doi=10.1007%2f978-981-16-0733-2_30&partnerID=40&md5=ae88663069b0bbdeb61b561dfaefd188

[23] M. Saritha and R. Milton, "A probabilistic logic approach to outcome prediction in team games using historical data and domain knowledge," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 5, pp. 5205–5214, 2021, cited By 0. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85085142665&doi=10.1007%2fs12652-020-01989-x&partnerID=40&md5=8322df4f0b40c8f13df2205fe15c936d

[24] A. Joseph, N. Fenton, and M. Neil, "Predicting football results using bayesian nets and other machine learning techniques," *Knowledge-Based Systems*, vol. 19, no. 7, pp. 544–553, 2006, cited By 74. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-33749652616&doi=10.1016%2fj.knosys.2006.04.011&partnerID=40&md5=bef2279c2481075d66b75d345f90ed9e

[25] G. Bilek and E. Ulas, "Predicting match outcome according to the quality of opponent in the english premier league using situational variables and team performance indicators," *International Journal of Performance

Analysis in Sport, vol. 19, no. 6, pp. 930–941, 2019, cited By 6. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85074564019&doi=10.1080%2f24748668.2019.1684773&partnerID=40&md5=e2f42f05dc39b97bb4daa0861f37ddd1

[26] A. Saricaoğlu, A. Aksoy, and T. Kaya, "Prediction of turkish super league match results using supervised machine learning techniques," *Advances in Intelligent Systems and Computing*, vol. 1029, pp. 273–280, 2020, cited By 1. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85069539581&doi=10.1007%2f978-3-030-23756-1_34&partnerID=40&md5=47c802f544fac6d36d8831e0e8ef25dd

[27] H. Xu, "Prediction on bundesliga games based on decision tree algorithm," 2021, pp. 234–238, cited By 0. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85104459153&doi=10.1109%2fICBAIE52039.2021.9389986&partnerID=40&md5=144f5be6630103d0ff985a9a6b04746f

[28] J. Domingues, B. Lopes, P. Mihaylova, and P. Georgieva, "Incremental learning for football match outcomes prediction," in *Pattern Recognition and Image Analysis*, A. Morales, J. Fierrez, J. S. Sánchez, and B. Ribeiro, Eds. Cham: Springer International Publishing, 2019, pp. 217–228.

[29] P. Robberechts and J. Davis, "Forecasting the fifa world cup – combining result- and goal-based team ability parameters," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11330 LNAI, pp. 16–30, 2019, cited By 2. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85064897352&doi=10.1007%2f978-3-030-17274-9_2&partnerID=40&md5=c0effdefac91a840892493ab6a48c5ae

[30] L. Galli, G. Galvan, T. Levato, C. Liti, V. Piccialli, and M. Sciandrone, "Football: Discovering elapsing-time bias in the science of success," *Chaos, Solitons and Fractals*, vol. 152, 2021, cited By 0. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85114379033&doi=10.1016%2fj.chaos.2021.111370&partnerID=40&md5=b5b36096b2c5836cb585305be70f77e4

[31] M. Rahman, A. Mustapha, R. Fauzi, and N. Razali, "Bayesian approach to classification of football match outcome," *International Journal of Integrated Engineering*, vol. 10, no. 6, pp. 155–158, 2018, cited By 2. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85059308430&partnerID=40&md5=7699b0c5288908ecdaf63f2bb34c6e32

[32] Q. Zhang, H. Xu, L. Zhou, and L. Wei, “Prediction of football match results based on model fusion,” vol. Part F148152, 2019, pp. 198–202, cited By 1. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85066496505&doi=10.1145%2f3319921.3319969&partnerID=40&md5=5cedac7ebeec5f7ad27b8cae1880ca28

[33] P. Pugsee and P. Pattawong, “Football match result prediction using the random forest classifier,” 2019, pp. 154–158, cited By 1. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85074871509&doi=10.1145%2f3358528.3358593&partnerID=40&md5=6abddc9ff4a512195c694ae7167371a9

[34] R. Musa, A. P. P. Abdul Majeed, M. Mohd Razman, and M. Shaharudin, “Match outcomes prediction of six top english premier league clubs via machine learning technique,” *Communications in Computer and Information Science*, vol. 1015, pp. 236–244, 2019, cited By 2. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85065102748&doi=10.1007%2f978-981-13-7780-8_20&partnerID=40&md5=5487e0a5b58dae27626f2229328b7af8

[35] S. Kusmakar, S. Shelyag, Y. Zhu, D. Dwyer, P. Gastin, and M. Angelova, “Machine learning enabled team performance analysis in the dynamical environment of soccer,” *IEEE Access*, vol. 8, pp. 90 266–90 279, 2020, cited By 3. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85085559808&doi=10.1109%2fACCESS.2020.2992025&partnerID=40&md5=13563f094ed522defc154e3aa7c5b6f5

[36] C. Miranda-Peña, H. Ceballos, L. Hervert-Escobar, and M. Gonzalez-Mendoza, “Predicting soccer results through sentiment analysis: A graph theory approach,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12747 LNCS, pp. 422–435, 2021, cited By 0. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85111372831&doi=10.1007%2f978-3-030-77980-1_32&partnerID=40&md5=54ab0d023b3d53104818c66983ceb060

[37] N. Razali, A. Mustapha, F. Yatim, and R. Ab Aziz, “Predicting football matches results using bayesian networks for english premier league (epl),” vol. 226, no. 1, 2017, cited By 13. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85028688247&doi=10.1088%2f1757-899X%2f226%2f1%2f012099&partnerID=40&md5=bf815aa260156264a50ec36333ad281a

[38] L. Hervert-Escobar, T. Matis, and N. Hernandez-Gress, “Prediction learning model for soccer matches outcomes,” 2018, pp. 63–69, cited By 0. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85092068838&doi=10.1109%2fMICAI46078.2018.00018&partnerID=40&md5=aade60ec43a7af06cc667cde5756781c

[39] A. Tsokos, S. Narayanan, I. Kosmidis, G. Baio, M. Cucuringu, G. Whitaker, and F. Király, “Modeling outcomes of soccer matches,” *Mach. Learn.*, vol. 108, no. 1, p. 77–95, jan 2019. [Online]. Available: https://doi.org/10.1007/s10994-018-5741-1

Appendix A

Examples Datasets

A.1 Football-Data

Div	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR	HTHG	HTAG	HTR
E0	17/08/2002	Blackburn	Sunderland	0	0	D	0	0	D
E0	17/08/2002	Charlton	Chelsea	2	3	A	2	1	H
E0	17/08/2002	Everton	Tottenham	2	2	D	1	0	H
E0	17/08/2002	Fulham	Bolton	4	1	H	3	1	H
E0	17/08/2002	Leeds	Man City	3	0	H	2	0	H

Referee	HS	AS	HST	AST	HF	AF	HC	AC	HY	AY	HR	AR	B365H	B365D
D Elleray	15	7	5	3	14	11	9	1	1	2	0	0	1.727	3.25
G Barber	5	21	5	12	10	12	3	6	0	3	1	0	2.8	3.25
N Barry	13	10	9	5	18	4	10	5	1	1	0	0	2.25	3.25
A Wiley	13	3	6	1	16	12	7	4	1	2	0	0	1.727	3.25
G Poll	13	18	8	10	13	13	2	7	1	1	0	0	1.667	3.4

B365A	GBH	GBD	GBA	IWH	IWD	IWA	LBH	LBD	LBA	SOH	SOD	SOA
4.333	1.6	3.4	4.5	1.8	3.1	3.8	1.615	3.25	5	1.83	3.4	3.75
2.2	2.96	3.15	2.25	2.9	3	2.2	2.8	3.2	2.2	2.75	3.25	2.25
2.75	2.39	3.11	2.81	2.3	3	2.7	2.25	3.2	2.75	2.38	3.2	2.63
4.333	1.83	3.26	4.05	1.8	3.1	3.8	1.833	3.2	3.75	1.91	3.25	3.5
G Poll	13	18	8	10	13	13	2	7	1	1	0	0

SBH	SBD	SBA	WHH	WHD	WHA	GB>2.5	GB<2.5	B365>2.5	B365<2.5
1.727	3.2	5	1.66	3.3	4.5	1.82	1.82		
3	3.25	2.2	2.75	3.1	2.3	1.87	1.77		
2.4	3.1	2.8	2.3	3.1	2.75	1.92	1.72		
1.8	3.2	4.333	1.72	3.2	4.33	1.82	1.82		
1.727	3.4	4.5	1.66	3.3	4.5	1.72	1.92		

A.2 Team Ratings

Team	ATT	MID	DEF	OVR
Liverpool	86	83	80	85
Man City	85	86	83	85
Tottenham	89	82	81	82
Chelsea	82	84	82	82
Man United	83	82	82	82
Arsenal	83	79	79	80
Leicester	80	81	79	80
Everton	81	79	80	79
Wolves	77	81	78	79
West Ham	79	79	77	78
Aston Villa	77	77	77	77
Crystal Palace	77	75	75	76
Southampton	77	77	76	76
Leeds	78	76	75	76
Newcastle	77	75	74	76
Burnley	75	76	77	76
Fulham	74	75	74	75
Brighton	75	76	76	75
West Brom	74	74	72	73
Sheffield United	71	74	73	73

A.3 Team Values

Team	Value
Chelsea	686.7
Man City	606.7
Tottenham	543.65
Arsenal	542
Man United	531.45
Liverpool	425.65
Everton	311.25
West Ham	273.05
Southampton	263.75
Crystal Palace	228.43
Leicester	218.65
Stoke	186
Watford	175.65
Swansea	157.1
Hull	135.85
Sunderland	132
Bournemouth	131.6
Middlesbrough	128.8
West Brom	123.2
Burnley	107.18

Appendix B

Random Forest and Ensembles

B.1 Ensembles using Two Algorithms

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS
RF	1	0	0	0	0.527	0.538	0.536	E-25	3	0	1	0	0.528	0.538	0.535
E-1	1	1	0	0	0.533	0.54	0.537	E-26	3	0	2	0	0.526	0.537	0.536
E-2	1	2	0	0	0.533	0.541	0.536	E-27	3	0	3	0	0.529	0.536	0.537
E-3	1	3	0	0	0.533	0.541	0.536	E-28	3	0	4	0	0.531	0.531	0.534
E-4	1	4	0	0	0.533	0.541	0.536	E-29	4	0	1	0	0.527	0.539	0.535
E-5	2	1	0	0	0.529	0.538	0.535	E-30	4	0	2	0	0.529	0.539	0.536
E-6	2	2	0	0	0.531	0.54	0.536	E-31	4	0	3	0	0.526	0.538	0.534
E-7	2	3	0	0	0.533	0.541	0.536	E-32	4	0	4	0	0.529	0.536	0.538
E-8	2	4	0	0	0.533	0.541	0.536	E-33	1	0	0	1	0.531	0.537	0.539
E-9	3	1	0	0	0.528	0.538	0.536	E-34	1	0	0	2	0.531	0.535	0.539
E-10	3	2	0	0	0.527	0.538	0.536	E-35	1	0	0	3	0.531	0.535	0.539
E-11	3	3	0	0	0.532	0.54	0.536	E-36	1	0	0	4	0.531	0.535	0.539
E-12	3	4	0	0	0.533	0.541	0.536	E-37	2	0	0	1	0.527	0.538	0.536
E-13	4	1	0	0	0.527	0.538	0.535	E-38	2	0	0	2	0.531	0.538	0.539
E-14	4	2	0	0	0.528	0.537	0.535	E-39	2	0	0	3	0.531	0.535	0.539
E-15	4	3	0	0	0.527	0.538	0.535	E-40	2	0	0	4	0.531	0.535	0.539
E-16	4	4	0	0	0.533	0.54	0.537	E-41	3	0	0	1	0.528	0.538	0.536
E-17	1	0	1	0	0.53	0.536	0.537	E-42	3	0	0	2	0.526	0.537	0.536
E-18	1	0	2	0	0.531	0.531	0.534	E-43	3	0	0	3	0.531	0.537	0.539
E-19	1	0	3	0	0.531	0.531	0.534	E-44	3	0	0	4	0.531	0.535	0.539
E-20	1	0	4	0	0.531	0.531	0.534	E-45	4	0	0	1	0.527	0.538	0.536
E-21	2	0	1	0	0.528	0.538	0.535	E-46	4	0	0	2	0.529	0.538	0.535
E-22	2	0	2	0	0.53	0.536	0.538	E-47	4	0	0	3	0.528	0.538	0.537
E-23	2	0	3	0	0.531	0.531	0.534	E-48	4	0	0	4	0.531	0.536	0.539
E-24	2	0	4	0	0.531	0.531	0.534								

B.2 Ensembles using Three Algorithms

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS
RF	1	0	0	0	0.527	0.538	0.536	E-89	3	3	1	0	0.532	0.54	0.537
E-49	1	1	1	0	0.532	0.54	0.535	E-90	3	3	2	0	0.531	0.54	0.536
E-50	1	1	2	0	0.532	0.536	0.537	E-91	3	3	3	0	0.531	0.54	0.536
E-51	1	1	3	0	0.531	0.531	0.534	E-92	3	3	4	0	0.531	0.539	0.536
E-52	1	1	4	0	0.531	0.531	0.534	E-93	3	4	1	0	0.534	0.54	0.536
E-53	1	2	1	0	0.535	0.541	0.536	E-94	3	4	2	0	0.533	0.541	0.536
E-54	1	2	2	0	0.532	0.54	0.536	E-95	3	4	3	0	0.531	0.54	0.537
E-55	1	2	3	0	0.531	0.537	0.537	E-96	3	4	4	0	0.534	0.54	0.536
E-56	1	2	4	0	0.531	0.531	0.534	E-97	4	1	1	0	0.529	0.539	0.537
E-57	1	3	1	0	0.533	0.541	0.536	E-98	4	1	2	0	0.527	0.539	0.536
E-58	1	3	2	0	0.534	0.54	0.536	E-99	4	1	3	0	0.529	0.54	0.537
E-59	1	3	3	0	0.533	0.54	0.535	E-100	4	1	4	0	0.531	0.54	0.535
E-60	1	3	4	0	0.532	0.537	0.537	E-101	4	2	1	0	0.528	0.539	0.535
E-61	1	4	1	0	0.533	0.541	0.536	E-102	4	2	2	0	0.529	0.539	0.536
E-62	1	4	2	0	0.533	0.541	0.536	E-103	4	2	3	0	0.531	0.54	0.535
E-63	1	4	3	0	0.535	0.54	0.536	E-104	4	2	4	0	0.531	0.539	0.535
E-64	1	4	4	0	0.533	0.54	0.536	E-105	4	3	1	0	0.531	0.539	0.537
E-65	2	1	1	0	0.529	0.54	0.535	E-106	4	3	2	0	0.531	0.54	0.535
E-66	2	1	2	0	0.53	0.54	0.536	E-107	4	3	3	0	0.53	0.54	0.536
E-67	2	1	3	0	0.533	0.537	0.537	E-108	4	3	4	0	0.531	0.54	0.536
E-68	2	1	4	0	0.531	0.531	0.534	E-109	4	4	1	0	0.533	0.541	0.537
E-69	2	2	1	0	0.532	0.541	0.536	E-110	4	4	2	0	0.533	0.541	0.535
E-70	2	2	2	0	0.531	0.54	0.536	E-111	4	4	3	0	0.531	0.541	0.536
E-71	2	2	3	0	0.532	0.54	0.536	E-112	4	4	4	0	0.532	0.54	0.535
E-72	2	2	4	0	0.531	0.537	0.537	E-113	1	1	0	1	0.533	0.54	0.535
E-73	2	3	1	0	0.534	0.541	0.536	E-114	1	1	0	2	0.533	0.537	0.539
E-74	2	3	2	0	0.532	0.54	0.537	E-115	1	1	0	3	0.531	0.535	0.539
E-75	2	3	3	0	0.532	0.54	0.536	E-116	1	1	0	4	0.531	0.535	0.539
E-76	2	3	4	0	0.533	0.54	0.536	E-117	1	2	0	1	0.534	0.54	0.537
E-77	2	4	1	0	0.533	0.541	0.536	E-118	1	2	0	2	0.533	0.54	0.537
E-78	2	4	2	0	0.535	0.541	0.536	E-119	1	2	0	3	0.533	0.537	0.538
E-79	2	4	3	0	0.531	0.541	0.536	E-120	1	2	0	4	0.531	0.535	0.539
E-80	2	4	4	0	0.532	0.541	0.536	E-121	1	3	0	1	0.533	0.541	0.536
E-81	3	1	1	0	0.527	0.538	0.536	E-122	1	3	0	2	0.535	0.541	0.536
E-82	3	1	2	0	0.529	0.539	0.536	E-123	1	3	0	3	0.533	0.54	0.536
E-83	3	1	3	0	0.533	0.54	0.535	E-124	1	3	0	4	0.533	0.537	0.538
E-84	3	1	4	0	0.533	0.536	0.537	E-125	1	4	0	1	0.533	0.541	0.536
E-85	3	2	1	0	0.529	0.54	0.537	E-126	1	4	0	2	0.533	0.541	0.536
E-86	3	2	2	0	0.531	0.54	0.535	E-127	1	4	0	3	0.535	0.541	0.535
E-87	3	2	3	0	0.531	0.54	0.536	E-128	1	4	0	4	0.534	0.54	0.536
E-88	3	2	4	0	0.532	0.54	0.535	E-129	2	1	0	1	0.531	0.539	0.536

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS
RF	1	0	0	0	0.527	0.538	0.536	E-170	4	3	0	2	0.534	0.54	0.536
E-130	2	1	0	2	0.534	0.54	0.536	E-171	4	3	0	3	0.532	0.54	0.536
E-131	2	1	0	3	0.533	0.537	0.538	E-172	4	3	0	4	0.534	0.54	0.536
E-132	2	1	0	4	0.531	0.535	0.539	E-173	4	4	0	1	0.534	0.54	0.535
E-133	2	2	0	1	0.533	0.541	0.535	E-174	4	4	0	2	0.534	0.54	0.536
E-134	2	2	0	2	0.533	0.54	0.536	E-175	4	4	0	3	0.533	0.541	0.536
E-135	2	2	0	3	0.533	0.54	0.537	E-176	4	4	0	4	0.533	0.54	0.536
E-136	2	2	0	4	0.532	0.537	0.538	E-177	1	0	1	1	0.53	0.535	0.54
E-137	2	3	0	1	0.535	0.54	0.536	E-178	1	0	1	2	0.532	0.535	0.54
E-138	2	3	0	2	0.534	0.54	0.536	E-179	1	0	1	3	0.531	0.535	0.539
E-139	2	3	0	3	0.533	0.54	0.535	E-180	1	0	1	4	0.531	0.535	0.539
E-140	2	3	0	4	0.534	0.54	0.537	E-181	1	0	2	1	0.531	0.535	0.539
E-141	2	4	0	1	0.533	0.541	0.536	E-182	1	0	2	2	0.531	0.535	0.54
E-142	2	4	0	2	0.535	0.54	0.536	E-183	1	0	2	3	0.532	0.535	0.54
E-143	2	4	0	3	0.534	0.54	0.535	E-184	1	0	2	4	0.531	0.535	0.539
E-144	2	4	0	4	0.533	0.54	0.536	E-185	1	0	3	1	0.531	0.531	0.534
E-145	3	1	0	1	0.528	0.537	0.535	E-186	1	0	3	2	0.531	0.534	0.539
E-146	3	1	0	2	0.531	0.54	0.536	E-187	1	0	3	3	0.532	0.535	0.539
E-147	3	1	0	3	0.534	0.539	0.536	E-188	1	0	3	4	0.532	0.535	0.54
E-148	3	1	0	4	0.533	0.537	0.538	E-189	1	0	4	1	0.531	0.531	0.534
E-149	3	2	0	1	0.53	0.54	0.536	E-190	1	0	4	2	0.531	0.531	0.534
E-150	3	2	0	2	0.533	0.54	0.536	E-191	1	0	4	3	0.53	0.535	0.538
E-151	3	2	0	3	0.533	0.54	0.536	E-192	1	0	4	4	0.532	0.535	0.539
E-152	3	2	0	4	0.533	0.54	0.536	E-193	2	0	1	1	0.531	0.536	0.538
E-153	3	3	0	1	0.533	0.541	0.535	E-194	2	0	1	2	0.531	0.535	0.539
E-154	3	3	0	2	0.533	0.541	0.536	E-195	2	0	1	3	0.532	0.535	0.54
E-155	3	3	0	3	0.534	0.541	0.536	E-196	2	0	1	4	0.531	0.535	0.539
E-156	3	3	0	4	0.533	0.539	0.537	E-197	2	0	2	1	0.531	0.535	0.54
E-157	3	4	0	1	0.534	0.54	0.535	E-198	2	0	2	2	0.532	0.535	0.539
E-158	3	4	0	2	0.534	0.54	0.535	E-199	2	0	2	3	0.531	0.535	0.54
E-159	3	4	0	3	0.534	0.54	0.535	E-200	2	0	2	4	0.531	0.535	0.54
E-160	3	4	0	4	0.534	0.54	0.535	E-201	2	0	3	1	0.531	0.535	0.539
E-161	4	1	0	1	0.527	0.538	0.535	E-202	2	0	3	2	0.532	0.535	0.54
E-162	4	1	0	2	0.527	0.539	0.535	E-203	2	0	3	3	0.531	0.535	0.539
E-163	4	1	0	3	0.532	0.54	0.536	E-204	2	0	3	4	0.531	0.535	0.54
E-164	4	1	0	4	0.533	0.54	0.535	E-205	2	0	4	1	0.531	0.531	0.534
E-165	4	2	0	1	0.528	0.538	0.534	E-206	2	0	4	2	0.531	0.535	0.538
E-166	4	2	0	2	0.53	0.54	0.537	E-207	2	0	4	3	0.532	0.535	0.539
E-167	4	2	0	3	0.533	0.54	0.536	E-208	2	0	4	4	0.531	0.535	0.539
E-168	4	2	0	4	0.532	0.54	0.535	E-209	3	0	1	1	0.528	0.538	0.536
E-169	4	3	0	1	0.531	0.54	0.536	E-210	3	0	1	2	0.531	0.536	0.539

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS
RF	1	0	0	0	0.527	0.538	0.536	E-226	4	0	1	2	0.53	0.54	0.536
E-211	3	0	1	3	0.532	0.535	0.54	E-227	4	0	1	3	0.531	0.537	0.538
E-212	3	0	1	4	0.532	0.535	0.54	E-228	4	0	1	4	0.531	0.535	0.54
E-213	3	0	2	1	0.53	0.536	0.537	E-229	4	0	2	1	0.53	0.539	0.537
E-214	3	0	2	2	0.532	0.535	0.54	E-230	4	0	2	2	0.531	0.537	0.538
E-215	3	0	2	3	0.531	0.535	0.539	E-231	4	0	2	3	0.531	0.535	0.539
E-216	3	0	2	4	0.531	0.535	0.539	E-232	4	0	2	4	0.531	0.536	0.539
E-217	3	0	3	1	0.532	0.535	0.539	E-233	4	0	3	1	0.531	0.536	0.537
E-218	3	0	3	2	0.532	0.535	0.538	E-234	4	0	3	2	0.532	0.535	0.539
E-219	3	0	3	3	0.532	0.535	0.539	E-235	4	0	3	3	0.531	0.535	0.539
E-220	3	0	3	4	0.531	0.535	0.54	E-236	4	0	3	4	0.531	0.535	0.54
E-221	3	0	4	1	0.531	0.535	0.539	E-237	4	0	4	1	0.53	0.535	0.539
E-222	3	0	4	2	0.531	0.535	0.539	E-238	4	0	4	2	0.532	0.535	0.539
E-223	3	0	4	3	0.531	0.536	0.54	E-239	4	0	4	3	0.531	0.535	0.539
E-224	3	0	4	4	0.531	0.535	0.54	E-240	4	0	4	4	0.531	0.535	0.539
E-225	4	0	1	1	0.528	0.538	0.536								

B.3 Ensembles using Four Algorithms

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS
RF	1	0	0	0	0.527	0.538	0.536	E-281	1	3	3	1	0.534	0.537	0.538
E-241	1	1	1	1	0.534	0.537	0.538	E-282	1	3	3	2	0.532	0.535	0.539
E-242	1	1	1	2	0.532	0.535	0.539	E-283	1	3	3	3	0.532	0.535	0.538
E-243	1	1	1	3	0.531	0.535	0.54	E-284	1	3	3	4	0.531	0.535	0.538
E-244	1	1	1	4	0.531	0.535	0.539	E-285	1	3	4	1	0.53	0.535	0.539
E-245	1	1	2	1	0.531	0.535	0.539	E-286	1	3	4	2	0.532	0.535	0.539
E-246	1	1	2	2	0.533	0.535	0.539	E-287	1	3	4	3	0.531	0.535	0.539
E-247	1	1	2	3	0.533	0.535	0.539	E-288	1	3	4	4	0.531	0.535	0.538
E-248	1	1	2	4	0.531	0.535	0.54	E-289	1	4	1	1	0.533	0.541	0.536
E-249	1	1	3	1	0.53	0.534	0.538	E-290	1	4	1	2	0.534	0.54	0.535
E-250	1	1	3	2	0.53	0.535	0.539	E-291	1	4	1	3	0.532	0.54	0.535
E-251	1	1	3	3	0.532	0.535	0.538	E-292	1	4	1	4	0.534	0.537	0.537
E-252	1	1	3	4	0.532	0.535	0.539	E-293	1	4	2	1	0.534	0.54	0.536
E-253	1	1	4	1	0.531	0.531	0.534	E-294	1	4	2	2	0.533	0.541	0.534
E-254	1	1	4	2	0.531	0.534	0.539	E-295	1	4	2	3	0.534	0.537	0.537
E-255	1	1	4	3	0.53	0.535	0.539	E-296	1	4	2	4	0.532	0.535	0.538
E-256	1	1	4	4	0.533	0.535	0.539	E-297	1	4	3	1	0.533	0.54	0.535
E-257	1	2	1	1	0.534	0.54	0.534	E-298	1	4	3	2	0.533	0.537	0.537
E-258	1	2	1	2	0.534	0.537	0.537	E-299	1	4	3	3	0.532	0.535	0.538
E-259	1	2	1	3	0.532	0.536	0.539	E-300	1	4	3	4	0.532	0.535	0.538
E-260	1	2	1	4	0.531	0.535	0.54	E-301	1	4	4	1	0.534	0.537	0.538
E-261	1	2	2	1	0.534	0.537	0.537	E-302	1	4	4	2	0.532	0.535	0.539
E-262	1	2	2	2	0.532	0.535	0.538	E-303	1	4	4	3	0.532	0.535	0.539
E-263	1	2	2	3	0.532	0.535	0.538	E-304	1	4	4	4	0.532	0.535	0.538
E-264	1	2	2	4	0.532	0.535	0.538	E-305	2	1	1	1	0.533	0.54	0.535
E-265	1	2	3	1	0.53	0.535	0.539	E-306	2	1	1	2	0.535	0.537	0.538
E-266	1	2	3	2	0.532	0.535	0.539	E-307	2	1	1	3	0.532	0.535	0.54
E-267	1	2	3	3	0.531	0.535	0.538	E-308	2	1	1	4	0.531	0.535	0.54
E-268	1	2	3	4	0.532	0.535	0.539	E-309	2	1	2	1	0.535	0.537	0.538
E-269	1	2	4	1	0.531	0.535	0.538	E-310	2	1	2	2	0.532	0.535	0.539
E-270	1	2	4	2	0.531	0.535	0.539	E-311	2	1	2	3	0.533	0.535	0.54
E-271	1	2	4	3	0.531	0.535	0.539	E-312	2	1	2	4	0.532	0.535	0.539
E-272	1	2	4	4	0.531	0.535	0.538	E-313	2	1	3	1	0.531	0.535	0.539
E-273	1	3	1	1	0.534	0.54	0.536	E-314	2	1	3	2	0.531	0.535	0.539
E-274	1	3	1	2	0.533	0.54	0.534	E-315	2	1	3	3	0.532	0.535	0.538
E-275	1	3	1	3	0.534	0.537	0.538	E-316	2	1	3	4	0.532	0.536	0.54
E-276	1	3	1	4	0.532	0.535	0.539	E-317	2	1	4	1	0.53	0.535	0.538
E-277	1	3	2	1	0.532	0.54	0.535	E-318	2	1	4	2	0.53	0.535	0.539
E-278	1	3	2	2	0.533	0.537	0.537	E-319	2	1	4	3	0.531	0.535	0.539
E-279	1	3	2	3	0.532	0.535	0.538	E-320	2	1	4	4	0.533	0.535	0.538
E-280	1	3	2	4	0.532	0.535	0.538	E-321	2	2	1	1	0.534	0.541	0.535

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS
RF	1	0	0	0	0.527	0.538	0.536	E-362	2	4	3	2	0.532	0.54	0.535
E-322	2	2	1	2	0.534	0.54	0.535	E-363	2	4	3	3	0.533	0.537	0.537
E-323	2	2	1	3	0.534	0.537	0.538	E-364	2	4	3	4	0.532	0.535	0.538
E-324	2	2	1	4	0.532	0.535	0.539	E-365	2	4	4	1	0.533	0.54	0.536
E-325	2	2	2	1	0.531	0.54	0.535	E-366	2	4	4	2	0.534	0.537	0.537
E-326	2	2	2	2	0.534	0.537	0.538	E-367	2	4	4	3	0.532	0.535	0.538
E-327	2	2	2	3	0.534	0.535	0.539	E-368	2	4	4	4	0.532	0.535	0.538
E-328	2	2	2	4	0.533	0.535	0.539	E-369	3	1	1	1	0.529	0.54	0.536
E-329	2	2	3	1	0.532	0.537	0.538	E-370	3	1	1	2	0.534	0.54	0.536
E-330	2	2	3	2	0.531	0.535	0.538	E-371	3	1	1	3	0.535	0.537	0.538
E-331	2	2	3	3	0.533	0.536	0.538	E-372	3	1	1	4	0.532	0.535	0.539
E-332	2	2	3	4	0.533	0.535	0.539	E-373	3	1	2	1	0.531	0.54	0.535
E-333	2	2	4	1	0.531	0.535	0.539	E-374	3	1	2	2	0.534	0.537	0.538
E-334	2	2	4	2	0.53	0.535	0.539	E-375	3	1	2	3	0.532	0.535	0.54
E-335	2	2	4	3	0.53	0.535	0.538	E-376	3	1	2	4	0.533	0.536	0.539
E-336	2	2	4	4	0.533	0.536	0.539	E-377	3	1	3	1	0.532	0.537	0.537
E-337	2	3	1	1	0.534	0.54	0.534	E-378	3	1	3	2	0.532	0.535	0.539
E-338	2	3	1	2	0.533	0.54	0.534	E-379	3	1	3	3	0.532	0.535	0.539
E-339	2	3	1	3	0.533	0.54	0.535	E-380	3	1	3	4	0.532	0.535	0.539
E-340	2	3	1	4	0.534	0.537	0.538	E-381	3	1	4	1	0.531	0.535	0.539
E-341	2	3	2	1	0.533	0.541	0.535	E-382	3	1	4	2	0.531	0.535	0.539
E-342	2	3	2	2	0.532	0.54	0.536	E-383	3	1	4	3	0.532	0.535	0.539
E-343	2	3	2	3	0.534	0.537	0.538	E-384	3	1	4	4	0.532	0.535	0.538
E-344	2	3	2	4	0.533	0.535	0.538	E-385	3	2	1	1	0.532	0.54	0.536
E-345	2	3	3	1	0.531	0.54	0.536	E-386	3	2	1	2	0.534	0.54	0.536
E-346	2	3	3	2	0.534	0.537	0.537	E-387	3	2	1	3	0.534	0.54	0.536
E-347	2	3	3	3	0.532	0.535	0.538	E-388	3	2	1	4	0.534	0.537	0.538
E-348	2	3	3	4	0.533	0.535	0.539	E-389	3	2	2	1	0.533	0.54	0.537
E-349	2	3	4	1	0.533	0.537	0.537	E-390	3	2	2	2	0.533	0.54	0.536
E-350	2	3	4	2	0.531	0.535	0.539	E-391	3	2	2	3	0.535	0.537	0.537
E-351	2	3	4	3	0.533	0.535	0.539	E-392	3	2	2	4	0.533	0.535	0.539
E-352	2	3	4	4	0.532	0.535	0.538	E-393	3	2	3	1	0.531	0.54	0.536
E-353	2	4	1	1	0.534	0.54	0.536	E-394	3	2	3	2	0.532	0.537	0.538
E-354	2	4	1	2	0.534	0.54	0.535	E-395	3	2	3	3	0.532	0.535	0.539
E-355	2	4	1	3	0.533	0.54	0.534	E-396	3	2	3	4	0.533	0.535	0.54
E-356	2	4	1	4	0.533	0.54	0.536	E-397	3	2	4	1	0.532	0.537	0.537
E-357	2	4	2	1	0.534	0.541	0.536	E-398	3	2	4	2	0.531	0.535	0.539
E-358	2	4	2	2	0.533	0.54	0.535	E-399	3	2	4	3	0.532	0.535	0.539
E-359	2	4	2	3	0.533	0.54	0.535	E-400	3	2	4	4	0.533	0.535	0.539
E-360	2	4	2	4	0.533	0.537	0.537	E-401	3	3	1	1	0.534	0.541	0.536
E-361	2	4	3	1	0.533	0.541	0.536	E-402	3	3	1	2	0.533	0.54	0.537

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F14FS
RF	1	0	0	0	0.527	0.538	0.536	E-450	4	2	1	2	0.531	0.54	0.536
E-403	3	3	1	3	0.533	0.54	0.536	E-451	4	2	1	3	0.534	0.54	0.536
E-404	3	3	1	4	0.534	0.54	0.536	E-452	4	2	1	4	0.533	0.539	0.536
E-405	3	3	2	1	0.531	0.54	0.535	E-453	4	2	2	1	0.532	0.54	0.536
E-406	3	3	2	2	0.532	0.54	0.536	E-454	4	2	2	2	0.532	0.54	0.536
E-407	3	3	2	3	0.534	0.541	0.536	E-455	4	2	2	3	0.532	0.54	0.536
E-408	3	3	2	4	0.534	0.537	0.537	E-456	4	2	2	4	0.534	0.537	0.538
E-409	3	3	3	1	0.533	0.54	0.536	E-457	4	2	3	1	0.531	0.541	0.535
E-410	3	3	3	2	0.532	0.54	0.536	E-458	4	2	3	2	0.534	0.54	0.535
E-411	3	3	3	3	0.535	0.538	0.537	E-459	4	2	3	3	0.535	0.537	0.537
E-412	3	3	3	4	0.533	0.535	0.538	E-460	4	2	3	4	0.531	0.535	0.539
E-413	3	3	4	1	0.531	0.54	0.536	E-461	4	2	4	1	0.532	0.54	0.535
E-414	3	3	4	2	0.532	0.537	0.537	E-462	4	2	4	2	0.533	0.537	0.538
E-415	3	3	4	3	0.531	0.535	0.539	E-463	4	2	4	3	0.531	0.535	0.539
E-416	3	3	4	4	0.533	0.535	0.539	E-464	4	2	4	4	0.532	0.535	0.539
E-417	3	4	1	1	0.532	0.54	0.535	E-465	4	3	1	1	0.533	0.541	0.537
E-418	3	4	1	2	0.533	0.54	0.536	E-466	4	3	1	2	0.534	0.54	0.537
E-419	3	4	1	3	0.534	0.54	0.534	E-467	4	3	1	3	0.532	0.54	0.536
E-420	3	4	1	4	0.533	0.54	0.535	E-468	4	3	1	4	0.534	0.539	0.536
E-421	3	4	2	1	0.534	0.54	0.537	E-469	4	3	2	1	0.531	0.541	0.536
E-422	3	4	2	2	0.533	0.54	0.535	E-470	4	3	2	2	0.533	0.54	0.535
E-423	3	4	2	3	0.533	0.54	0.535	E-471	4	3	2	3	0.534	0.54	0.535
E-424	3	4	2	4	0.533	0.539	0.535	E-472	4	3	2	4	0.534	0.54	0.534
E-425	3	4	3	1	0.533	0.541	0.536	E-473	4	3	3	1	0.531	0.54	0.535
E-426	3	4	3	2	0.532	0.54	0.536	E-474	4	3	3	2	0.532	0.541	0.536
E-427	3	4	3	3	0.533	0.54	0.535	E-475	4	3	3	3	0.531	0.54	0.536
E-428	3	4	3	4	0.534	0.537	0.538	E-476	4	3	3	4	0.535	0.538	0.537
E-429	3	4	4	1	0.531	0.54	0.536	E-477	4	3	4	1	0.531	0.54	0.535
E-430	3	4	4	2	0.531	0.54	0.535	E-478	4	3	4	2	0.531	0.54	0.535
E-431	3	4	4	3	0.534	0.537	0.538	E-479	4	3	4	3	0.533	0.537	0.538
E-432	3	4	4	4	0.531	0.535	0.538	E-480	4	3	4	4	0.532	0.535	0.539
E-433	4	1	1	1	0.525	0.538	0.535	E-481	4	4	1	1	0.534	0.541	0.537
E-434	4	1	1	2	0.53	0.54	0.536	E-482	4	4	1	2	0.533	0.54	0.535
E-435	4	1	1	3	0.533	0.54	0.536	E-483	4	4	1	3	0.533	0.54	0.536
E-436	4	1	1	4	0.535	0.537	0.537	E-484	4	4	1	4	0.534	0.54	0.535
E-437	4	1	2	1	0.529	0.54	0.535	E-485	4	4	2	1	0.532	0.54	0.536
E-438	4	1	2	2	0.53	0.54	0.536	E-486	4	4	2	2	0.533	0.54	0.535
E-439	4	1	2	3	0.534	0.537	0.539	E-487	4	4	2	3	0.534	0.54	0.535
E-440	4	1	2	4	0.532	0.535	0.54	E-488	4	4	2	4	0.534	0.54	0.536
E-441	4	1	3	1	0.532	0.54	0.536	E-489	4	4	3	1	0.531	0.54	0.536
E-442	4	1	3	2	0.533	0.537	0.538	E-490	4	4	3	2	0.531	0.54	0.535
E-443	4	1	3	3	0.532	0.535	0.539	E-491	4	4	3	3	0.534	0.541	0.536
E-444	4	1	3	4	0.532	0.535	0.54	E-492	4	4	3	4	0.534	0.54	0.536
E-445	4	1	4	1	0.532	0.537	0.538	E-493	4	4	4	1	0.532	0.54	0.536
E-446	4	1	4	2	0.531	0.535	0.539	E-494	4	4	4	2	0.533	0.54	0.535
E-447	4	1	4	3	0.531	0.535	0.539	E-495	4	4	4	3	0.532	0.54	0.535
E-448	4	1	4	4	0.532	0.535	0.539	E-496	4	4	4	4	0.534	0.537	0.538
E-449	4	2	1	1	0.531	0.539	0.536								

Appendix C

XGBoost and Ensembles

C.1 Ensembles using Two Algorithms

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS
XGB	0	1	0	0	0.533	0.541	0.504	E-25	0	3	1	0	0.533	0.541	0.504
E-1	1	1	0	0	0.531	0.54	0.504	E-26	0	3	2	0	0.533	0.541	0.504
E-2	2	1	0	0	0.529	0.538	0.504	E-27	0	3	3	0	0.534	0.537	0.504
E-3	3	1	0	0	0.527	0.538	0.504	E-28	0	3	4	0	0.531	0.531	0.504
E-4	4	1	0	0	0.529	0.538	0.503	E-29	0	4	1	0	0.533	0.541	0.504
E-5	1	2	0	0	0.533	0.541	0.504	E-30	0	4	2	0	0.533	0.541	0.504
E-6	2	2	0	0	0.533	0.541	0.504	E-31	0	4	3	0	0.533	0.541	0.504
E-7	3	2	0	0	0.528	0.538	0.504	E-32	0	4	4	0	0.534	0.537	0.504
E-8	4	2	0	0	0.527	0.538	0.503	E-33	0	1	0	1	0.533	0.537	0.504
E-9	1	3	0	0	0.533	0.541	0.504	E-34	0	1	0	2	0.531	0.535	0.504
E-10	2	3	0	0	0.533	0.541	0.504	E-35	0	1	0	3	0.531	0.535	0.504
E-11	3	3	0	0	0.531	0.541	0.504	E-36	0	1	0	4	0.531	0.535	0.504
E-12	4	3	0	0	0.528	0.539	0.504	E-37	0	2	0	1	0.533	0.541	0.504
E-13	1	4	0	0	0.533	0.541	0.504	E-38	0	2	0	2	0.533	0.537	0.504
E-14	2	4	0	0	0.533	0.541	0.504	E-39	0	2	0	3	0.531	0.535	0.504
E-15	3	4	0	0	0.533	0.541	0.504	E-40	0	2	0	4	0.531	0.535	0.504
E-16	4	4	0	0	0.533	0.54	0.504	E-41	0	3	0	1	0.533	0.541	0.504
E-17	0	1	1	0	0.534	0.537	0.504	E-42	0	3	0	2	0.533	0.541	0.504
E-18	0	1	2	0	0.531	0.531	0.504	E-43	0	3	0	3	0.533	0.537	0.504
E-19	0	1	3	0	0.531	0.531	0.504	E-44	0	3	0	4	0.531	0.535	0.504
E-20	0	1	4	0	0.531	0.531	0.504	E-45	0	4	0	1	0.533	0.541	0.504
E-21	0	2	1	0	0.533	0.541	0.504	E-46	0	4	0	2	0.533	0.541	0.504
E-22	0	2	2	0	0.534	0.537	0.504	E-47	0	4	0	3	0.533	0.541	0.504
E-23	0	2	3	0	0.531	0.531	0.504	E-48	0	4	0	4	0.533	0.537	0.504
E-24	0	2	4	0	0.531	0.531	0.504								

C.2 Ensembles using Three Algorithms

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS
XGB	0	1	0	0	0.533	0.541	0.504	E-89	3	3	1	0	0.532	0.54	0.504
E-49	1	1	1	0	0.531	0.54	0.504	E-90	3	3	2	0	0.532	0.54	0.504
E-50	1	1	2	0	0.532	0.536	0.504	E-91	3	3	3	0	0.533	0.54	0.504
E-51	1	1	3	0	0.531	0.531	0.504	E-92	3	3	4	0	0.532	0.54	0.504
E-52	1	1	4	0	0.531	0.531	0.504	E-93	4	3	1	0	0.529	0.538	0.504
E-53	2	1	1	0	0.529	0.539	0.504	E-94	4	3	2	0	0.531	0.541	0.504
E-54	2	1	2	0	0.532	0.54	0.504	E-95	4	3	3	0	0.532	0.54	0.504
E-55	2	1	3	0	0.533	0.537	0.504	E-96	4	3	4	0	0.531	0.54	0.504
E-56	2	1	4	0	0.531	0.531	0.504	E-97	1	4	1	0	0.533	0.541	0.504
E-57	3	1	1	0	0.527	0.538	0.504	E-98	1	4	2	0	0.533	0.541	0.504
E-58	3	1	2	0	0.529	0.539	0.504	E-99	1	4	3	0	0.536	0.54	0.504
E-59	3	1	3	0	0.531	0.54	0.504	E-100	1	4	4	0	0.532	0.54	0.504
E-60	3	1	4	0	0.533	0.536	0.504	E-101	2	4	1	0	0.533	0.541	0.504
E-61	4	1	1	0	0.528	0.539	0.504	E-102	2	4	2	0	0.534	0.54	0.504
E-62	4	1	2	0	0.528	0.537	0.504	E-103	2	4	3	0	0.533	0.541	0.504
E-63	4	1	3	0	0.529	0.539	0.504	E-104	2	4	4	0	0.533	0.54	0.504
E-64	4	1	4	0	0.529	0.54	0.504	E-105	3	4	1	0	0.534	0.54	0.504
E-65	1	2	1	0	0.535	0.54	0.504	E-106	3	4	2	0	0.532	0.541	0.504
E-66	1	2	2	0	0.532	0.54	0.504	E-107	3	4	3	0	0.533	0.54	0.504
E-67	1	2	3	0	0.532	0.537	0.504	E-108	3	4	4	0	0.533	0.54	0.504
E-68	1	2	4	0	0.531	0.531	0.504	E-109	4	4	1	0	0.531	0.541	0.504
E-69	2	2	1	0	0.531	0.54	0.504	E-110	4	4	2	0	0.531	0.541	0.504
E-70	2	2	2	0	0.534	0.54	0.504	E-111	4	4	3	0	0.532	0.54	0.504
E-71	2	2	3	0	0.531	0.54	0.504	E-112	4	4	4	0	0.531	0.54	0.504
E-72	2	2	4	0	0.532	0.537	0.504	E-113	1	1	0	1	0.532	0.54	0.504
E-73	3	2	1	0	0.529	0.54	0.504	E-114	1	1	0	2	0.532	0.537	0.504
E-74	3	2	2	0	0.532	0.54	0.504	E-115	1	1	0	3	0.531	0.535	0.504
E-75	3	2	3	0	0.531	0.54	0.504	E-116	1	1	0	4	0.531	0.535	0.504
E-76	3	2	4	0	0.531	0.54	0.504	E-117	2	1	0	1	0.531	0.54	0.504
E-77	4	2	1	0	0.528	0.537	0.504	E-118	2	1	0	2	0.533	0.54	0.504
E-78	4	2	2	0	0.529	0.539	0.504	E-119	2	1	0	3	0.533	0.537	0.504
E-79	4	2	3	0	0.53	0.54	0.504	E-120	2	1	0	4	0.531	0.535	0.504
E-80	4	2	4	0	0.532	0.54	0.504	E-121	3	1	0	1	0.528	0.539	0.504
E-81	1	3	1	0	0.533	0.541	0.504	E-122	3	1	0	2	0.532	0.54	0.504
E-82	1	3	2	0	0.534	0.54	0.504	E-123	3	1	0	3	0.532	0.539	0.504
E-83	1	3	3	0	0.531	0.54	0.504	E-124	3	1	0	4	0.533	0.537	0.504
E-84	1	3	4	0	0.532	0.537	0.504	E-125	4	1	0	1	0.529	0.538	0.504
E-85	2	3	1	0	0.534	0.54	0.504	E-126	4	1	0	2	0.528	0.539	0.504
E-86	2	3	2	0	0.534	0.54	0.504	E-127	4	1	0	3	0.531	0.539	0.504
E-87	2	3	3	0	0.532	0.54	0.504	E-128	4	1	0	4	0.533	0.54	0.504
E-88	2	3	4	0	0.532	0.54	0.504	E-129	1	2	0	1	0.534	0.54	0.504

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS
XGB	0	1	0	0	0.533	0.541	0.504	E-170	3	4	0	2	0.534	0.54	0.504
E-130	1	2	0	2	0.533	0.54	0.504	E-171	3	4	0	3	0.535	0.541	0.504
E-131	1	2	0	3	0.534	0.537	0.504	E-172	3	4	0	4	0.534	0.54	0.504
E-132	1	2	0	4	0.531	0.535	0.504	E-173	4	4	0	1	0.534	0.54	0.504
E-133	2	2	0	1	0.533	0.54	0.504	E-174	4	4	0	2	0.533	0.54	0.504
E-134	2	2	0	2	0.533	0.54	0.504	E-175	4	4	0	3	0.533	0.54	0.504
E-135	2	2	0	3	0.533	0.54	0.504	E-176	4	4	0	4	0.534	0.54	0.504
E-136	2	2	0	4	0.533	0.537	0.504	E-177	0	1	1	1	0.531	0.535	0.504
E-137	3	2	0	1	0.533	0.54	0.504	E-178	0	1	1	2	0.531	0.535	0.504
E-138	3	2	0	2	0.532	0.54	0.504	E-179	0	1	1	3	0.531	0.535	0.504
E-139	3	2	0	3	0.533	0.54	0.504	E-180	0	1	1	4	0.531	0.535	0.504
E-140	3	2	0	4	0.534	0.54	0.504	E-181	0	1	2	1	0.532	0.535	0.504
E-141	4	2	0	1	0.527	0.537	0.504	E-182	0	1	2	2	0.531	0.535	0.504
E-142	4	2	0	2	0.532	0.539	0.504	E-183	0	1	2	3	0.531	0.535	0.504
E-143	4	2	0	3	0.533	0.54	0.504	E-184	0	1	2	4	0.531	0.535	0.504
E-144	4	2	0	4	0.533	0.539	0.504	E-185	0	1	3	1	0.531	0.531	0.504
E-145	1	3	0	1	0.533	0.541	0.504	E-186	0	1	3	2	0.532	0.535	0.504
E-146	1	3	0	2	0.534	0.54	0.504	E-187	0	1	3	3	0.531	0.535	0.504
E-147	1	3	0	3	0.533	0.54	0.504	E-188	0	1	3	4	0.531	0.535	0.504
E-148	1	3	0	4	0.533	0.537	0.504	E-189	0	1	4	1	0.531	0.531	0.504
E-149	2	3	0	1	0.534	0.54	0.504	E-190	0	1	4	2	0.531	0.531	0.504
E-150	2	3	0	2	0.534	0.54	0.504	E-191	0	1	4	3	0.532	0.535	0.504
E-151	2	3	0	3	0.534	0.54	0.504	E-192	0	1	4	4	0.531	0.535	0.504
E-152	2	3	0	4	0.533	0.54	0.504	E-193	0	2	1	1	0.534	0.537	0.504
E-153	3	3	0	1	0.533	0.541	0.504	E-194	0	2	1	2	0.531	0.535	0.504
E-154	3	3	0	2	0.534	0.541	0.504	E-195	0	2	1	3	0.531	0.535	0.504
E-155	3	3	0	3	0.534	0.54	0.504	E-196	0	2	1	4	0.531	0.535	0.504
E-156	3	3	0	4	0.534	0.539	0.504	E-197	0	2	2	1	0.531	0.535	0.504
E-157	4	3	0	1	0.531	0.539	0.504	E-198	0	2	2	2	0.531	0.535	0.504
E-158	4	3	0	2	0.532	0.54	0.504	E-199	0	2	2	3	0.531	0.535	0.504
E-159	4	3	0	3	0.533	0.54	0.504	E-200	0	2	2	4	0.531	0.535	0.504
E-160	4	3	0	4	0.532	0.54	0.504	E-201	0	2	3	1	0.532	0.535	0.504
E-161	1	4	0	1	0.533	0.541	0.504	E-202	0	2	3	2	0.531	0.535	0.504
E-162	1	4	0	2	0.533	0.541	0.504	E-203	0	2	3	3	0.531	0.535	0.504
E-163	1	4	0	3	0.534	0.54	0.504	E-204	0	2	3	4	0.531	0.535	0.504
E-164	1	4	0	4	0.534	0.54	0.504	E-205	0	2	4	1	0.531	0.531	0.504
E-165	2	4	0	1	0.533	0.541	0.504	E-206	0	2	4	2	0.532	0.535	0.504
E-166	2	4	0	2	0.534	0.54	0.504	E-207	0	2	4	3	0.531	0.535	0.504
E-167	2	4	0	3	0.534	0.54	0.504	E-208	0	2	4	4	0.531	0.535	0.504
E-168	2	4	0	4	0.534	0.54	0.504	E-209	0	3	1	1	0.533	0.541	0.504
E-169	3	4	0	1	0.534	0.54	0.504	E-210	0	3	1	2	0.534	0.537	0.504

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS
XGB	0	1	0	0	0.533	0.541	0.504	E-226	0	4	1	2	0.533	0.541	0.504
E-211	0	3	1	3	0.531	0.535	0.504	E-227	0	4	1	3	0.534	0.537	0.504
E-212	0	3	1	4	0.531	0.535	0.504	E-228	0	4	1	4	0.531	0.535	0.504
E-213	0	3	2	1	0.534	0.537	0.504	E-229	0	4	2	1	0.533	0.541	0.504
E-214	0	3	2	2	0.532	0.535	0.504	E-230	0	4	2	2	0.534	0.537	0.504
E-215	0	3	2	3	0.531	0.535	0.504	E-231	0	4	2	3	0.532	0.535	0.504
E-216	0	3	2	4	0.531	0.535	0.504	E-232	0	4	2	4	0.531	0.535	0.504
E-217	0	3	3	1	0.531	0.535	0.504	E-233	0	4	3	1	0.534	0.537	0.504
E-218	0	3	3	2	0.531	0.535	0.504	E-234	0	4	3	2	0.532	0.535	0.504
E-219	0	3	3	3	0.531	0.535	0.504	E-235	0	4	3	3	0.532	0.535	0.504
E-220	0	3	3	4	0.531	0.535	0.504	E-236	0	4	3	4	0.531	0.535	0.504
E-221	0	3	4	1	0.532	0.535	0.504	E-237	0	4	4	1	0.531	0.535	0.504
E-222	0	3	4	2	0.531	0.535	0.504	E-238	0	4	4	2	0.531	0.535	0.504
E-223	0	3	4	3	0.531	0.535	0.504	E-239	0	4	4	3	0.531	0.535	0.504
E-224	0	3	4	4	0.531	0.535	0.504	E-240	0	4	4	4	0.531	0.535	0.504
E-225	0	4	1	1	0.533	0.541	0.504								

C.3 Ensembles using Four Algorithms

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS
XGB	0	1	0	0	0.533	0.541	0.504	E-281	3	1	3	1	0.533	0.537	0.504
E-241	1	1	1	1	0.534	0.538	0.504	E-282	3	1	3	2	0.532	0.535	0.504
E-242	1	1	1	2	0.532	0.535	0.504	E-283	3	1	3	3	0.533	0.535	0.504
E-243	1	1	1	3	0.531	0.535	0.504	E-284	3	1	3	4	0.531	0.535	0.504
E-244	1	1	1	4	0.531	0.535	0.504	E-285	3	1	4	1	0.53	0.535	0.504
E-245	1	1	2	1	0.531	0.535	0.504	E-286	3	1	4	2	0.53	0.536	0.504
E-246	1	1	2	2	0.533	0.536	0.504	E-287	3	1	4	3	0.532	0.535	0.504
E-247	1	1	2	3	0.532	0.535	0.504	E-288	3	1	4	4	0.531	0.535	0.504
E-248	1	1	2	4	0.531	0.535	0.504	E-289	4	1	1	1	0.528	0.538	0.504
E-249	1	1	3	1	0.531	0.534	0.504	E-290	4	1	1	2	0.529	0.54	0.504
E-250	1	1	3	2	0.53	0.535	0.504	E-291	4	1	1	3	0.533	0.54	0.504
E-251	1	1	3	3	0.532	0.536	0.504	E-292	4	1	1	4	0.534	0.537	0.504
E-252	1	1	3	4	0.532	0.535	0.504	E-293	4	1	2	1	0.528	0.539	0.504
E-253	1	1	4	1	0.531	0.531	0.504	E-294	4	1	2	2	0.533	0.54	0.504
E-254	1	1	4	2	0.53	0.534	0.504	E-295	4	1	2	3	0.534	0.537	0.504
E-255	1	1	4	3	0.531	0.535	0.504	E-296	4	1	2	4	0.533	0.535	0.504
E-256	1	1	4	4	0.533	0.535	0.504	E-297	4	1	3	1	0.533	0.54	0.504
E-257	2	1	1	1	0.532	0.54	0.504	E-298	4	1	3	2	0.534	0.537	0.504
E-258	2	1	1	2	0.535	0.537	0.504	E-299	4	1	3	3	0.533	0.535	0.504
E-259	2	1	1	3	0.532	0.535	0.504	E-300	4	1	3	4	0.532	0.535	0.504
E-260	2	1	1	4	0.531	0.535	0.504	E-301	4	1	4	1	0.533	0.537	0.504
E-261	2	1	2	1	0.533	0.537	0.504	E-302	4	1	4	2	0.532	0.535	0.504
E-262	2	1	2	2	0.533	0.535	0.504	E-303	4	1	4	3	0.531	0.535	0.504
E-263	2	1	2	3	0.532	0.536	0.504	E-304	4	1	4	4	0.532	0.535	0.504
E-264	2	1	2	4	0.532	0.535	0.504	E-305	1	2	1	1	0.533	0.54	0.504
E-265	2	1	3	1	0.53	0.535	0.504	E-306	1	2	1	2	0.533	0.537	0.504
E-266	2	1	3	2	0.531	0.535	0.504	E-307	1	2	1	3	0.532	0.535	0.504
E-267	2	1	3	3	0.531	0.535	0.504	E-308	1	2	1	4	0.531	0.535	0.504
E-268	2	1	3	4	0.533	0.535	0.504	E-309	1	2	2	1	0.533	0.537	0.504
E-269	2	1	4	1	0.531	0.535	0.504	E-310	1	2	2	2	0.532	0.535	0.504
E-270	2	1	4	2	0.531	0.535	0.504	E-311	1	2	2	3	0.532	0.535	0.504
E-271	2	1	4	3	0.531	0.535	0.504	E-312	1	2	2	4	0.532	0.535	0.504
E-272	2	1	4	4	0.531	0.535	0.504	E-313	1	2	3	1	0.531	0.535	0.504
E-273	3	1	1	1	0.531	0.539	0.504	E-314	1	2	3	2	0.531	0.535	0.504
E-274	3	1	1	2	0.533	0.54	0.504	E-315	1	2	3	3	0.531	0.535	0.504
E-275	3	1	1	3	0.534	0.538	0.504	E-316	1	2	3	4	0.532	0.535	0.504
E-276	3	1	1	4	0.533	0.535	0.504	E-317	1	2	4	1	0.53	0.534	0.504
E-277	3	1	2	1	0.534	0.54	0.504	E-318	1	2	4	2	0.531	0.535	0.504
E-278	3	1	2	2	0.533	0.537	0.504	E-319	1	2	4	3	0.532	0.535	0.504
E-279	3	1	2	3	0.532	0.535	0.504	E-320	1	2	4	4	0.531	0.535	0.504
E-280	3	1	2	4	0.533	0.535	0.504	E-321	2	2	1	1	0.533	0.541	0.504

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS
XGB	0	1	0	0	0.533	0.541	0.504	E-362	4	2	3	2	0.532	0.54	0.504
E-322	2	2	1	2	0.534	0.54	0.504	E-363	4	2	3	3	0.534	0.537	0.504
E-323	2	2	1	3	0.534	0.537	0.504	E-364	4	2	3	4	0.532	0.535	0.504
E-324	2	2	1	4	0.532	0.535	0.504	E-365	4	2	4	1	0.531	0.54	0.504
E-325	2	2	2	1	0.531	0.539	0.504	E-366	4	2	4	2	0.533	0.537	0.504
E-326	2	2	2	2	0.534	0.537	0.504	E-367	4	2	4	3	0.534	0.535	0.504
E-327	2	2	2	3	0.533	0.535	0.504	E-368	4	2	4	4	0.531	0.535	0.504
E-328	2	2	2	4	0.532	0.535	0.504	E-369	1	3	1	1	0.534	0.54	0.504
E-329	2	2	3	1	0.532	0.537	0.504	E-370	1	3	1	2	0.533	0.54	0.504
E-330	2	2	3	2	0.531	0.535	0.504	E-371	1	3	1	3	0.533	0.537	0.504
E-331	2	2	3	3	0.533	0.536	0.504	E-372	1	3	1	4	0.532	0.535	0.504
E-332	2	2	3	4	0.532	0.535	0.504	E-373	1	3	2	1	0.533	0.54	0.504
E-333	2	2	4	1	0.531	0.535	0.504	E-374	1	3	2	2	0.534	0.537	0.504
E-334	2	2	4	2	0.53	0.535	0.504	E-375	1	3	2	3	0.532	0.535	0.504
E-335	2	2	4	3	0.53	0.535	0.504	E-376	1	3	2	4	0.531	0.535	0.504
E-336	2	2	4	4	0.532	0.536	0.504	E-377	1	3	3	1	0.534	0.537	0.504
E-337	3	2	1	1	0.532	0.54	0.504	E-378	1	3	3	2	0.532	0.535	0.504
E-338	3	2	1	2	0.534	0.54	0.504	E-379	1	3	3	3	0.532	0.535	0.504
E-339	3	2	1	3	0.533	0.539	0.504	E-380	1	3	3	4	0.531	0.535	0.504
E-340	3	2	1	4	0.535	0.537	0.504	E-381	1	3	4	1	0.531	0.535	0.504
E-341	3	2	2	1	0.532	0.54	0.504	E-382	1	3	4	2	0.532	0.535	0.504
E-342	3	2	2	2	0.532	0.54	0.504	E-383	1	3	4	3	0.531	0.535	0.504
E-343	3	2	2	3	0.535	0.537	0.504	E-384	1	3	4	4	0.531	0.535	0.504
E-344	3	2	2	4	0.533	0.535	0.504	E-385	2	3	1	1	0.533	0.54	0.504
E-345	3	2	3	1	0.533	0.54	0.504	E-386	2	3	1	2	0.533	0.539	0.504
E-346	3	2	3	2	0.532	0.537	0.504	E-387	2	3	1	3	0.533	0.541	0.504
E-347	3	2	3	3	0.532	0.535	0.504	E-388	2	3	1	4	0.534	0.537	0.504
E-348	3	2	3	4	0.533	0.536	0.504	E-389	2	3	2	1	0.534	0.54	0.504
E-349	3	2	4	1	0.533	0.537	0.504	E-390	2	3	2	2	0.532	0.54	0.504
E-350	3	2	4	2	0.532	0.535	0.504	E-391	2	3	2	3	0.534	0.537	0.504
E-351	3	2	4	3	0.531	0.535	0.504	E-392	2	3	2	4	0.533	0.535	0.504
E-352	3	2	4	4	0.533	0.535	0.504	E-393	2	3	3	1	0.532	0.54	0.504
E-353	4	2	1	1	0.529	0.54	0.504	E-394	2	3	3	2	0.534	0.537	0.504
E-354	4	2	1	2	0.532	0.54	0.504	E-395	2	3	3	3	0.532	0.535	0.504
E-355	4	2	1	3	0.534	0.54	0.504	E-396	2	3	3	4	0.533	0.535	0.504
E-356	4	2	1	4	0.533	0.54	0.504	E-397	2	3	4	1	0.533	0.537	0.504
E-357	4	2	2	1	0.532	0.539	0.504	E-398	2	3	4	2	0.531	0.535	0.504
E-358	4	2	2	2	0.532	0.54	0.504	E-399	2	3	4	3	0.532	0.535	0.504
E-359	4	2	2	3	0.533	0.54	0.504	E-400	2	3	4	4	0.531	0.535	0.504
E-360	4	2	2	4	0.534	0.537	0.504	E-401	3	3	1	1	0.533	0.54	0.504
E-361	4	2	3	1	0.531	0.54	0.504	E-402	3	3	1	2	0.534	0.54	0.504

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F1FS
XGB	0	1	0	0	0.533	0.541	0.504	E-450	2	4	1	2	0.534	0.54	0.504
E-403	3	3	1	3	0.533	0.54	0.504	E-451	2	4	1	3	0.534	0.54	0.504
E-404	3	3	1	4	0.533	0.54	0.504	E-452	2	4	1	4	0.534	0.54	0.504
E-405	3	3	2	1	0.532	0.54	0.504	E-453	2	4	2	1	0.534	0.54	0.504
E-406	3	3	2	2	0.533	0.54	0.504	E-454	2	4	2	2	0.533	0.541	0.504
E-407	3	3	2	3	0.534	0.54	0.504	E-455	2	4	2	3	0.532	0.54	0.504
E-408	3	3	2	4	0.534	0.537	0.504	E-456	2	4	2	4	0.534	0.537	0.504
E-409	3	3	3	1	0.531	0.54	0.504	E-457	2	4	3	1	0.533	0.54	0.504
E-410	3	3	3	2	0.531	0.54	0.504	E-458	2	4	3	2	0.533	0.54	0.504
E-411	3	3	3	3	0.534	0.537	0.504	E-459	2	4	3	3	0.533	0.537	0.504
E-412	3	3	3	4	0.533	0.536	0.504	E-460	2	4	3	4	0.532	0.535	0.504
E-413	3	3	4	1	0.531	0.54	0.504	E-461	2	4	4	1	0.531	0.54	0.504
E-414	3	3	4	2	0.532	0.537	0.504	E-462	2	4	4	2	0.535	0.537	0.504
E-415	3	3	4	3	0.531	0.535	0.504	E-463	2	4	4	3	0.532	0.535	0.504
E-416	3	3	4	4	0.533	0.536	0.504	E-464	2	4	4	4	0.531	0.535	0.504
E-417	4	3	1	1	0.533	0.539	0.504	E-465	3	4	1	1	0.534	0.54	0.504
E-418	4	3	1	2	0.534	0.54	0.504	E-466	3	4	1	2	0.533	0.54	0.504
E-419	4	3	1	3	0.534	0.54	0.504	E-467	3	4	1	3	0.534	0.54	0.504
E-420	4	3	1	4	0.533	0.539	0.504	E-468	3	4	1	4	0.533	0.54	0.504
E-421	4	3	2	1	0.531	0.54	0.504	E-469	3	4	2	1	0.533	0.54	0.504
E-422	4	3	2	2	0.532	0.54	0.504	E-470	3	4	2	2	0.532	0.539	0.504
E-423	4	3	2	3	0.534	0.54	0.504	E-471	3	4	2	3	0.533	0.54	0.504
E-424	4	3	2	4	0.533	0.54	0.504	E-472	3	4	2	4	0.533	0.54	0.504
E-425	4	3	3	1	0.532	0.54	0.504	E-473	3	4	3	1	0.532	0.54	0.504
E-426	4	3	3	2	0.532	0.539	0.504	E-474	3	4	3	2	0.534	0.54	0.504
E-427	4	3	3	3	0.533	0.54	0.504	E-475	3	4	3	3	0.532	0.54	0.504
E-428	4	3	3	4	0.534	0.538	0.504	E-476	3	4	3	4	0.534	0.537	0.504
E-429	4	3	4	1	0.532	0.54	0.504	E-477	3	4	4	1	0.531	0.54	0.504
E-430	4	3	4	2	0.531	0.54	0.504	E-478	3	4	4	2	0.532	0.54	0.504
E-431	4	3	4	3	0.533	0.537	0.504	E-479	3	4	4	3	0.534	0.537	0.504
E-432	4	3	4	4	0.532	0.536	0.504	E-480	3	4	4	4	0.532	0.535	0.504
E-433	1	4	1	1	0.533	0.541	0.504	E-481	4	4	1	1	0.533	0.541	0.504
E-434	1	4	1	2	0.534	0.54	0.504	E-482	4	4	1	2	0.533	0.542	0.504
E-435	1	4	1	3	0.532	0.54	0.504	E-483	4	4	1	3	0.534	0.54	0.504
E-436	1	4	1	4	0.534	0.537	0.504	E-484	4	4	1	4	0.533	0.54	0.504
E-437	1	4	2	1	0.534	0.54	0.504	E-485	4	4	2	1	0.532	0.54	0.504
E-438	1	4	2	2	0.533	0.54	0.504	E-486	4	4	2	2	0.534	0.541	0.504
E-439	1	4	2	3	0.533	0.537	0.504	E-487	4	4	2	3	0.533	0.54	0.504
E-440	1	4	2	4	0.532	0.535	0.504	E-488	4	4	2	4	0.533	0.54	0.504
E-441	1	4	3	1	0.532	0.54	0.504	E-489	4	4	3	1	0.531	0.54	0.504
E-442	1	4	3	2	0.534	0.537	0.504	E-490	4	4	3	2	0.531	0.54	0.504
E-443	1	4	3	3	0.532	0.535	0.504	E-491	4	4	3	3	0.534	0.54	0.504
E-444	1	4	3	4	0.531	0.535	0.504	E-492	4	4	3	4	0.534	0.54	0.504
E-445	1	4	4	1	0.534	0.537	0.504	E-493	4	4	4	1	0.531	0.54	0.504
E-446	1	4	4	2	0.532	0.535	0.504	E-494	4	4	4	2	0.532	0.54	0.504
E-447	1	4	4	3	0.532	0.535	0.504	E-495	4	4	4	3	0.531	0.54	0.504
E-448	1	4	4	4	0.531	0.535	0.504	E-496	4	4	4	4	0.535	0.537	0.504
E-449	2	4	1	1	0.534	0.54	0.504								

Appendix D

Logistic Regression and Ensembles

D.1 Ensembles using Two Algorithms

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS
LR	0	0	1	0	0.531	0.531	0.535	E-25	1	0	3	0	0.531	0.531	0.535
E-1	0	1	1	0	0.534	0.537	0.536	E-26	2	0	3	0	0.531	0.531	0.535
E-2	0	2	1	0	0.533	0.541	0.536	E-27	3	0	3	0	0.53	0.536	0.537
E-3	0	3	1	0	0.533	0.541	0.536	E-28	4	0	3	0	0.528	0.538	0.534
E-4	0	4	1	0	0.533	0.541	0.536	E-29	1	0	4	0	0.531	0.531	0.535
E-5	0	1	2	0	0.531	0.531	0.535	E-30	2	0	4	0	0.531	0.531	0.535
E-6	0	2	2	0	0.534	0.537	0.536	E-31	3	0	4	0	0.531	0.531	0.535
E-7	0	3	2	0	0.533	0.541	0.536	E-32	4	0	4	0	0.529	0.535	0.537
E-8	0	4	2	0	0.533	0.541	0.536	E-33	0	0	1	1	0.531	0.535	0.538
E-9	0	1	3	0	0.531	0.531	0.535	E-34	0	0	1	2	0.531	0.535	0.537
E-10	0	2	3	0	0.531	0.531	0.535	E-35	0	0	1	3	0.531	0.535	0.537
E-11	0	3	3	0	0.534	0.537	0.536	E-36	0	0	1	4	0.531	0.535	0.537
E-12	0	4	3	0	0.533	0.541	0.536	E-37	0	0	2	1	0.531	0.531	0.535
E-13	0	1	4	0	0.531	0.531	0.535	E-38	0	0	2	2	0.531	0.535	0.538
E-14	0	2	4	0	0.531	0.531	0.535	E-39	0	0	2	3	0.531	0.535	0.537
E-15	0	3	4	0	0.531	0.531	0.535	E-40	0	0	2	4	0.531	0.535	0.537
E-16	0	4	4	0	0.534	0.537	0.536	E-41	0	0	3	1	0.531	0.531	0.535
E-17	1	0	1	0	0.53	0.536	0.537	E-42	0	0	3	2	0.531	0.531	0.535
E-18	2	0	1	0	0.528	0.538	0.535	E-43	0	0	3	3	0.531	0.534	0.538
E-19	3	0	1	0	0.527	0.537	0.535	E-44	0	0	3	4	0.531	0.535	0.537
E-20	4	0	1	0	0.529	0.539	0.534	E-45	0	0	4	1	0.531	0.531	0.535
E-21	1	0	2	0	0.531	0.531	0.535	E-46	0	0	4	2	0.531	0.531	0.535
E-22	2	0	2	0	0.529	0.536	0.537	E-47	0	0	4	3	0.531	0.531	0.535
E-23	3	0	2	0	0.528	0.538	0.534	E-48	0	0	4	4	0.531	0.534	0.538
E-24	4	0	2	0	0.529	0.539	0.534								

D.2 Ensembles using Three Algorithms

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS
LR	0	0	1	0	0.531	0.531	0.535	E-89	1	3	3	0	0.532	0.54	0.535
E-49	1	1	1	0	0.533	0.54	0.535	E-90	2	3	3	0	0.531	0.54	0.534
E-50	2	1	1	0	0.53	0.539	0.536	E-91	3	3	3	0	0.532	0.54	0.535
E-51	3	1	1	0	0.528	0.539	0.534	E-92	4	3	3	0	0.532	0.54	0.534
E-52	4	1	1	0	0.527	0.538	0.535	E-93	1	4	3	0	0.534	0.541	0.535
E-53	1	2	1	0	0.535	0.541	0.534	E-94	2	4	3	0	0.532	0.54	0.534
E-54	2	2	1	0	0.532	0.541	0.534	E-95	3	4	3	0	0.533	0.54	0.534
E-55	3	2	1	0	0.529	0.54	0.535	E-96	4	4	3	0	0.532	0.541	0.534
E-56	4	2	1	0	0.528	0.538	0.535	E-97	1	1	4	0	0.531	0.531	0.535
E-57	1	3	1	0	0.533	0.541	0.536	E-98	2	1	4	0	0.531	0.531	0.535
E-58	2	3	1	0	0.534	0.541	0.534	E-99	3	1	4	0	0.532	0.537	0.537
E-59	3	3	1	0	0.531	0.541	0.534	E-100	4	1	4	0	0.531	0.54	0.535
E-60	4	3	1	0	0.53	0.54	0.535	E-101	1	2	4	0	0.531	0.531	0.535
E-61	1	4	1	0	0.533	0.541	0.536	E-102	2	2	4	0	0.532	0.537	0.537
E-62	2	4	1	0	0.533	0.541	0.536	E-103	3	2	4	0	0.533	0.54	0.536
E-63	3	4	1	0	0.534	0.54	0.535	E-104	4	2	4	0	0.532	0.54	0.535
E-64	4	4	1	0	0.531	0.541	0.535	E-105	1	3	4	0	0.532	0.537	0.536
E-65	1	1	2	0	0.532	0.537	0.536	E-106	2	3	4	0	0.533	0.54	0.535
E-66	2	1	2	0	0.531	0.54	0.535	E-107	3	3	4	0	0.533	0.54	0.536
E-67	3	1	2	0	0.531	0.539	0.535	E-108	4	3	4	0	0.531	0.54	0.535
E-68	4	1	2	0	0.527	0.538	0.535	E-109	1	4	4	0	0.532	0.54	0.534
E-69	1	2	2	0	0.533	0.54	0.535	E-110	2	4	4	0	0.532	0.54	0.535
E-70	2	2	2	0	0.532	0.54	0.535	E-111	3	4	4	0	0.532	0.54	0.535
E-71	3	2	2	0	0.531	0.54	0.535	E-112	4	4	4	0	0.533	0.54	0.535
E-72	4	2	2	0	0.531	0.539	0.535	E-113	0	1	1	1	0.531	0.535	0.535
E-73	1	3	2	0	0.535	0.54	0.534	E-114	0	1	1	2	0.531	0.535	0.537
E-74	2	3	2	0	0.533	0.54	0.534	E-115	0	1	1	3	0.531	0.535	0.537
E-75	3	3	2	0	0.532	0.541	0.535	E-116	0	1	1	4	0.531	0.535	0.537
E-76	4	3	2	0	0.531	0.54	0.535	E-117	0	2	1	1	0.534	0.537	0.535
E-77	1	4	2	0	0.533	0.541	0.536	E-118	0	2	1	2	0.531	0.535	0.535
E-78	2	4	2	0	0.535	0.541	0.534	E-119	0	2	1	3	0.531	0.535	0.537
E-79	3	4	2	0	0.533	0.54	0.534	E-120	0	2	1	4	0.531	0.535	0.537
E-80	4	4	2	0	0.531	0.541	0.534	E-121	0	3	1	1	0.533	0.541	0.536
E-81	1	1	3	0	0.531	0.531	0.535	E-122	0	3	1	2	0.534	0.537	0.535
E-82	2	1	3	0	0.532	0.537	0.536	E-123	0	3	1	3	0.531	0.535	0.535
E-83	3	1	3	0	0.529	0.54	0.535	E-124	0	3	1	4	0.531	0.535	0.537
E-84	4	1	3	0	0.53	0.539	0.536	E-125	0	4	1	1	0.533	0.541	0.536
E-85	1	2	3	0	0.533	0.537	0.536	E-126	0	4	1	2	0.533	0.541	0.536
E-86	2	2	3	0	0.532	0.541	0.536	E-127	0	4	1	3	0.534	0.537	0.535
E-87	3	2	3	0	0.53	0.54	0.535	E-128	0	4	1	4	0.531	0.535	0.535
E-88	4	2	3	0	0.531	0.54	0.534	E-129	0	1	2	1	0.532	0.535	0.537

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS
LR	0	0	1	0	0.531	0.531	0.535	E-170	0	3	4	2	0.531	0.535	0.536
E-130	0	1	2	2	0.531	0.535	0.535	E-171	0	3	4	3	0.531	0.535	0.536
E-131	0	1	2	3	0.531	0.535	0.537	E-172	0	3	4	4	0.531	0.535	0.535
E-132	0	1	2	4	0.531	0.535	0.537	E-173	0	4	4	1	0.531	0.535	0.536
E-133	0	2	2	1	0.531	0.535	0.536	E-174	0	4	4	2	0.531	0.535	0.536
E-134	0	2	2	2	0.531	0.535	0.535	E-175	0	4	4	3	0.531	0.535	0.536
E-135	0	2	2	3	0.531	0.535	0.535	E-176	0	4	4	4	0.531	0.535	0.535
E-136	0	2	2	4	0.531	0.535	0.537	E-177	1	0	1	1	0.531	0.535	0.536
E-137	0	3	2	1	0.534	0.537	0.535	E-178	1	0	1	2	0.532	0.535	0.537
E-138	0	3	2	2	0.532	0.535	0.536	E-179	1	0	1	3	0.531	0.535	0.537
E-139	0	3	2	3	0.531	0.535	0.535	E-180	1	0	1	4	0.531	0.535	0.537
E-140	0	3	2	4	0.531	0.535	0.535	E-181	2	0	1	1	0.531	0.537	0.535
E-141	0	4	2	1	0.533	0.541	0.536	E-182	2	0	1	2	0.531	0.535	0.536
E-142	0	4	2	2	0.534	0.537	0.535	E-183	2	0	1	3	0.532	0.535	0.537
E-143	0	4	2	3	0.532	0.535	0.536	E-184	2	0	1	4	0.531	0.535	0.537
E-144	0	4	2	4	0.531	0.535	0.535	E-185	3	0	1	1	0.528	0.539	0.534
E-145	0	1	3	1	0.531	0.531	0.535	E-186	3	0	1	2	0.529	0.537	0.535
E-146	0	1	3	2	0.532	0.535	0.537	E-187	3	0	1	3	0.531	0.535	0.536
E-147	0	1	3	3	0.531	0.535	0.535	E-188	3	0	1	4	0.532	0.535	0.537
E-148	0	1	3	4	0.531	0.535	0.537	E-189	4	0	1	1	0.528	0.539	0.534
E-149	0	2	3	1	0.532	0.535	0.537	E-190	4	0	1	2	0.528	0.538	0.534
E-150	0	2	3	2	0.531	0.535	0.536	E-191	4	0	1	3	0.53	0.537	0.536
E-151	0	2	3	3	0.531	0.535	0.535	E-192	4	0	1	4	0.531	0.535	0.536
E-152	0	2	3	4	0.531	0.535	0.535	E-193	1	0	2	1	0.53	0.535	0.537
E-153	0	3	3	1	0.531	0.535	0.536	E-194	1	0	2	2	0.532	0.535	0.536
E-154	0	3	3	2	0.531	0.535	0.536	E-195	1	0	2	3	0.532	0.535	0.537
E-155	0	3	3	3	0.531	0.535	0.535	E-196	1	0	2	4	0.531	0.535	0.537
E-156	0	3	3	4	0.531	0.535	0.535	E-197	2	0	2	1	0.53	0.535	0.536
E-157	0	4	3	1	0.534	0.537	0.535	E-198	2	0	2	2	0.531	0.535	0.536
E-158	0	4	3	2	0.532	0.535	0.536	E-199	2	0	2	3	0.532	0.535	0.536
E-159	0	4	3	3	0.532	0.535	0.536	E-200	2	0	2	4	0.531	0.535	0.537
E-160	0	4	3	4	0.531	0.535	0.535	E-201	3	0	2	1	0.531	0.537	0.535
E-161	0	1	4	1	0.531	0.531	0.535	E-202	3	0	2	2	0.532	0.535	0.535
E-162	0	1	4	2	0.531	0.531	0.535	E-203	3	0	2	3	0.532	0.535	0.536
E-163	0	1	4	3	0.532	0.535	0.537	E-204	3	0	2	4	0.532	0.535	0.537
E-164	0	1	4	4	0.531	0.535	0.535	E-205	4	0	2	1	0.528	0.538	0.536
E-165	0	2	4	1	0.531	0.531	0.535	E-206	4	0	2	2	0.531	0.537	0.536
E-166	0	2	4	2	0.532	0.535	0.537	E-207	4	0	2	3	0.533	0.535	0.535
E-167	0	2	4	3	0.531	0.535	0.536	E-208	4	0	2	4	0.531	0.535	0.536
E-168	0	2	4	4	0.531	0.535	0.535	E-209	1	0	3	1	0.531	0.531	0.535
E-169	0	3	4	1	0.532	0.535	0.537	E-210	1	0	3	2	0.531	0.535	0.537

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS
LR	0	0	1	0	0.531	0.531	0.535	E-226	1	0	4	2	0.531	0.531	0.535
E-211	1	0	3	3	0.531	0.535	0.537	E-227	1	0	4	3	0.531	0.535	0.537
E-212	1	0	3	4	0.532	0.535	0.537	E-228	1	0	4	4	0.531	0.535	0.536
E-213	2	0	3	1	0.531	0.535	0.537	E-229	2	0	4	1	0.531	0.531	0.535
E-214	2	0	3	2	0.531	0.536	0.537	E-230	2	0	4	2	0.531	0.535	0.538
E-215	2	0	3	3	0.531	0.535	0.536	E-231	2	0	4	3	0.532	0.535	0.537
E-216	2	0	3	4	0.531	0.535	0.536	E-232	2	0	4	4	0.531	0.535	0.537
E-217	3	0	3	1	0.531	0.535	0.535	E-233	3	0	4	1	0.531	0.535	0.538
E-218	3	0	3	2	0.531	0.535	0.535	E-234	3	0	4	2	0.532	0.535	0.537
E-219	3	0	3	3	0.532	0.535	0.536	E-235	3	0	4	3	0.531	0.535	0.536
E-220	3	0	3	4	0.532	0.535	0.537	E-236	3	0	4	4	0.532	0.535	0.536
E-221	4	0	3	1	0.531	0.537	0.535	E-237	4	0	4	1	0.53	0.535	0.535
E-222	4	0	3	2	0.532	0.535	0.535	E-238	4	0	4	2	0.532	0.535	0.536
E-223	4	0	3	3	0.532	0.535	0.535	E-239	4	0	4	3	0.531	0.535	0.536
E-224	4	0	3	4	0.532	0.535	0.535	E-240	4	0	4	4	0.532	0.535	0.536
E-225	1	0	4	1	0.531	0.531	0.535								

D.3 Ensembles using Four Algorithms

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS
LR	0	0	1	0	0.531	0.531	0.535	E-281	3	3	1	1	0.534	0.541	0.535
E-241	1	1	1	1	0.535	0.537	0.536	E-282	3	3	1	2	0.533	0.541	0.534
E-242	1	1	1	2	0.533	0.535	0.537	E-283	3	3	1	3	0.534	0.54	0.535
E-243	1	1	1	3	0.531	0.535	0.537	E-284	3	3	1	4	0.533	0.54	0.535
E-244	1	1	1	4	0.531	0.535	0.537	E-285	4	3	1	1	0.532	0.54	0.534
E-245	2	1	1	1	0.533	0.539	0.535	E-286	4	3	1	2	0.534	0.54	0.536
E-246	2	1	1	2	0.535	0.537	0.535	E-287	4	3	1	3	0.533	0.54	0.534
E-247	2	1	1	3	0.532	0.535	0.536	E-288	4	3	1	4	0.533	0.54	0.537
E-248	2	1	1	4	0.531	0.535	0.537	E-289	1	4	1	1	0.533	0.541	0.536
E-249	3	1	1	1	0.529	0.54	0.535	E-290	1	4	1	2	0.534	0.54	0.535
E-250	3	1	1	2	0.532	0.54	0.535	E-291	1	4	1	3	0.532	0.54	0.535
E-251	3	1	1	3	0.534	0.537	0.535	E-292	1	4	1	4	0.533	0.537	0.536
E-252	3	1	1	4	0.533	0.535	0.536	E-293	2	4	1	1	0.534	0.54	0.534
E-253	4	1	1	1	0.528	0.538	0.534	E-294	2	4	1	2	0.534	0.541	0.534
E-254	4	1	1	2	0.531	0.539	0.535	E-295	2	4	1	3	0.533	0.54	0.534
E-255	4	1	1	3	0.533	0.54	0.535	E-296	2	4	1	4	0.532	0.54	0.535
E-256	4	1	1	4	0.535	0.537	0.536	E-297	3	4	1	1	0.534	0.54	0.534
E-257	1	2	1	1	0.533	0.54	0.535	E-298	3	4	1	2	0.534	0.54	0.535
E-258	1	2	1	2	0.534	0.537	0.535	E-299	3	4	1	3	0.533	0.54	0.535
E-259	1	2	1	3	0.532	0.535	0.536	E-300	3	4	1	4	0.534	0.54	0.536
E-260	1	2	1	4	0.531	0.535	0.537	E-301	4	4	1	1	0.534	0.541	0.536
E-261	2	2	1	1	0.534	0.54	0.536	E-302	4	4	1	2	0.534	0.54	0.534
E-262	2	2	1	2	0.533	0.54	0.535	E-303	4	4	1	3	0.533	0.54	0.535
E-263	2	2	1	3	0.534	0.537	0.535	E-304	4	4	1	4	0.534	0.54	0.535
E-264	2	2	1	4	0.532	0.535	0.536	E-305	1	1	2	1	0.531	0.535	0.536
E-265	3	2	1	1	0.531	0.54	0.534	E-306	1	1	2	2	0.532	0.535	0.536
E-266	3	2	1	2	0.534	0.54	0.534	E-307	1	1	2	3	0.532	0.535	0.536
E-267	3	2	1	3	0.533	0.54	0.536	E-308	1	1	2	4	0.531	0.535	0.537
E-268	3	2	1	4	0.534	0.537	0.535	E-309	2	1	2	1	0.533	0.537	0.536
E-269	4	2	1	1	0.529	0.54	0.535	E-310	2	1	2	2	0.532	0.535	0.536
E-270	4	2	1	2	0.531	0.539	0.534	E-311	2	1	2	3	0.533	0.535	0.536
E-271	4	2	1	3	0.535	0.54	0.536	E-312	2	1	2	4	0.532	0.535	0.536
E-272	4	2	1	4	0.532	0.54	0.536	E-313	3	1	2	1	0.533	0.54	0.534
E-273	1	3	1	1	0.534	0.54	0.534	E-314	3	1	2	2	0.533	0.537	0.535
E-274	1	3	1	2	0.532	0.541	0.534	E-315	3	1	2	3	0.532	0.535	0.536
E-275	1	3	1	3	0.534	0.537	0.535	E-316	3	1	2	4	0.533	0.536	0.536
E-276	1	3	1	4	0.532	0.535	0.536	E-317	4	1	2	1	0.53	0.54	0.535
E-277	2	3	1	1	0.533	0.54	0.534	E-318	4	1	2	2	0.532	0.54	0.534
E-278	2	3	1	2	0.533	0.54	0.537	E-319	4	1	2	3	0.533	0.537	0.536
E-279	2	3	1	3	0.534	0.54	0.534	E-320	4	1	2	4	0.532	0.535	0.536
E-280	2	3	1	4	0.534	0.537	0.536	E-321	1	2	2	1	0.534	0.537	0.535

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS
LR	0	0	1	0	0.531	0.531	0.535	E-362	3	4	2	2	0.534	0.54	0.535
E-322	1	2	2	2	0.531	0.535	0.536	E-363	3	4	2	3	0.533	0.54	0.534
E-323	1	2	2	3	0.531	0.535	0.536	E-364	3	4	2	4	0.533	0.54	0.535
E-324	1	2	2	4	0.532	0.535	0.536	E-365	4	4	2	1	0.531	0.54	0.535
E-325	2	2	2	1	0.53	0.539	0.535	E-366	4	4	2	2	0.534	0.54	0.536
E-326	2	2	2	2	0.534	0.537	0.536	E-367	4	4	2	3	0.533	0.54	0.535
E-327	2	2	2	3	0.534	0.535	0.536	E-368	4	4	2	4	0.534	0.54	0.536
E-328	2	2	2	4	0.532	0.535	0.536	E-369	1	1	3	1	0.531	0.535	0.537
E-329	3	2	2	1	0.531	0.54	0.536	E-370	1	1	3	2	0.531	0.535	0.535
E-330	3	2	2	2	0.532	0.54	0.535	E-371	1	1	3	3	0.532	0.536	0.537
E-331	3	2	2	3	0.534	0.537	0.535	E-372	1	1	3	4	0.531	0.535	0.536
E-332	3	2	2	4	0.533	0.535	0.535	E-373	2	1	3	1	0.53	0.535	0.536
E-333	4	2	2	1	0.531	0.54	0.535	E-374	2	1	3	2	0.531	0.535	0.536
E-334	4	2	2	2	0.532	0.54	0.534	E-375	2	1	3	3	0.532	0.535	0.535
E-335	4	2	2	3	0.532	0.54	0.536	E-376	2	1	3	4	0.532	0.536	0.536
E-336	4	2	2	4	0.534	0.537	0.536	E-377	3	1	3	1	0.532	0.537	0.536
E-337	1	3	2	1	0.533	0.54	0.535	E-378	3	1	3	2	0.531	0.535	0.535
E-338	1	3	2	2	0.533	0.537	0.535	E-379	3	1	3	3	0.532	0.535	0.535
E-339	1	3	2	3	0.531	0.535	0.536	E-380	3	1	3	4	0.531	0.535	0.536
E-340	1	3	2	4	0.532	0.535	0.536	E-381	4	1	3	1	0.531	0.54	0.534
E-341	2	3	2	1	0.533	0.541	0.535	E-382	4	1	3	2	0.533	0.537	0.536
E-342	2	3	2	2	0.533	0.54	0.535	E-383	4	1	3	3	0.532	0.535	0.536
E-343	2	3	2	3	0.534	0.537	0.536	E-384	4	1	3	4	0.532	0.535	0.536
E-344	2	3	2	4	0.533	0.535	0.535	E-385	1	2	3	1	0.531	0.535	0.536
E-345	3	3	2	1	0.532	0.54	0.535	E-386	1	2	3	2	0.532	0.535	0.536
E-346	3	3	2	2	0.533	0.541	0.536	E-387	1	2	3	3	0.531	0.535	0.535
E-347	3	3	2	3	0.534	0.54	0.535	E-388	1	2	3	4	0.532	0.535	0.536
E-348	3	3	2	4	0.534	0.537	0.535	E-389	2	2	3	1	0.532	0.537	0.535
E-349	4	3	2	1	0.531	0.54	0.536	E-390	2	2	3	2	0.531	0.535	0.536
E-350	4	3	2	2	0.532	0.54	0.536	E-391	2	2	3	3	0.533	0.535	0.537
E-351	4	3	2	3	0.534	0.54	0.534	E-392	2	2	3	4	0.532	0.535	0.536
E-352	4	3	2	4	0.534	0.54	0.535	E-393	3	2	3	1	0.531	0.54	0.535
E-353	1	4	2	1	0.534	0.54	0.534	E-394	3	2	3	2	0.533	0.537	0.536
E-354	1	4	2	2	0.533	0.541	0.536	E-395	3	2	3	3	0.531	0.535	0.537
E-355	1	4	2	3	0.533	0.537	0.535	E-396	3	2	3	4	0.533	0.536	0.535
E-356	1	4	2	4	0.532	0.535	0.536	E-397	4	2	3	1	0.532	0.54	0.535
E-357	2	4	2	1	0.533	0.54	0.534	E-398	4	2	3	2	0.533	0.54	0.536
E-358	2	4	2	2	0.534	0.54	0.534	E-399	4	2	3	3	0.535	0.538	0.535
E-359	2	4	2	3	0.533	0.54	0.534	E-400	4	2	3	4	0.531	0.535	0.536
E-360	2	4	2	4	0.533	0.537	0.535	E-401	1	3	3	1	0.534	0.537	0.535
E-361	3	4	2	1	0.534	0.541	0.535	E-402	1	3	3	2	0.532	0.535	0.536

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F10FS
LR	0	0	1	0	0.531	0.531	0.535	E-450	1	2	4	2	0.531	0.535	0.536
E-403	1	3	3	3	0.532	0.535	0.536	E-451	1	2	4	3	0.531	0.535	0.536
E-404	1	3	3	4	0.531	0.535	0.535	E-452	1	2	4	4	0.531	0.535	0.535
E-405	2	3	3	1	0.532	0.54	0.534	E-453	2	2	4	1	0.531	0.535	0.536
E-406	2	3	3	2	0.534	0.537	0.535	E-454	2	2	4	2	0.53	0.535	0.536
E-407	2	3	3	3	0.532	0.535	0.536	E-455	2	2	4	3	0.529	0.535	0.536
E-408	2	3	3	4	0.533	0.535	0.537	E-456	2	2	4	4	0.532	0.535	0.536
E-409	3	3	3	1	0.532	0.54	0.534	E-457	3	2	4	1	0.532	0.537	0.535
E-410	3	3	3	2	0.531	0.54	0.535	E-458	3	2	4	2	0.531	0.535	0.536
E-411	3	3	3	3	0.534	0.537	0.536	E-459	3	2	4	3	0.532	0.535	0.536
E-412	3	3	3	4	0.533	0.535	0.536	E-460	3	2	4	4	0.531	0.535	0.536
E-413	4	3	3	1	0.529	0.54	0.535	E-461	4	2	4	1	0.531	0.54	0.535
E-414	4	3	3	2	0.531	0.54	0.536	E-462	4	2	4	2	0.533	0.537	0.536
E-415	4	3	3	3	0.533	0.54	0.536	E-463	4	2	4	3	0.532	0.535	0.535
E-416	4	3	3	4	0.535	0.538	0.535	E-464	4	2	4	4	0.532	0.535	0.536
E-417	1	4	3	1	0.532	0.54	0.535	E-465	1	3	4	1	0.531	0.535	0.536
E-418	1	4	3	2	0.534	0.537	0.535	E-466	1	3	4	2	0.532	0.535	0.536
E-419	1	4	3	3	0.532	0.535	0.536	E-467	1	3	4	3	0.531	0.535	0.536
E-420	1	4	3	4	0.532	0.535	0.536	E-468	1	3	4	4	0.531	0.535	0.535
E-421	2	4	3	1	0.533	0.54	0.536	E-469	2	3	4	1	0.533	0.537	0.535
E-422	2	4	3	2	0.533	0.54	0.535	E-470	2	3	4	2	0.53	0.535	0.536
E-423	2	4	3	3	0.533	0.537	0.535	E-471	2	3	4	3	0.532	0.535	0.536
E-424	2	4	3	4	0.531	0.535	0.536	E-472	2	3	4	4	0.532	0.535	0.536
E-425	3	4	3	1	0.532	0.54	0.535	E-473	3	3	4	1	0.531	0.54	0.535
E-426	3	4	3	2	0.533	0.54	0.535	E-474	3	3	4	2	0.531	0.537	0.535
E-427	3	4	3	3	0.532	0.54	0.535	E-475	3	3	4	3	0.531	0.535	0.536
E-428	3	4	3	4	0.534	0.537	0.535	E-476	3	3	4	4	0.533	0.536	0.536
E-429	4	4	3	1	0.532	0.54	0.535	E-477	4	3	4	1	0.531	0.54	0.536
E-430	4	4	3	2	0.53	0.54	0.535	E-478	4	3	4	2	0.531	0.54	0.536
E-431	4	4	3	3	0.534	0.541	0.537	E-479	4	3	4	3	0.533	0.537	0.536
E-432	4	4	3	4	0.533	0.54	0.535	E-480	4	3	4	4	0.532	0.535	0.536
E-433	1	1	4	1	0.531	0.531	0.535	E-481	1	4	4	1	0.534	0.537	0.535
E-434	1	1	4	2	0.53	0.535	0.537	E-482	1	4	4	2	0.532	0.535	0.536
E-435	1	1	4	3	0.53	0.535	0.536	E-483	1	4	4	3	0.532	0.535	0.535
E-436	1	1	4	4	0.532	0.535	0.536	E-484	1	4	4	4	0.532	0.535	0.536
E-437	2	1	4	1	0.531	0.535	0.537	E-485	2	4	4	1	0.533	0.54	0.535
E-438	2	1	4	2	0.53	0.535	0.536	E-486	2	4	4	2	0.534	0.537	0.536
E-439	2	1	4	3	0.531	0.535	0.537	E-487	2	4	4	3	0.531	0.535	0.536
E-440	2	1	4	4	0.531	0.535	0.535	E-488	2	4	4	4	0.532	0.535	0.536
E-441	3	1	4	1	0.53	0.535	0.536	E-489	3	4	4	1	0.532	0.539	0.535
E-442	3	1	4	2	0.531	0.536	0.537	E-490	3	4	4	2	0.53	0.54	0.535
E-443	3	1	4	3	0.531	0.535	0.535	E-491	3	4	4	3	0.534	0.537	0.535
E-444	3	1	4	4	0.532	0.535	0.536	E-492	3	4	4	4	0.532	0.535	0.536
E-445	4	1	4	1	0.533	0.537	0.535	E-493	4	4	4	1	0.533	0.54	0.535
E-446	4	1	4	2	0.533	0.535	0.535	E-494	4	4	4	2	0.531	0.54	0.535
E-447	4	1	4	3	0.533	0.535	0.535	E-495	4	4	4	3	0.532	0.54	0.535
E-448	4	1	4	4	0.532	0.536	0.536	E-496	4	4	4	4	0.534	0.537	0.537
E-449	1	2	4	1	0.531	0.535	0.537								

Appendix E

Support Vector Machine and Ensembles

E.1 Ensembles using Two Algorithms

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS
SVM	0	0	0	1	0.531	0.535	0.54	E-25	0	0	1	3	0.531	0.535	0.54
E-1	0	1	0	1	0.533	0.537	0.54	E-26	0	0	2	3	0.531	0.535	0.54
E-2	0	2	0	1	0.533	0.541	0.537	E-27	0	0	3	3	0.531	0.534	0.54
E-3	0	3	0	1	0.533	0.541	0.537	E-28	0	0	4	3	0.531	0.531	0.535
E-4	0	4	0	1	0.533	0.541	0.537	E-29	0	0	1	4	0.531	0.535	0.54
E-5	0	1	0	2	0.531	0.535	0.54	E-30	0	0	2	4	0.531	0.535	0.54
E-6	0	2	0	2	0.533	0.537	0.54	E-31	0	0	3	4	0.531	0.535	0.54
E-7	0	3	0	2	0.533	0.541	0.537	E-32	0	0	4	4	0.531	0.535	0.54
E-8	0	4	0	2	0.533	0.541	0.537	E-33	1	0	0	1	0.531	0.537	0.54
E-9	0	1	0	3	0.531	0.535	0.54	E-34	2	0	0	1	0.528	0.538	0.536
E-10	0	2	0	3	0.531	0.535	0.54	E-35	3	0	0	1	0.528	0.538	0.534
E-11	0	3	0	3	0.533	0.537	0.54	E-36	4	0	0	1	0.527	0.539	0.535
E-12	0	4	0	3	0.533	0.541	0.537	E-37	1	0	0	2	0.531	0.535	0.54
E-13	0	1	0	4	0.531	0.535	0.54	E-38	2	0	0	2	0.531	0.537	0.539
E-14	0	2	0	4	0.531	0.535	0.54	E-39	3	0	0	2	0.528	0.538	0.534
E-15	0	3	0	4	0.531	0.535	0.54	E-40	4	0	0	2	0.528	0.538	0.535
E-16	0	4	0	4	0.533	0.537	0.54	E-41	1	0	0	3	0.531	0.535	0.54
E-17	0	0	1	1	0.531	0.535	0.54	E-42	2	0	0	3	0.531	0.535	0.54
E-18	0	0	2	1	0.531	0.531	0.535	E-43	3	0	0	3	0.53	0.537	0.538
E-19	0	0	3	1	0.531	0.531	0.535	E-44	4	0	0	3	0.529	0.537	0.534
E-20	0	0	4	1	0.531	0.531	0.535	E-45	1	0	0	4	0.531	0.535	0.54
E-21	0	0	1	2	0.531	0.535	0.54	E-46	2	0	0	4	0.531	0.535	0.54
E-22	0	0	2	2	0.531	0.535	0.54	E-47	3	0	0	4	0.531	0.535	0.54
E-23	0	0	3	2	0.531	0.531	0.535	E-48	4	0	0	4	0.531	0.536	0.54
E-24	0	0	4	2	0.531	0.531	0.535	E-24	0	0	4	2	0.531	0.531	0.535

E.2 Ensembles using Three Algorithms

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS
SVM	0	0	0	1	0.531	0.535	0.54	E-89	0	3	1	3	0.531	0.535	0.54
E-49	0	1	1	1	0.531	0.535	0.54	E-90	0	3	2	3	0.531	0.535	0.54
E-50	0	1	2	1	0.532	0.535	0.54	E-91	0	3	3	3	0.531	0.535	0.54
E-51	0	1	3	1	0.531	0.531	0.535	E-92	0	3	4	3	0.531	0.535	0.54
E-52	0	1	4	1	0.531	0.531	0.535	E-93	0	4	1	3	0.534	0.537	0.539
E-53	0	2	1	1	0.534	0.537	0.539	E-94	0	4	2	3	0.532	0.535	0.54
E-54	0	2	2	1	0.531	0.535	0.54	E-95	0	4	3	3	0.532	0.535	0.54
E-55	0	2	3	1	0.532	0.535	0.54	E-96	0	4	4	3	0.531	0.535	0.54
E-56	0	2	4	1	0.531	0.531	0.535	E-97	0	1	1	4	0.531	0.535	0.54
E-57	0	3	1	1	0.533	0.541	0.537	E-98	0	1	2	4	0.531	0.535	0.54
E-58	0	3	2	1	0.534	0.537	0.539	E-99	0	1	3	4	0.531	0.535	0.54
E-59	0	3	3	1	0.531	0.535	0.54	E-100	0	1	4	4	0.531	0.535	0.54
E-60	0	3	4	1	0.532	0.535	0.54	E-101	0	2	1	4	0.531	0.535	0.54
E-61	0	4	1	1	0.533	0.541	0.537	E-102	0	2	2	4	0.531	0.535	0.54
E-62	0	4	2	1	0.533	0.541	0.537	E-103	0	2	3	4	0.531	0.535	0.54
E-63	0	4	3	1	0.534	0.537	0.539	E-104	0	2	4	4	0.531	0.535	0.54
E-64	0	4	4	1	0.531	0.535	0.54	E-105	0	3	1	4	0.531	0.535	0.54
E-65	0	1	1	2	0.531	0.535	0.54	E-106	0	3	2	4	0.531	0.535	0.54
E-66	0	1	2	2	0.531	0.535	0.54	E-107	0	3	3	4	0.531	0.535	0.54
E-67	0	1	3	2	0.532	0.535	0.54	E-108	0	3	4	4	0.531	0.535	0.54
E-68	0	1	4	2	0.531	0.531	0.535	E-109	0	4	1	4	0.531	0.535	0.54
E-69	0	2	1	2	0.531	0.535	0.54	E-110	0	4	2	4	0.531	0.535	0.54
E-70	0	2	2	2	0.531	0.535	0.54	E-111	0	4	3	4	0.531	0.535	0.54
E-71	0	2	3	2	0.531	0.535	0.54	E-112	0	4	4	4	0.531	0.535	0.54
E-72	0	2	4	2	0.532	0.535	0.54	E-113	1	1	0	1	0.533	0.54	0.537
E-73	0	3	1	2	0.534	0.537	0.539	E-114	2	1	0	1	0.532	0.539	0.536
E-74	0	3	2	2	0.532	0.535	0.54	E-115	3	1	0	1	0.528	0.538	0.535
E-75	0	3	3	2	0.531	0.535	0.54	E-116	4	1	0	1	0.528	0.539	0.535
E-76	0	3	4	2	0.531	0.535	0.54	E-117	1	2	0	1	0.534	0.54	0.538
E-77	0	4	1	2	0.533	0.541	0.537	E-118	2	2	0	1	0.534	0.54	0.537
E-78	0	4	2	2	0.534	0.537	0.539	E-119	3	2	0	1	0.532	0.539	0.537
E-79	0	4	3	2	0.532	0.535	0.54	E-120	4	2	0	1	0.529	0.539	0.533
E-80	0	4	4	2	0.531	0.535	0.54	E-121	1	3	0	1	0.533	0.541	0.537
E-81	0	1	1	3	0.531	0.535	0.54	E-122	2	3	0	1	0.535	0.54	0.537
E-82	0	1	2	3	0.531	0.535	0.54	E-123	3	3	0	1	0.533	0.54	0.539
E-83	0	1	3	3	0.531	0.535	0.54	E-124	4	3	0	1	0.531	0.54	0.536
E-84	0	1	4	3	0.532	0.535	0.54	E-125	1	4	0	1	0.533	0.541	0.537
E-85	0	2	1	3	0.531	0.535	0.54	E-126	2	4	0	1	0.533	0.541	0.537
E-86	0	2	2	3	0.531	0.535	0.54	E-127	3	4	0	1	0.534	0.54	0.537
E-87	0	2	3	3	0.531	0.535	0.54	E-128	4	4	0	1	0.534	0.54	0.538
E-88	0	2	4	3	0.531	0.535	0.54	E-129	1	1	0	2	0.533	0.537	0.539

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS
SVM	0	0	0	1	0.531	0.535	0.54	E-170	2	3	0	4	0.533	0.54	0.538
E-130	2	1	0	2	0.532	0.54	0.538	E-171	3	3	0	4	0.533	0.54	0.537
E-131	3	1	0	2	0.53	0.538	0.538	E-172	4	3	0	4	0.533	0.54	0.536
E-132	4	1	0	2	0.528	0.539	0.535	E-173	1	4	0	4	0.534	0.54	0.537
E-133	1	2	0	2	0.534	0.54	0.537	E-174	2	4	0	4	0.534	0.54	0.537
E-134	2	2	0	2	0.534	0.54	0.538	E-175	3	4	0	4	0.533	0.54	0.537
E-135	3	2	0	2	0.533	0.541	0.537	E-176	4	4	0	4	0.534	0.54	0.537
E-136	4	2	0	2	0.533	0.54	0.537	E-177	1	0	1	1	0.531	0.535	0.539
E-137	1	3	0	2	0.534	0.54	0.537	E-178	2	0	1	1	0.531	0.536	0.536
E-138	2	3	0	2	0.534	0.54	0.537	E-179	3	0	1	1	0.53	0.537	0.534
E-139	3	3	0	2	0.533	0.541	0.537	E-180	4	0	1	1	0.529	0.538	0.533
E-140	4	3	0	2	0.533	0.54	0.538	E-181	1	0	2	1	0.531	0.535	0.539
E-141	1	4	0	2	0.533	0.541	0.537	E-182	2	0	2	1	0.531	0.535	0.539
E-142	2	4	0	2	0.534	0.541	0.537	E-183	3	0	2	1	0.53	0.537	0.538
E-143	3	4	0	2	0.535	0.54	0.537	E-184	4	0	2	1	0.528	0.539	0.534
E-144	4	4	0	2	0.534	0.54	0.537	E-185	1	0	3	1	0.531	0.531	0.535
E-145	1	1	0	3	0.531	0.535	0.54	E-186	2	0	3	1	0.53	0.535	0.54
E-146	2	1	0	3	0.534	0.537	0.54	E-187	3	0	3	1	0.531	0.535	0.539
E-147	3	1	0	3	0.533	0.54	0.537	E-188	4	0	3	1	0.531	0.536	0.537
E-148	4	1	0	3	0.531	0.54	0.537	E-189	1	0	4	1	0.531	0.531	0.535
E-149	1	2	0	3	0.533	0.537	0.539	E-190	2	0	4	1	0.531	0.531	0.535
E-150	2	2	0	3	0.534	0.54	0.537	E-191	3	0	4	1	0.531	0.535	0.538
E-151	3	2	0	3	0.533	0.54	0.538	E-192	4	0	4	1	0.531	0.535	0.538
E-152	4	2	0	3	0.533	0.54	0.537	E-193	1	0	1	2	0.531	0.535	0.54
E-153	1	3	0	3	0.533	0.54	0.538	E-194	2	0	1	2	0.532	0.536	0.54
E-154	2	3	0	3	0.533	0.54	0.538	E-195	3	0	1	2	0.531	0.536	0.537
E-155	3	3	0	3	0.534	0.54	0.537	E-196	4	0	1	2	0.528	0.538	0.536
E-156	4	3	0	3	0.533	0.54	0.538	E-197	1	0	2	2	0.532	0.535	0.54
E-157	1	4	0	3	0.536	0.54	0.537	E-198	2	0	2	2	0.531	0.535	0.54
E-158	2	4	0	3	0.534	0.54	0.538	E-199	3	0	2	2	0.531	0.535	0.539
E-159	3	4	0	3	0.534	0.541	0.537	E-200	4	0	2	2	0.53	0.537	0.539
E-160	4	4	0	3	0.534	0.54	0.537	E-201	1	0	3	2	0.531	0.535	0.538
E-161	1	1	0	4	0.531	0.535	0.54	E-202	2	0	3	2	0.531	0.535	0.54
E-162	2	1	0	4	0.531	0.535	0.54	E-203	3	0	3	2	0.531	0.535	0.54
E-163	3	1	0	4	0.533	0.537	0.539	E-204	4	0	3	2	0.531	0.535	0.539
E-164	4	1	0	4	0.533	0.54	0.537	E-205	1	0	4	2	0.531	0.531	0.535
E-165	1	2	0	4	0.531	0.535	0.54	E-206	2	0	4	2	0.531	0.535	0.539
E-166	2	2	0	4	0.533	0.537	0.539	E-207	3	0	4	2	0.531	0.535	0.54
E-167	3	2	0	4	0.533	0.539	0.537	E-208	4	0	4	2	0.532	0.536	0.539
E-168	4	2	0	4	0.533	0.54	0.537	E-209	1	0	1	3	0.531	0.535	0.54
E-169	1	3	0	4	0.533	0.537	0.54	E-210	2	0	1	3	0.532	0.535	0.541

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS
SVM	0	0	0	1	0.531	0.535	0.54	E-226	2	0	1	4	0.531	0.535	0.54
E-211	3	0	1	3	0.531	0.535	0.541	E-227	3	0	1	4	0.532	0.535	0.54
E-212	4	0	1	3	0.53	0.536	0.538	E-228	4	0	1	4	0.532	0.535	0.54
E-213	1	0	2	3	0.532	0.535	0.54	E-229	1	0	2	4	0.531	0.535	0.54
E-214	2	0	2	3	0.531	0.535	0.54	E-230	2	0	2	4	0.532	0.535	0.541
E-215	3	0	2	3	0.532	0.535	0.539	E-231	3	0	2	4	0.531	0.535	0.539
E-216	4	0	2	3	0.531	0.535	0.539	E-232	4	0	2	4	0.531	0.535	0.54
E-217	1	0	3	3	0.53	0.535	0.54	E-233	1	0	3	4	0.532	0.535	0.541
E-218	2	0	3	3	0.532	0.535	0.539	E-234	2	0	3	4	0.532	0.535	0.539
E-219	3	0	3	3	0.531	0.535	0.54	E-235	3	0	3	4	0.533	0.535	0.538
E-220	4	0	3	3	0.531	0.535	0.539	E-236	4	0	3	4	0.532	0.535	0.539
E-221	1	0	4	3	0.531	0.534	0.539	E-237	1	0	4	4	0.531	0.535	0.54
E-222	2	0	4	3	0.532	0.535	0.54	E-238	2	0	4	4	0.531	0.535	0.54
E-223	3	0	4	3	0.532	0.535	0.54	E-239	3	0	4	4	0.531	0.535	0.54
E-224	4	0	4	3	0.532	0.535	0.541	E-240	4	0	4	4	0.531	0.535	0.54
E-225	1	0	1	4	0.531	0.535	0.54								

E.3 Ensembles using Four Algorithms

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS
SVM	0	0	0	1	0.531	0.535	0.54	E-281	1	3	3	1	0.534	0.537	0.54
E-241	1	1	1	1	0.535	0.537	0.539	E-282	2	3	3	1	0.533	0.54	0.536
E-242	2	1	1	1	0.533	0.54	0.536	E-283	3	3	3	1	0.532	0.539	0.537
E-243	3	1	1	1	0.53	0.539	0.535	E-284	4	3	3	1	0.531	0.54	0.537
E-244	4	1	1	1	0.528	0.539	0.534	E-285	1	3	4	1	0.531	0.535	0.54
E-245	1	1	2	1	0.53	0.535	0.54	E-286	2	3	4	1	0.533	0.537	0.54
E-246	2	1	2	1	0.534	0.537	0.539	E-287	3	3	4	1	0.531	0.54	0.539
E-247	3	1	2	1	0.533	0.54	0.537	E-288	4	3	4	1	0.531	0.54	0.538
E-248	4	1	2	1	0.529	0.541	0.534	E-289	1	4	1	1	0.533	0.541	0.537
E-249	1	1	3	1	0.531	0.535	0.539	E-290	2	4	1	1	0.534	0.54	0.538
E-250	2	1	3	1	0.531	0.535	0.539	E-291	3	4	1	1	0.533	0.54	0.537
E-251	3	1	3	1	0.532	0.537	0.539	E-292	4	4	1	1	0.534	0.541	0.538
E-252	4	1	3	1	0.532	0.54	0.536	E-293	1	4	2	1	0.534	0.541	0.537
E-253	1	1	4	1	0.531	0.531	0.535	E-294	2	4	2	1	0.534	0.54	0.537
E-254	2	1	4	1	0.53	0.535	0.539	E-295	3	4	2	1	0.534	0.541	0.537
E-255	3	1	4	1	0.53	0.535	0.539	E-296	4	4	2	1	0.531	0.54	0.538
E-256	4	1	4	1	0.532	0.537	0.539	E-297	1	4	3	1	0.533	0.54	0.537
E-257	1	2	1	1	0.533	0.54	0.538	E-298	2	4	3	1	0.534	0.54	0.538
E-258	2	2	1	1	0.533	0.54	0.537	E-299	3	4	3	1	0.531	0.541	0.536
E-259	3	2	1	1	0.533	0.54	0.537	E-300	4	4	3	1	0.532	0.54	0.537
E-260	4	2	1	1	0.53	0.54	0.536	E-301	1	4	4	1	0.534	0.537	0.54
E-261	1	2	2	1	0.534	0.537	0.54	E-302	2	4	4	1	0.532	0.54	0.537
E-262	2	2	2	1	0.53	0.54	0.537	E-303	3	4	4	1	0.531	0.54	0.538
E-263	3	2	2	1	0.532	0.541	0.537	E-304	4	4	4	1	0.531	0.539	0.538
E-264	4	2	2	1	0.531	0.54	0.537	E-305	1	1	1	2	0.531	0.535	0.54
E-265	1	2	3	1	0.531	0.535	0.54	E-306	2	1	1	2	0.534	0.537	0.54
E-266	2	2	3	1	0.533	0.537	0.54	E-307	3	1	1	2	0.532	0.54	0.537
E-267	3	2	3	1	0.531	0.54	0.537	E-308	4	1	1	2	0.53	0.54	0.537
E-268	4	2	3	1	0.531	0.54	0.535	E-309	1	1	2	2	0.533	0.535	0.541
E-269	1	2	4	1	0.531	0.535	0.538	E-310	2	1	2	2	0.531	0.535	0.54
E-270	2	2	4	1	0.53	0.535	0.54	E-311	3	1	2	2	0.532	0.537	0.539
E-271	3	2	4	1	0.532	0.537	0.54	E-312	4	1	2	2	0.532	0.54	0.537
E-272	4	2	4	1	0.531	0.54	0.536	E-313	1	1	3	2	0.53	0.535	0.54
E-273	1	3	1	1	0.535	0.54	0.537	E-314	2	1	3	2	0.531	0.535	0.54
E-274	2	3	1	1	0.533	0.541	0.538	E-315	3	1	3	2	0.532	0.535	0.541
E-275	3	3	1	1	0.533	0.54	0.538	E-316	4	1	3	2	0.533	0.537	0.539
E-276	4	3	1	1	0.532	0.541	0.538	E-317	1	1	4	2	0.531	0.534	0.539
E-277	1	3	2	1	0.533	0.54	0.538	E-318	2	1	4	2	0.531	0.535	0.54
E-278	2	3	2	1	0.534	0.54	0.538	E-319	3	1	4	2	0.531	0.535	0.54
E-279	3	3	2	1	0.533	0.54	0.537	E-320	4	1	4	2	0.532	0.535	0.54
E-280	4	3	2	1	0.531	0.541	0.539	E-321	1	2	1	2	0.534	0.537	0.539

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS
SVM	0	0	0	1	0.531	0.535	0.54	E-362	2	4	3	2	0.532	0.54	0.537
E-322	2	2	1	2	0.533	0.54	0.536	E-363	3	4	3	2	0.533	0.54	0.538
E-323	3	2	1	2	0.533	0.541	0.537	E-364	4	4	3	2	0.532	0.54	0.537
E-324	4	2	1	2	0.531	0.54	0.536	E-365	1	4	4	2	0.532	0.535	0.54
E-325	1	2	2	2	0.531	0.535	0.54	E-366	2	4	4	2	0.533	0.537	0.54
E-326	2	2	2	2	0.535	0.537	0.539	E-367	3	4	4	2	0.532	0.54	0.538
E-327	3	2	2	2	0.533	0.54	0.537	E-368	4	4	4	2	0.531	0.54	0.538
E-328	4	2	2	2	0.53	0.54	0.537	E-369	1	1	1	3	0.531	0.535	0.54
E-329	1	2	3	2	0.531	0.535	0.541	E-370	2	1	1	3	0.532	0.536	0.54
E-330	2	2	3	2	0.531	0.535	0.54	E-371	3	1	1	3	0.534	0.537	0.54
E-331	3	2	3	2	0.533	0.538	0.539	E-372	4	1	1	3	0.534	0.54	0.538
E-332	4	2	3	2	0.533	0.54	0.536	E-373	1	1	2	3	0.532	0.535	0.539
E-333	1	2	4	2	0.531	0.535	0.54	E-374	2	1	2	3	0.533	0.535	0.54
E-334	2	2	4	2	0.53	0.535	0.54	E-375	3	1	2	3	0.531	0.535	0.539
E-335	3	2	4	2	0.531	0.535	0.54	E-376	4	1	2	3	0.534	0.537	0.539
E-336	4	2	4	2	0.533	0.537	0.538	E-377	1	1	3	3	0.532	0.535	0.54
E-337	1	3	1	2	0.533	0.54	0.537	E-378	2	1	3	3	0.532	0.535	0.539
E-338	2	3	1	2	0.533	0.54	0.537	E-379	3	1	3	3	0.532	0.535	0.539
E-339	3	3	1	2	0.535	0.54	0.537	E-380	4	1	3	3	0.532	0.535	0.54
E-340	4	3	1	2	0.534	0.54	0.537	E-381	1	1	4	3	0.531	0.535	0.54
E-341	1	3	2	2	0.533	0.537	0.54	E-382	2	1	4	3	0.531	0.535	0.54
E-342	2	3	2	2	0.532	0.54	0.538	E-383	3	1	4	3	0.531	0.536	0.539
E-343	3	3	2	2	0.534	0.54	0.539	E-384	4	1	4	3	0.533	0.535	0.54
E-344	4	3	2	2	0.532	0.54	0.535	E-385	1	2	1	3	0.532	0.535	0.54
E-345	1	3	3	2	0.532	0.535	0.54	E-386	2	2	1	3	0.534	0.537	0.541
E-346	2	3	3	2	0.533	0.537	0.54	E-387	3	2	1	3	0.534	0.54	0.537
E-347	3	3	3	2	0.531	0.54	0.537	E-388	4	2	1	3	0.533	0.54	0.537
E-348	4	3	3	2	0.532	0.54	0.536	E-389	1	2	2	3	0.532	0.536	0.539
E-349	1	3	4	2	0.532	0.535	0.541	E-390	2	2	2	3	0.533	0.535	0.539
E-350	2	3	4	2	0.53	0.535	0.54	E-391	3	2	2	3	0.535	0.537	0.539
E-351	3	3	4	2	0.532	0.537	0.539	E-392	4	2	2	3	0.533	0.54	0.536
E-352	4	3	4	2	0.532	0.54	0.537	E-393	1	2	3	3	0.531	0.535	0.54
E-353	1	4	1	2	0.534	0.54	0.537	E-394	2	2	3	3	0.533	0.535	0.54
E-354	2	4	1	2	0.534	0.54	0.537	E-395	3	2	3	3	0.532	0.535	0.539
E-355	3	4	1	2	0.533	0.54	0.538	E-396	4	2	3	3	0.534	0.537	0.539
E-356	4	4	1	2	0.534	0.54	0.537	E-397	1	2	4	3	0.531	0.535	0.541
E-357	1	4	2	2	0.533	0.54	0.538	E-398	2	2	4	3	0.53	0.535	0.54
E-358	2	4	2	2	0.533	0.54	0.537	E-399	3	2	4	3	0.532	0.535	0.539
E-359	3	4	2	2	0.533	0.54	0.537	E-400	4	2	4	3	0.532	0.535	0.539
E-360	4	4	2	2	0.533	0.54	0.537	E-401	1	3	1	3	0.534	0.537	0.54
E-361	1	4	3	2	0.533	0.537	0.54	E-402	2	3	1	3	0.533	0.54	0.536

ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS	ALG	RF	XGB	LR	SVM	CFS	PCCFS	F12FS
SVM	0	0	0	1	0.531	0.535	0.54	E-450	2	2	1	4	0.532	0.535	0.54
E-403	3	3	1	3	0.533	0.539	0.537	E-451	3	2	1	4	0.533	0.537	0.54
E-404	4	3	1	3	0.533	0.54	0.535	E-452	4	2	1	4	0.534	0.539	0.537
E-405	1	3	2	3	0.532	0.535	0.54	E-453	1	2	2	4	0.532	0.535	0.539
E-406	2	3	2	3	0.533	0.537	0.539	E-454	2	2	2	4	0.533	0.535	0.539
E-407	3	3	2	3	0.533	0.54	0.537	E-455	3	2	2	4	0.533	0.535	0.539
E-408	4	3	2	3	0.533	0.54	0.537	E-456	4	2	2	4	0.534	0.537	0.539
E-409	1	3	3	3	0.532	0.535	0.54	E-457	1	2	3	4	0.531	0.535	0.539
E-410	2	3	3	3	0.531	0.535	0.54	E-458	2	2	3	4	0.532	0.535	0.539
E-411	3	3	3	3	0.534	0.537	0.54	E-459	3	2	3	4	0.534	0.536	0.54
E-412	4	3	3	3	0.533	0.54	0.537	E-460	4	2	3	4	0.532	0.535	0.54
E-413	1	3	4	3	0.531	0.535	0.54	E-461	1	2	4	4	0.531	0.535	0.54
E-414	2	3	4	3	0.533	0.535	0.541	E-462	2	2	4	4	0.532	0.535	0.54
E-415	3	3	4	3	0.531	0.535	0.54	E-463	3	2	4	4	0.532	0.535	0.538
E-416	4	3	4	3	0.533	0.537	0.54	E-464	4	2	4	4	0.532	0.535	0.539
E-417	1	4	1	3	0.533	0.54	0.537	E-465	1	3	1	4	0.532	0.535	0.539
E-418	2	4	1	3	0.533	0.54	0.536	E-466	2	3	1	4	0.535	0.537	0.539
E-419	3	4	1	3	0.533	0.54	0.537	E-467	3	3	1	4	0.533	0.54	0.537
E-420	4	4	1	3	0.534	0.54	0.536	E-468	4	3	1	4	0.534	0.54	0.537
E-421	1	4	2	3	0.534	0.537	0.54	E-469	1	3	2	4	0.532	0.535	0.538
E-422	2	4	2	3	0.532	0.54	0.538	E-470	2	3	2	4	0.533	0.535	0.54
E-423	3	4	2	3	0.533	0.54	0.537	E-471	3	3	2	4	0.534	0.537	0.539
E-424	4	4	2	3	0.533	0.54	0.536	E-472	4	3	2	4	0.534	0.539	0.537
E-425	1	4	3	3	0.532	0.535	0.54	E-473	1	3	3	4	0.531	0.535	0.54
E-426	2	4	3	3	0.533	0.537	0.54	E-474	2	3	3	4	0.532	0.535	0.539
E-427	3	4	3	3	0.532	0.54	0.537	E-475	3	3	3	4	0.533	0.535	0.539
E-428	4	4	3	3	0.533	0.54	0.538	E-476	4	3	3	4	0.535	0.537	0.541
E-429	1	4	4	3	0.532	0.535	0.54	E-477	1	3	4	4	0.531	0.535	0.54
E-430	2	4	4	3	0.532	0.535	0.54	E-478	2	3	4	4	0.532	0.535	0.54
E-431	3	4	4	3	0.533	0.537	0.54	E-479	3	3	4	4	0.532	0.535	0.54
E-432	4	4	4	3	0.531	0.54	0.537	E-480	4	3	4	4	0.534	0.535	0.539
E-433	1	1	1	4	0.531	0.535	0.54	E-481	1	4	1	4	0.534	0.537	0.54
E-434	2	1	1	4	0.531	0.535	0.54	E-482	2	4	1	4	0.533	0.54	0.537
E-435	3	1	1	4	0.532	0.535	0.539	E-483	3	4	1	4	0.533	0.54	0.538
E-436	4	1	1	4	0.535	0.537	0.54	E-484	4	4	1	4	0.534	0.54	0.536
E-437	1	1	2	4	0.531	0.535	0.54	E-485	1	4	2	4	0.532	0.535	0.54
E-438	2	1	2	4	0.532	0.535	0.539	E-486	2	4	2	4	0.534	0.537	0.539
E-439	3	1	2	4	0.533	0.535	0.541	E-487	3	4	2	4	0.533	0.54	0.537
E-440	4	1	2	4	0.532	0.535	0.54	E-488	4	4	2	4	0.533	0.54	0.536
E-441	1	1	3	4	0.532	0.535	0.54	E-489	1	4	3	4	0.532	0.535	0.54
E-442	2	1	3	4	0.532	0.536	0.54	E-490	2	4	3	4	0.532	0.535	0.54
E-443	3	1	3	4	0.532	0.535	0.539	E-491	3	4	3	4	0.534	0.537	0.538
E-444	4	1	3	4	0.531	0.535	0.54	E-492	4	4	3	4	0.533	0.54	0.538
E-445	1	1	4	4	0.532	0.536	0.54	E-493	1	4	4	4	0.531	0.535	0.54
E-446	2	1	4	4	0.532	0.536	0.538	E-494	2	4	4	4	0.532	0.535	0.54
E-447	3	1	4	4	0.531	0.535	0.539	E-495	3	4	4	4	0.531	0.535	0.54
E-448	4	1	4	4	0.532	0.535	0.54	E-496	4	4	4	4	0.534	0.537	0.54
E-449	1	2	1	4	0.531	0.535	0.54								