**Predicting the performance of business partners, using issue data of the iSense system**

Mapping a perception to data using machine learning

**Master thesis**

**Dennis Muller**

University of Twente supervisors:
Maurice van Keulen & Bart Nieuwenhuizen

Nedap supervisor:
Jaap Zaal

Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente
Netherlands
05-02-2018

# Abstract

Nedap retail helps retailers in their diverse needs in loss prevention, stock management and store monitoring. These solutions are being used for monitoring, identifying and detecting tagged products in stores. Retailers use Nedap's products to protect themselves against loss prevention, to manage their stocks and to monitor their stores. Nedap is selling these devices to retailers in many countries across the globe. Nedap's policy is to outsource specific activities like installation and maintenance of their devices to local business partners. Nedap does not have to enrol employees abroad and business partners are familiar with local legislation. iSense is a new alarm pedestal that detects and identifies goods passing the entrance of a retail store. Currently, the market is making a shift to the new iSense system. The research question is how to use emerging big data analysis to extract business partner performance information from the iSense messages. To answer this question, we use supervised machine learning with a continuous output. We used a questionnaire to obtain input labels and interviews with various experts to obtain the candidate features. We proof that the strongest mapping of perception of the experts to the features is the gradient boosting regressor. These features are reduced using principle component analysis in order to fit on the sparse data. Our model does predict the perception of the experts, however insights in the data show that these perceptions are not always correct. These insights provide Nedap with information to assist business partners in improving their performance and understand their problems in making the shift to the new iSense system. Using the opinions of the relevant parties as input values for a machine learning algorithm proved valuable to address problems and obtain insights. We believe that our approach can be generalized to other cases.

# Preface

This thesis has been conducted on behalf of Nedap Retail, at Nedap headquarters in Groenlo. Nedap Retail helped with supporting this thesis wherever needed in order to achieve the goal of the research, which is to gain insight in the performance of her business partners. I would like to thank entire Nedap Retail for this. Particularly, I would like to thank Jaap Zaal for his help with getting in contact with the interviewees from all over the world and the technical operations and services team with their help in understanding the data and various edge cases in the systems. Furthermore, I would like to thank my supervisors from the University of Twente: Maurice van Keulen and Bart Nieuwenhuizen. The thorough discussions helped increasing the quality of this thesis and directed me in the right direction when necessary.

# Contents

**List of Figures** **iv**

**List of Tables** **v**

**1 Introduction** **1**
1.1 Objectives . . . . . 2
1.2 Approach . . . . . 6
1.3 Contributions . . . . . 8
1.4 Structure . . . . . 8

**2 Background** **10**
2.1 Big Data . . . . . 10
2.1.1 What is big data . . . . . 10
2.1.2 Benefits of big data analysis . . . . . 12
2.1.3 Barriers in big data . . . . . 13
2.2 Data Mining . . . . . 14
2.2.1 Machine Learning . . . . . 15
2.2.2 Classification . . . . . 16
2.2.3 Regression . . . . . 17
2.2.4 Clustering . . . . . 17
2.2.5 Association Rule Learning . . . . . 18
2.3 Feature Engineering . . . . . 19
2.4 Summary . . . . . 20

**3 Context** **21**
3.1 Nedap Retail . . . . . 21
3.2 Business Partners . . . . . 22
3.3 Summary . . . . . 22

**4 Candidate Features** **23**
4.1 Data Exploration . . . . . 23
4.1.1 Issue categories . . . . . 23
4.1.2 Responsibility . . . . . 26
4.1.3 Severity Category . . . . . 27
4.1.4 Issue duration . . . . . 28
4.1.5 Issue count . . . . . 28
4.1.6 Stores, devices and gates . . . . . 29

4.1.7 Summary . . . 30
4.2 Interviews . . . 31
4.2.1 Goal . . . 31
4.2.2 Approach . . . 31
4.2.3 Interviewees . . . 32
4.2.4 Results . . . 33
4.2.4.1 Communication . . . 33
4.2.4.2 Training . . . 34
4.2.4.3 Global vs Local . . . 35
4.2.4.4 Performance time window . . . 35
4.2.5 Summary . . . 36
4.3 Questionnaire . . . 36
4.3.1 Goal & Approach . . . 36
4.3.2 Data . . . 38
4.3.3 Results . . . 38
4.3.4 Summary . . . 39
4.4 Data preparation . . . 40
4.5 Conclusion . . . 40

**5 The Model** **42**
5.1 Approach . . . 42
5.1.1 Summary . . . 44
5.2 Models . . . 45
5.2.1 Data scaling . . . 45
5.2.2 Multiple perceptions . . . 45
5.2.3 Long-list . . . 46
5.2.4 Criteria . . . 47
5.2.5 Short-list . . . 48
5.2.5.1 Regression tree . . . 49
5.2.5.2 Gradient Boosting Regressor . . . 49
5.2.6 Summary . . . 51
5.3 Feature selection . . . 51
5.4 Final model . . . 53
5.4.1 Accuracy . . . 54
5.4.2 Consistency . . . 55
5.4.3 Results . . . 55
5.4.4 Best time window . . . 58
5.4.4.1 Monthly interval . . . 58

5.4.4.2 Yearly interval . . . . . . . . . . . . . . . . . . . 59
5.4.4.3 Quarterly interval . . . . . . . . . . . . . . . . . 60
5.4.4.4 Moving time window . . . . . . . . . . . . . . . 60
5.5 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . 60

**6 Validation of the predictions 61**
6.1 Goal . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 61
6.2 Approach . . . . . . . . . . . . . . . . . . . . . . . . . . . . 62
6.3 Observations . . . . . . . . . . . . . . . . . . . . . . . . . . 62
6.4 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 64
6.5 Subjective mapping . . . . . . . . . . . . . . . . . . . . . . . 66
6.6 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66

**7 Evaluation 67**
7.1 iSense vs OST . . . . . . . . . . . . . . . . . . . . . . . . . 67
7.2 Feature importance . . . . . . . . . . . . . . . . . . . . . . . 68
7.3 Global priority . . . . . . . . . . . . . . . . . . . . . . . . . 69
7.4 Offline stores . . . . . . . . . . . . . . . . . . . . . . . . . . 71

**8 Conclusion 73**
8.1 Generalization . . . . . . . . . . . . . . . . . . . . . . . . . 76
8.2 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . 77
8.3 Future Work & Recommendations . . . . . . . . . . . . . . . . 78

**A Appendix A 87**
A.1 Interviews . . . . . . . . . . . . . . . . . . . . . . . . . . . 87
A.1.1 First interview . . . . . . . . . . . . . . . . . . . . . . 87
A.1.2 Second interview . . . . . . . . . . . . . . . . . . . . . 88
A.1.3 Third interview . . . . . . . . . . . . . . . . . . . . . . 90
A.1.4 Fourth interview . . . . . . . . . . . . . . . . . . . . . 91
A.1.5 Fifth interview . . . . . . . . . . . . . . . . . . . . . . 91
A.1.6 Sixth interview . . . . . . . . . . . . . . . . . . . . . . 92
A.1.7 Seventh interview . . . . . . . . . . . . . . . . . . . . . 93
A.1.8 Eighth interview . . . . . . . . . . . . . . . . . . . . . 94
A.1.9 Ninth interview . . . . . . . . . . . . . . . . . . . . . . 95
A.1.10 Tenth interview . . . . . . . . . . . . . . . . . . . . . . 96

# List of Figures


1 An iSense system with gates at a store . . . . . . . . . . . . . . . . 3
2 An iSense system that is overhead at a store, the system is attached to the roof instead of a gate . . . . . . . . . . . . . . . . . . . . . . 3
3 The hierarchy of Nedap . . . . . . . . . . . . . . . . . . . . . . . . 4
4 How the model is created and used to make predictions . . . . . . 5
5 The CRISM-DM process . . . . . . . . . . . . . . . . . . . . . . . 6
6 The division of machine learning techniques and the algorithms . 15
7 A simple classification example . . . . . . . . . . . . . . . . . . . 16
8 A simple clustering example . . . . . . . . . . . . . . . . . . . . . 17
9 A simple association example . . . . . . . . . . . . . . . . . . . . 19
10 Total amount of issues above the duration on the X-label . . . . . 29
11 Visualization of overfitting and underfitting . . . . . . . . . . . . . 44
12 An example how a MinMax-scalar scales the data . . . . . . . . . 46
13 A simple regression tree . . . . . . . . . . . . . . . . . . . . . . . 50
14 A simple gradient boosting regressor . . . . . . . . . . . . . . . . 50
15 The regression tree with friedmans MSE as error function, the top two splits can be seen . . . . . . . . . . . . . . . . . . . . . . . . 70
16 The regression tree with MAE as error function, the top two splits can be seen . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 70
17 The regression tree with MSE as model, the top two splits can be seen . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 70

# List of Tables

1 Issue type with the statistics and category of each type . . . . . . 25
2 Responsibility Matrix, responsibility against issue type with how much impact an issue type has . . . . . . . . . . . . . . . . . . . 27
3 Issue type with the statistics and category of each type . . . . . . 37
4 Business partners ratings from questionnaire and interviews . . . . 39
5 An overview of all candidate features, these are repeated for local retailers and global retailers . . . . . . . . . . . . . . . . . . . . . . 41
6 Accuracy percentages by different cases . . . . . . . . . . . . . . 48
7 Accuracy percentages by different cases . . . . . . . . . . . . . . 55
8 The different gradient boosting regressors and parameters . . . . . 56
9 The different regression trees and parameters . . . . . . . . . . . 57
10 The different error percentages of the different models, the columns show the iteration number and the rows show the error percentage per model . . . . . . . . . . . . . . . . . . . . . . . . . . . 57
11 The predicted rating of a business partner with the average perception of the interviewees . . . . . . . . . . . . . . . . . . . . . . 64
12 The principle components and their eigen vectors with the values of the different features . . . . . . . . . . . . . . . . . . . . . . . 71

# 1 Introduction

Big data and data mining are buzzwords currently making their way in the field of research and practice. The opportunities stored in these large data sets are immense and the value likewise [34]. Many of these opportunities are not chased due to time and manpower, but the opportunities are understood within a company. Nedap N.V. is one of these companies that stores large data sets in different fields, for example; retail, livestock, healthcare [41]. This research is conducted at Nedap Retail part of Nedap N.V. [40]. Nedap Retail (from now on Nedap) has two products in the market to support retailers with their daily business. Loss prevention is one of the products Nedap offers to retailers the other being stock management. Stock management are systems that help the retailers with their inventory and tracking of sales. Loss prevention can be a gate system as can be seen in figure 1 or an overhead detection system as can be seen in figure 2. These systems are used to detect theft and register articles that have left the store. Within loss prevention, Nedap has two systems in the market to help retailers. The first system is the old system called OST. This system is not intelligent when it comes to issue handling, it just tells the client that something is wrong with an error-log on what is wrong with the system. The second type is called iSense, which can be both the overhead systems and the gate system. The huge difference between iSense and OST is that iSense is an intelligent system that will analyse itself and come back with a conclusion on what is wrong with the system. The systems are reporting issues, which range from hardware related problems to detection problems. Based on this information the client can easily find the problem and solve this in order to have their system function at maximum performance. Nedap is currently in the middle of making the shift to iSense systems and away from the old OST systems, therefore the focus of this thesis lays on iSense.

Nedap does not install the systems themselves. Nedap has a global network of business partners, which install Nedap's solutions at the retailers. Nedap is active in over 127 countries and each country has at least one business partner. Each business partner has its own region in which they are active and responsible for the systems. This is not only installation, but also servicing the retailers after installing the systems to ensure maximum quality. The way this hierarchy works can be seen in figure 3 Importantly Nedap stores all data of these systems and create platforms for the retailers, business partners and themselves to see issues of the systems and allow for remote connections to solve them. This means that the business partners are directly responsible for ensuring that the systems are installed correctly, issues that come up are resolved and configuration of the system is done correctly. The business partners are directly responsible for the quality of systems, since it is their responsibility to execute the installation of the systems and the servicing of these systems. Which brings us to main question Nedap has, how are our business partners performing?

Currently Nedap wants to improve their insight in the performance of their business partners. The iSense systems report issues and these are stored by Nedap. This data contains information about when they occur, when these are solved and what type of issue is reported. This information shows the up-time of systems, what problems the system had and how long it took a business partner to solve the problems. Based on this issue data it should be possible to determine the performance of a business partner. The question is whether the current perception of performance that exists within Nedap can be related to data or that the perception is based on unknown factors.

## 1.1 Objectives

This research aims to address the perception of performance that was mentioned in previous section. There is a need for insight in the performance of business partners in order to achieve scalability and ensure the quality of the systems throughout the world. These insights should help improve the existing performance of business partners, by knowing their strength and weaknesses and allowing Nedap to improve the quality of their business partners. This research combines feature engineering, data preparation and machine learning to rate business partners according to the perception of the performance that exists within Nedap. The goal of this research is to provide a model that rates business partners of Nedap that use

Figure 1: An iSense system with gates at a store

Figure 2: An iSense system that is overhead at a store, the system is attached to the roof instead of a gate

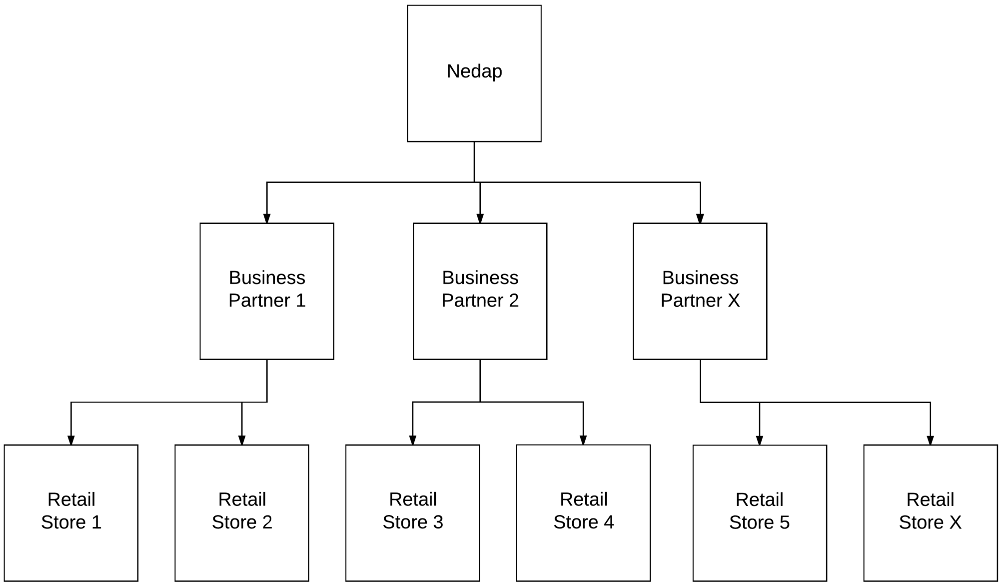



Figure 3: The hierarchy of Nedap

the iSense system, based on the features obtained by feature engineering that impact the performance of the business partner. This brings us to following problem statement.

## Problem statement

How can the performance of a business partner be determined, based on data from the issues provided by the iSense system?

This problem can be addressed by answering four research questions which are discussed below.

The performance of a business partner has two parts: the first part being what the performance is based on and the second part what the current performance of the business partners is. The important difference is that one question is to determine what this performance is based on, where the other question is to identify the perception of performance that exists within Nedap.

RQ1: What features define the performance of a business partner?

RQ2: What is the performance of a business partner?

Based on these features a prediction model is built and data needs to be enriched to support these features. Figure 4 shows the way the previous research questions contribute to the model.

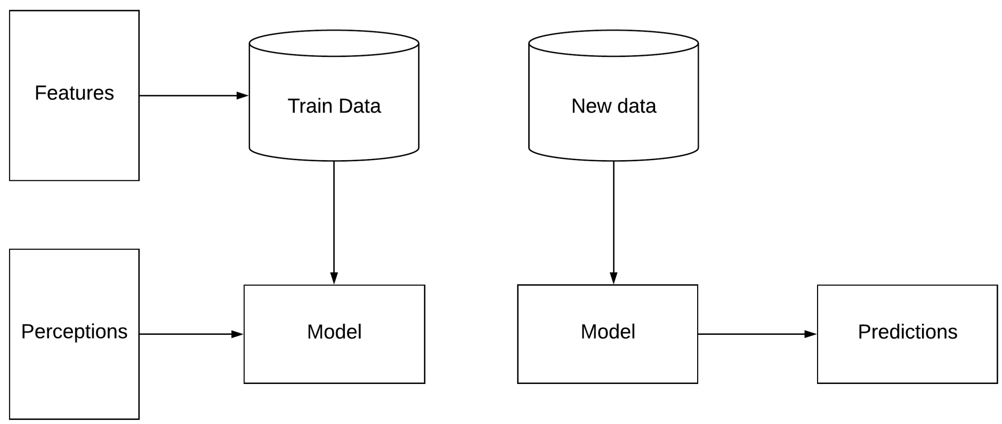



Figure 4: How the model is created and used to make predictions

RQ2 obtains the perceptions used to train the model. RQ1 finds out what features the performance of a business partner is based on, which is the definition of feature engineering. What this model looks like is currently unknown therefore, the following research question is defined:

RQ3: What is the best model to rate business partners based on issue data of the iSense system?

This model predicts ratings for all business partners based on the features, however not every indicator correlates with the rating of a business partner. For this the last research question is defined:

RQ4: What insights does the produced model give about business partners?

## 1.2 Approach

The purpose of this research is to create a model that predicts the performance of business partners and gives insight into the strenghts and weaknesses of the business partners. The long term goal is to allow Nedap to manage their business partners and improve their overall performance (RQ). To keep a steady structure in this thesis, we use CRISP-DM, which is a data mining model that describes the commonly process, used by data mining experts to tackle data mining problems [55]. Figure 5 shows the CRISP-DM process as described in literature.

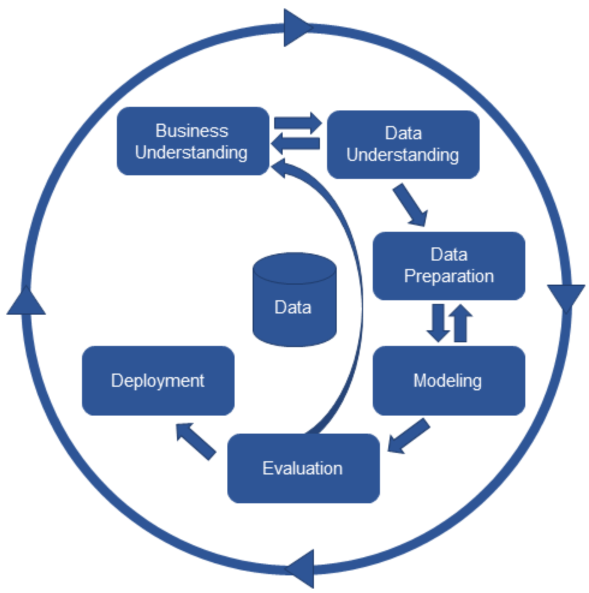



Figure 5: The CRISM-DM process

Oscar Marban et al call CRISP-DM the ”de facto standard” and it was the most used method for doing data mining over multiple years according to surveys [37]. The model was officially released as version 1 in 2000 and has remained the same over the years [10]. Kurgan describes in his work that CRISP is a strong industrial support in the data mining area [30]. Currently CRISP-DM is still the most commonly used process for data mining and analysis in the field of research as it was a couple years ago as claimed by Azevedo [3], even though there has not been a statistical analysis on this. This thesis closely follows the CRISP-DM process as figure 5 shows. As the technique perfectly matches the process that needs to be conducted in this thesis.

To enable a model to rate business partners, we first need to conduct a literature review on what the current situation is within Nedap. To achieve this we analyse what a business partner is and does for Nedap, this can be related to the business understanding part of the CRISP-DM cycle. This is followed by analysing what techniques in data analysis are commonly used for problems in this area of research.

Next, we analyse what features have an impact on the performance of a business partner (RQ1). These features come forward from literature, data exploration and interviews, this step is considered the data understanding part of CRISP-DM. Van der Spoel concluded in his research that besides looking at literature, looking at the organization and talking with experts impacts the features and new features are found [58]. To achieve this, experts are being asked about what potential features could be. Besides asking experts, the data is also being explored to find potential features of the performance, these can be validated by asking experts about their opinion on the features. The result of this is a list with features that are used by the model. However to allow the model to use these features data needs to be prepared and enriched, which is the data preparation step of the CRISP-DM cycle.

To achieve the mapping of the perception to data, the ”golden reference” needs to be known. To achieve this, we need to obtain information about the current performance of business partners within Nedap (RQ2). The data needs to be mapped to this perception of the truth if possible, to see if the current perception is close to the truth.

Subsequently, the model is created based on the list of features (RQ3). This process is considered the modeling step of the CRISP-DM cycle. This model will evaluate the importance of features and their relevance to the rating of a business partner. The model learns which features are important and how they relate to the performance of a business partner. To achieve this different models are tested and the best model is chosen. The definition of the best model is based on how accurate the model is in predicting the ratings.

Finally, the model is predicting the rating of all business partners. We research what these ratings imply about the performance of a business partner and if these predictions correctly display the truth (RQ4). This requires validation of the prediction with different experts that are interviewed in earlier stages of the research.

The way this is done, is by training the model on training data. Once the model is trained, it predicts the ratings of business partners over a new period of data, which the model did not see yet. These predictions are discussed with experts to see how well the model predicts and what insights this model gives. We consider this step as evaluation in the CRISP-DM cycle. The last step is deployment, which follows if the model is correct, however falls outside the scope of this thesis.

## 1.3 Contributions

This thesis contributes by giving Nedap insight in the performance of business partners installing the iSense system. These insights should give understanding in which business partners are performing up to standards and which business partners need assistance in the transition that is currently in progress. Additionally, these insights could provide new projects to further increase data understanding and readability throughout the company.

For research in the field of data understanding and determining partner performance there were also some contributions. The methodology described in this thesis can be used to gain insight in data in most fields of research. The approach to use opinions of interviewees to train a model showed efficient in showing relevant parties what important features are and how important these features are. For research in performance of partners this thesis is a good example on what to expect in a similar situation and what problems can arise.

## 1.4 Structure

First, chapter 2 reviews different techniques in big data and data mining, which are considered for the model that is going to be build. These techniques are all commonly used in machine learning and their advantages and disadvantages are listed in this chapter.

After this, chapter 3 lists contextual information that is necessary for understanding the current situation at Nedap and the complication of the situation. This information consists of: what Nedap does, the definition of a business partner, complications that came forward in the pre-research and ways to tackle these complications.

Following this, chapter 4 describes the features received from: interviews, the questionnaire, literature and data exploration. These features are used in chapter 5 as features for the machine learning model. chapter 5 also explains the choice of machine learning model and how this choice has been made. Chapter 6 discussed the validation of the model. For this the predictions the model is making are compared to the perceptions given in the interviews during the validation. Chapter 7.4 discusses the results of the model, what insights this model provided and what these insights mean for Nedap.

Finally, chapter 8.3 concludes this thesis by answering the research questions, section 8.2 discusses the limitations and strengths of this thesis and finally section 8.3 explains what future research can be done on this project and other projects that came forward in the research.

# 2 Background

This chapter reviews the concept of big data. It starts off by explaining what big data is, followed by the benefits and challenges in big data. The second part of the literature review describes the different techniques of data mining and machine learning. This is continued by a brief description of feature engineering and concludes with a short summary.

## 2.1 Big Data

Over the last few years, volumes of data have increased significantly. The amount of data in 2012 is expected to have grown by 700 percent in 2018 [63]. Big data is a term for data sets that are so large and/or complex that traditional data processing software cannot properly deal with this. Where in the past big data was considered a problem, today it is seen as a huge opportunity to gain more insights into application and business information. Which leads to a new view on storing data, analyze which fields are meaningful and store as much data about these as possible. According to Zakir et al 60 percent of the respondents said that they should focus on data and analysis of this data [63]. The main goals for this would be to generate insights on customers, segmentation and targeting to improve the overall performance of the company [63]. The large amount of data stored by companies also allows for predictive analysis. Predictive analysis is the use of historical data to forecast on customer behavior and trends. The methods used to achieve predictive analysis could be by using statistical models or machine learning algorithms in order to identify patterns and to learn from this data [63]. John Walker claims in his book that many businesses use forecasting and predictive analysis in order to gain a competitive advantage [29]. He believes that the structure of an entire industry will be reshaped based on the change big data analysis will provide.

### 2.1.1 What is big data

Russom claims that big data is defined by three V's, which Chen et al support [47][11]. These three V's are variety, volume and velocity. Kwom et al also characterize big data by the three V's and mentioned that big data analytics is believed to be the next 'blue ocean' in business opportunities, meaning it can redefine businesses as they are currently known [31]. Their definition of big data analytics is: "all technologies and techniques that a company can employ to analyze large scale, complex data for various applications to augment firm performance". These claims have recently been reviewed for the current market and situation by Gandomi et al and concluded that the opportunities described in the past have not been fully exploited, however many are trying to do so [20].

As mentioned, the commonly used definition of big data is the three V's. The first V is volume, volume can be defined by a variety of aspects such as counting records, transactions, tables, or files. In order for data to be considered big data the volume has to be massive, which is the case when standard processing processing software cannot deal with it anymore [59]. Laney claims that as data grows the value of an individual record decreases [32], however once the data becomes large enough the value increases since big data analytics will become possible [42]. SAP has surveyed small and middle sized companies and the results showed that 76% of the companies see big data as an opportunity [48].

One of the differences between data analysis and big data analysis is that big data analysis requires technologies that support high-velocity data capture, storage and analysis of this data. Which is the second V, velocity. Where data analysis can also be done on small data sets with simple technologies to achieve the wanted results, big data requires technologies that can handle high-velocity data capturing, storage and analysis of this data, such as; noSQL, machine learning and map-reduce [47] [20][59]. Big data offers a lot of possibilities when it comes to analysis. Since there is so much data it is significantly easier to detect trends and occurrences that might seem random at first, but appear to be a trend [38].

And the last V is variety. When data is received from only a single instance the amount of data can still be large, however it would still be considered data instead of big data, since the variety is small. The challenge of big data is that the data is received from many different sources and the types are different making it impossible to store them in the same database normally. This means that big data is frequently unstructured which makes it harder to do analysis on [47] [38]. Data is considered big data, when one or more V's are present, which leads to the claim of Ward that standard processing application cannot deal with it anymore [59].

Gandomi et al mention in their paper that some parties have defined big data as more than the three V's and tried adding some others [20]. One of the mentioned

V's is veracity. IBM claims that besides the accepted three V's they believe veracity should be added [64]. Veracity is perceived by the unreliability to include some sources of data. For example customers often speak their minds on social media and therefore this contains a lot of valuable information. But the data is very uncertain and hard to mine. SAS sees variability and complexity as another V [49]. SAS mentions in an example that when asking two persons to measure a plant, one returns with one meter while the other says 100 centimeters. Both answers are similar yet they are described differently. This definitely could be a challenge when receiving data from many sources. Oracle supports SAS that variety should be seen as a V and adds another V, value [42]. Value should be considered as an important aspect of big data according to Oracle, since the data is of low value density, however when analyzed in large volumes it becomes worth a lot.

### 2.1.2 Benefits of big data analysis

Since big data has been gaining ground in the business sector it is important to know the reason businesses apply big data analysis. According to Russom any business that has involvement with customers could benefit from big data analytics on the following points [47]:

Business will have better-targeted social-influence marketing. Social-influence marketing is a new approach when it comes to marketing and this focuses on individuals rather than an entire group. These individuals are approached and get compensated for promoting the respective business. The marketing will indirectly reach an entire group that follow the individual. [47]

Not only marketing will become easier according to Russom, but customer-base segmentation will be more complete since based on this large stack of data, customers are more easily grouped in segments and categories.

The final benefit of using big data analytics is that analytic applications are likely to benefit from the large amount of available data [47]. A few examples of these applications are fraud detection, quantification of risks or automation of decision making for real-time business processes.

This thesis is an example of gaining insight into market opportunities. Because of the large data set and variety of data, there are a lot of opportunities, which can be found within the data. These opportunities might not seem clear at first but when digging deeper into the data the opposite can be claimed true. Since big data has advanced a lot over the years it is nowadays far easier to store this data structured in a way that allows analysis to be far more effective than before [39], not only that but Michael Ketina also supports Russoms claim that the main reason businesses are doing analysis is to gain insights into customers, market-direction and to gain new insights. These new insight can range from forecasting to analyzing the root cause of costs to fraud detection [47].

### 2.1.3 Barriers in big data

While the opportunities are immense, there are also some barriers and challenges in big data analytics. Russoms says that inadequate staffing and skills are the leading barriers to big data analytics [47]. McAfee supports this claim by saying that there are too few data scientists in general [38]. After all, many organizations are still new to big data analytics and often correlation is being mistaken for causation which has the effect that misleading patterns are found in data and perceived as true.

Besides inadequate staff, businesses often do not support big data analysis as a program due the large concerns behind the analytics. These range from privacy concerns to cultural challenges. Michael Ketina supports these claims in his paper while adding to this that businesses need to make choices in what data to store, because otherwise the amount of data stored will grow out of control [39]. He also mentions the issue of privacy being a large risk, since the more data stored with CCTV, on the work floor and in general about the customers could give large insights in every activity that a person is doing. Privacy is something that needs to be taken in account, as business partners might not be happy that Nedap uses the data to analyze their performance.

Variety and complexity is seen as a challenge in big data. Oracle and SAS both see challenge in the variety of data, since the input streams are so different [49] [42]. The challenge that this brings is that there are a couple steps that need to be taken. These steps are data preparation and could need some of the following steps: connecting, matching, cleansing and transforming the data from many different sources. Once these steps have been completed the data can be used in analysis.

The final point Michael Ketina is making, that is important in regards to this thesis is what is done with the results. Analysis is favored by many businesses, but it could happen that the results found can be an issue to the affected parties [39]. As explained with privacy, a business partner can fear their position if their performance is under standards. If this happens to be the case, caution is important and what is done with the results might need to change from what was initially planned.

## 2.2 Data Mining

Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner.

Hand defines data mining as the analysis of large observational data sets to find relationships and to summarize data in understandable and useful ways [23]. Larose supports this definition and names a few technologies that could be used [33]. Linoff even calls it a business process to find meaningful patterns and rules in large data sets [35]. There are two common goals for businesses to do data mining [16]:

1. Descriptive analysis, to understand what the data means and what information is stored in data.
2. Predictive analysis, to predict trends and gain competitive advantage over competitors.

A combination of the two goals is what most businesses do, as predictions are only useful if they can be described and explained. Since identifying individual customers is too time consuming, data mining techniques are often used in data analysis of customer data. There are many techniques in data science to achieve the above goals, below is a list of the most commonly known techniques when it comes to analyzing customer data.

### 2.2.1 Machine Learning

Machine learning is the technique of finding patterns, making predictions and obtaining descriptive information on a data set without specifying how the computer needs to do this. There are many different models each with it's different strengths and weaknesses [7]. Most programming languages support a machine learning library and each implementation should give the same results. Machine learning is split up in two categories based on the principle the underlying algorithm is using. This division can be seen in figure 6.

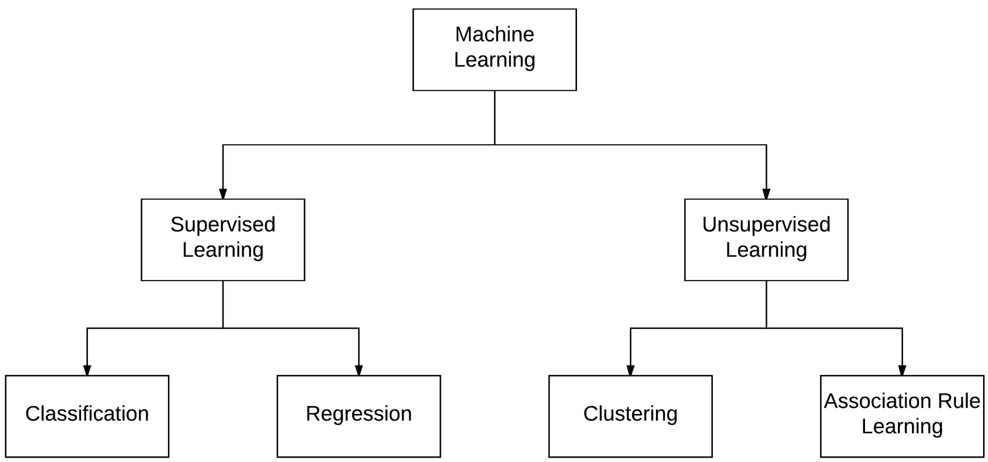



Figure 6: The division of machine learning techniques and the algorithms

The difference between supervised and unsupervised learning is that for supervised learning you need a truth. The different models try to map the input features to the truth that has been given. These features are the indicators mentioned in the earlier sections. A feature is a data column that has a potential relation to the truth. The models adjust the different weights of the features to try to map the given input to its prediction. Unsupervised learning only need the input features and tries to find patterns and correlations between the different features, the model will try to find correlations in the data and that is the strength of unsupervised learning. The techniques used for supervised and unsupervised learning differ and the next sections describe the difference between the techniques.

### 2.2.2 Classification

Classification is a technique that given labeled data constructs classes to assign samples in these predetermined groups [56]. The labeled data consists of many records and each record is unique. In order to classify data in groups a classification model is used, these can have many different forms such as a set of rules, neural networks, decision trees and many more. A classification model trains itself on training data and constructs a model based on what it has learned. Once it has been trained it can be used to predict new samples by putting the new samples into the model, which allocates them to the defined classes.

Figure 7 shows a simple classification example. In this example there is a large pile of fruits that needs to be classified. This large stack needs to be split into four predefined categories: apples, oranges, bananas and grapes. The model identifies each fruit based on whether they are round or not. When that split has been made it can split once more on color, which splits the different fruits up in their respective classes.

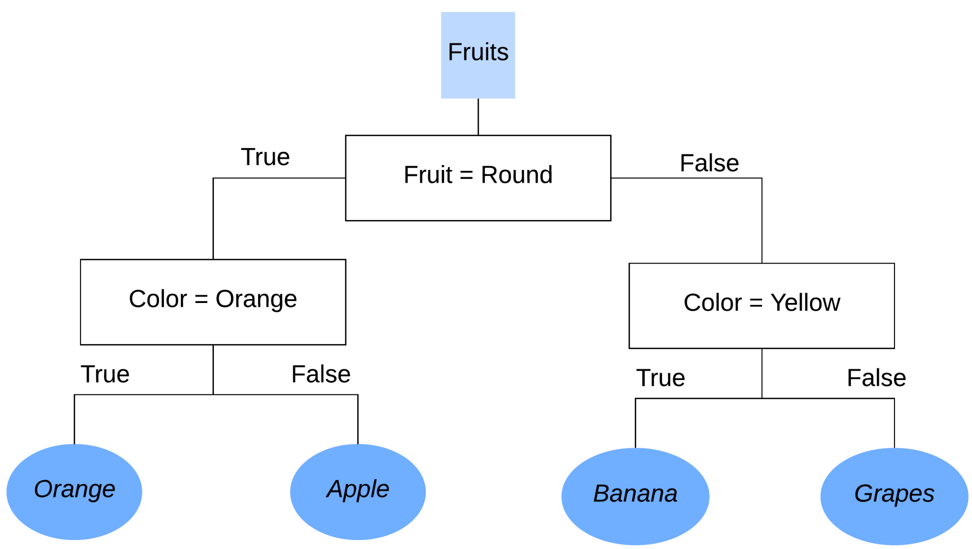



Figure 7: A simple classification example

### 2.2.3 Regression

Regression is a supervised learning method [7]. There are many different algorithms that work with regression models, the biggest difference compared to classification models is that regression models do not have a categorical output. This means that the prediction made by regression-models is continuous and does not limit itself to pre-defined classes (discrete) [7]. When the decision is made on supervised learning the only question that remains is to determine whether the wanted output is continuous or categorical.

### 2.2.4 Clustering

Clustering is often confused with classification. The key difference between clustering and classification is that clustering is an unsupervised method. Gan et al describe in their book that Data Clustering is a method of creating groups of objects (called clusters) in a way that all objects in a cluster are very similar to each other [19]. They are still different but share enough similarities to be considered in the same cluster. One of the key differences between clustering and classification is that the user defines what the clustering is going to be by choosing a similarity function [56]. There are common similarity functions such as k-means, k-median and min-sum [6], however the user can define its own similarity function since this is different for each domain and based on what the user assumes from the data [61].

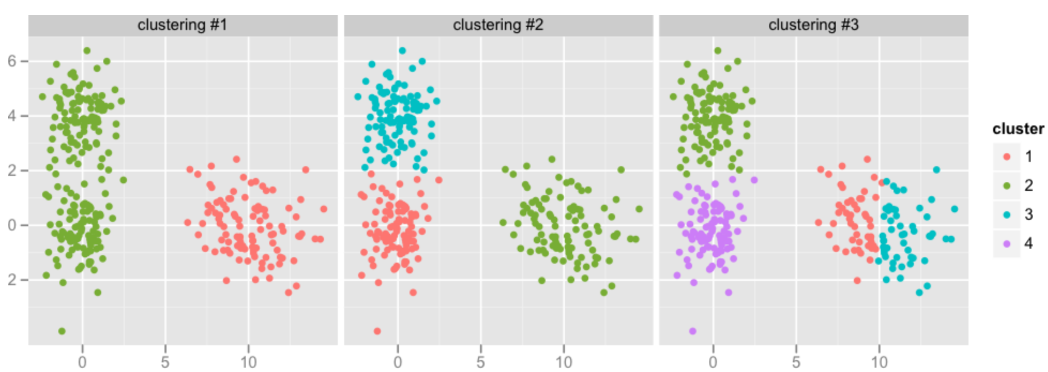


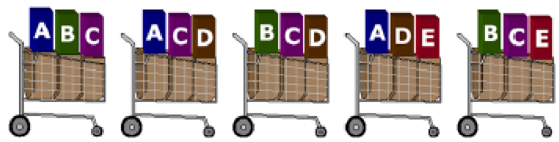

Figure 8: A simple clustering example

To simplify clustering again, the same example will be used as before about a stack of foods arriving. The food is not classified like before, but instead the model is used to determine samples that share features. Based on the similarity function the user defines which clusters show the best relation in the data. This could be based on clustering whether it is a vegetable or fruit or based on the colour of the food. Based on the amount of clusters and similarity function the model clusters the data. Figure 8 shows the difference the user can make by defining the amount of clusters. Each individual clustering has a different amount of clusters. The user can now look at the graphs and determine which amount of clusters best represents the samples, this example uses the same similarity function, however the user could also have tried different similarity functions.

### 2.2.5 Association Rule Learning

In 1993 association rule mining was introduced by Agrawal et al [2]. Association mining is the technique where relationships in the data set are built, so called associations [19]. An association is a rule which assumes there is a likelihood of a specific pattern reoccurring in the data. These patterns are defined in the form of implications such as $X \Rightarrow Y$ where X and Y are items within the data set. The rule should be read as when X occurs in the data set there is a high likelihood of Y appearing in the data set [25]. This likelihood of the rule applying to a case is called confidence. Besides confidence there is another statistic that is important for association mining, which is support. Support is the amount of times X appears in the entire data set. This statistic is a measurement of how often the rule might apply and how strong the rule is. The more often it occurs the more valuable the association rule is.

Figure 9 shows how an association rule works. The figure shows a case where five shopping carts are filled with type of products. Based on these products association rules are made. One of the association rules is that when a customer buys product A they also buy product D. The figure shows that the support is two out of five, since two carts have product A and D and there is a total of 5 carts. The confidence shows that there are two cases where the rule is correct and one case where it is not.




| Rule | Support | Confidence |
|---|---|---|
| $A \Rightarrow D$ | 2/5 | 2/3 |
| $C \Rightarrow A$ | 2/5 | 2/4 |
| $A \Rightarrow C$ | 2/5 | 2/3 |
| $B \& C \Rightarrow D$ | 1/5 | 1/3 |

Figure 9: A simple association example

## 2.3 Feature Engineering

To allow data mining techniques to work, features need to be defined. A feature is an attribute of a data sample that is used by the different models. The process of defining features for data mining is called feature engineering. The term feature engineering in data mining is not a formally agreed definition, more a broadly accepted procedure of steps [62]. This process allows the data mining algorithm to have input of a set of features, which are based on knowledge of the domain, the data and assumptions. This includes steps such as transforming the data into another format, such as a date into day of the week [44]. The knowledge of the domain can be obtained through expert-interviews, surveys, literature and previous research. Data exploration shows the structure of the data and the potential information stored within the data. A data expert can obtain features from the data based on previous analysis, domain knowledge and assumptions. Once these steps are completed a list with features can be made. These features are used as potential indicators for a data mining algorithm, which on its turn determines the relevance of said features, this process is called feature selection. Guyon et al explain that there are many benefits to feature selection, which include data understanding, data visualization and improve prediction performance [21]. Dash et al mention there are various techniques to choose relevant features, as some of the defined features might only cause noise [12]. This process is called feature selection. There are many different techniques to do feature selection and there is no defined correct method as each case is unique as mentioned by Dash et al [12].

## 2.4 Summary

This section discusses relevant literature with regards to big data, data mining and feature engineering. The big data sub-section explains the meaning, the challenges and the benefits of big data. Following this, data mining is discussed with different techniques commonly used in data mining. The last section describes feature engineering and how it is related to the previous sub-section. The next section discusses Nedap, what a business partner is and does, the business model of a business partner and concludes with potential complications in this thesis.

# 3 Context

## 3.1 Nedap Retail

Nedap retail is a company, that has its headquarters located in Groenlo in the Netherlands and is a business unit of Nedap [41]. Nedap Retail works around the globe to deliver industry-leading products, services and solutions for their customers' diverse needs in loss prevention, stock management and store monitoring. Their inventive thinking and collaborative spirit allows them to deliver tailor-made solutions for the fast paced retail sector. Below is their philosophy and text as mentioned in their manual:

"We simplify retail management while improving your customers' shopping experience. By taking most recurring tasks off your hands, we create time for you to devote to your customers. And that is what retail is all about. Whether you run a small local store or a large international chain, you will benefit from our broad range of products, ideas and services.

Nedap solutions are built upon 40 years of global experience, market expertise and close cooperation with leading retailers. Our worldwide operations are supported by a flexible network of certified partners across the globe. Nedap systems are future-proof (RFID-ready), cost-efficient and Eco-friendly. Our mission is simply to make sure your customers maintain the best shopping experience whilst we help you protect your profits. Our philosophy: "Merchandise simply available."" [45]

When it comes to loss prevention Nedap has two systems currently in the market. The first system is the old system called OST. This system is not intelligent when it comes to issue handling, it just tells the client that something is wrong with an error-log. Based on this error-log the application analyze the issue that the system is reporting. The second type is called iSense [46]. The difference between iSense and OST is that iSense is an intelligent system that will analyse itself and come back with a conclusion on what is wrong with the system. Based on this information the client can easily find the problem and solve this to have their system function at maximum capacity. Since iSense is the new system from Nedap and because the old system is being phased-out this research only looks at the iSense system.

Nedap does not install the system them self and is not directly responsible for everyday problems. This is what Nedap has business partners for, what a business partner is and does is described in section **??**. Nedap stores the data and has dashboards available for their business partners and their retailers to provide information about the issues arise. The information stored contains time-stamps, issue-types and duration of the issues. This is why Nedap wants to have insight in the performance of her business partners.

## 3.2 Business Partners

This section has been removed for public view.

## 3.3 Summary

This section described what Nedap is and does, why Nedap has business partners and what these business partners do. Next section describes the techniques used to obtain indicators (features) of performance for the model.

# 4 Candidate Features

This chapter describes the different techniques used to find the features for the model. The techniques used to determine the different features are: data exploration, interviews and the questionnaire. Each section overviews a technique used, the goal of the technique and the results of the technique. The chapter concludes with an overview of all features that are included in the model. The features obtained in different techniques are validated through expert opinions and discussions with colleagues.

## 4.1 Data Exploration

This section will take a closer look at the data set, which provides the main sources of information in this thesis. We describe what information is stored in the database, what the different fields mean, which are relevant for this thesis and conclude the section with a summary of the features that came forward. Section 2.1 discussed big data. The data used in this thesis is considered big data due to the variety and volume of the data. The data comes from several streams and databases and the volume is large as every five minutes each of the systems is sending its metrics to the servers. The challenges and barriers mentioned in the literature review are taken into account in the following stages of this research.

### 4.1.1 Issue categories

The issue data stored from the iSense system has a label field. This label specifies what type of issue the system is reporting. These labels can be categorized in different categories that the system is having problems with.

1. Configuration, an issue occurred that is related to the configuration of the system. Can be solved remotely.
2. Hardware, an issue occurred with the hardware of the system, either a cable is disconnected or part of the system has broken down. Requires physical support at the retail shop.

3. Health, an issue that occurs when the system has problems performing. This issue might require physical support, but can sometimes be solved remotely.

4. Integration, an issue that requires Nedap to solve. This has often to do with connection to the database or the systems supporting iSense.

5. Network, an issue of this type means something is wrong with the network at the retail shop. This requires physical support to solve.

These categories show what kind of tasks need to be done in order to solve an issue. Some issues can be solved remotely, some issues require the business partner to physically visit the retail shop and a couple issues require Nedap to solve them. The list of issue types and their average, trimmed average, mean, category and count can be found in table 1. Three issue types have already been filtered out in this list, since Nedap has the responsibility for these issues and are not related to scoring business partner and can therefore be excluded in the performance of a business partner. There were a lot of issues resolved within five minutes that could not be resolved by human action. These issues have been filtered out in order to get a good view of the statistics of the data. The amount of issues reported by the system are over the last year.

| Issue Type [1] | Average | Trimmed Average | Mean | Category | Count |
|---|---|---|---|---|---|
| type a | 15014 | 227 | 10 | configuration | 73 |
| type b | 12857 | 304 | 648 | configuration | 53743 |
| type c | 7766 | 25 | 14 | configuration | 597 |
| type d | 2795 | 725 | 785 | configuration | 789 |
| type e | 792 | 11 | 5 | configuration | 35176 |
| type f | 4046 | 915 | 611 | configuration | 1258 |
| type g | 10155 | 29 | 14 | configuration | 1197 |
| type h | 394 | 124 | 101 | configuration | 21317 |
| type i | 3825 | 1548 | 1449 | hardware | 26228 |
| type j | 1615 | 8 | 5 | hardware | 4449 |
| type k | 608 | 9 | 4 | hardware | 25321 |
| type l | 345 | 30 | 30 | hardware | 17246 |
| type m | 96 | 31 | 25 | hardware | 41370 |
| type n | 764 | 14 | 10 | hardware | 469 |
| type o | 3942 | 213 | 18 | hardware | 26450 |
| type p | 130 | 12 | 10 | hardware | 13627 |
| type q | 90 | 10 | 10 | hardware | 250 |
| type r | 127 | 10 | 10 | hardware | 16247 |
| type s | 233 | 10 | 10 | hardware | 4595 |
| type t | 2007 | 10 | 10 | hardware | 1639 |
| type u | 309 | 10 | 10 | hardware | 49837 |
| type v | 561 | 388 | 445 | health | 1917 |
| type w | 246 | 75 | 12 | health | 1233779 |
| type x | 309 | 33 | 19 | integration | 50453 |
| type y | 239 | 60 | 34 | integration | 163039 |
| type z | 147 | 201 | 20 | network | 3968230 |

Table 1: Issue type with the statistics and category of each type

[1]Labels have been anonymized due to intellectual property

### 4.1.2 Responsibility

In order to get some information who is responsible for an issue type a responsibility matrix has been made. Table 2 shows the responsibilities by the different parties and if physical support is needed to solve an issue type. The issue types have been made anonymous due to intellectual property. The main reason for doing this is that some issue types depend on help of the retailer and are most likely less important when scoring a business partner. Physical support is also an interesting field that can be taken in account. If a business partner needs to physically visit a store in order to solve an issue, travel distance might need to be taken in account. and supports business partner when asked to. Whether this entire list is complete and correct will also be validated in the interviews that will be held.

| Issue Type | Business Partner | Retail Store | Physical Support | Severity Category |
|---|---|---|---|---|
| type a | Always | No | No | Medium |
| type b | Always | No | No | Low |
| type c | Combined with Nedap | No | No | Medium |
| type d | No | Always | Always | Low |
| type e | Always | No | No | Medium |
| type f | Always | No | No | Medium |
| type g | Combined with Nedap | No | No | High |
| type h | Always | Sometimes | Sometimes | Medium |
| type i | Always | No | Always | Low |
| type j | Always | No | No | Medium |
| type k | Always | No | No | Medium |
| type l | Always | No | No | Medium |
| type m | Always | No | No | Medium |
| type n | Always | No | No | Medium |
| type o | Always | Sometimes | Sometimes | High |
| type p | Always | No | No | Medium |
| type q | Always | No | Always | Medium |
| type r | Always | No | No | Medium |
| type s | Always | No | No | Medium |
| type t | Always | No | No | Medium |
| type u | Always | No | No | Medium |
| type v | Combined with Nedap | No | No | Medium |
| type w | No | Always | No | Low |
| type x | Always | Always | No | Low |
| type y | Always | Always | No | Low |
| type z | Always | Always | No | High |

Table 2: Responsibility Matrix, responsibility against issue type with how much impact an issue type has

### 4.1.3 Severity Category

Every issue is labeled with a type of problem that is reported by iSense. The label of an issue shows how severe an issue is. The severity can easily be categorized in three categories, ranging from barely impacting the system to total system failure. Speaking with multiple experts on these issue types a list came forward with each issue type and its respective severity category. The categories with the amount of types from table 2 and the total amount of issues reported can be found in table 1. One issue type is not represented in the severity categories, which is type B. The reason for not being included in the model is that this issue type can be triggered by multiple parties to temporarily mute the system to fix existing issues. According to experts, these issues are impossible to track whether it was done to prevent issues from occurring or to solve problems. The impact of this issue type is little as it is most often used to deploy a firmware update or configure the system at installation.

#### 4.1.4 Issue duration

The table that stores the issues has two fields: ”reported_at” and ”resolved_at”. By using these two fields the duration of issue can be calculated. The combination of the severity category and the issue duration show the impact the reported issue had on the system. Without issue duration it is impossible to relate the issue data to the performance of a business partner, as their main responsibility is ensuring the system performs optimally. The exploration showed there were a large portion of issues that were resolved within ten minutes, which can be seen in figure 10. Important to notice is that there are a lot of issues that are under ten minutes.

Initially the idea was to exclude issues below ten minutes, since these would not require any manual effort. However, in the pre-research an issue came forward with the framework analyzing the issues. Because of this, issues could be resolved when the issue still existing and then re-opened. This means that issues below ten minutes could actually be a long list of issues happening consecutively. Therefore the conclusion is that all issues are summarized as the total issue duration.

#### 4.1.5 Issue count

The data exploration showed devices that have very low issue durations. However when looking closer, these devices have a significant increase in the amount of issues. This does not necessarily imply that long durations are bad, since many short issues also imply the system is down for a long duration. Therefore issue

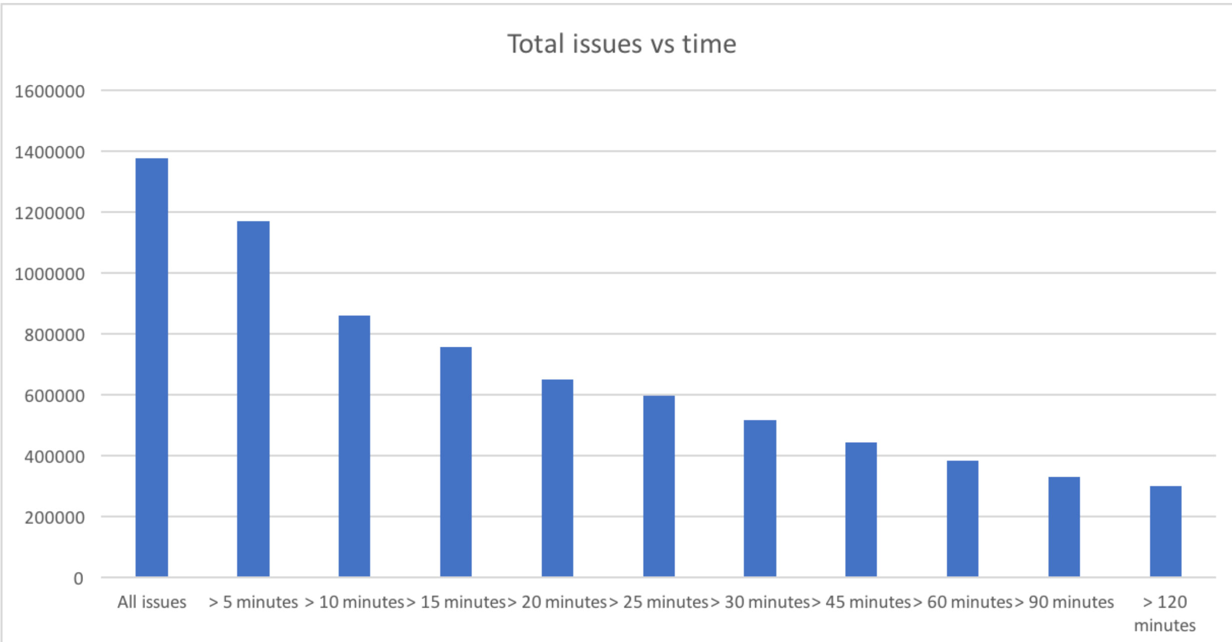



Figure 10: Total amount of issues above the duration on the X-label

duration is included as an indicator on how a business partner is performing.

### 4.1.6 Stores, devices and gates

We discussed issues in the previous sections and how these fall in different categories. Each of these issues can be related to three severity levels as mentioned in table 2. Figure **??** shows that a business partner manages stores. These stores can have one or multiple devices running the infrastructure. In this diagram the devices are iSense devices. iSense can have a small system consisting of two gates, or a complex system containing as many as the retailer wishes for.

A system which contains more gates will report more issues than a small system when configured properly. The existing infrastructure at Nedap does not take into account the severity of an issue. This means that when one gate reports a ”type A” issue, this could mean one gate is down, or all gates are down. For some issue types it is more common that one gate could cause the issue, where some issues are related to the device. Since the amount of stores, devices and gates differs between business partners, issue durations and counts need to take this in account. It is clear that a business partner with more stores, will have more issues and a longer total duration. Therefore the amount and duration of an issue is related to the amount of gates. This assumes that more gates will cause more issues, which sounds logical and according to colleagues is likely true. Therefore, the amount of gates is considered a feature.

Discussion with account managers and sales representatives brought up the relevance of the amount of stores. Business partners with many stores need more personnel to solve issues that require physical presence. They also need more personnel in general to handle issues that come up and pose a higher risk of having unemployed personnel, since recruitment continues to grow for these business partners. A possible scenario is that a business partner will have long unresolved issues, which could be related to lack of personnel to handle this. Therefore we include the amount of stores as a feature. There is no data available on the amount of personnel a business partner has, this could eventually be obtained and in future added to the model to show this. For now we include the amount of stores as an indication on how large a business partner is.

#### 4.1.7 Summary

This section discussed the data exploration. Based on the data exploration a few features came forward which are the following:

- Severity category - categorical (high/medium/low)
- Issue duration - integer of total minutes
- Issue count - integer
- Amount of stores - integer
- Amount of devices - integer
- Amount of gates - integer

These features are discussed in their sections with arguments as to why they are included. Next section describes the interviews that have been conducted.

## 4.2 Interviews

This section describes the goal of conducting interviews for this thesis, the approach used in the interviews, who has been interviewed, the result of the interviews and a discussion. The section concludes with a list of indicators that have been added to the results from previous section.

### 4.2.1 Goal

During the data exploration many candidate features came forward, which could not be confirmed to have an impact on the performance of a business partner due to lack of knowledge on the subject. One of the goals of the interview is to analyze the assumptions with the interviewees. If an assumption has influence on the performance of a business partner, said assumption is included in the list of indicators. The second goal of the interview is to have a large diversity of opinions to have a complete view on the performance of a business partner. The third goal of the interview is to create discussion within Nedap on what indicators have impact on the performance of a business partner. The way this is done is by validating opinions of previous interviews in the next interview. The main reason to validate opinions is to ensure the indicators are general indicators and not specific for a function or person.

### 4.2.2 Approach

Interviews are one of the most familiar techniques to collect qualitative data [13]. The chosen approach for interviews is an unstructured interview. The main reason for using an unstructured interview is that it allows interviewees to freely give their opinion and not be directed into answers. An unstructured interview is a misleading term, which could imply that it is without a structure [14]. However the interview has themes which the interviewee needs to answer. These question are formed openly so that the interviewee is free to talk and give their opinion. Based on these answers new questions can come forward to elaborate their opinion or why an opinion of another interviewee is incorrect. Jacob et al describe in their paper a list of tips when conducting interviews [28]. These tips are taken into consideration when preparing for the interviews. Each interview consists of the parts, which are defined as following:

- Introduction
- Discussion and validation
- Actual performance

Each interview starts with a brief introduction into our research, following into an open question asking what they believe indicates the performance of a business partner. During the interview questions are asked to let interviewees elaborate or to create a discussion with opinions from previous interviews. The interview ends with the same set of questions each time, which is to rate 15 business partners on a scale from 1 to 10. These ratings can be elaborated during the interviews, to get an indication on why some business partners are getting higher or lower ratings than others. These ratings are used by the model as input data and are expanded by conducting a questionnaire, which is discussed in the next section. The way these 15 business partners have been chosen is elaborated in the section of the questionnaire.

### 4.2.3 Interviewees

As mentioned in previous sections the diversity of the experts that are interviewed is important to ensure all parts of performances have been discussed. To achieve this we conduct interviews with many different teams and from different continents. Different continents is important to ensure that the opinion of Nedap US and Nedap China is similar and that there are no indicators that are only relevant to a specific region. The interviewees are from the following teams:

- Account managers
- Developer services team
- Sales director Asia
- Technical operations
- Pre-sales
- Business partner manager
- Sales team
- Sales director USA

Since technical operations has regular contact with the business partner on issues, assist business partners if necessary and work with the data used in this thesis, we interview two members of the technical operations team separately to see if they both agree with each other. All interviewees are currently working within Nedap except for the business partner manager, who recently moved to another company. The teams that are not represented in the interviews are R&D and hardware on iSense. The reason for not including them in the interviews is that when briefly speaking with them, they mentioned they have no insight in performance of a business partner and only focus on the performance of iSense as a system.

### 4.2.4 Results

The transcripts of the interviews can be found in appendix A. The diversity of the interviewees resulted in many different aspects that might be relevant as features for the model. In order for a feature to be included in the list of features, a feature had to meet the following criteria:

1. Related to iSense
2. Can be found in data
3. Validated by other interviewees

If a feature cannot be found in data, it could still be a great addition to the system in the future, these features are discussed in the future work section. Similarly, features not related to iSense can be additions when a similar model is being created for the other systems within Nedap. These features are also discussed in the section future work. Below is a description of each feature, why it is relevant for this thesis and what the arguments of the interviewees are to include the mentioned feature.

#### 4.2.4.1 Communication

A feature that has been mentioned in almost every interview is the communication of a business partner. This communication is a two-way communication. The business partner needs to keep the retailer informed about problems with their systems and that they are working on this. Secondly, the business partner needs to have an open communication with Nedap about problems with the systems when they cannot solve the issues, have problems recurring and have feedback about the system. The communication between the business partner and the retailer is external and no data is available to include this as a feature. The communication towards Nedap does exist in a database, which is called Freshdesk [18]. Freshdesk store all tickets that are submitted by business partners, with priorities, type of question and many other fields. These tickets indicate how much a business partner is communicating with Nedap. The interviewees mentioned that communication is neither bad or good, some business partners communicate a lot to give feedback, which is good. However, there are also business partners that communicate a lot, but ask simple question which should be known information. The way this is currently stored in Freshdesk, it cannot be related to performance yet, however improvements to this are being worked on. Ideally, the feature would be included as a categorical field, which indicates how well the communication of a business partner is. Since this requires significant changes to the existing system and transformation of the existing data, it is excluded from this thesis.

#### 4.2.4.2 Training

Multiple interviewees mentioned that the training employees of a business partners need to follow is very important as an indicator of performance. They have often noticed that a business partners' performance is increasing when they recently had a training and slowly deteriorates. All offices of Nedap store data on when a business partner followed a physical training. Besides this, there is e-learning. These are courses Nedap offers to business partners which can be done remotely at their own office. Nedap stores the amount of e-training a business partner has completed. This data includes the training that they completed, when they completed it and which employee completed this. Interviewees mentioned that it shows dedication from the business partner to train their engineers and could be seen as an indicator of performance. Therefore training is included in the model as a feature. Nedap recommends her business partners to have a physical training every year to stay up to date on the technology and ways to resolve issues. Based on this, the feature is defined as ”X years since last physical training”. The physical training is only followed by the head-engineer of the business partner, who is responsible for the systems. This head-engineer trains the other employees with practical knowledge as some of the information is only relevant for him. The other employees can follow e-learning to further train their knowledge. Based on this, the features e-learnings and physical training are included.

#### 4.2.4.3 Global vs Local

As discussed in chapter 3, a business partner has two different type of retailers that they serve. On the one hand are the globals, which are given to the business partner by Nedap. On the other hand are retailers that they find themselves. Interviewees were asked to answer a question on whether global retailers are just as important as a local store for Nedap. Many interviewees said that every store should be equal. However, when asked if two retailers have an issue at the same time who should be helped first, many interviewees said that obviously the global should have priority. We had the assumption this was the case and during the interviews it became clear that global retailers are treated differently. Based on the input that came forward in the interviews, retailers are separated in two groups. The first group are all global retailers of Nedap, the second group are the ”local heroes” that the business partner found them self. These groups each have the same amount of indicators, but are separated under the assumption that business partners treat the globals different.

#### 4.2.4.4 Performance time window

Every interview ended with the set of questions, to score a business partner on a scale from one to ten. In order to make these ratings meaningful, we asked each interviewee whether the performance of a business partner was stable over a period of a year. Each interviewee said that the performance should be close to steady over a year. They mentioned a business partner can slightly improve or get worse, but generally they can see a steady performance. To see which model best represents the perception that lives within Nedap, different time windows are tested. And the time window that best represents the perception is going to be used. Eventually, Nedap wants to see the performance of a business partner over the last three months to determine whether actions they took had influence on the performance of a business partner. The time windows that are tested by the models are:

- Yearly interval

- Three month interval
- Monthly interval

### 4.2.5 Summary

This section started by explaining the goal of the interviews, the approach, who the interviewees are and concluded by explaining which features were added to the list and their arguments. The feature list has been extended with the following features:

- Physical training X years ago - integer
- E-learning - integer - amount of e-learnings completed
- Split between global and local - all existing features are split up in two groups
- Different time windows

Next section describes the questionnaire that has been conducted and it elaborates how the last part of the interview came to its definition as it is the questionnaire, but with the opportunity to elaborate the ratings.

## 4.3 Questionnaire

This section describes the questionnaire that has been conducted. The section starts with an explanation on the goal of the questionnaire, followed by explaining the questions of the questionnaire and concludes with the result of the questionnaire.

### 4.3.1 Goal & Approach

The goal of the questionnaire is to obtain training data for the model. The input data for the model consists of ratings from 1 to 10 of business partners. The model trains on these ratings and attempts to predict the ratings of business partners that have not been rated on new, more recent, data.

The questionnaire asks experts to score a set of 15 business partners on a scale of 1 to 10. These 15 business partners are selected on the following criteria:

- More than 50% of their systems using iSense
- More than 20 devices
- Multiple regions

The first criterion is important, since it allows us to relate the ratings to iSense rather than it most likely being related to the old OST system. The second criterion ensures that a rating can be given on a business partner. 20 devices indicates that a business partner has had a few stores for a while now. The last criterion is to ensure that the inputs received verify the assumption that came forward during the interviews, which is that region is irrelevant. During the interviews it became clear that while region often matters, in this case the performance of a business partner should not depend on the region. To verify this, the business partners asked in the questionnaire are from different regions. The model should also predict the ratings of these business partners correctly based on the training data.

| Statistic | Result |
|---|---|
| Number of business partners with more than 50% iSense | 20 |
| Number of business partners with more than 20 devices and 50% of their systems iSense | 15 |
| Total Number of iSense devices within this group | 1260 |
| Total Number of devices in this group | 1794 |
| Percentage of iSense devices compared to the total | 33.77% |

Table 3: Issue type with the statistics and category of each type

The questionnaire allows the participant to not answer a question if they deem their knowledge insufficient to rate said business partner. This is to ensure that a guess is done rather than the truth according to them. This does not affect the training data, because the group of participants is broad enough to obtain enough ratings for each partner. In combination with the interviews, the goal is to get input from 20 different parties. Under the assumption that each participant on average rates half of the business partners, we would have 150 input ratings for the model to train on.

### 4.3.2 Data

The questionnaire asked interviewees to rate business partners over the last year. The performance of a business partner is steady according to interviewees over a year, with a few peaks and dips. The questionnaire should be conducted every three months to obtain more data for the models to train on. This means for the validation of this research only one questionnaire is conducted. As mentioned there were 148 responses, which is the input set for this thesis. In future every three months a questionnaire or other method should be conducted to keep increasing the input-set for a model to train on.

All features use data from the specified period. The model determines which time-period is the best, which is done in a later section. The features found in this section that are related to issues (count and duration) are only from the period given to the model. The amount of stores, devices, gates are based on data from the start of the system, however within the period this amount could show an increment. Physical training is updated on the day a physical training has been completed, therefore can fall within a new period. E-learnings is the total amount of e-learnings completed that are currently active. Some e-learnings are outdated and are updated in the system as outdated. The feature excludes e-learnings that are no longer relevant for the time-period.

### 4.3.3 Results

Each interviewee has been asked to fill in the questionnaire during the interview, so they could elaborate their ratings to obtain a little more contextual information. Besides the interviewees, another ten participants have been asked to fill in the questionnaire. The total amount of ratings obtained is 148 and the business partner that has been rated the least is 4 times. The results of the questionnaire can be seen in table 4.

| Business [2] Partner | Minimum Rating | Maximum Rating | Average Rating | Res-ponses | Standard Deviation |
|---|---|---|---|---|---|
| Business partner 1 | 3 | 7 | 5,86 | 7 | 1,491 |
| Business partner 2 | 7 | 9 | 7,73 | 13 | 0,720 |
| Business partner 3 | 5 | 7 | 5,70 | 5 | 0,872 |
| Business partner 4 | 5 | 9 | 7,50 | 14 | 1,082 |
| Business partner 5 | 5 | 8,5 | 6,70 | 10 | 1,003 |
| Business partner 6 | 4 | 8 | 6,42 | 12 | 1,068 |
| Business partner 7 | 5 | 8 | 6,62 | 13 | 0,842 |
| Business partner 8 | 4 | 8 | 5,96 | 12 | 1,130 |
| Business partner 9 | 5 | 8 | 6,50 | 12 | 0,891 |
| Business partner 10 | 5 | 9 | 7,33 | 6 | 1,724 |
| Business partner 11 | 5 | 9 | 7,67 | 6 | 1,247 |
| Business partner 12 | 2 | 7 | 5,29 | 12 | 1,388 |
| Business partner 13 | 5 | 6 | 5,25 | 4 | 0,433 |
| Business partner 14 | 2 | 7 | 5,08 | 12 | 1,622 |
| Business partner 15 | 5 | 8 | 7,14 | 11 | 0,867 |

Table 4: Business partners ratings from questionnaire and interviews

While some business partners are rated fairly similar with a potential outlier, there are some business partners which opinions are fairly different, for example business partner 9, where the average is precisely in the middle of both the minimum and maximum. This is one of the main challenges of the model, since it needs to determine which of the rating is correct, if at all correct.

### 4.3.4 Summary

This section described the questionnaire that has been conducted. We discussed what the goal of the questionnaire was, what the questions in the questionnaire were and the results that came forward. Next section concludes the features found in this chapter.

[2] Business partners have been anonymized due to privacy concerns

## 4.4 Data preparation

In order to include some of the features that come forward in this section, data needed to be prepared. The amount of gates is a data row that was recently added to the system. Due to the firmware some systems did not report the amount of gates to Nedap and this means there are a few devices that have no amount of gates reported. Discussion with experts about the amount of gates resulted in taking two gates as a default value. The minimal amount of gates a system needs to have is two, the assumption made here is that the retail stores that do not have the right firmware are likely the smaller retail stores. The average amount of gates in the field was 2.65 and the experts that the larger systems have reported their gates. Which is why two gates would best represent the situation.

Currently, global customers are not stored in a database. However, sales know the names of the retailers that are global retailers. To make sure the data can be split in the two categories an additional column has been added to the local data set that matches the string names of the global retailers.

The amount of stores and devices also needed to be limited. In the data set are also devices that are no longer active or stores that have been blocked. These needed to be filtered out to only show the relevant devices and stores. Secondly, the data set also contained OST systems. Both these situation have been filtered out in the data set by using the existing infrastructure and fields to filter them.

## 4.5 Conclusion

This section discussed three different techniques to obtain feature definitions. In table 5 an overview of all features can be seen with their types. These features are used in the model that is created in next section. The features are identical for both global and local retailers, except they are divided in the two categories to determine whether business partners treat either better than the other. The issue duration and count are repeated three times for each of the mentioned severity category.

| Feature | Type | Value Definition |
|---|---|---|
| Amount of stores | Integer | Total amount of iSense stores |
| Amount of devices | Integer | Total amount of iSense devices |
| Amount of gates | Integer | Total amount of iSense gates |
| Physical Training | Integer | Years since last training |
| E-learning | Integer | E-learnings completed |
| Issue duration high severity | Integer | Total minutes a system is down |
| Issue count high severity | Integer | Total amount of issues |
| Issue duration medium severity | Integer | Total minutes a system is down |
| Issue count medium severity | Integer | Total amount of issues |
| Issue duration low severity | Integer | Total minutes a system is down |
| Issue count low severity | Integer | Total amount of issues |

Table 5: An overview of all candidate features, these are repeated for local retailers and global retailers

Next section is going to describe the approach in creating a model to predict the performance of a business partner, the techniques that are considered, the criteria for choosing a model and concludes with a description of the chosen model.

# 5 The Model

This section describes the choices made in creating the model. First, we discuss the approach that is used to create the model, the reason for using this approach and what the pitfalls of this approach are. Following this, we discuss the different models within this method, the advantages and disadvantages of each model and conclude by elaborating the chosen model.

## 5.1 Approach

The goal of this thesis is to predict the performance of business partners based on a set of indicators. Chapter 2 discussed machine learning, which is the model that is going to be used. The reason for using machine learning is that the goal to predict the performance of a business partner is best executed by machine learning as discussed in chapter 2. There are many languages that support machine learning, however for this thesis we use Scikit, as the models have been validated in other researches and Python is a known language to us [43]. Previous section discussed the questionnaire that was conducted to obtain ratings of a set of business partners, these ratings are needed to predict new ratings for all business partners. Since predicting is one of the goals of this thesis, supervised machine learning is used. Supervised machine learning requires labeled data to work and the input of the questionnaire can directly be used as labels for the model.

Chapter 4 described the indicators that came forward during data exploration and interviews. These potential features are used in the machine learning model. Chapter 2 discussed feature engineering and the steps that fall under feature engineering. Machine learning determines, which of the features give insights in the performance of a business partner.

As mentioned by Fayyad et al, besides predictive analysis, the second goal is to understand the prediction and show how a business partner can improve their rating [16]. In section 2, supervised and unsupervised machine learning was discussed. Since the goal of this thesis is to map the perception within Nedap to a model, supervised learning is used. Supervised machine learning can be split in two separate categories: black-box and white-box. A black-box model shows no understanding in how it came to its predictions, whereas a white-box model gives insight in how it got to its predictions. For Nedap a prediction is irrelevant when it cannot be explained, but the accuracy of the rating still has to be high. To not limit the model to only white-box models, a dashboard is created which shows the rating of the business partner in question and all features with their values, relative position (ranking compared to other business partners) and the time-window. This is to ensure that if the model were to make an incorrect prediction, it would show the used data and the rating can be ignored. This dashboard also shows the respective weights of each feature as determined by the machine learning model. This allows us to freely choose between white-box models and black-box models.

The last choice made on the method of machine learning is the choice between a classification problem and a regression problem. During the interviews it came forward that differences between business partners can be small. Classification would mean that business partners are grouped in pre-defined classes. Since differences between business partners can be very small, we chose for regression. Regression models give a decimal precision prediction, making them more valuable than classification. Secondly, since the ratings used as labels differ a lot, as can be seen in table 3, a prediction in decimals will give a better representation of the truth.

The largest pitfall of this thesis is that we are translating opinions of employees within Nedap to data. There is a chance that the ratings given in the questionnaire are based on a gut-feeling and not on facts. When this is the case, the model is accurately predicting the gut-feeling. Secondly, since there is a limited set of business partners and a limited set of employees that have a perception on the performance of a business partner, the data set that the model is going to learn from is not immense. The issue data is millions of records, but since these need to be aggregated on a business partner, the issue duration and count are still singular values. While a small amount of data is not necessarily a problem, caution needs to be taken in using too many features and overfitting and under-fitting needs to be taken in account. Hawkins describes overfitting as following: if two features can capture the relationship in a data set, no more should be used [24]. Babyak goes a bit further saying that when taking it to the extreme, if the number of unknown in a model is equal to the amount of observations, the model will always fit perfectly [4]. Eric Cai mentions that an overfitted model captures noise instead of the correct predictors [9]. These claims are supported in the book by Burnham [8], who goes in more depth in explaining how to avoid overfitting. Underfitting is the opposite of overfitting and is the risk of using too little predictors. This results in the model not being able to capture the underlying trend [9]. Based on the literature and research conducted in this area, the definitions for overfitting and underfitting are as follows:

- Overfitting occurs when the model has too many variables or too little data to capture the underlying pattern
- Underfitting occurs when the model has too little variables or features are too generic to capture the underlying pattern

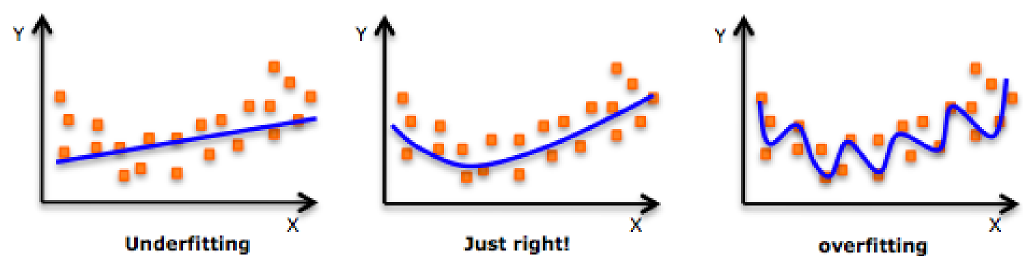



Figure 11: Visualization of overfitting and underfitting

It is important to have a correct fit. Overfitting and underfitting has been visualized in figure 11. The figure shows what a machine learning model does when it underfits and overfits. Ensuring a correct fit is a challenge that is going to be tackled by using different models and parameterizing the models well. Besides this, there are methods that help ensuring a correct fit such as feature reduction (reducing the set of features to exclude noise). These are discussed in later stages of this thesis.

### 5.1.1 Summary

This section discussed the choice of machine learning, the choice of not excluding black-box models, the reason for choosing a regression model instead of a classification model and lastly discussed some pitfalls that need to be taken in account when working with machine learning in this context. Next section is going to describe which machine learning models are used in order to predict the performance of a business partner.

## 5.2 Models

As discussed in last section, the model should be a regression model. This limits the amount of possible models. First, we determine a long-list which is a list with models applicable to this case. Following this, a short-list is determined on a set of criteria. The shortlist candidates are thoroughly tested and optimized. To find the best model, we use accuracy and consistency. The accuracy of the candidate models describe how well it can predict the test data and consistency describes the performance over multiple iterations. The consistency is good when the model is stable and not related to the received test/train data.

### 5.2.1 Data scaling

For regression models to work, it is important to scale data so that the model does not base its conclusions on the feature with most variance [50]. This is important, since the range of the defined features differ a lot. Some of the proposed techniques filter outliers. Filtering outliers is not wished for in our case, since if a business partner is under performing this business partner needs to be rated lower on that feature. The choice in this thesis has fallen on a MinMax scalar. A MinMax scalar transforms all data to the given range. This technique compresses the data in the center and outliers are still on the outer edges. The way the MinMax-scalar scales the data can be seen in figure 12. In the figure the range is 0 to 1. In our model a range of -1 to 1 is used.

### 5.2.2 Multiple perceptions

As mentioned in the questionnaire section, there are multiple different perceptions for a single business partner. Each of these perception is coupled to the same training data. The challenge for the machine learning algorithm is to find a pattern in the data that best matches the different perceptions. This was also mentioned as one of the biggest pitfalls of this thesis. The impact this has on the accuracy of a model is, is that the rating a model predicts will rarely match the expected value. The accuracy section goes into more depth on this aspect.

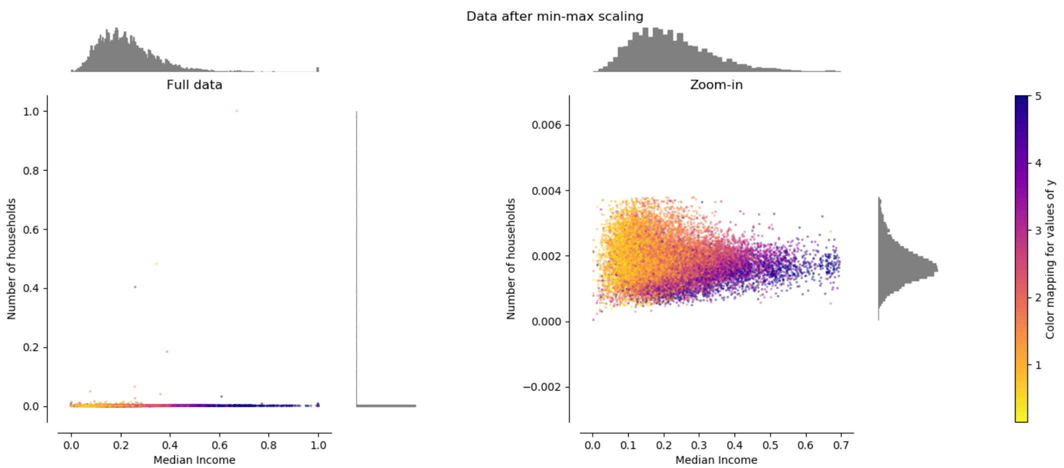



Figure 12: An example how a MinMax-scalar scales the data

### 5.2.3 Long-list

As mentioned before, Scikit is used in this thesis [43]. Scikit has Sklearn available in python, which is a library that has validated implementation of many machine learning algorithms. The long-list, that this section describes is the complete list of supervised regression models in Sklearn [52] as can be seen below:

- Generalized linear model
- Linear and quadratic discriminant analysis
- Kernel ridge regression
- Support vector machines
- Stochastic gradient descent
- Gaussian Processes
- Cross decomposition
- Naive Bayes
- Decision trees
- Ensemble methods
- Multi-class and multi-label algorithms
- Isotonic regression
- Neural networks

Each of the models mentioned above have multiple implementations. The scope of this thesis is to choose a model that meets the criteria necessary for Nedap. The ensemble methods mentioned consists of multiple algorithms, all of these use another classifier and run this multiple times. This means that if any of the models is able to fit, this model will also be able to fit. Since implementation of a model is very simple in Sklearn, every model is implemented and checked on whether it meets a set of criterion. These criteria are discussed below.

### 5.2.4 Criteria

The first criterion is that the model can fit itself with the amount of data that is available. Some models require a large amount of data in order to converge properly. These models cannot be included in this thesis, since the amount of data that has been rated is sparse. To detect that a model did not converge, two methods have been found. The first method is that the model will constantly predict the same value or close to the same value. The second method is that the models requires a very large amount of iterations, that the amount of data provided cannot ensure that each iteration has unique data. When a model sees the same data multiple times, it is very likely to overfit and therefore excluded.

The second criterion that is considered in choosing a model is the output of the model. Models are not restricted in their prediction. Some models fit the test data accordingly, however when the test data is not in range of the train data they predict values far out of the wanted range of ratings. These models cannot be included, because the output of a model needs to be a rating between one and ten.

If a model does not meet these criteria then it is excluded from the long-list. The way this is tested if by conducting an experiment to determine which models meet the criteria. For this Python and Sklearn are used and the model of the shortlist are implemented. For this, a data set is used from a period not used by model (but with the same features). This data is split up in two parts: the training data and the testing data. The models are going to fit on the training data and try to predict the values of the testing data. These predictions are looked at and if a model did not succeed to fit or predicts outside the range of output (not within 1 to 10) they are excluded. Next sub-section shows the result of this by showing an overview on what criterion the models were excluded.

### 5.2.5 Short-list

The biggest batch of models did not make the short-list, because they could not fit properly. These models are known to require a lot of data to fit. Since, the sparsity of data exists in this thesis these models are excluded. Some models have problems predicting values within a given range and gave results with large negative or positive values. Therefore, these models are also excluded since they did not meet the second criterion. Table 6 shows on what criterion the different models failed to pass.

| Model | Able to fit | Predict within range |
|---|---|---|
| Generalized linear model | X | |
| Linear and quadratic discriminant analysis | X | |
| Kernel ridge regression | X | |
| Support vector machines | | X |
| Stochastic gradient descent | | |
| Gaussian Processes | X | |
| Cross decomposition | | X |
| Naive Bayes | X | |
| **Decision trees** | **X** | **X** |
| **Ensemble methods** | **X** | **X** |
| Multi-class and multi-label algorithms | | X |
| Isotonic regression | | X |
| Neural networks | | X |

Table 6: Accuracy percentages by different cases

After exclusions based on both criteria, only two models remained, which are the regression tree and the gradient boosting regressor. The gradient boosting regressor is an ensemble method, which is commonly used. Both models have an underlying algorithm of a decision tree. Logically, a decision tree predicts within the range of values as it is based on the training data. Both models show promising results in understanding the underlying trend and predicting new ratings. These models are discussed in the next sections.

#### 5.2.5.1 Regression tree

A regression tree is the same as a decision tree, except that it allows its predictions are a continuous output (regression) allowing it to predict values that have not been specified in the training data [51]. A regression tree works by making rulings, to which each set of features is compared. A ruling always looks like the following:

$$\text{feature name} <= \text{value of feature}$$

Based on the features each individual sample is given a value true or false. True values continue left in the tree, where false values continue to the right side. This can be see in figure 13, where the ruling that has been made is that ”medInc” is lower or equal to the value 5.04. If this is the sample, the set of features will continue left, otherwise they go the right side of the tree. Eventually a regression tree reaches the bottom where a value is assigned to the given sample. Figure 13 clearly shows that when the feature ”medInc” is high the sample is given a high score. These value assignments are called leafs and is one of the main methods to prevent overfitting in a regression tree. In Scikit there is an option to define the minimum amount of samples a regression tree needs to have in order to split into a leaf.

The way a regression tree predicts new samples is that the data of the new sample is held against the trained tree and the new samples will eventually reach a leaf, which assigns a rating to the samples. For a regression tree it is very important that data is always in the same range, otherwise the rulings will not work as the dimensions changed.

#### 5.2.5.2 Gradient Boosting Regressor

A gradient boosting regressor is an additive mode, which has a regression tree as underlying algorithm [53]. The difference between a gradient boosting regressor and a regression tree is that a gradient boosting regressor creates a regression tree and based on the loss, makes new attempts. Figure 14 shows how a gradient boosting regressor comes to its result. The figure shows an example where a gradient boosting regressor consists of 3 regression trees as can be seen in figure 13. Each of the regression trees has an error rate on how well it performed. Tree 1 gives this error rate to tree 2 on which it will try to perform better. This results

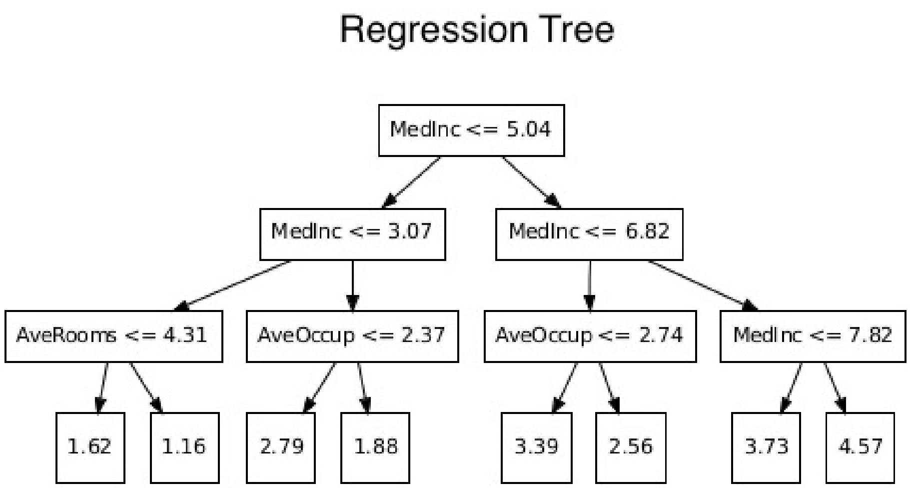



Figure 13: A simple regression tree

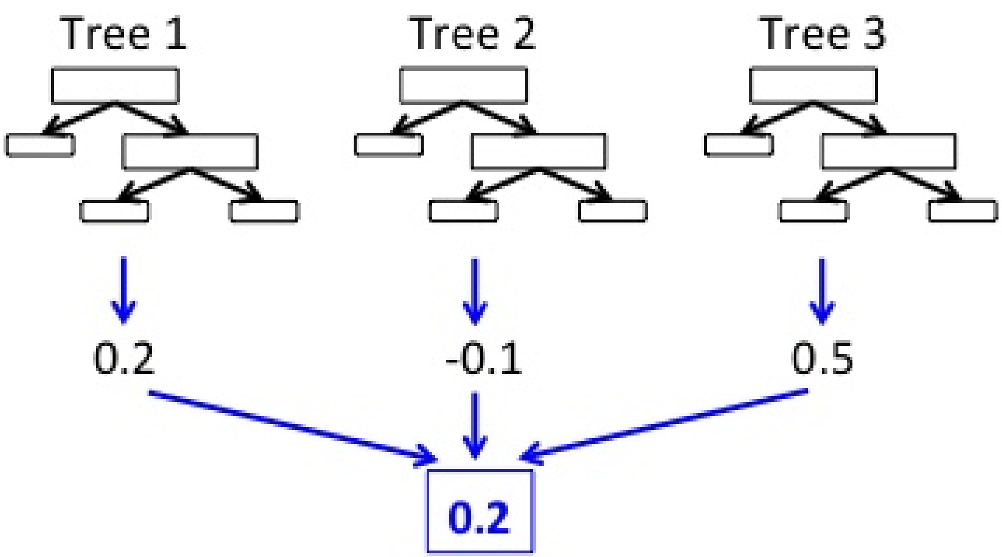



Figure 14: A simple gradient boosting regressor

in a different output and a new error rate. This new attempt is continued until the specified amount of attempts have been reached or the loss becomes larger multiple times in a row, in which case it stops early. When the model completed his fitting, it will use the average of all trees as output. Which is why a gradient boosting regressor is known to be very robust to overfitting as it continuously maps the loss to the next tree.

Since the underlying algorithm of a gradient boosting regressor is a regression tree, its predictions work the same except that it consists of the result of many regression trees. While this model looks very similarly to the regression tree, the predictions the model makes are very different.

#### 5.2.6 Summary

This section discussed the choice of scaling method in this thesis and why data needs to be scaled. Following this, the different models that are applicable to this thesis are discussed and two candidates came forward that met the defined criteria, regression tree and gradient boosting regressor. Next section discusses what criteria determine the final model and which of the candidates is chosen.

### 5.3 Feature selection

Before the choice of a model can be made, the candidate features need to be analyzed and determine whether each feature is relevant for the models to determine predictions on. The data consists of the features described in section 4. Since there might be correlations between the different features and the data is sparse for the amount of features. To allow the models to fit on the sparse data, the features are reduced this is done using Principle Component Analysis (PCA). Two definitions of PCA are as follows:

- ”PCA is a multivariate technique that analyzes a data table in which observations are described by several inter-correlated quantitative dependent variables” [1]
- ”PCA in many ways forms the basis for multivariate data analysis. PCA provides an approximation of a data table” [60]

PCA is used in thesis to reduce the amount of features to components that capture the maximum variance within these features. By reducing the amount of features, the amount of possible combination within the data is reduced, allowing models to fit easier. To ensure that the model is not under-fitted or overfitted the aim is to chose the right amount of components PCA gives. Jackson discusses stopping rules in his research and mentions that capturing 95% of the variance is very high [27]. The stopping rule that is often used is to have the PCA components capture between 80% and 90% of the variance. This thesis follows this stopping rule to reduce the chance of under-fitting and over-fitting.

The initial attempts of PCA to create new components indicated that some features were not helpful in generating a model to predict the performance of business partners, it just created noise. To determine which features created noise and need to be excluded, an experiment was created to exclude these features. The experiment used a regression tree to see patterns in what features were used and the accuracy determined whether a model was performing better with the exclusion of a feature.

For the experiment it is important to use a white-box algorithm so that the features are insightful. Therefore, a decision tree is used to gain insight in the features and which show correlation to the perceptions. The experiment starts by using all features and then printing a decision tree, in this decision tree the features should have impact on the output and increase the accuracy of the model. A forward approach is used, meaning that the model starts with no features and features are added and the results of the model are determined [36]. If features show to have strong impact on the probability distribution they are relevant, if the distribution does not change it is better to exclude them [22]. If the added feature is removed and shows no change in the probability distribution then a feature is irrelevant if the distribution shows difference and the accuracy increase a feature is considered relevant. This experiment is conducted in Python using the Sklearn framework. The Python script consisted of a decision tree with a maximum depth of 5. The maximum depth means that the decision tree is allowed to make 5 layers, the choice of 5 is made to keep the model insightful. Of the candidate features a total of four features were excluded, which are as follows:

- Amount of stores
- Amount of gates
- Amount of devices
- E-learnings completed

When adding these features the model did not show a difference in the probability distribution. The explanation for excluding these features is logical as well. In the current state these features heavily favor business partners that have a lot of stores, devices, gates, e-training completed. Since the amount of employees of a business partner is not known, there is no method to reduce these features to scale them according to the size of a business partner. In future including these features might be useful to determine the efficiency of a business partner. The model shows low accuracy when these features were included, since these features were not reduced by using the size of a business partner. When these features were excluded the model was able to differentiate the input to separate classes and the accuracy improved. Since scaling the feature is not possible in the current state, these features were excluded and the final model is using the following features:

- Physical training X years ago
- Issue duration low severity
- Issue count low severity
- Issue duration medium severity
- Issue count medium severity
- Issue duration high severity
- Issue count high severity

## 5.4 Final model

Previous section discussed the two candidate techniques on the short-list. This section discusses the two models and their performance. The performance of the algorithms is tested by splitting the testing data into a training set and a test set. The training set is used to fit the model on the data, the process of fitting is learning the samples by knowing the prediction. These predictions are based on the perceptions that have been received during the interviews and questionnaire. The models attempt to match these predictions to the data as good as possible.

To make the decision on what model performs best, two criteria are defined which are as following:

- Accuracy
- Consistency

The model that performs best on both criterion is going to be the model to rate the performance of a business partner.

### 5.4.1 Accuracy

In order to determine the accuracy of a model, the input data is split into two parts. This split is 80% training data and 20% test data. Each model is allowed to train itself on the training data until it fits according to its standards. Afterwards it tries to predict the unseen test data. Each of the predictions is then compared to the actual values, which are the perceived truths obtained from interviews and questionnaires. The predictions and actual values are compared based on the following formula, which calculates the error rate of a model:

$$Errorrate = \frac{100\%}{N} \sum_{k=1}^{N} \frac{|predictedvalue - actualvalue|}{9}$$

The absolute difference is taken, as it does not matter for a predictions whether it is higher or lower than the actual value. These should both be treated as equals when it comes to accuracy. As long as the difference is equal the predicted value should not matter in the accuracy. This difference is divided by 9, which is the maximum difference a model can predict as the minimum rating is 1 and the maximum rating is 10. The main reason for choosing this is that Floris and Tofallis mentioned MAPE (Mean Absolute Percentage Error) is biased to lower predictions [17][57]. Since the distribution is equal and always within the same range the difference between the minimum and maximum rating is best for this case. This difference is summed and divided by the total amount of test samples to get an average error rate. The accuracy is then calculated as 100% minus the error rate. Logically, the models should predict close to the actual value, so a high accuracy is expected. Necessarily this does not mean the models are predicting correctly, this accuracy is merely used to choose the best model of the proposed models. How well the chosen model is performing, is determined in the validation section by conducting another set of interviews. Table 7 shows the percentages given by the accuracy model. The chosen accuracy model always predicts the error margin from 0% to 100% as can be seen in table 7.

| Predicted value | Actual value | Error rate |
|---|---|---|
| 5.0 | 5.0 | 0% |
| 6.0 | 5.0 | 11.11% |
| 5.0 | 6.0 | 11.11% |
| 10.0 | 5.5 | 50.0% |
| 10.0 | 1.0 | 100% |

Table 7: Accuracy percentages by different cases

### 5.4.2 Consistency

In order to ensure a model is consistent in predicting values, multiple iterations with random states are done. A random state is the generation of random numbers in the models that are used in machine learning, this is called a seed [54]. By keeping a random state identical, the seeds the model is getting also remains the same. This is useful to optimize models and test different parameters. Once the models are optimized, the random state is not specified anymore meaning that every iteration is unique.

To check how consistent the models are multiple iterations are done. For each iteration, the accuracy is checked and the predictions the models are giving on not-labeled data. The model that shows the highest accuracy and the most consistent predictions is the model that is the most optimal for predicting the ratings of business partners. Nedap wants to predict the performance of a business partner and needs to have a clear definition of a rating. If a model is inconsistent in predicting ratings, there is no approach in letting a business partner know what they can improve.

### 5.4.3 Results

To make the final decision on the mentioned criteria, a total of 15 models were implemented in Python. These 15 models have each been optimized to the best possible result by keeping the random state the same. By keeping the random state the same parameters can be tweaked easily to determine the best setup for each model. Regression trees were significantly easier to optimize, since they are less

| Model | Loss function | Learning Rate |
|---|---|---|
| 1. Gradient Boosting Regressor | LS | 0.1 |
| 2. Gradient Boosting Regressor | LS | 0.05 |
| 3. Gradient Boosting Regressor | LS | 0.025 |
| 4. Gradient Boosting Regressor | huber | 0.1 |
| 5. Gradient Boosting Regressor | huber | 0.05 |
| 6. Gradient Boosting Regressor | huber | 0.025 |
| 7. Gradient Boosting Regressor | lad | 0.1 |
| 8. Gradient Boosting Regressor | lad | 0.05 |
| 9. Gradient Boosting Regressor | lad | 0.025 |

Table 8: The different gradient boosting regressors and parameters

complex than the gradient boosting regressor.

For the regression tree the interesting parameters were minimum sample leaf and maximum depth. Minimum sample leaf is the minimum amount of samples needed to create a leaf. Maximum depth is the maximum amount of splits a decision tree is allowed to make. It is common to not specify the tree depth, but limit the regression tree with the minimum sample leaf.

The gradient boosting regressor took more attempts and still there were enough parameters to not limit the gradient booster to a single implementation. Both methods have three different loss functions. A loss function is the method that an algorithm uses to learn the data set and punish itself on incorrect predictions. The algorithm tries to reduce the loss function as good as possible and when it has achieved its lowest possible loss it is considered optimized. Since the complexity of these three loss functions are severe and are hard to judge on a single case, we take all three loss functions in consideration. The gradient boosting regressor had another interesting parameter, which was hard to judge on a single case, which is the learning rate. The learning rate is how fast an algorithm is allowed to draw conclusion based on the given input, a higher learning rate allows a model to explore more in the beginning, which allows the algorithm to make harsher decision. To determine if our data requires a slow or fast learning rate needs to be determined over several iterations, therefore we also include three learning rates for the gradient boosting regressor. This brings a total of 15 models that are implemented and they have the following parameters:

| Model | Loss function | Minimum sample leaf |
|---|---|---|
| 10. Regression Tree | mse | 10 |
| 11. Regression Tree | mse | 15 |
| 12. Regression Tree | friedman_mse | 10 |
| 13. Regression Tree | friedman_mse | 15 |
| 14. Regression Tree | mae | 10 |
| 15. Regression Tree | mae | 15 |

Table 9: The different regression trees and parameters

| # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 9.80 | 12.88 | 10.34 | 10.53 | 10.80 | 9.54 | 11.11 | 10.12 | 11.40 | 10.64 | 10.72 |
| 2 | 9.85 | 12.84 | 10.39 | 10.41 | 10.85 | 9.49 | 10.96 | 9.98 | 11.45 | 10.51 | 10.67 |
| 3 | 9.88 | 12.85 | 10.33 | 10.51 | 10.86 | 9.54 | 11.10 | 10.09 | 11.41 | 10.63 | 10.72 |
| **4** | **9.75** | **12.65** | **10.10** | **10.31** | **10.97** | **9.37** | **10.99** | **10.17** | **11.16** | **10.60** | **10.61** |
| **5** | **9.75** | **12.66** | **10.06** | **10.47** | **10.72** | **9.44** | **10.97** | **10.19** | **11.19** | **10.48** | **10.59** |
| **6** | **9.86** | **12.67** | **10.10** | **10.57** | **10.89** | **9.54** | **10.95** | **10.19** | **11.24** | **10.52** | **10.65** |
| 7 | 10.79 | 12.91 | 10.04 | 11.62 | 11.48 | 9.14 | 11.16 | 10.36 | 11.78 | 11.05 | 11.03 |
| 8 | 10.80 | 12.90 | 9.94 | 11.47 | 11.48 | 9.02 | 11.06 | 10.31 | 11.67 | 10.82 | 10.95 |
| 9 | 10.69 | 12.89 | 9.88 | 11.50 | 11.42 | 9.05 | 10.94 | 10.27 | 11.75 | 10.94 | 10.93 |
| 10 | 9.83 | 12.85 | 10.37 | 10.40 | 10.88 | 9.54 | 11.04 | 10.04 | 11.32 | 10.54 | 10.68 |
| 11 | 9.82 | 12.83 | 10.37 | 10.50 | 10.73 | 10.15 | 11.15 | 9.88 | 11.48 | 10.52 | 10.74 |
| 12 | 9.83 | 12.85 | 10.37 | 10.40 | 10.88 | 9.54 | 11.04 | 10.04 | 11.32 | 10.54 | 10.68 |
| 13 | 9.82 | 12.83 | 10.37 | 10.50 | 10.73 | 10.15 | 11.15 | 9.88 | 11.48 | 10.52 | 10.74 |
| 14 | 10.65 | 12.82 | 10.14 | 11.25 | 10.83 | 9.31 | 10.97 | 10.28 | 11.76 | 10.83 | 10.88 |
| 15 | 10.83 | 12.78 | 10.14 | 11.90 | 10.19 | 9.31 | 10.97 | 10.28 | 11.39 | 10.97 | 10.87 |

Table 10: The different error percentages of the different models, the columns show the iteration number and the rows show the error percentage per model

Table 8 shows the different parameters for the gradient boosting regressor and table 9 shows the different parameters tested for the regression trees. The number on the x-axis shows the model that corresponds with table 8 and table 9. Table 10 shows the error percentage over 10 iterations. Each iteration is unique with a new set of seeds. The three averages that are bold are the have the lowest average error percentages. Interestingly, these are all gradient boosting regressor with huber as error algorithm [26]. The different learning rates have little impact, but a normal learning rate seems to have the highest consistency.

Looking at the lowest overall average error percentage the gradient boosting regressor, that uses huber as loss function, shows to be better than the regression trees. And since the average error rate is based on multiple iterations this shows that this model is most consistent in predicting performance. The huber algorithm is known to be very robust as it shows in the data, therefore the final model is a gradient boosting regressor with the huber algorithm as loss function [26]. The learning rate did not have much impact on the results, but since the error rate of the average learning rate showed the best error rate, this learning rate is used.

The error percentage of 10.59% means that if a perception by a random employee within Nedap is held against the prediction of the model, the model would be 10.59% off. So if the perception is a 6.0, the model would on average predict between a 5.4 and a 6.6. Next section discusses validation interviews are held, where new perceptions are held against predictions of the model to see how well the model performs on new data.

#### 5.4.4 Best time window

Section 4 discussed the choice of interval time periods. These interval time periods are discussed with experts, data rows from the same time period are compared and the desired output Nedap wants are taken into account. For each of the time periods the advantages and disadvantages are discussed and how it would affect the model that is created.

##### 5.4.4.1 Monthly interval

The first proposed method is to use data gathered every month. The advantages are that ratings change frequently and ratings are based on only the most recent events. The disadvantage of using this is that a business partner is only judged on the most recent events and that a bad month would directly result in a bad rating. Recent events could be very negative for a business partner, such as many new installations would result in a potential harder month and would be punished for this. Looking at the data showed many empty features for business partners. The advantages and disadvantages were discussed with the partner manager and technical operation team from Nedap and based on their input it was concluded to exclude this proposed time period. The main reasons given are as follows:

- A month could mean there is insufficient data for a feature
- Recent events are important, but not only recent events
- Everyone can have a bad month, as long as it goes up afterwards

#### 5.4.4.2 Yearly interval

The second proposed method is to take data from the previous year. The advantages of this is that every feature for all business partners definitely has a value. Besides that, it would show the long term performance of a business partner. The disadvantages are that if a business partner had 1 really bad month, the business partner would be rated low based on that month for an entire year. If an upward trend happens, it would be harder to detect and the past would still affect the business partner. Additionally, Nedap wants to predict the performance of business partners every three months, this would result in re-using 75% of the data on each prediction. The proposed time-period was also discussed with the same parties and it was concluded to exclude this time period as well. The reasons for excluding this time period are the following:

- Desired output is more frequently
- A lot of data re-usage
- Bad months would affect the business partner too long
- Change needs to be seen early

#### 5.4.4.3 Quarterly interval

The last proposed method is to take quartiles as time period. The advantage of using quartiles as time period is that there is a high certainty that every feature is filled for a business partner and the values of the features change frequently. This proposed method combines the yearly and monthly time periods. The disadvantage is that business partner with very little stores would still have too little data. These advantages and the disadvantage were discussed with the experts and it was concluded that this time period is the best representation of the performance. Their reasons were:

- Data shows a longer period, but still short enough that recent events have influence
- The smallest business partners can be discarded until they grow more
- There is little data re-usage

#### 5.4.4.4 Moving time window

Lastly, a discussion was brought up to use a moving time window of quartiles. This would mean that the performance of a business partner could be updated more frequently. We had a discussion about this and concluded it would be strange to use a moving time window. It was argued that it would skewer the data a lot, since the changes in the data would be little compared to the reused data. Therefore, the final choice has fallen on quarterly intervals.

## 5.5 Summary

This section discussed the choice of model. The section started off by describing the approach in choosing a final model. Following this, the relevant techniques were discussed and a short-list has been made with candidates for the final model. The section concluded by analyzing the short-list and come up with the best model to predict the performance of a business partner. The final choice is a gradient boosting regressor with a huber error loss algorithm. Next section discusses how the model is going to be validated and if applicable what adjustments have been made on the final model.

# 6 Validation of the predictions

This section describes the validation of the predictions made by the model created in previous section. First, the goal and the approach of the validation is discussed. Following this, the results of the validations are discussed by explaining to how they impact the created model.

## 6.1 Goal

In order to find out whether the predictions the model is given are a close representation of the truth, validation of the predictions are necessary. There are three main goals in the validation step of this research that are important to ensure the quality of the predictions. The goals are as following:

- How are the predictions compared to the perception of the truth of the 15 selected business partners?
- How are the predictions of business partners that were not in the questionnaire?
- If the rating does not match the perception, what does the data show?

These three questions are the main goals in the validation step. The first question translates to how well the model fitted. If the model is predicting values very different from the input, it could mean that the data was too sparse for the model to fit. However, if a few predictions are off it could mean that the model successfully found the underlying trend and correctly rates a business partner based on the data. The second question is to ensure that the model can also predict business partners that have not been included in the training data for the model. This step ensures the model did not learn to predict 15 chosen partners, but predicts every business partner according to the perception. The last question determines if a perception does not match the prediction, whether this is because the perception is wrong or because the model failed to learn the underlying pattern.

## 6.2 Approach

The validation step is conducted as following, another set of interviews are conducted. The goal is to determine whether the predictions of the model are correct according to interviewees. The interviews are structured interviews [28], where the interviewee is asked about business partners they have contact with and how their performance is in their expert opinion. First, the interviewee gives their rating of the performance of said business partner, based on three months prior to the interview. These ratings are then being compared to what our model is predicting. Once the interviewee has given about 10 ratings on business partners (business partners that were mentioned in the questionnaire and not mentioned business partners), the differences are discussed with the interviewee. The way this is conducted is by asking the interviewee if the model would give the business partner a rating of X, whether they think this could be correct or that they are surprised by the rating. These comments are used to look into the data given to the model and a check is done to see whether these concerns are correct or that the model has given a correct prediction.

Another advantage of doing the previous step is that another set of input ratings are generated. These new ratings are added to the model in future to improve its predictive value. This increases the amount of data that is available, allowing more and different models to fit.

## 6.3 Observations

A total amount of nine validation interviews have been conducted. Each interviewee was asked about business partners they have intensive contact. The interviewees are also asked to re-rate the business partners they rated in the previous interviews, but now over the last three months. Each interviewee can freely respond to predictions of the model and discuss their opinion about said prediction. In the result section, these events are hopefully shown to have an impact on the performance or the data validates that these events had no impact on the performance and the business partner handled correctly.

The first observation mentioned was that the amount of samples for a business partner has big influence on the rating the model predicts. Business partners with very little stores could have no issues reported in the data. The interviewees mentioned it would probably be better to exclude business partners that have less than 10 stores due to the scarcity of the data. This observation is taken into account for the results by not including business partners that had less than 10 stores. A business partner who has a few stores cannot be predicted yet, as several features can be empty. The model has been tuned to exclude business partners with less than 10 stores.

The second observation made is that business partners with a lot of stores are rated lower. Discussing this observation with the interviewee, we concluded that there were several business partners with a high store count that were in the top of the ratings. The logical relationship between size and amount of stores is that, a business partner with a high store count has more personnel to maintain the situation. Currently it is not possible to know the amount of employees a business partner has, but this could be a great addition in the future to determine the efficiency of a business partner.

The third observation by an interviewee from the technical team was that the complexity of each system might impact the results. He says that the amount of gates are not always filled in should have a default value. This had already been taken into account when creating the model, however the default value chosen was slightly low according to him. A discussion about this concluded that the amount of gates a system should have as default should be 3. This was discussed with the interviewees from earlier who mentioned the default value of two gates and they also concluded that 3 gates might be a better default value.

A fourth observation was by an interviewee from the US office. He mentioned that it could be interesting to see only new stores in the model, as the steady retail stores should be considered optimally. The discussion following this concluded with that it could definitely be worth looking into, however due to the scope of this thesis excluded. The scope of this thesis is to determine the performance of a business partner on all aspects and not only the new installations.

There were several observations about certain business partners either scoring too high or too low. These observation have been written down with the argumentation on why the interviewee deemed the rating wrong. The data of these business partners was shown to the interviewees after they rated the business partner and it was discussed whether the data showed in line with their opinion. It was seen in their data that their perception was often based on the combination performance of iSense and OST and not only iSense. This observation is discussed further in next sub-section.

## 6.4 Results

As mentioned in previous section, the interviewees were asked to give a rating to the performance of a business partner over the last three months. Once they have given their rating, we ask the interviewees about the prediction made by the model. The model has predicted the performance of all business partners during the three months leading up to the first set of interviews and ratings. The interviewees were asked whether the prediction of the model could be correct or that they believe it is totally wrong. If an interviewee said that it was totally wrong then the data is being looked at to see whether this is correct.

| Business Partner [3] | Predicted Value | Average Value | Responses | Error Rate |
|---|---|---|---|---|
| Business partner A-1 | 7.17 | 6.93 | 3 | 3.35% |
| Business partner A-2 | 5.54 | 6.12 | 4 | 10.47% |
| Business partner A-3 | 7.55 | 7.18 | 4 | 5.03% |
| Business partner A-4 | 6.83 | 6.00 | 3 | 12.15% |
| Business partner A-5 | 6.72 | 6.96 | 4 | 3.57% |
| Business partner A-6 | 6.16 | 6.43 | 4 | 6.70% |
| Business partner A-7 | 5.90 | 6.33 | 3 | 7.29% |
| Business partner A-8 | 5.50 | 5.33 | 3 | 3.09% |
| Business partner A-9 | 7.08 | 7.00 | 3 | 1.12% |
| Business partner A-10 | 6.71 | 6.52 | 3 | 2.83% |
| Business partner A-11 | 8.05 | 7.38 | 2 | 8.32% |
| Business partner A-12 | 7.47 | 7.50 | 2 | 0.40% |
| Business partner A-13 | 6.05 | 4.78 | 5 | 20.99% |
| Business partner A-14 | 6.19 | 5.13 | 4 | 17.12% |
| Business partner A-15 | 6.44 | 7.00 | 3 | 8.70% |
| Averages | 6.62 | 6.44 | 3.33 | 7.41% |

Table 11: The predicted rating of a business partner with the average perception of the interviewees

Interestingly, the predictions by the model showed very in line with the feelings from the interviewees. The results of the validation interviews can be seen in table 11. Since the input ratings ranged from two to nine, the model classifies within this range. The way this affects the results is that the range the model could predict in is smaller, therefore the accuracy is slightly worse on this scale than on a scale from one to ten. The average error percentage of 7.41% is better than the error percentage received in the results of section 5. This is likely caused by the outliers that could have been used in the training or test data. The model learned to filter these outliers, but if for the calculation of the error percentage these outliers were used in the test data, the model would be off a lot. This negatively affects the error percentage. An error percentage of 7.41% means that if a perception would be a 6.0 then the model would predict between 5.56 and 6.44.

The data definitely showed that the prediction was basing it on perception that exists within Nedap, however for some business partners this meant that the rating they were given does not match the actual situation. There were a few business partners were the interviewee mentioned that he/she thinks the rating was either too low or too high, such as business partner A-13 and A-14. This caused discussion between various interviewees whether the perception is actually true or that the business partner was actually better or worse than expected. The question that was asked various times was if it would be possible to create a model that predicts the actual situation and not on the perceptions. The way this would be done is by showing the features and their values to experts without knowing the business partner. By rating the rows based on the data, a machine learning model can be trained to use these rating as input data instead of the perceptions. This research has gained a lot of insight in what determines the performance of a business partner and the transition to a model that predicts based on the data should not be too hard, but falls out of the scope of this thesis. There were a few business partners that only received one or two responses, the complexity was that not every interviewee could answer what rating each business partner would get over the last three months.

[3]The business partners have been anonymized due to privacy, important to see is that the business partners from earlier do not match these business partners

## 6.5 Subjective mapping

The technique used in this thesis is to map data onto perceptions. The machine learning model has been trained to predict perceptions of the performance of a business partner. The reason to use this method for data understanding, is to evaluate the perceptions that exist. CRISP-DM has been discussed in the introduction and is a process that is repeated once understanding has been created. The predictions are used to create a model that correctly shows the performance of business partners. This method has advantages and disadvantages. The biggest advantage is that it creates discussion and shows employees that the perception might not always be correct. The biggest disadvantage is that it costs time to create a model that maps perceptions and is replaced by another model based on the obtained features. Personally, the advantages outweigh the disadvantages when time is not an issue. However, since normally time is expensive it might be better to perform feature engineering using the steps in this research and based on these ask interviewees to rate the data, rather than the business partner.

## 6.6 Summary

This section discussed the validation of the model. To achieve this interviews were conducting to get another set of input ratings and to discuss the predictions the model is currently giving. The interviewees had some observations on the predictions and a few small adjustments are made to the model. Overall the model was predicting in line with the opinions of the interviewees and if not it caused a discussion about whether the perception is actually true or not. Next section concludes the results that came forward in this thesis.

# 7 Evaluation

This section discusses the discussion that were raised on the results and the interviews following this. In CRISP-DM this step is transforming evaluation into business understandings. Each sub-section discusses an insight that helped Nedap to know more about their business partners, how it helped Nedap and what impact it has for Nedap and her business partners.

## 7.1 iSense vs OST

Currently business partners are making the transition from OST to iSense, but most business partners still install OST most often and handle issues related to the OST systems. This transition is upcoming and the iSense systems are making use of a new technology, RFID instead of RF. This shift in technology is the first insight that came forward in the validation section. The perception that exists of some business partners is mostly based on OST. Some business partners that were expected to be really bad, had more to do with the expected performance based on OST (as of this business partner most systems were OST). When the rating of iSense was a lot higher, the interviewees looked at the data and their own dashboards and came to the realization that the business partner was performing well on iSense systems. This proves the necessity of dashboard to give insight in all data available of a business partner, since currently the perception is not always related to the actual situation.

The second insight the discussion between iSense and OST provides is whether every business partner is ready for the transition to iSense. Some partners, which are scoring very high according to the perceptions are not performing that well. Discussion with various experts concluded that the partners are around for a long time and always implemented systems very well. However the new iSense system is very different and requires a shift from the mechanics to work with the new system. This shift can result in a business partner lacking engineers that know how to work with RFID systems and have problems understand the complexity of this. These business partners believe they understand the system well, but in practice these business partners lack basic knowledge about the differences and how to handle these. This insight caused a discussion within Nedap, if some partners are up to the challenge of working with iSense systems and can provide the required quality for the global retailers. If this were not to be the case, Nedap has to find another business partner in the country to work with the new RFID system and let the other partner serve the old OST system. This discussion is currently being investigated and information regarding performance on the iSense systems is very wished for in combination with information of the OST systems.

The way this affects business partners is that the information shows a business partner is struggling with iSense, while performing well on OST. These business partners need to improve and by gaining insight in their problem it should be easier to help them overcome the problems that they currently have.

## 7.2 Feature importance

During the interviews many features were mentioned as potential indicators of the performance of a business partner. The model indicates that some of the mentioned features are not related to the performance of a business partner. One of the features that turned out to have no impact on the perception of the performance was the physical training. During the interviews it was mentioned that the performance of a business partner should go up once they had a physical training. A physical training should refresh the knowledge a business partner has and be informed about the recent adjustments to the systems. However, during the validation the perception of the performance of some business partners was that they should score low, since they recently had a physical training. This indicates that a physical training does not necessarily mean an increased performance if the business partner does not understand the relevance of the discussed techniques. When speaking with the experts on this, they mentioned that sometimes a business partner attends physical training and the experts giving the training see that the engineers from the business partner have no idea how it exactly works. They still believe that training helps a business partner improve his performance, however it is not guaranteed by following a training.

Another feature that did not impact the performance of a business partner in its current state was the amount of e-learnings completed by a business partner. Currently it is not known how many engineers a business partner has, therefore the larger business partners are far more likely to have completed more e-learnings. One of the known larger business partners completed less e-learnings than others. According to experts, this could be related to the fact that they have more head-engineers to give trainings to the local engineers. Therefore e-learnings become slightly less useful, since experience in the field is always different to an e-learning. This invoked a discussion on how e-learnings could show an indication on the knowledge of a business partner. First, it would be important to know the amount of engineers a business partner has, which can be a task from the sales team to find out. Secondly, e-learnings could be seen as an indication of how many effort a business partner wants to put into knowing the iSense systems. Necessarily this means it does not affect the performance of a business partner, just the time they are willing to put into better understanding the systems of Nedap.

A business partner is advised to follow a physical training every year. Nedap wants to offer these trainings more often, since it helps business partners with the implementation stage. Members of the technical operation team say that once a business partner recently followed a physical training, they receive less tickets about issues that are occurring on systems. According to them these issues are resolved by the head-engineer, since his knowledge is up to date. Since tickets could be used as a feature due to the complexity of this system and the required changes, there is no data to prove this. However, according to technical operation members this can definitely be noticed in the quality of tickets they receive in Freshdesk. It would be interesting to see the impact of physical training on a feature related to communication, specially since communication was mentioned a lot during the interviews. However, stand-alone physical training showed no direct correlation to the included features and the components barely used this feature.

## 7.3 Global priority

To determine patterns that the model was basing its predictions on, the regression trees were analyzed. The best performing model was a gradient boosting regressor, which is a forest of regression trees. A common technique data scientist use to get insights in the machine learning models is to create a decision tree and analyze the patterns that come forward in a decision tree. The regression trees discussed in section 5 were close to the accuracy of the gradient boosting regressor. The patterns the regression trees showed were that principle components 1 and 2 determine the high-level splits as can be seen in the figure 15, figure 16 and figure 17.

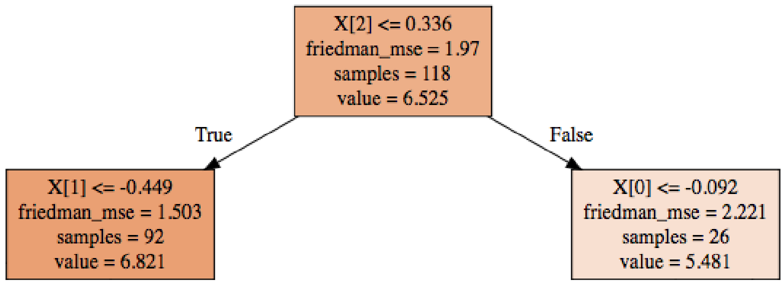



Figure 15: The regression tree with friedmans MSE as error function, the top two splits can be seen

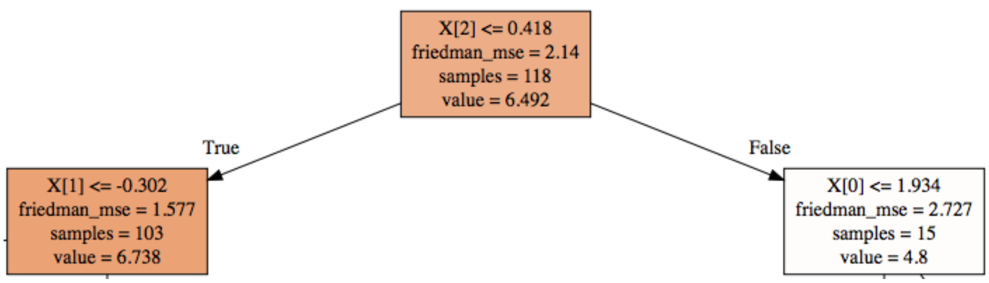



Figure 16: The regression tree with MAE as error function, the top two splits can be seen

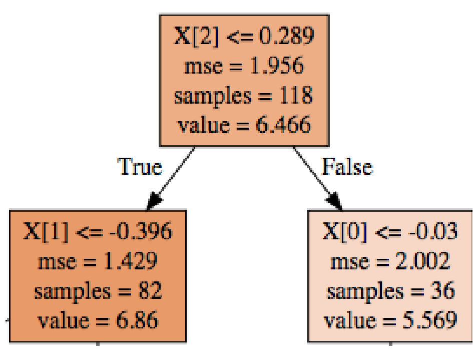



Figure 17: The regression tree with MSE as model, the top two splits can be seen

The corresponding principle components were analyzed by looking at the eigen vectors of each component. ”An eigen vector is by definition a transformation that represents a 1-D invariant subspace of the vector space in which the transformation is applied”, according to Balakrishnama [5]. These eigen vectors show how much a features had impact on the principle component that was used by the model. These eigen vectors can be seen in table 12.

| PCA Component | Physical Training | Low Issues Local | Medium Issues Local | High Issues Local | Low Issues Global | Medium Issues Global | High Issues Global |
|---|---|---|---|---|---|---|---|
| 0 | 0.177 | -0.049 | -0.043 | 0.736 | 0.452 | 0.220 | -0.099 |
| 1 | -0.284 | 0.109 | 0.091 | -0.270 | 0.094 | 0.655 | 0.728 |
| 2 | -0.187 | -0.249 | -0.131 | 0.399 | 0.340 | 0.467 | 0.413 |

Table 12: The principle components and their eigen vectors with the values of the different features

Principle component 2, which is used as the primary split on all three models, is determined by the high severity category of non-globals and all three categories of global retailers. Principle component 1, which is the secondary split is only determined by the medium and high severity of global retailers. The feature that many of the interviewees mentioned, that global retailers have priority, showed in the regression trees to be true. The components that determined the largest splits in the trees were related to the components that impacted the features of global retailers. Principle component 2 is also affected by the high severity issues of non-global retailers, logically these are the severest issues that can happen and they need to be low. Both components are impacted by the performance of a business partner on a global retailer and all three models determined that these components are the top splits for the regression tree. For Nedap this shows that the perceptions the interviewees gave are according to the model based on the performance of the business partner on the global retailers’ stores.

## 7.4 Offline stores

During the creation of the model a few business partners showed to have a lot of issues, which were not expected. When looking at these business partners and which issues were not resolved the longest, it showed some retail stores were purposely set to offline. These retail stores only use the iSense systems to have theft detection (which still works), but data is not reported to Nedap. Secondly, while most functionality can work there is no guarantee that this is the case, since the systems do not sent their metrics to Nedap. Offline stores were never mentioned during the interviewees and therefore have not been included in the model as a separate feature. However, when this was brought up with several teams they mentioned that it would be really interesting knowing which business partners have a lot of offline stores. Ideally, every stores is reporting data and it is believed some business partners might not know certain systems are offline. By bringing this information to the business partners and retailers these systems can be brought online to ensure the quality of the systems. For business partners this could mean additional work in getting offline systems to communicate with Nedap. To include this feature in a future mode Nedap does not need to change anything to their infrastructure. Currently the data has a reported_at and resolved_at field, issues that are open for longer than 3 months could be considered offline and these stores could be summarized in a separate feature to determine the offline duration of their devices.

# 8 Conclusion

This section concludes this thesis by first reviewing the research questions that were posed in the introduction. Following this, section 8.1 discusses how this thesis could contribute to other companies and researches. Lastly, section 8.2 discusses the strength and weaknesses and section 8.3 discusses the recommendations for future work.

This thesis posed four research questions. Each of the research questions are concluded individually below:

**RQ1: What indicators define the performance of a business partner?**

This thesis did a data exploration and conducted interviews that provided candidate features that indicate the performance of a business partner. The list of candidate features is used in an experiment to determine which of the features provided insight in the performance of business partners and which features only caused noise. Based on these candidate features there is an understanding what the perception of the performance is potentially determined by. Some features could be excluded by the model if they show no correlation to the underlying perception. The list of features was the following:

- Physical training X years ago
- Issue duration low severity
- Issue count low severity
- Issue duration medium severity
- Issue count medium severity
- Issue duration high severity
- Issue count high severity

Following this, the actual perception needed to be addressed to get the perceptions from various teams within Nedap. For this the following research question was posed:

**RQ2: What is the performance of a business partner?**

A questionnaire was conducted in order to determine the performance of a business partner, the questionnaire asked employees from various teams to rate 15 business partners on a scale from 1 to 10. These business partners in the questionnaire were carefully selected based on the amount of iSense systems they have installed and what percentage of their systems are iSense. A total of 148 responses were obtained by the questionnaire, these ratings can be seen in table 4. These responses in combination with the candidate features provide the input for the model. To find out what the best model is the following research question was posed:

**RQ3: What is the best model to rate business partners based on issue data of the iSense system?**

Based on the candidate features and perceptions a model was created. Many different methods were analyzed and criteria were defined to determine the best model for this case. First, the choice has been made to go for a supervised learning model with a continuous output. The reason for choosing this is that the responses of the questionnaire are labels that can be used to predict performance of business partners. Secondly, the choice on a continuous output model is that the difference between business partners could be very small as came forward during the interviews. To make a decision on what the best model is a set of criteria were determined to which a supervised machine learning algorithm had to suffice. These were able to fit on sparse data and predicting ratings within the wished range of one to ten. Of all algorithms that were implemented, the regression tree and the gradient boosting regressor remained. These two models were implemented and scored on two aspects to determine a final model, which were accuracy and consistency. The final choice is a gradient boosting regressor, which is a ”forest of regression trees”, meaning that it uses a regression tree as underlying algorithm and creates many instances of this regression tree to find the average optimal tree. Following the creation of the model, validation were held to determine how accurate the predictions of the model were. Section 6 discussed this validation by asking interviewees again to rate business partners. These business partners were not entirely the same as in the questionnaire, since there were two goals. The first goal was to validate whether the model could predict business partners that were not seen before. The second goal was to ensure that for a new time period the model would predict the same business partners correct. For this the model was given a time-period of three months that was not used before. Based on this period interviewees were asked to rate a business partner over the last three months and the results were compared to the predictions of the model. This concluded that the model that was created predicted the ratings of business partners very close to the perception. However, while it is very close to the perception it came forward during observations that based on the data some business partners would deserve to score higher or lower than the prediction the model gave.

**RQ4: What insights does the produced model give about business partners?**

The model was able to map data onto the perceptions of interviewees and this generated insights for Nedap. There were a couple insights that came forward during the validations and these insights are discussed individually below.

Currently business partners often install the old OST systems, but the transition to the new iSense systems is steadily increasing. This means for a different kind of implementation and the engineers that install the systems require a different set of skills to get the systems optimally running. This transition is important to understand the difference in the perception and the truth as a business partner that performed well for years might not understand the iSense systems. This insight provided Nedap with the discussion whether every business partner is capable of installing iSense systems and that they might need to look for additional business partners.

During the interviews and data exploration many features came forward as potential indicators of the performance, however the model showed that not each of these features contribute to the performance. Some features simply were not ready to be added to the model, such as e-learning, since the systems supporting these features did not contain all necessary information. Some other features eventually showed to have no impact on the performance of a business partner, either because the feature had no correlation to a rating or because the feature did not say anything about the performance. These features were related to the size of a business partner (amount of stores, devices and gates) and had no impact on the perception of performance, because a larger partner did not imply good performance, neither bad performance. The included features provided Nedap with information on what is important for a business partner to perform on and also what changes need to be done to include more features in an updated model.

When exploring the context of business partner it already came forward that business partners have two categories of retailers; global retailers and local retailers. The difference between these two was discussed in section 3 and during the interviews the decision was made to duplicate the features for globals and locals. The model showed in the principle component analysis that the features that provided the biggest splits in the regression trees were related to the performance of a business partner on the global retailers. This insight supported the interviewees who mentioned that ratings are likely related to the performance of business partners on the global retailers. For Nedap this means that business partners need to perform well on the global retailers, while for business partners themselves the information on how well they perform on local retailers could be useful.

These insights provide new research goals, possibilities and projects which are discussed in section 8.3.

## 8.1 Generalization

This thesis is conducted at Nedap and the results of this thesis are applicable to Nedap's situation. However, the research produced various results that can help other companies, researchers and other parties to achieve similar results. This subsection discusses several parts of this thesis and how the techniques used could contribute in other companies, fields and researches.

The main technique that was used in this thesis was using opinions of employees as input for a supervised machine learning algorithm. The features that were found, relevant to the case in this thesis, were used by the supervised machine learning algorithm to predict new situations. The machine learning algorithm was able to predict opinions of unknown situation in most cases. Importantly, the cases were the model was unable to predict the opinion brought up a discussion with the person as to if his/her opinion was correct. Most interviewees tried validating their opinion by viewing the data of the features and comparing to this to a case they felt was better or worse. If the data then showed the opposite they started to think about if their opinions are actual correct or that it is more a feeling. This creates a discussion through the interviewees and produces relevance of features and whether some features actually have influence on the researched situation. Based on these discussions experts can better estimate the performance of a business partner on data that was previously not available or they did not think was relevant for the performance of a business partner. This causes experts to be more in line with each other.

Feature engineering by doing interviews proved very useful in companies with several teams. By interviewing various teams all features from different views come forward. These features might not all be applicable to the scope of the research, however they show what shortcomings currently exists within the company. Additionally, they create coherence among the different teams. The different teams all want to gain insight in the data and discuss what features should be included and excluded from a model. Besides the inclusion and exclusion of feature, the input of the model is based on opinions of the interviewees. Since a discussion is brought up about the various inputs received, the different teams start thinking why others have different opinions and whether their opinion is actually correct. This showed in line with research done by Sjoerd van der Poel to use domain expertise in combination with literature in data analysis [58]. Since in practise situations are different from what can be found in literature.

## 8.2 Discussion

Reviewing this thesis, strengths and limitations can be identified. First limitation was that there was no ”golden reference”. The research took place over a period of 6 months and in this period input data was generated, however this was merely of one moment with a validation step following this. The model would have been more accurate and precise when there were additional data gathering moments, but due to lack of time the data remained sparse. With more data more models would most likely be able to fit properly, allowing for a larger choice of models.

Another limitation was the way Freshdesk data was currently stored. The Freshdesk data would have provided an additional feature about communication, which was deemed an important aspect by multiple interviewees. As restructuring this data and updating all previous data would cost far too much time for this thesis, this data was excluded.

The final limitation was the Esper framework [15]. This framework is used to analyze the metrics reported by the iSense systems. In the pre-research it was discovered this framework had a bug, which resolves and reports issues at incorrect intervals according to documentation. This affects all issues in ways that there are issues reported faster than should be possible. As it affects all issues, this has not necessarily been a problem, however it needs to be resolved to get an optimal indication of the issues.

One of the strengths of this thesis was the large diversity in the interviewees. By taking a large diverse group from multiple teams across the world, the amount of information obtained about business partner performance was immense. This information, while not all included in this thesis, provides insights for future projects and how different teams rate business partners. Each team scores business partners partially on the same aspects, which were used in the model as a generic business partner rating. However, the aspects that were only relevant for a certain team could provide useful information for that specific team.

A second strength is that the CRISP-DM cycle was followed in this thesis. The thesis provided its results and based on these results many new projects can be done with the business knowledge obtained. The goal of the CRISP-DM cycle is to keep iterating and increase the quality of the started cycle which comes forward in this thesis. The methodology and executing of this thesis therefore followed the correct steps which ensured optimal use of the time available.

## 8.3 Future Work & Recommendations

During the process of the research a couple recommendations came forward. The first recommendation is a generic one, currently Nedap has a lot of data and much of this data is currently not insightful for most of the employees. This data has great value and should be made visible to Nedap, partners and retailers. The following recommendations go more in depth on what data could be the following development steps.

This thesis had the focus on iSense, however recommended is looking at the performance of business partners on the old OST system as well. Section 7.4 showed that there was a big difference in performance of partners on OST and iSense. This difference is important to have insightful, since a business partner that is very good on iSense could be a really good partner for Nedap in the transition, while a partner excelling at OST systems, might need additional training. This information is not only handy for Nedap, but also for her business partners as they currently have no overview on how well they are performing.

Sales employees and account managers mentioned in interviews that information on how business partners handle their global retailers is very important. The focus of this thesis was to get a general performance of a business partner, however for these parties the same model could be applied to their global retailer. The development step for this is relative small, since it would mean scoping the global retailers to the global retailer in question. This information would, besides the mention parties, also be very relevant for the global retailer themselves to confirm the performance is up to their standards. This information could also be used in future sales with big retailers to show that quality has been reached on similar retailers.

The model that was created can predict the performance of business partners based on the perceptions. Section 6 showed that the model succeeded in predicting the perception, but that the actual performance differs from the perception. Business partners can be trained and educated on their mistakes once they are factual, therefore it is recommended that a model is created that is trained on facts, rather than the perceptions used in this thesis. The way this would be done is by showing the features to interviewees and then asking them to rate the rankings without knowing the name of the business partner. The allows the interviewees to have a better perception of the performance. The model that comes forward from this method, would be able to predict performance of business partners on better estimations.

As mentioned in the limitations, a bug was detected in the Esper framework. As it affects every issue, it is recommended to find a solution for this bug so that issues are correctly being resolved and reported at consistent intervals. This would require the framework to be adjusted in a way that the current infrastructure still works. A solution has been discussed during the research, but due to the severity of the change it was postponed till after the research.

Section 4 discussed the communication as a feature. For this the way data is stored in Freshdesk needs to be changed. This project was too large for the scope of this thesis, but the information gain by adding this feature is huge. Therefore, it is recommended the changes to Freshdesk are done, so that this data can be included in the model. Additionally, this feature could also be used for other purposes including the current application used to determine how well the knowledge of business partners is.

The last recommendation is to store information on how the systems are configured. Each iSense system has configurations on how far the range of the sensors are and more. This data could be used to determine a standard configuration for stores with a certain layout once this is know. Additionally, this data could provide more information about optimal system performance. Business partners could have very high standards when it comes to detecting articles that the retailers have. This also means that the chance for a system to have issues is higher as the range in which the system is detecting is increased. This information can be taken in account in a future model, once this information has been gathered.

# References


[1] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.

[2] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of items in large databases. In *Acm sigmod record*, volume 22, pages 207–216. ACM, 1993.

[3] Ana Azevedo, Manuel Filipe Santos, et al. Kdd, semma and crisp-dm: a parallel overview. In *IADIS European Conf. Data Mining*, volume 8, pages 182–185, 2008.

[4] Michael A Babyak. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic medicine*, 66(3):411–421, 2004.

[5] Suresh Balakrishnama and Aravind Ganapathiraju. Linear discriminant analysis-a brief tutorial. *Institute for Signal and information Processing*, 18, 1998.

[6] Maria-Florina Balcan, Avrim Blum, and Santosh Vempala. Clustering via similarity functions: Theoretical foundations and algorithms. In *Proceedings of the 40 th ACM Symposium on Theory of Computing (STOC)*, 2008.

[7] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[8] Kenneth P Burnham and David R Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2003.

[9] Eric Cai. Underfitting and Overfitting. `http://www.statsblogs.com/2014/03/20/machine-learning-lesson-of-the-day\-overfitting-and-underfitting/`, 2014. [Online; accessed 15 November 2017].

[10] Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, and Rudiger Wirth. *CRISP-DM 1.0 Step-by-step data mining guide*, 2000.

[11] Hsinchun Chen, Roger HL Chiang, and Veda C Storey. Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 36(4), 2012.

[12] Manoranjan Dash and Huan Liu. Feature selection for classification. *Intelligent data analysis*, 1(1-4):131–156, 1997.

[13] Barbara DiCicco-Bloom and Benjamin F Crabtree. The qualitative research interview. *Medical education*, 40(4):314–321, 2006.

[14] Owen Doody and Maria Noonan. Preparing and conducting interviews to collect data. *Nurse researcher*, 20(5):28–32, 2013.

[15] Esper. Esper official homepage. `http://www.espertech.com/esper/`, 2017. [Online; accessed 18 September 2017].

[16] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11):27–34, 1996.

[17] Benito E Flores. A pragmatic view of accuracy measurement in forecasting. *Omega*, 14(2):93–98, 1986.

[18] Freshdesk. Freshdesk Homepage. `https://freshdesk.com`, 2017. [Online; accessed 13 November 2017].

[19] Guojun Gan, Chaoqun Ma, and Jianhong Wu. *Data clustering: theory, algorithms, and applications*. SIAM, 2007.

[20] Amir Gandomi and Murtaza Haider. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2):137–144, 2015.

[21] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.

[22] Mark Andrew Hall. Correlation-based feature selection for machine learning. *Department of Computer Science*, 1999.

[23] David J Hand, Heikki Mannila, and Padhraic Smyth. *Principles of data mining*. MIT press, 2001.

[24] Douglas M Hawkins. The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1–12, 2004.

[25] Jochen Hipp, Ulrich Güntzer, and Gholamreza Nakhaeizadeh. Algorithms for association rule mining—a general survey and comparison. *ACM sigkdd explorations newsletter*, 2:58–64, 2000.

[26] Peter J Huber et al. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

[27] Donald A Jackson. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology*, 74(8):2204–2214, 1993.

[28] Stacy A Jacob and S Paige Furgerson. Writing interview protocols and conducting interviews: Tips for students new to the field of qualitative research. *The Qualitative Report*, 17(42):1–10, 2012.

[29] Saint John Walker. Big data: A revolution that will transform how we live, work, and think, 2014.

[30] LUKASZ A. KURGAN and PETR MUSILEK. A survey of knowledge discovery and data mining process models. *The Knowledge Engineering Review*, 21(1):1–24, 2006.

[31] Ohbyung Kwon, Namyeon Lee, and Bongsik Shin. Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, 34(3):387–394, 2014.

[32] Doug Laney. 3d data management: Controlling data volume, velocity and variety. *META Group Research Note*, 6:70, 2001.

[33] Daniel T Larose. *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons, 2014.

[34] Steve LaValle, Eric Lesser, Rebecca Shockley, Michael S Hopkins, and Nina Kruschwitz. Big data, analytics and the path from insights to value. *MIT sloan management review*, 52(2):21, 2011.

[35] Gordon S Linoff and Michael JA Berry. *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons, 2011.

[36] Huan Liu and Lei Yu. Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on knowledge and data engineering*, 17(4):491–502, 2005.

[37] Óscar Marbán, Gonzalo Mariscal, and Javier Segovia. A data mining & knowledge discovery process model. In *Data Mining and Knowledge Discovery in Real Life Applications*. InTech, 2009.

[38] Andrew McAfee, Erik Brynjolfsson, Thomas H Davenport, et al. Big data: the management revolution. *Harvard business review*, 90(10):60–68, 2012.

[39] Katina Michael and Keith W Miller. Big data: New opportunities and new challenges [guest editors' introduction]. *Computer*, 46(6):22–24, 2013.

[40] NEDAP. NEDAP Retail. `http://www.nedap-retail.com`, 2017. [Online; accessed 09 September 2017].

[41] Nedap N.V. Nedap Groenlo. `http://nedap.com`, 2017. [Online; accessed 10 September 2017].

[42] Oracle. Oracle's definition of big data. `https://www.oracle.com/big-data/index.html`, 2017. [Online; accessed 17 September 2017].

[43] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(Oct):2825–2830, 2011.

[44] Sunil Ray. Data exploration and feature engineering. `https://www.analyticsvidhya.com/blog/2016/01/guide-data-exploration/`, 2017. [Online; accessed 23 October 2017].

[45] Nedap Retail. Device management manual. Internal document, 2017. [Accessed 10 October 2017].

[46] NEDAP retail. iSense. `http://www.nedap-retail.com/solutions/isense-lumen-rf-rfid-hybrid/`, 2017. [Online; accessed 09 September 2017].

[47] Philip Russom et al. Big data analytics. *TDWI best practices report, fourth quarter*, 19:40, 2011.

[48] SAP. Sap sees opportunities in big data. http://global.sap.com/news-reader/index.epx?PressID=19188, 2017. [Online; accessed 15 September 2017].

[49] SAS. Sas variety as another v. http://blogs.sas.com/content/iml/2014/11/19/coefficient-of-variation.html, 2017. [Online; accessed 15 September 2017].

[50] Scikit. Comparing effects of different scaling methods. http://scikit-learn.org/stable/auto_examples/preprocessing/plot_all_scaling.html, 2017. [Online; accessed 20 November 2017].

[51] Scikit. Decision Tree Regressor. http://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html#sklearn.tree.DecisionTreeRegressor, 2017. [Online; accessed 21 November 2017].

[52] Scikit. Different supervised learning methods in machine learning. http://scikit-learn.org/stable/supervised_learning.html, 2017. [Online; accessed 20 November 2017].

[53] Scikit. Gradient Boosting Regressor. http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html, 2017. [Online; accessed 21 November 2017].

[54] Scipy. Numpy random state explanation. https://docs.scipy.org/doc/numpy-1.13.0/reference/generated/numpy.random.RandomState.html, 2017. [Online; accessed December 1 2017].

[55] Colin Shearer. The crisp-dm model: the new blueprint for data mining. *Journal of data warehousing*, 5(4):13–22, 2000.

[56] Pang-Ning Tan et al. *Introduction to data mining*. Pearson Education India, 2006.

[57] Chris Tofallis. A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, 66(8):1352–1362, 2015.

[58] Sjoerd van der Spoel, Chintan Amrit, and Jos van Hillegersberg. Predictive analytics for truck arrival time estimation: a field study at a european distribution centre. *International journal of production research*, 55(17):5062–5078, 2017.

[59] Jonathan Stuart Ward and Adam Barker. Undefined by data: a survey of big data definitions. *arXiv preprint arXiv:1309.5821*, 2013.

[60] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

[61] Rui Xu and Donald Wunsch. Survey of clustering algorithms. *IEEE Transactions on neural networks*, 16(3):645–678, 2005.

[62] Zdenek Zabokrtsky. Feature engineering in machine learning. *Institute of Formal and Applied Linguistics, Charles University in Prague*, 2015.

[63] Jasmine Zakir, Tom Seymour, and Kristi Berg. Big data analytics. *Issues in Information Systems*, 16(2), 2015.

[64] Paul Zikopoulos, Chris Eaton, et al. *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media, 2011.

# A Appendix A

## A.1 Interviews

This section is going to describe the information obtained in each interview. When all interviews are conducted a conclusion will be drawn on what features will be included in the initial model and the training set of the model will partially be filled with data from the interviews.

### A.1.1 First interview

The first interview was with a sales director of Nedap Asia and Pacific. In Asia they mainly work with iSense as an EAS (Emergency Alert System). At the start of the interview he mentioned that the performance of a business partner is based on the amount of complaints they receive form their customers and mainly their global client. Each global client has their own requirements, however these are the same for each country worldwide, making them identical for each business partner. The amount of complaints these global clients register on a business partner could indirectly imply the performance of said business partner in that country. He also mentioned that a client has specifications on how the system should look like and how it needs to be installed and if they do this well, there should be no complaints.

When speaking about contextual influences, he mentions that in Asia it is not uncommon for cables to interfere with systems, making them perform worse. He says that the business partner cannot do much to change this, however is not sure whether this should be considered for the model, it is more of a hunch. To add to his earlier statement, he believes that global clients should be prioritized by the business partner to ensure that the performance of said global client is as good as it can be. It does not mean that smaller clients do not matter, however when a choice needs to be made between the two, the global client should have priority. Lastly large stores have a higher chance of breaking down, since more customers means more random occurring events.

When asked if duration of an issue or the amount of issues is more important, the answer was that they are both important, however the first matters on how the business partner communicates with the retailer. If the business partner lets the retailer know he is working on it, it is deemed more acceptable to have a longer duration. Many issues mean that the system is never really stable. How the business partner handles the communication with the retailer is one of the most important aspects according to him. What makes a business partner perform bad is when they do not understand the escalation model. Nedap has a system that when the engineers cannot solve the problem with the system they can call Nedap, however in practise this happens more than it should, since business partners find it easy when Nedap solves their problems.

### A.1.2 Second interview

The second interview was with a member of the technical operation team of Nedap Retail. Technical operations are responsible for the following aspects:

1. After-sales
2. Education en communication
3. Testing development and firmware with the R&D team
4. Give business partners support if necessary
5. Share knowledge with other teams

The huge difference with the first interview is that this person knows a lot more about the technical aspects of the iSense system. This means some technical questions will be asked, which will help creating the model. When asked what defines the performance of a business partner he says that the following points define the performance of a business partner:

1. Response time towards clients
2. Knowledge level of the different systems
3. How willing they are to learn new techniques and how they get these to the market

4. Thinking about future-proof solutions
5. Give feedback about the system
6. Ask for training when they lack the skills
7. Properly install systems without to much issues

According to him every retailers should be equal, but in practise the opposite is true. Culture plays a big role there, because some countries are very strict while others do not mind this. He also says that global customers of Nedap should have a small priority, however business partners need their local customers in order to stay alive, therefore it is important that both are satisfied with the performance of the business partner.

Since technical operations also work on the system, some more detailed questions were asked for the severity of the different issue types, how these could should be scored and whether the responsibility matrix was correct. He says that it would be nearly impossible to score individual issue types and categories should definitely be used to make it easier. What the definition of these categories are is not certain yet, but discussion about this will follow later. He says that three categories should be sufficient for the scoring of the issues, which are:

1. Low impact
2. Medium impact
3. High impact

He believes that the size of a retailer should not matter in the performance, however sometimes it might be the case that small retailers do not want to pay for servicing and business partner will neglect these retailers most of the time, making for high issue durations. He finishes the interview with saying that according to him, the most important feature for a good performance is the level of knowledge. The better they can handle issues and the less help they need from Nedap the better.

### A.1.3 Third interview

The third interview was with a member of the sales and global account manager team. This team is directly responsible for contracts, negotiations and prices of the system. They have the final say in contracts and agreements. They use the business partners of Nedap in order to achieve the execution of the contracts. The interviewee is mainly responsible for one of the biggest retailers of Nedap and this retailer recently made the transition from OST to iSense, the newer system.

From his perspective the most important aspect of a business partner is that they keep the retailers satisfied. What satisfies a retailer is very different and hard to get to paper, as some wants to pay a lot for top quality services and some want to pay as little as possible and still have these services. Not only this but often a retailer expects business partners to do stuff that falls completely out of their scope. He says that the retailer does not know too much about the technical aspects of the system, as long as it does not bother their customers, they do not mind. Issues that have a bad influence on their customers are horrendous.

He says that global customers should definitely be separated and prioritized. The so called local heroes are important for the business partner, but far less for Nedap. The global customers drive the market and if these are served well, more retailers will make the transition to Nedap. Therefore these should be more important. A change of business partners could be necessary if a business partner cannot meet the standards Nedap wants.

Where a business partner is stationed is also very important, since some countries are lazy with cables and interference. Every culture is different, that in some cultures business partners have it far easier to make the retailers happy than in others. According to him business partners are bad when they treat their retailers bad, want high margins and service prices etc. How long issues stand open is definitely important, since this shows the experience and service of the business partner. Some of the business partners are so used to OST that they need to make the transition to iSense and lack training and experience, even when not asking for help. Communication is very important for this and letting a retailer know an issue is being worked on is professionalism. Communication is one of the key aspects of performance and secondly living up to the realistic promises being made.

### A.1.4 Fourth interview

The fourth interview was with the old business partner manager of Nedap, which is probably the most experienced person when it comes to the performance of business partners. According to her it is important that installations of the systems, their sales vision (how much of they sell new systems and not fall back to the old systems) and how they handle the after-sales (issue handling of the systems). She mentions that every store is equal and the size of stores do not matter, because small stores (customer-wise) might have significant higher valued products. It is important that a business partner strives for perfection and that he is not only interested in getting his money, but also keeping his retailers satisfied. Communication towards the business partner is important. The retailer is always aware that something is wrong with the system, however if they communicate this to the business partner is a problem. She also says that for Nedap global client are more important, however looking at the future this should not be case. Reputation of the business partner in the local market is just as important.

Furthermore she agrees with a lot of the points being made by the previous interviews and says it depends a lot from what perspective you are looking, but according to her after-sales, installation and communication are the most important parts. She adds that a business partner will most likely be scored badly when they do not communicate with Nedap when they have a problem. She believes some business partners do not ask questions often enough and lack training. Therefore a lot of the systems still break.

### A.1.5 Fifth interview

The fifth interview was another member of the sales team. She started off by giving a large summary of what she believed was important for my research by my initial mail, so the structure of the interview was a little different.

She started off by mentioning that besides asking different employees of Nedap I should also speak with Business Partners and retailers. This is because they probably have a different view than all the teams together. From sales out she is trying to get new clients and if they are global clients they will be passed on to an account manager. Based on this they try to make a standard. The business partner will execute the made standard and based on this handle service and installation.

Afterwards there is barely communication with the clients anymore, since this is passed on to business partner and retailer. Every part of the world has different standards making it harder to have a general standard, at least for now. In some cases the installation of iSense is just a side-job for a business partner and their largest part of income is from other systems unrelated to Nedap. Retailers only complain at Nedap when it is too late for a business partner to fix and she often goes to meetings with retailers (important ones). She asks here if they are happy or not and often when the business partner is at the meeting they are more positive compared to when the business partner is not at the meeting. She recommends speaking with someone from Nedap who tags along at some installation of business partners (important clients and in the beginning of a business partner), since he can give more details to what is important for the business partner.

She says there is a huge difference in technical supports of business partners. Some take responsibility for the system, which means that they perform well, whereas others do not mind if a system does not work well. She says that some business partner try to get the performance of the system is also based on the amount of detection the system does. Some business partner push for really high percentages making them detect problems more often, which can be a good thing. She says business partners sell on expectation management, when a business partner says the system will work 80% the retailer accepts this and then this standard is fine. Which makes it harder to detect what is good and not, since this is not stored. Every store should be treated equally according to her, specially from a sales perspective. However contradicting this statement she says that the business partner should focus on global clients, since these are very important for Nedap. She says that by doing this the business partner also has advantage from Nedap, since they are giving them clients and this happens a lot lately. She does not know whether the system is working optimally and says that a dialog with a business partner could prove useful to detect whether they can work with the system or see massive failures that cause certain issues to be open longer.

### A.1.6 Sixth interview

The sixth interview was with the sales guy in the US. He is a big believer of how more problems the better the performance of a business partner. According to him retailers are more satisfied with a business partner when they see mistakes being solved and know that they are being valued. When no issues arise they tend to believe that something is wrong with the system or that the system is not installed correctly. A big part is how to handle the solving of these issues, communication is key here. He says first in, first out should be the system for business partners to solve issues, however currently global clients will have priority due to importance. He does say that for a scalable future this system needs to be first in, first out. He says that store-size does not matter and each store should be equal. He adds that interference is more common in some countries, which makes the location of a business partner important. According to him the most important aspect of performance is that they are true to their word and SLAs. He would not mind business partners not communicating if they solve issues really fast, according to him communication is important, but if they can solve issues a lot faster with communication this should be the priority. A business partner would score bad if the retailer feels under-valued. This could be caused by letting them know issues are being resolved, while in practise this is not the case.

### A.1.7 Seventh interview

The seventh interview was also with someone of the technical operations team. We asked him to not discuss answers with his colleague in order to prevent bias. According to him performance of a business partner is defined by the amount of product they install for Nedap, the amount of times they ask questions and whether these are repetitive or not and if they handle issues themselves or ask external parties. Amount of installations is related to the amount of stores and questions to the amount of tickets in FreshDesk. However the last aspect is immeasurable according to him. A lot of the business partners are still used to the old systems and not to IT-centered applications, which is why he thinks they perform less.

When asked about culture he believes it definitely matters how the business partner reacts to situations, however according to him it should not matter. A business partners needs to perform at his best and meet the agreements he made with his retailers. These agreements might differ a lot depending on the country. Business partners also handle situation differently based on culture, in Europe they let systems escalate first before asking for help, in Asia they only say yes we can handle it when they cannot and in the US they do not think out of the box.

He says that while some issues are not primarily the responsibility of the business partner to solve, it is still important that if they are standing open too long a business partner should contact the retailer to solve these and if necessary assist the retailer. For his team it is not relevant how the contact between retailer and business partner is, as long as issues are being solved within reasonable time. He says that in theory every retailer should be treated equally and that first in, first out needs to be the technique. However in practise it really depends on the retailer, some retailers do not mind slow servicing if they pay less, while others want to pay the highest amount to have an up-time close to 100%. This depends on the agreed upon SLAs. The general performance on handling issues is one of the most important aspect according to him. He would rather see issues that are recurring to be registered as one issue instead of multiple. How more pro-active a business partner is the better and if business partners are not learning from their mistakes it makes them bad.

### A.1.8 Eighth interview

The eight interview was with an account manager of two large global clients of Nedap. He started the interview by saying that the most important aspect of the performance of a business partner is their knowledge level both on sales and technical. The technical aspect he defined here as the installation of the systems and the after services on how they solve the issues that come up. He says that the level of speaking English is a big part of the performance of a business partner. The Nedap installation wizards are in English and if the technician does not understand the wizard it will most likely not be installed correctly. When he looks at a business partner he takes in account the amount of employees they have, so their capacity to solve problems and install new systems. He says that the portal Nedap has, offers a possibility to check the amount of technical staff that did e-learning, which could be an indicator on how well their knowledge level is.

He strongly believes that each business partner should be scored equally no matter the location of the business partner. Performance is performance and even though some factors are less important in certain regions. The size of a store should not matter as every store is just as important. The same should go for the global clients, however in practise most business partners will prioritize these global clients. He says that the input Nedap is getting, is mainly from the global clients since these are in straight communication with Nedap, where many local stores only communicate with the business partner. He also says that communication is very important if not the most important aspect of the work of a business partner. When the retailer know their issues are being resolved they are satisfied even if it might take longer. He also believes business partners prioritize flagship stores, since these need to perform perfectly and are often the stores that complain the most, where smaller stores from the same retailer might not complain.

### A.1.9 Ninth interview

The ninth interview was with a fellow colleague from the services team. They provide the infrastructure and systems that store the data that is used in this research and also allow business partners and retailers to use their platforms for information regarding the systems. When asked what defines the performance of a business partner he says that from a technical level it is based on how well issues are handled, how often the issues come up, how many gates are in a system and how many devices which is influenced by the gates. According to him it does not matter where the business partner is stationed as the systems do not change based on the country. There might be different circumstances, but that is something the business partner should have experience with coping. And he believes that every retail store is just as important, the systems are similar in every store and the same issues should appear in each store.

He says that a good performance is when a business partner is solving issues fast and they do not come back. Issues should not appear in the first place and a good installation of the system should help with not having issues. A bad performance is defined by the opposite of this, issues that are not resolved for a long time or a system that has a lot of issues. Issues that are not resolved for a long time indicate that either the business partner has not seen the issues or cannot resolve the issues. There is support available in case an issue is complicated, so there should be no reason for issues that stand open for long. When asking more information about the technical parts of the system he says that he agrees with technical operations about the three categories that exist in the system, except that he disagrees with the issues that are currently being placed in a certain category. He says that he believes that store sizes do not matter, however for business partners they might, but this should not affect the overall performance. He does not know if the amount of customers has an impact on the issues that appear in the system, but could be taken into account.

### A.1.10 Tenth interview

The tenth interview was with a member of the pre-sales team. The pre-sales team is partially responsible for ensuring installation of systems goes well and if necessary assist locally at the retailer. He says that for a business partner it is important they know what their technicians need to do and what makes a good technician. He says that environmental problems are very common and one of the main complications of a technician. A technician cannot be good if the partner only has a few stores, since it is something you learn over time and get better at every time. He says that the trainings Nedap offer should be done at least every two years, since it gives a business partner the required knowledge that they need. Shortly before a large installation they have the option to a speed-course (as a conference call) to be up to date on techniques for the installation.

A business partner need to have analytic insight, communicate with Nedap and the retailer. A retailer always needs to be up to date on the state of the system and how any problems will be tackled. Nedap never hears about problems with local retailers, only the business partner hears about this. Main-technicians need a significant higher knowledge level, since they give training to the local technicians that assist the main-technician. A business partner treats every retailer differently, since the income of every retailer is different and so are the requirements. Based on these required a business partner is treated. He strongly believes that a global retailer gets different treatment than a local retailer and a split should be really suggested. He does say that there are some issues that are close to impossible to solve by the business partner due to a fault in the system. These issues are still reported in the issue table and in future need a system to not be included in the model.