UNIVERSITY OF TWENTE.

Faculty of Behavioural, Management and Social Sciences

Forecasting patient arrival at an emergency department

Bastiaan Kauffeld

M.Sc. Thesis
5-22-2023

Supervisors:
dr. A. Asadi
dr.ir. L.L.M. van der Wegen

Industrial Engineering and Management
Faculty of Behavioural, Management
and Social Sciences
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

This report is intended as a report for the execution of a Master's thesis as part of the Master Industrial Engineering and Management at the University of Twente.

University of Twente, Industrial Engineering and Management, PO Box 217, 7500 AE Enschede, tel. +31(0)534899111

Author

Bastiaan Kauffeld

Student number

S2011549

UT supervisor

dr. A. Asadi

Second supervisor

dr.ir. L.L.M. van der Wegen

External orginization

Diakonessenhuis Utrecht

External supervisor

drs. H. Bos

Date

22-5-2023

Management summary

Introduction

The healthcare sector is facing increasing demand due to a growing and aging population. This presents a challenge in maintaining high-quality care while keeping healthcare affordable and accessible. To address this challenge, the healthcare sector must work more efficiently, including the efficient utilization of resources such as staff and equipment.

The ED is a crucial part of any hospital. Patients who arrive at the ED often require immediate medical attention. To provide timely and effective treatment, it is essential to have an adequate number of staff members on hand. However, staffing can be expensive, so it is important to deploy personnel efficiently. Forecasting can help ensure that staff is available when needed, by aligning ED demand with staff availability.

Research goal

The goal of our research is answer the following question:

What are suitable forecasting models, for patient arrivals at the ED at Diakonessenhuis Utrecht, with error measures as low as possible?

We provided the hospital with a forecasting tool, which makes daily predictions of ED patient arrivals.

Methods

The hospital provided a dataset of daily patient arrivals consisting of patient arrivals on 1155 days, in the period from 1-1-2017 to 29-2-2020. We use this data to train and compare several ED arrivals prediction models we found by conducting a literature review.

We selected a group of (penalized) linear models, as they are intuitive, easy to implement and are amongst the best performing models in the field of ED forecasting. Additionally, we used a Random Forest to predict ED arrivals, which is one of the popular machine learning methods used by more recent research on ED forecasting.

Results

The RF model outperforms other models with a MAPE of 11.21, representing an 8.79% improvement over the baseline prediction. This baseline prediction is an estimate of the hospital's current forecasting method. The penalized linear models, including Lasso, Ridge, and Elastic Net, have similar performance to the RF model with MAPEs of 11.24, 11.24, and 11.25 respectively. The MLR model has a slightly higher MAPE of 11.67.

The most predictive variables for ED arrivals are weekdays and months. Mondays, Fridays, and Saturdays tend to be busier than other days of the week. Amongst the months, August and July have the strongest predictive power, with a decrease in arrivals during these summer months. Public holidays also affect ED arrivals. Similar to the postponement effect observed on Sundays, people tend to delay their hospital visits on public holidays. This results in a decrease in arrivals on public holidays and an increase on the following day.

Conclusion

Our developed prediction model outperforms the hospital's current forecasting method with a relative improvement of 8.79%. Although the RF model has slightly better performance, we recommend using the Lasso model due to its intuitiveness and ease of implementation.

We recommend integrating the prediction model into the hospital's existing KPI dashboard. This can be achieved through an API that takes input from the hospital database and outputs daily predictions. Alternatively, a simpler option is to use the forecasting tool we provided to generate predictions for a longer period and manually insert them into the dashboard application.

We also recommend that the hospital periodically updates the forecasting model with new data by refitting the coefficients of the Lasso model. If an API is built, this model can be updated continuously without much effort. For the simpler implementation option, we recommend updating the model once per year.

Acknowledgements

This thesis finalizes my time at the University of Twente. During this time I have learned a lot, both professionally and personally. I am grateful for this period and are looking forward to what comes next.

I would like to thank my supervisors for guiding me through my thesis process. First of all, I want to thank my lead supervisor, Amin. At the start of my thesis you told me I was the first student you would supervise, which was an advantage and disadvantage. The disadvantage being you are still getting to know all the details of the thesis process at the UT, but the advantage being you were very excited to be a supervisor. Afterwards I can tell you I only experienced the advantage of your enthusiasm. I enjoyed working together as well as getting to know you. Your feedback was very useful in guiding my thesis.

Secondly, I would like to thank my secondary supervisor, Leo. After working for a couple of years as a student assistant in one of your courses, you offered to supervise my thesis. I am happy that I accepted your offer, as I enjoyed working with you and your critical view has improved my thesis greatly.

Also, I would like to thank my supervisor at Diakonessenhuis, Hayo. I am lucky that my external supervisor is also an expert in the field, which lead to many nice discussions and ideas to improve my thesis. Your feedback on my thesis has been very clear and useful. Additionally, I would like to thank the other members of the Integral Capacity Management team for their involvement and interest in my thesis.

Lastly, I would like to thank my friends, family and my girlfriend for supporting me through this process. In both ups and downs you have been there for me and I am very grateful for that.

Bastiaan Kauffeld

Enschede, 5-22-2023

Contents

1 Introduction

In this chapter, we briefly discuss the organization in Section 1.1 and give a problem description in Section 1.2. In Section 1.3, we provide our research goal, followed by research questions in Section 1.4. Finally, in Section 1.5 we define the scope of the research.

1.1 Organization

This research is conducted at Diakonessenhuis, which is a hospital in Utrecht, with subsidiary locations in Zeist and Doorn. In our research, we focus on the location in Utrecht, which has around 500 beds and over 2800 employees. The main involved departments are the emergency department (ED), and the Integral Capacity Management Team (ICM).

1.2 Problem description

The healthcare sector is dealing with increases in demand due to an increasing and aging population. Therefore, it becomes challenging to maintain the high quality of care while keeping healthcare affordable and accessible. As a result, the healthcare sector is forced to work more efficiently, including efficient utilization of resources such as staff or equipment.

The ED is a vital department in a hospital. As the urgency of patients arriving at the ED is high, fast treatment is often of great importance. To be able to provide fast treatment, sufficient staff should be present at the ED. However, staff is costly and therefore important to be deployed well. Effective resource utilization in the ED is crucial. Underutilization can result in increased costs, while overutilization can compromise the quality of care, increase patient wait times, and lead to staff dissatisfaction. The main challenge in managing resources at the ED is the randomness of ED patient arrivals. Unexpected peaks in demand can have major consequences on the performance of the system. Better understanding of the underlying arrival process can improve matching supply and demand at the ED. For instance, if we can predict the number of arrivals at the ED, we can deploy staff accordingly. Therefore, forecasting patient arrivals can help decision-making regarding resource utilization easier.

The hospital currently does not have a formal forecasting method. Instead, rough predictions are made based on patterns from historical data and based on experience. Mostly the day of the week determines the prediction of ED arrivals. In some cases this prediction changes, for example if major events are held in Utrecht, or on special days such as new years evening.

To improve the current forecasting method, the hospital is seeking to use a data-driven approach to forecast arrivals at the ED. Two types of forecasts can be of value. The hospital indicated forecasts can assist in both tactical planning as well as operational planning. To assist tactical planning, forecasts should be available well in advance. To assist operational planning, forecasts need to be available only a couple of days in advance, giving the opportunity to include variables that would not be available otherwise. However, the results of our research shows variables that would be included in a short term model, such as weather variables or information about public events, fail to improve the model. Therefore, we develop one model, which included variables that are known in advance. This model serves for both short term and long term forecasts. Furthermore, the hospital seeks to gain insights into the contributing factors of ED patient arrivals.

1.3 Research goal

The main goal of our research is to answer the following question: What are suitable forecasting models, for patient arrivals at the ED at Diakonessenhuis Utrecht, with error measures as low as possible?

Suitable here means the extent to which the models are of practical use as the hospital intends to implement the forecasts into their existing planning tools. Decision-makers have to rely on these forecasts, which means the interpretability of the forecasting models is essential.

By achieving this goal we contribute both to theory and practice. The theoretical contribution is adding another case study to the existing literature, further strengthening the used methods. Also, while a lot of literature exists on the subject, ED patient arrival forecasting is not well researched specific to the Dutch Healthcare system. While most variables that we research have been researched before, the inclusion of a variable that indicates extreme weather conditions is a new contribution to theory.

The impact of our thesis will be mostly to practice, as Diakonessenhuis Utrecht will use our results. The tool we developed will be added to their existing KPI dashboard at the ED. The information from the forecasts can be used to improve resource allocation. Furthermore, the forecasts on ED arrivals can be used to predict downstream care in the hospital. In addition, the insights into variable importance are of value to the hospital.

1.4 Research questions

In this section, we specify a total of 6 research questions. For each question, we explain the method of answering the question.

Question 1

What does the ED arrival process look like and what kind of patient arrival forecasting is used to support it?

Hospital staff is interviewed to identify the current situation.

Question 2

What ED patient arrival forecasting models exist in literature?

An extensive literature study will be performed to find suitable forecasting models. In addition, appropriate performance measures will be identified.

Question 3

What is the performance of the suitable models found in literature?

We will compare the performance of a selection of suitable models found in literature. Answering this research question includes model building, variable selection, and validation. Historical data from the hospital is used for model development.

Question 4

What are the contributing variables for ED patient arrival, and what is their importance?

We answer this question with the outcomes of variable selection from question 3. In addition, depending on the selected models from question 4 we use model coefficients to identify importance of variables.

Moreover, we provide descriptive statistics on the contribution of variables. This increases the insights in variable importance, and serves as additional validation for the models.

Question 5

How can the developed forecasting models assist decision-making in the hospital?

In the last step of the research, we aim to relate the outcomes of forecasting models to actions to be taken in decision-making. For instance, advice on the number of staff needed or the expected bed occupation. Moreover, we provide recommendations on the implementation of the forecasting models.

1.5 Scope

In this research, we focus on predicting the ED patient arrivals at the Diakonessenhuis hospital in Utrecht. We use their historical data on patient arrivals to build and validate forecasting models. Data is available from 2017 up to 2022. While the data is specific to one hospital, the research results are likely to be generalizable to other hospitals in the Netherlands. In addition, while the trends in other cultures and climates are probably different, the research results are likely still relevant for hospitals outside the Netherlands.

The forecasts will be less applicable under special circumstances, such as pandemics or natural disasters. Hence, we exclude events like the global pandemic of Covid-19. The historical data on patient arrivals after March 2020 is heavily influenced by the Covid-19 pandemic. During the pandemic ED arrivals rise due to Covid patients, but the number of regular patients decrease due to measurements such as lockdowns. As the circumstances after March 2020 are clearly different from the period before, we only include data until March 2020.

2 Context description

In this chapter we answer the first research question:

What does the ED arrival process look like and what kind of patient arrival forecasting is used to support it?

We answer this question by providing a process description in Section 2.1. In addition, we give an overview of stakeholders in Section 2.2.

2.1 Process description

In this section we describe the ED process and the current forecasting method.

ED process

The ED treats patients with a need for urgent care. Patients most frequently enter the ED directly via an ambulance or via referral from the general practitioners post. Alternatively patients can come in at the ED via a different hospital department, or in rare cases for check-ups.

Once admitted to the ED, the patient has to wait before a doctor sees the patient. This waiting time depends on the severeness of the patient's injury and the crowdedness at the ED. After waiting the doctor sees the patient and takes diagnostic steps. The patient is then treated at the ED, admitted to the hospital or discharged. Figure 1 shows this process.

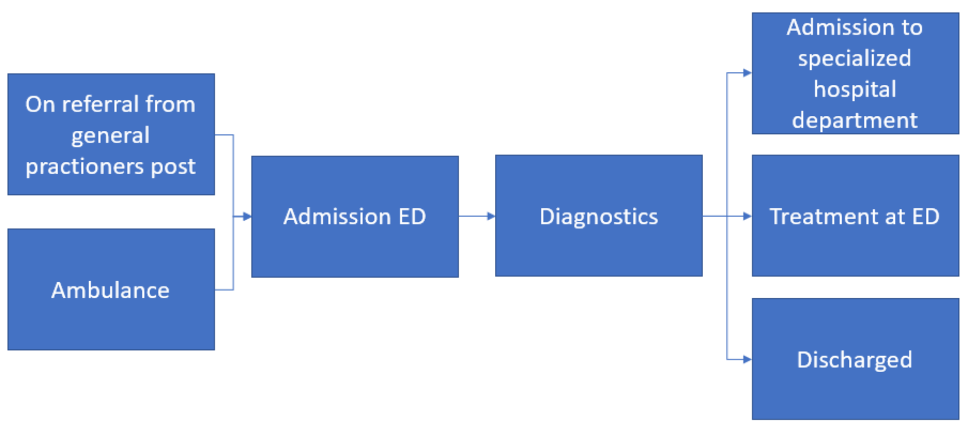

Figure 1: ED process

Current forecasting method

The most crucial resources at the ED are the doctors and nurses. Currently the hospital allocates staff based on expected patient arrival. However, the hospital does not use any scientific forecasting method. Instead the hospital creates rough forecasts partly based on the observation of historical data, but mainly based on intuition. Weekdays and months are main factors in currently determining a prediction. Additionally special circumstances like new years evening or large events are taken into account. A detailed explanation of the current forecasting method is provided in Section 5.2.

Similar forecasts are being used to predict bed occupation throughout the hospital departments. Part of bed occupation is predictable because of planned surgeries and other medical procedures requiring hospital admission. However, the part of occupied beds followed by ED admissions is harder to predict. Each admitted patient requires a specific kind of care and is therefore admitted to a specific department. A large number of emergency patients admissions causes problems, as every hospital department only has a limited number of available beds. To a certain extent patients can be admitted to a different department, but this is undesirable, as the personnel and equipment is often not ideal to treat a patient. Timely knowledge of bed occupancy allows bed managers to allocate their resources more efficiently.

2.2 Stakeholders

The main stakeholder is the ED, as it is directly influenced by ED forecasting. Because the ED serves as an entry point for many patients who visit the hospital, many departments are affected by the ED. We briefly discuss the other stakeholders.

During diagnostics, other departments are frequently involved to perform tests (e.g. X-rays, MRI scans etc). These departments have a mix of planned patients and emergency patients. The inflow of patients from the emergency department has an impact on the performance of these auxiliary processes, as these departments have a trade-off between the utilization of resources and the availability of resources in case of emergency patient arrivals.

In some cases emergency patients need surgery, which involves the surgery department. Similar to departments performing diagnostic tests, a trade-off exists between utilization and availability of operating rooms. However, high availability is more crucial at the surgery department.

Part of the patients is admitted to the hospital. These patients receive a bed at one of the nursing departments, depending on the patient type. As is the case with many hospital departments, the nursing departments also have a mix between planned patients (for example before and after a planned surgery) and emergency patients. Hospital beds are scarce, because of the equipment, space and personnel that is involved.

The ICM team is concerned with capacity issues within the hospital and initiated the forecasting project. Ideally the forecasting model proposed will be part of an existing dashboard the ICM team provides for the ED.

3 Literature review

We provide a systematic literature review (SLR) in this chapter. Appendix A gives an overview of the article selection used in the SLR. The main goal is to answer the second research question:

What ED patient arrival forecasting models exist in literature?

We start by introducing ED patient arrival forecasting in Section 3.1. In Section 3.2 we discuss available forecasting models from literature. In Section 3.3 we discuss which variables are included in literature. Finally, we provide a conclusion on the literature review in Section 3.4.

3.1 ED patient arrival forecasting

ED overcrowding has a negative impact on quality of care and can lead to medical errors (Ong et al., 2009; Sudarshan et al., 2021; Zhang et al., 2022). An increase in the length of stay of patients is associated with increased risk of mortality (Mataloni et al., 2019). In addition, ED overcrowding leads to increased waiting times and decreased staff productivity (Savage et al., 2015). Forecasting ED patient arrivals is a key factor in designing strategies aimed to prevent ED overcrowding (Kadri et al., 2018)

Developing ED patient arrival forecasting models usually consists of 3 steps (Wargon et al., 2009). In step 1 patient arrival data is collected over a period of time. The obtained sample is then split into test data and training data. In step 2 a prediction model is developed, involving finding the optimal values of model parameters. Then patient arrival is predicted using the model. In step 3 the predictions from step 2 are compared to the data from the test set, as part of validation.

With ED patient arrivals forecasts, managers can prepare high level care activities, optimize internal resources and predict downstream care services (Harrou et al., 2016). In addition, accurate forecasting is essential for efficient management of resources (Zhang et al., 2022), leading to reduced patient waiting time and length of stay (Harrou et al., 2020; Jilani et al., 2019). K. Xu and Chan (2016) considered that using forecasts, even when they are noisy, is still helpful in managing the ED. Furthermore, insights into the importance of contributing variables is of value, since understanding what drives demand is a major factor in improving efficiency (Kudyba, 2018; Rema & Sikdar, 2021).

3.2 Overview of models

Extensive research is done on the topic of ED patient arrival forecasting. One of the simpler methods is the naive method, which takes the observation with the same week number and day number from a year ago as prediction(Rocha & Rodrigues, 2021). However, this model is not suitable for forecasting and only serves as a baseline method to compare with other models. Another relatively simple method is exponential smoothing, which can include seasonal components, but also weights more recent observations more heavily than earlier observations (Rocha & Rodrigues, 2021; Rema & Sikdar, 2021).

Some researchers use queueing theory, which uses mathematical equations to describe arrival patterns, often using Poisson distributions (Wiler et al., 2011; Savage et al., 2015). However, queuing models tend to oversimplify the underlying system, resulting in less realistic results compared to alternatives (Hu et al., 2018).

Another group of prediction methods considers a relationship between predictor variables and response variables. The strength of the variables are measured by regression coefficients. The weighted sum of predictor variables and their regression coefficients determine the value of the prediction (M. Xu et al., 2013). Multiple linear regression (MLR) is a well known method, which assumes a linear relationship between predictor variables and response variables. While linear methods are considered to be relatively simple, they often are at least as good or better as more sophisticated methods like complex machine learning methods (Vollmer et al., 2021). This paper uses penalized linear models, which is an extension to regular linear models. Penalizing the use of coefficients in MLR serves as part of variable selection and to find a balance between overfitting and underfitting. Among the most popular penalized linear models are lasso, ridge and elastic net (Xing & Zhang, 2022). These methods are also widely applied for prediction models outside the field of ED forecasting, for example in finance (Liu & Guo, 2022), health economics (Wester et al., 2021) and chemistry (Binns & Ayub, 2021).

Time series models are among the most popular models for predicting ED arrivals (Harrou et al., 2020). The (Seasonal) Autoregressive Integrated Moving Average ((S)ARIMA) is the most widely used in ED forecasting (Harrou et al., 2016; Carvalho-Silva et al., 2018; Lin & Chia, 2017; de Brito et al., 2019). However, in comparison to the remaining available models, time series approaches are challenging in model development and require above average computational resources. (Wiler et al., 2011).

A more recent development in the field of ED forecasting is the use of deep learning and machine learning models. Several state-of-the-art methods are used in ED forecasting, such as decision trees, Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Naïve Bayes (Gul & Celik, 2020). Harrou et al. (2020) find deep learning models to be promising. M. Xu et al. (2013) show ANN outperform MLR, although their dataset is limited to only one year of data. However, there are also limitations, as these models are complex and require computational power (Sudarshan et al., 2021). In addition, the performance of deep/machine learning methods is data dependent, since other papers show traditional methods outperform deep/machine learning methods (Jones et al., 2008; Vollmer et al., 2021; Choudhury & Urena, 2020).

Jones et al. (2008) argues no universally superior forecasting method exists. Therefore, it is best to test multiple forecasting methods to find out which method fits the data in the best way. Also, the underlying arrival process can differ per hospital and country, which is why Batal et al. (2001) mentions that the generalizability of ED patient arrival forecasting methods among other clinical facilities is limited. Thus, each institution can use the overall principles, but should aim to develop their own model.

3.3 Variables

All models used in literature use seasonal factors as predictors variables which are weekday, month and sometimes week number. Some work includes additional variables. Sudarshan et al. (2021) has shown meteorological variables can help in predicting ED arrivals. However, Batal et al. (2001) shows while including meteorological variables does have an impact, the improvement in predicting power is minimal. Other research suggests meteorological data fails to improve model performance (Wargon et al., 2009). The lack of consensus on the inclusion of meteorological variables emphasizes the difference in the underlying arrival processes, which can be explained by the difference in cultures and climates. Another variable that can contribute is an influenza outbreak (M. Xu et al., 2013). Furthermore, the ED arrivals can be affected by public holidays or a day after a public holiday (Q. Xu et al., 2016). The variables we use in our research are months, weekdays, week numbers, extreme weather warnings and FC Utrecht matches (nearby big soccer matches). We discuss these variables more deeply in Section 4.1.

3.4 Conclusion

As mentioned by many researchers, ED forecasting is of crucial importance in improving ED efficiency. In the literature we find a variety of ED forecasting models, which can be broadly categorized into ARIMA models, queueing models, MLR models, and deep learning models. We find MLR models to be most suitable for the hospital. While linear models can be potentially outperformed by more complex models, its ease of model development, implementation and interpretability make it a preferred model.

As we have seen, the generalizability of existing ED forecasting models is limited, since the data used in earlier research is very specific to a care institution. This makes model development an important step, which is discussed in the next chapter.

4 Model development

We discuss the development of forecasting models in this chapter. In Section 4.1 we conduct exploratory data analysis on the hospital data. Section 4.2 discusses several regression models used for forecasting patient arrivals. In Section 4.3 we show how the data is corrected for trends. We discuss indicators for model performance in Section 4.4. We conclude on our findings in Section 4.5.

4.1 Exploratory data analysis

The dataset used in this thesis consists of patients arrivals on 1155 days, in the period from 1-1-2017 to 2-29-2020. In the following section we provide information on the characteristics of the data.

Patient arrivals

Figure 2 shows the daily patient arrivals in the data. The plot shows a minor decreasing trend over the years. Fitting a trendline to the data yields an intercept of 76.48 and a slope of -0.00425. Furthermore, there is a clear decrease in arrivals during the summer months.

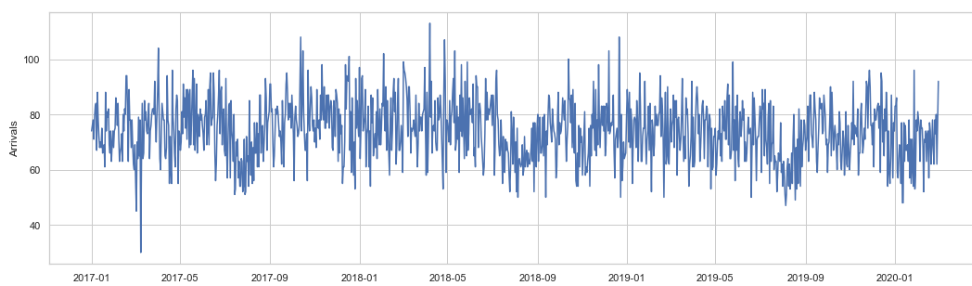

Figure 2: Daily patient arrivals (period 1-1-2017 to 2-29-2020, n = 1155, source: Diakonessenhuis Utrecht).

Figure 3 shows the distribution of daily arrivals. We assume the daily arrivals to follow a normal distribution. Appendix B shows statistical evidence for this assumption.

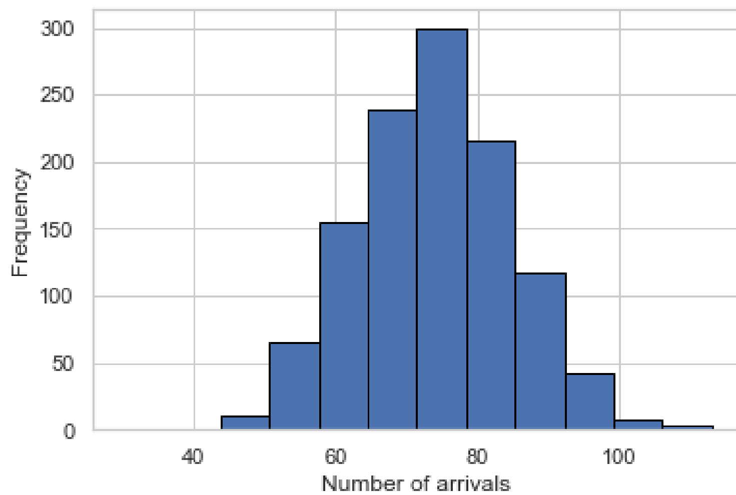

Figure 3: Histogram of daily patient arrivals (period 1-1-2017 to 2-29-2020, n = 1155, source: Diakonessenhuis Utrecht).

Table 1 shows descriptive statistics on the number of daily arrivals.

Table 1: Descriptive statistics of daily patient arrivals (period 1-1-2017 to 2-29-2020, n = 1155, source: Diakonessenhuis Utrecht).

Mean	74,02
Std	10,92
Min	30
Q_1	66
Q_2	74
Q_3	81
Max	113

Variables

An overview of the variables that are included in our research is given in Table 2. All included variables are categorical variables. For example, the weekday variable can take 7 values, one for each day of the week. Due to variable encoding, the weekday variable is split into 7 variables. Each separate variable is a boolean, indicating true/false if it is a Monday, for example. This gives a total of 80 variables.

Table 2: Overview of variables included to predict ED arrivals. Between brackets are the number of variables that are associated with each group of variables.

Variables
Weekdays (7)
Week numbers (53)
Months (12)
Public holiday (True/False)
Day after Public holiday (True/False)
Weather warnings (True/False)
FC Utrecht matches (True/False)

Figures 4-9 utilize violin plots as a data visualization tool to analyze the behavior of the variables under consideration. These plots display a separate distribution for each possible value of the variable being studied, with the width of each distribution representing the frequency of the data points falling within that range. Additionally, an interquartile range in black is displayed within each distribution, with a white dot representing the mean. For example, in Figure 4, the violin plot for Monday's arrivals demonstrates that the majority of data points cluster around the mean value of 78, with the highest and lowest arrivals being around 110 and 40, respectively.

The most common variables to include are calendar related (Jones et al., 2008; M. Xu et al., 2013). Figure 4 and Figure 5 shows the arrivals per weekday and month, respectively. While the distributions for weekdays are relatively similar, Monday, Friday and Saturday are the busiest days. Friday and Saturday can be explained by the weekend effect, where people are more active than usual by participating in outdoor activities. On Mondays catching up takes place, because people tend to postpone their hospital visits on Sundays. The distributions of months are similar, except for July and August, which are the months most people take holiday breaks . In this period a lot of people are on holidays abroad (*Schoolvakanties*, 2023), resulting in less ED arrivals as shown in Figure 5.

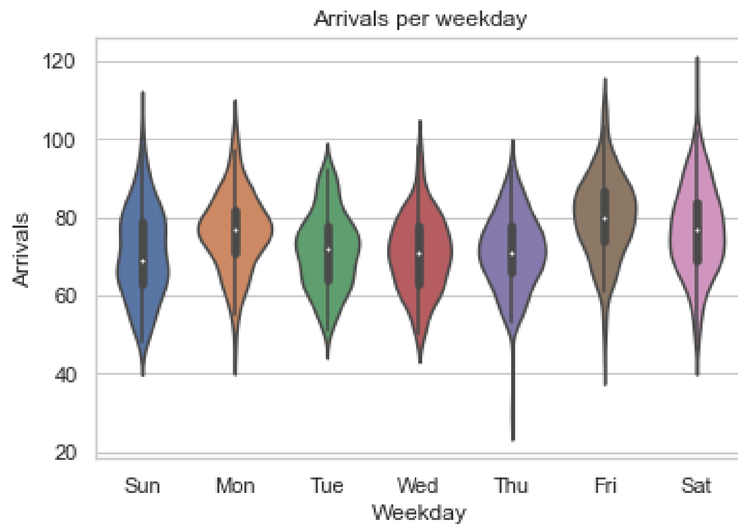

Figure 4: The distributions of arrivals per weekday (period 1-1-2017 to 2-29-2020, n = 1155, source: Diakonessenhuis Utrecht).

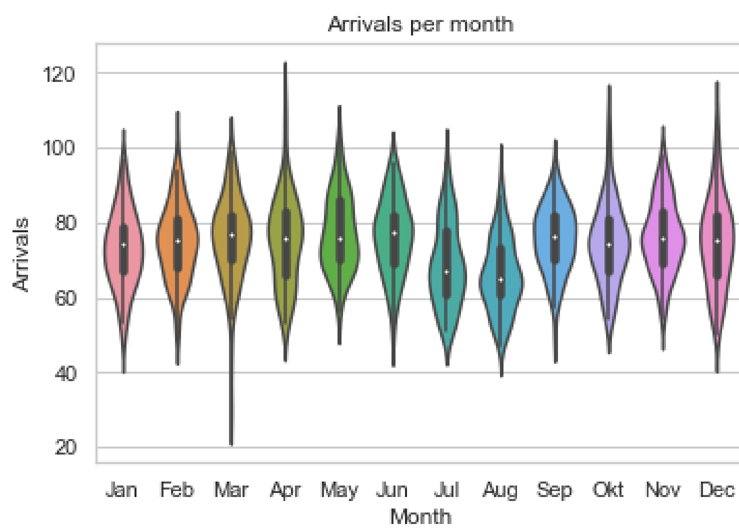

Figure 5: The distributions of arrivals per month (period 1-1-2017 to 2-29-2020, n = 1155, source: Diakonessenhuis Utrecht).

Public holidays are also known to have an effect on ED arrivals, as people typically postpone an ED on these days, resulting in fewer arrivals on a holiday, but more arrivals the day after (Q. Xu et al., 2016). Figure 6 shows the distributions of arrivals on public holidays versus “regular” days. Figure 7 shows the arrivals the day after a public holiday versus “regular” days. Each separate public holiday occurs once a year, making the data points that have a specific public holiday scarce. To cope with this effect we cluster several public holidays into one variable, which indicates if a public holiday occurs. We include public holidays in which the majority of the Dutch people have a day off. In the Netherlands some public holidays have a two day duration. For these holidays (i.e. Easter, Christmas and Pentecost) we only take the second day as a public holiday in our model, since that is when the postponement effect is the highest. Furthermore, Kingsday and New years day are excluded. We assume they differ from regular holidays because the day before and on the day itself a lot of celebratory activities take place, which result in ED arrivals.

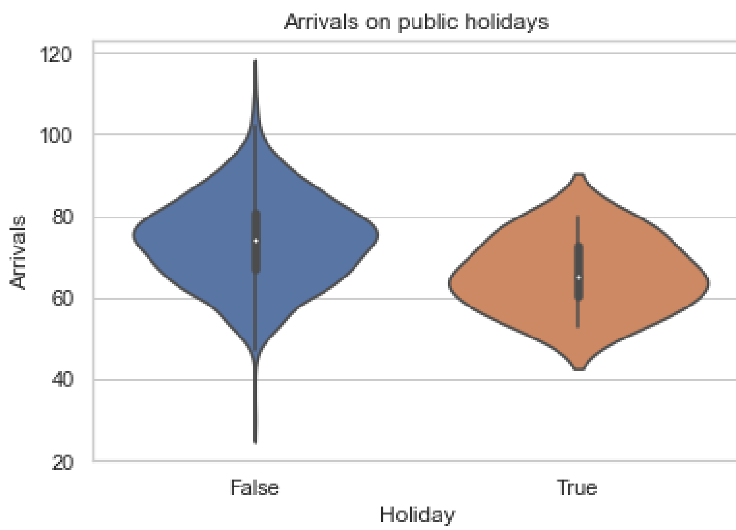

Figure 6: The distribution of arrivals on public holidays and for comparison the distribution of the remaining days (period 1-1-2017 to 2-29-2020, n = 1155, source: Diakonessenhuis Utrecht).

Figure 7: The distribution of arrivals on the day after a public holiday and for comparison the distribution of the remaining days. (period 1-1-2017 to 2-29-2020, n = 1155, source: Diakonessenhuis Utrecht).

Meteorological variables can have an influence on ED arrivals Sudarshan et al. (2021). However, this influence is often small and depends strongly on the climate in which the data is collected Batal et al. (2001); Wargon et al. (2009). One possible contributor to ED arrivals are extreme weather conditions, as people get into accidents more likely. We use color code warnings as an indicator for extreme weather conditions. In the Netherlands these color code warnings are issued at least 24 hours in advance (*KNMI Waarschuwingen*, 2022). Code Green functions as a base case, yellow means to be alert, orange means a high likelihood of extreme weather, and red means an extreme weather condition is about to happen which will have high impact on society. Since the number of days extreme weather conditions occur is low, we group the separate causes together in one variable, which indicates if a weather warning was issued. In our model only orange and red warnings are considered as an extreme weather condition. Figure 8 shows the arrivals during extreme weather conditions, separating the cause of the warning. Figure 9 shows arrivals in case we group weather warnings into one variable. No large differences are seen in the figures, which is also in line with their 95% confidence intervals, (73.44, 74.71) for days without a weather warning and (66.05, 76.15) for days with a weather warning, respectively. However, the arrivals appear to be slightly lower than usual on days with weather warnings. We expect more people to get injured by the weather conditions, but at the same time a lot of people are being cautious and/or stay at home, which reduces the chance of ED arrivals.

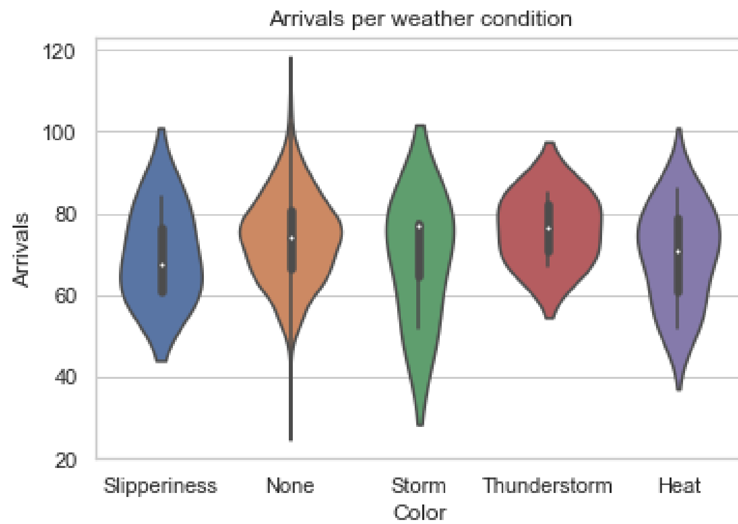

Figure 8: The distributions of arrivals on days with weather warnings (period 1-1-2017 to 2-29-2020, n = 1155, source: Diakonessenhuis Utrecht).

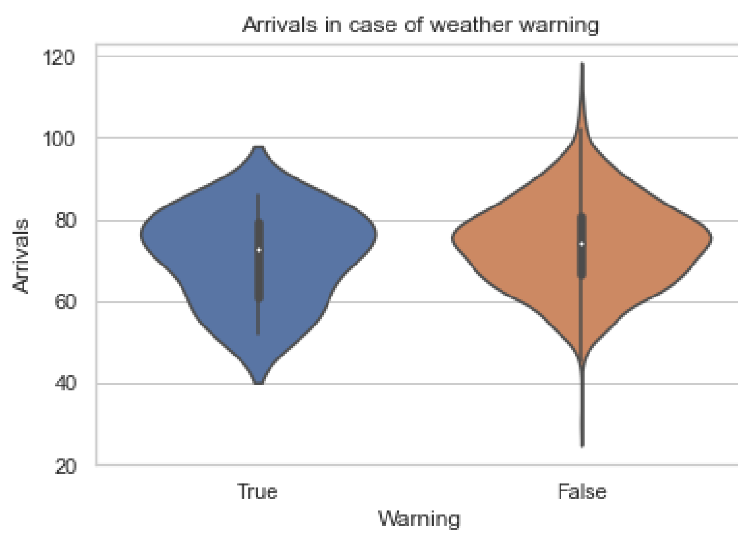

Figure 9: The distributions of arrivals on days with weather warnings (period 1-1-2017 to 2-29-2020, n = 1155, source: Diakonessenhuis Utrecht).

4.2 Prediction models

In our research we compare multiple methods for predicting ED arrivals. In the following section we discuss these prediction models. We start with multiple linear regression, followed by penalized linear models and end with Random Forest.

Multiple linear regression (MLR)

We start by introducing notation:

y_i = value of observation i

x_{ij} = value of variable j in observation i

ϵ = residuals, assumed to be distributed according to $N(0, \sigma^2)$

β_0 = intercept

β_j = coefficient of variable j

$\hat{y}_i$ = estimate for observation i

In case of n observations with and p variables, the multiple linear regression formula is:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \epsilon \tag{1}$$

Then estimates for the predictions are calculated by:

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij} \tag{2}$$

The coefficients are fitted by minimizing the squared error:

$$SE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{3}$$

By substituting (2) into (3), we can obtain the estimates in the following way:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\hat{\beta}} \{ \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \sum_{j=1}^{p} \hat{\beta}_j x_{ij})^2 \} \tag{4}$$

Lasso regression

A downside of MLR is a risk of overfitting, as every variable gets a coefficient. This means insignificant variables also get coefficients, which leads to overfitting.

Additionally, (partially) correlated variables cause problems, since the model may not be able to distribute the coefficients over correlated variables. For example, a large positive coefficient can be cancelled out by a large negative coefficient in case they are correlated.

Lasso regression adds a penalty term to regular MLR formula as shown in Equation 4, which penalizes the use of coefficients. The model now chooses only the variables that are most important in predicting. The model is forced to shrink as the penalty parameter increases, i.e., it sets coefficients to zero, which serves as variable selection. The lasso is shown in (5) (Hastie et al., 2009).

λ = penalty parameter

$$\hat{\boldsymbol{\beta}}^{\boldsymbol{lasso}} = \arg\min_{\hat{\beta}}\{\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \sum_{j=1}^{p}\hat{\beta}_j x_{ij})^2 + \lambda\sum_{j=1}^{p}|\hat{\beta}_j|\} \tag{5}$$

The penalty term in the lasso takes the absolute value of the coefficients, making sure both negative and positive coefficients are penalized. Note that only coefficients of variables are penalized, meaning the intercept $\hat{\beta}_0$ is left out.

Ridge regression

Ridge regression is similar to lasso regression, except the coefficients in the penalty term are now squared. The formulation is shown in (6) (Hastie et al., 2009).

$$\hat{\boldsymbol{\beta}}^{\boldsymbol{ridge}} = \arg\min_{\hat{\beta}}\{\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \sum_{j=1}^{p}\hat{\beta}_j x_{ij})^2 + \lambda\sum_{j=1}^{p}\hat{\beta}j^2\} \tag{6}$$

Although ridge and lasso are very similar, their penalty terms serve for different objectives. Lasso regression aims to reduce model complexity by reducing the number of non-zero coefficients. Ridge regression aims to reduce problems related to correlation of variables, but does not set coefficients to 0. The difference between lasso and ridge is best explained by formulating them into their equivalent optimization problem (Hastie et al., 2009):

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{\boldsymbol{lasso}} = \arg\min_{\hat{\beta}}\{\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \sum_{j=1}^{p}\hat{\beta}_j x_{ij})^2\}, \\ \text{subject to} \sum_{j=1}^{p}|\hat{\beta}_j| \leq t \end{aligned} \tag{7}$$

$$\hat{\boldsymbol{\beta}}^{\boldsymbol{ridge}} = \arg\min_{\hat{\beta}}\{\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \sum_{j=1}^{p}\hat{\beta}_j x_{ij})^2\},$$
$$\text{subject to} \sum_{j=1}^{p}\hat{\beta}j^2 \leq t \tag{8}$$

where t has a one-to-one correspondence to λ. Figure 10 shows the two dimensional case of the ridge and lasso. In this case lasso has a constraint region $\beta_1 + \beta_2 \leq t$ and ridge has a constraint region $\beta_1^2 + \beta_2^2 \leq t$. Due to the diamond shape (rhomboid in multidimensional case) in the lasso optimal solutions occur at corner points, meaning certain coefficients are set to 0. In case of the left plot in Figure 10 β_1 is set to 0, meaning this variable is dropped. The disk shape of the ridge does not share this behavior and instead only shrinks the coefficients.

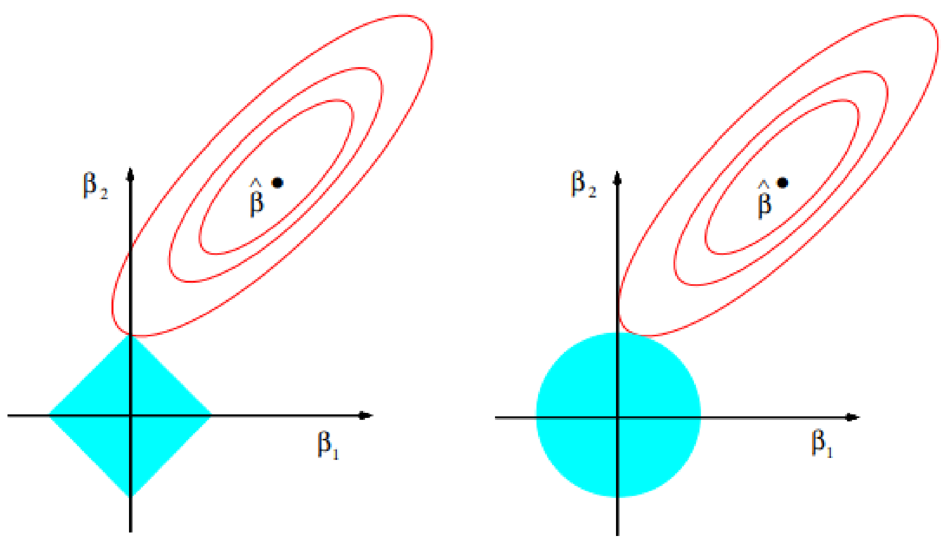

Figure 10: Figure from (Hastie et al., 2009). Estimations of lasso (left) and ridge (right) in two dimensional case. Blue shows the constraint regions and the red ellipses show the contours of the least squares error.

Elastic net

The elastic net is a combination of ridge regression and lasso regression. We introduce a new parameter and then formulate the elastic net in Equation (9). The elastic net benefits both shrinking the model, as well as reducing problems imposed by correlated variables.

α = fraction of ridge used in penalty term

$(1 - \alpha)$ = fraction of lasso used in penalty term

$$\hat{\boldsymbol{\beta}}^{\boldsymbol{elastic}} = \arg\min_{\hat{\beta}} \{ \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \sum_{j=1}^{p} \hat{\beta}_j x_{ij})^2 + \lambda \sum_{j=1}^{p} (\alpha \hat{\beta}_j^{\,2} + (1 - \alpha) |\hat{\beta}_j|) \} \quad (9)$$

Random forest

Although the penalized linear methods are intuitive, the linearity assumption often does not hold. For this reason we introduce a classical machine learning method called Random Forest (RF). This method averages over the outcome of multiple unique decision trees (Hastie et al., 2009).

A RF consists of decision trees. A decision tree splits the data into many partitions, based on input variables. At each split, called a node, a variable j and a splitting point s are chosen to split the data. Figure 11 shows and example of a decision tree. In the first node of the example tree the variable weekday Friday is chosen. The split point in this case is 0.5, since weekday Friday is a binary variable and only takes values 0 and 1. Each node of the tree, called terminal node, corresponds to a region in the data. An observation is classified into a region, which determines its prediction. In case of regression, the value of the prediction is the average of observations that lie within that region.

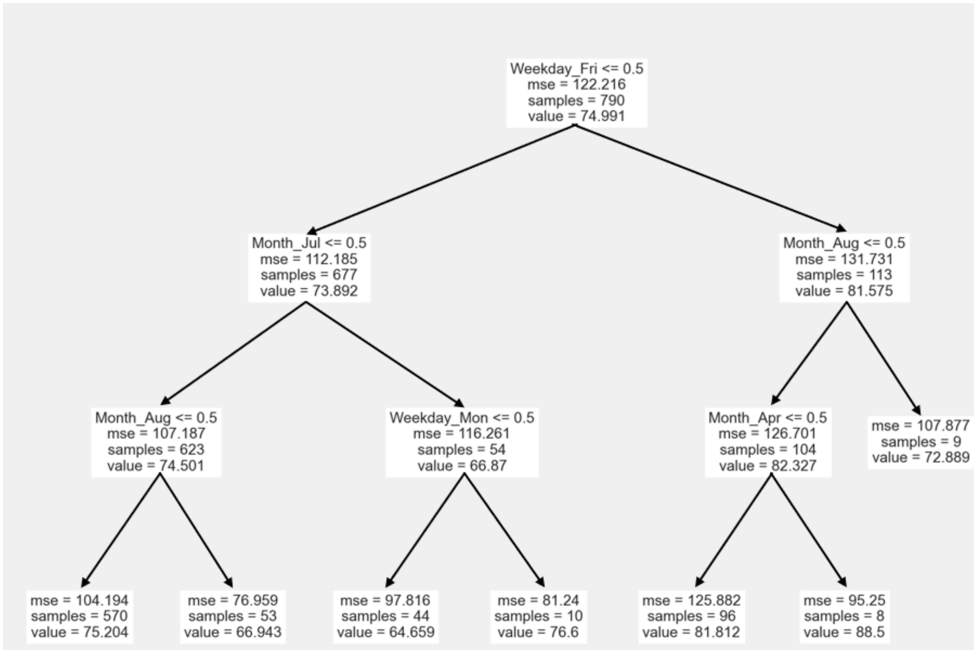

Figure 11: Example of decision tree.

More formally, in case of regression we have a set of N observations, that is (x_i, y_i), $i = 1, 2, ..., N$, with y_i being the value of observation i and $\boldsymbol{x_i} = (x_{i1}, x_{i2}, ..., x_{ip})$ being the vector of input variables for observation i which has p variables. Given an observation with vector of input variables $\boldsymbol{x}$ and the data is split into M regions, the prediction is shown in Equation (10).

$$\hat{y}(\boldsymbol{x}) = \sum_{m=1}^{M} c_m I(\boldsymbol{x} \in R_m) \tag{10}$$

Here c_m corresponds to a constant for region m and $I(x \in R_m)$ is an indicator function which equals to 1, if the vector of input variables $\boldsymbol{x}$ falls in region R_m, 0 otherwise. The estimate for each c_m is the average of observations that fall within region R_m:

$$\hat{c}_m = \text{average}(y_i | \boldsymbol{x_i} \in R_m) \tag{11}$$

The goal is to determine the regions $R_1, R_2, ..., R_M$ in such a way the MSE is minimized. As finding all optimal regions simultaneously is computationally challenging, a greedy approach is used to find the regions. This approach splits the data, one node at a time. Starting with all data, the data is split into two regions by choosing a variable j and a splitting point s:

$$R_1(j,s) = \{X|X_j \leq s\} \quad \text{and} \quad R_2(j,s) = \{X|X_j > s\} \tag{12}$$

We want to choose s and j in such a way that the total MSE of the remaining regions $R_1(j,s)$ and $R_2(j,s)$ is minimized, as shown in Equation (13):

$$\min_{j,s}\{\min_{c_1} \sum_{\boldsymbol{x_i} \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{\boldsymbol{x_i} \in R_2(j,s)} (y_i - c_2)^2\} \tag{13}$$

Where estimates for c_1 and c_2 are again calculated by using the average of observations that fall within region R_1 and R_2:

$$\hat{c}_1 = \text{ave}(y_i|\boldsymbol{x_i} \in R_1(j,s)) \quad \text{and} \quad \hat{c}_2 = \text{ave}(y_i|\boldsymbol{x_i} \in R_2(j,s)) \tag{14}$$

After splitting the data into two regions, the process is repeated on the two remaining regions, and repeated afterwards until all regions all determined, resulting in a fully grown decision tree.

In the RF algorithm, a total of B decision trees are created. For each decision tree T_b a bootstrap sample (random sample with replacement) of size n is used to fit the tree. In case of the RF algorithm a random selection of m out of the total p variables is considered in each node. The best variable out of these m is chosen to split the data, which creates two new child nodes. This process is repeated iteratively until the minimum node size n_{min} is reached, meaning splitting the data even further would mean less than or equal to n_{min} data points end up in a child node.

Finally the prediction is determined by averaging over the total of B created trees. The prediction for a vector of input variables x is shown in Equation (15).

$$\hat{y}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \tag{15}$$

Where $T_b(x)$ is the predicted value of tree b using the input vector x.

4.3 Trend correction

As mentioned in Section 4.1, a negative trend in the number of ED arrivals exists. With this negative trend, the model is likely to overestimate the number of ED arrivals. To illustrate, consider splitting the data into a train and test set. If the average of ED arrivals is higher in the training set, the predictions on the test set will likely be too high. In addition, when the hospital uses the prediction model new trends in future data can cause the models to perform poorly.

Multiple sophisticated methods exist for trend correction, such as the Baxter King filter (Baxter & King, 1999) or the band pass filter (Christiano & Fitzgerald, 2003). We propose a simple method which serves as a practical solution for dealing with trends. In this method we use a moving average of the observations of the most recent year to correct the prediction of the model.

We let n be the number of observations used to train the model and N the total number of observations. This means the observations $y_1, y_2, ..., y_n$ are used to fit the model and $y_{n+1}, y_{n+2}, ..., y_N$ are future observations.

We let μ be the average of observations in the data used to fit the model.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{16}$$

MA is the moving average of the last k observations. Note that in our model we set k to 365.

$$MA = \frac{1}{k}\sum_{i=N-k+1}^{N} y_i \tag{17}$$

We then correct our predictions with the remainder of $MA - \mu$. This way if $MA > \mu$, meaning the average of recent observations are greater than the observations the prediction model is trained on, we increase our prediction. Vice versa, if $MA < \mu$ we decrease our prediction.

If we have an vector of input variables x and a prediction model $f(x)$, the final prediction is as follows:

$$\hat{y}(x) = f(x) + MA - \mu \tag{18}$$

4.4 Performance indicators

To evaluate the performance of a prediction model, error measures are used. The most common method to compare predictive models is the Root Mean Squared Error (RMSE) (Kuhn et al., 2013). The RMSE is defined as :

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{19}$$

Another common error measure is the Mean Absolute Percentage Error (MAPE). The MAPE is defined as (Chopra & Meindl, 2016):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \tag{20}$$

A variant to the MAPE is the Mean Absolute Error (MAE). The MAE is similar to MAPE, except it gives an absolute error instead of a percentage, which can be more intuitive for decision makers. The MAE is defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{21}$$

4.5 Conclusion

We include a total of 80 variables, including weekdays, week numbers, months, public holidays, the day after public holidays, weather warnings and FC Utrecht matches. From exploratory data analysis we observed the calendar related variables impact the ED arrivals. Also, from the exploratory analysis we see that on public holidays people tend to postpone their hospital visit by one day. Weather warnings and FC Utrecht matches have minimal impact, according to exploratory analysis.

We use penalized linear models (lasso, ridge and elastic net) as an extension to MLR to reduce model complexity and avoid correlation between variables. In addition, we use RF, which is one of the state of the art machine learning models for predictions.

We corrected a slight negative trend in the data, with a simple method that corrects the data using the moving average of observations of the most recent year. As performance indicators to assess model performance we choose MAPE, MAE and MSE.

In the next chapter we apply the data to the models we described. We then analyse the results of each method and reflect on the results.

5 Results

In this chapter we answer two research questions. We start with answering:

What is the performance of the suitable models found in literature?

Followed by:

What are the contributing variables for ED patient arrival, and what is their importance?

In Section 5.1 we discuss the methods we use for hyperparameter tuning. To compare the performance of the prediction models to the current situation at the hospital we provide a baseline prediction in Section 5.2. In Section 5.3 we discuss the performance of each model we tested. To gain insights in the underlying process of ED arrivals, we discuss variable importance in Section 5.4.

5.1 Hyperparameter tuning

Most prediction models contain hyperparameters, which are model parameters that need to be set before developing a model. For example, the penalty parameter λ in the penalized linear models from Section 4.2. The choice of hyperparameters influences the performance of prediction models, therefore they need to be determined carefully. The process of finding values for hyperparameters is called hyperparameter tuning.

To build the models we split the data into a training set and a validation set. The validation set contains one year of data. To avoid overfitting, we tune hyperparameters of each prediction method only using the training data. To find the optimal hyperparameter values we use resampling methods, meaning we split the training data again into training and test sets, which has shown to find better hyperparameters (Kuhn et al., 2013).

A common resampling technique is k-fold cross validation. This method splits the data into k folds (roughly equal sized groups) and then uses k - 1 folds for model building and 1 fold to validate the models. In practice the number of folds k = 5 or k = 10 is often used (Kuhn et al., 2013). We use k-fold cross validation with k equals to 5, as our test data set becomes too small when we set k = 10, due to the limited amount of data. This means we split the training data into 5 folds, for a total of 5 splits. In each split, 4 out the 5 folds are used as training data and the remaining fold is used as validation data for tuning.

For each combination of splitting the data into test and training folds, we try a range of values for hyperparameters. Afterwards, we return the values of hyperparameters that have the best mean score (lowest MSE) over all combinations of test and training folds. We use randomized folds to ensure each fold contains a mix of data points. This way, within each fold the data contains different months, weeks, etc. Figure 12 shows a visual representation of the cross validation process.

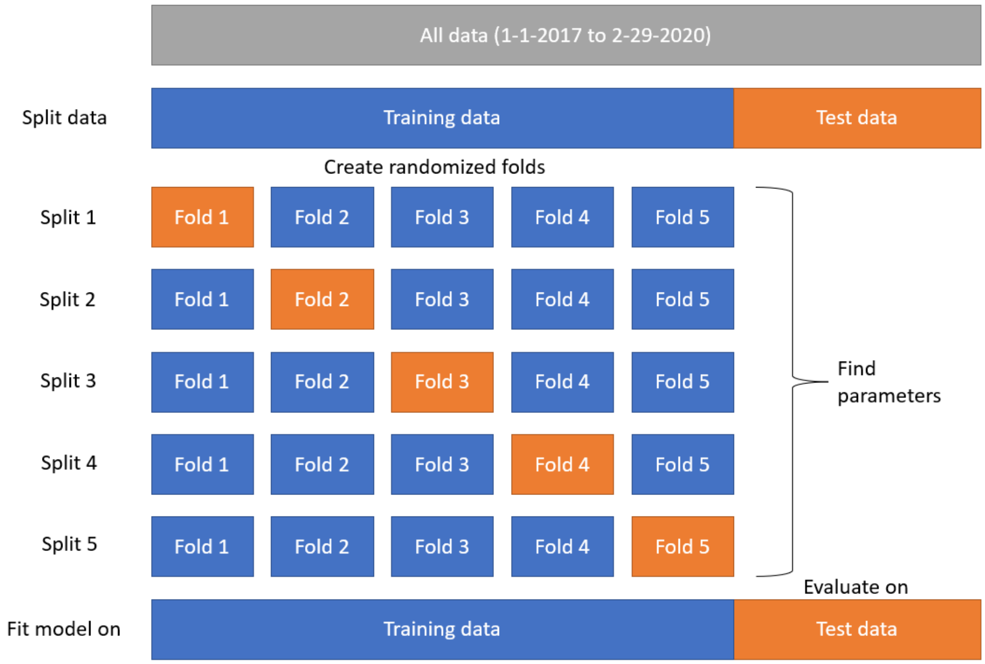

Figure 12: Visual representation of the cross validation process.

Linear models

In the case of lasso, ridge and elastic net the penalty parameter λ needs tuning. In case of elastic net, the parameter α also needs tuning, which balances the amount of lasso and ridge used. Table 3 shows the range of values considered for tuning the lasso, ridge and elastic net. In case of the elastic net, which has two hyperparameters, combinations of λ and α are considered. The upper and lower bounds for the hyperparameters are set by trial and error. The concave shapes in Figures 13 and 15 show evidence that the optimal values lie within our set range of hyperparameters.

Table 3: Range of values used in hyperparameter tuning for the linear models.

Parameter	**Range**	**Increment**	**Best value**
λ ; **lasso**	0,1	0.01	0.04
λ ; **ridge**	0,20	0.2	9.2
α ; **elastic net**	0.01,1	0.01	0.39
λ ; **elastic net**	0,15	0.01	0.02

Figure 13 shows the behavior of the MSE when tuning parameters in the Lasso. The plot shows the typical bias-variance trade-off which exists in machine learning problems. As λ increases, the number of variables used in the lasso decreases, which is shown in Figure 14. At λ equals 0 the lasso is equals to regular MLR, meaning all variables are included. Slight overfitting occurs when including all variables, since the MSE decreases when increasing λ, to an optimum of λ equals to 0.04. When increasing λ further, underfitting takes place, since too few variables are included to make meaningful predictions.

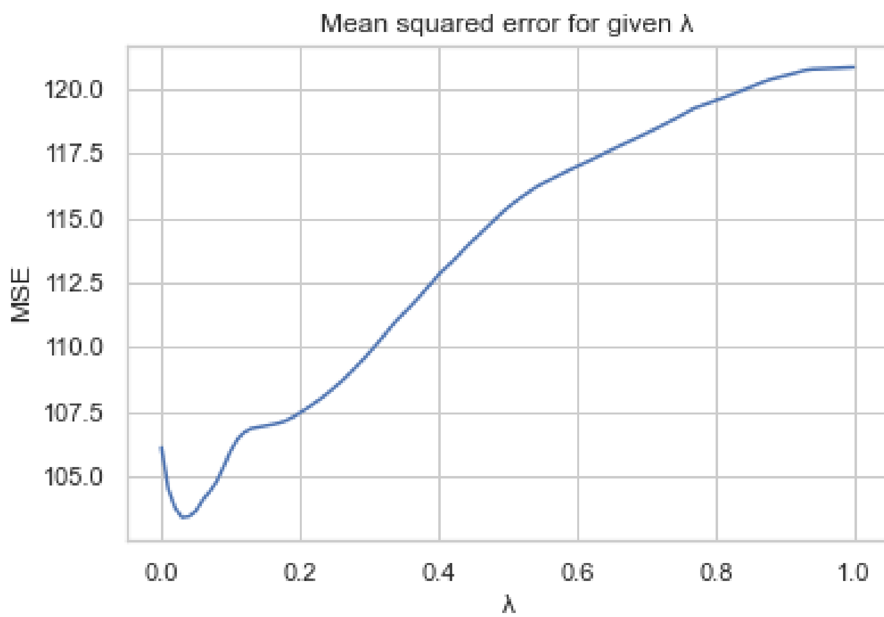

Figure 13: MSE for tested hyperparameters for lasso.

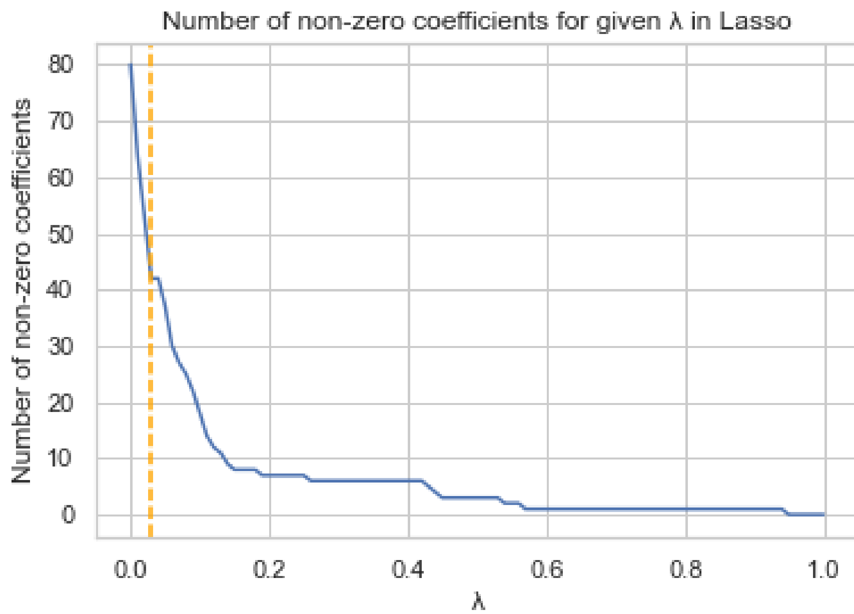

Figure 14: Number of variables included for tested hyperparameters for lasso. The yellow dotted line indicates the optimal λ found.

Figure 15 shows the behavior of the MSE when tuning the penalty parameter in the ridge. Similar to tuning the lasso penalty parameter, a trade-off exists between over and under penalization. While the lasso aims to reduce the number of variables used, the ridge aims at dealing with correlated variables by introducing bias into the model. The optimum of the ridge penalty parameter lies around $\lambda = 5$, which indicates some correlation exists between variables. One obvious correlation is between months and weeks. For example, the first 4 weeks of the year are always in January. Also partial correlation between variables exists. For example, around half of the soccer matches of FC Utrecht are on a sunday, meaning they are partly correlated.

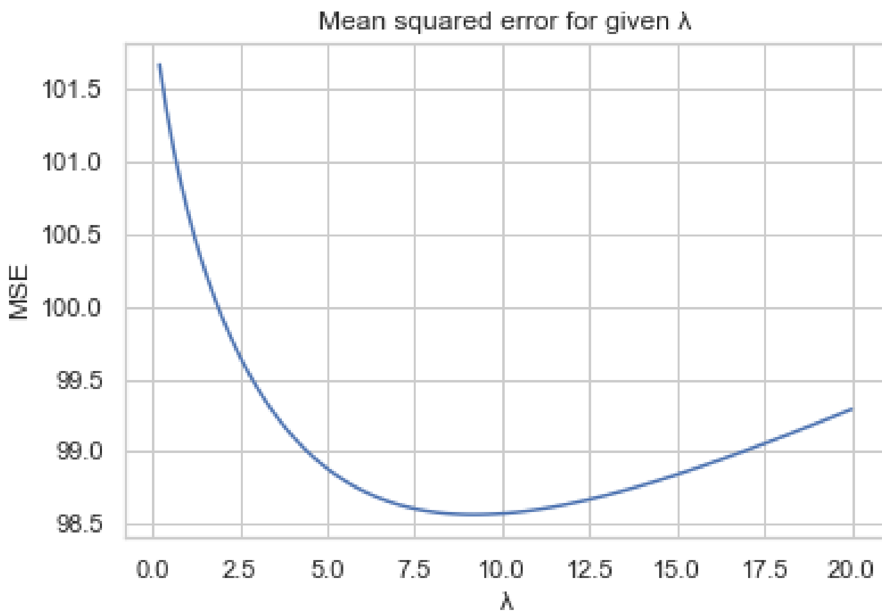

Figure 15: MSE for tested hyperparameters for the ridge.

Figure 16 shows the MSE for each tested α (ratio between lasso and ridge used) in elastic net regression. The MSE shown in Figure 16 is the lowest MSE amongst the values of λ tested in the optimization process. The MSE of all tested values of α are relatively close to eachother. The lowest MSE occurs at $\alpha = 0.39$

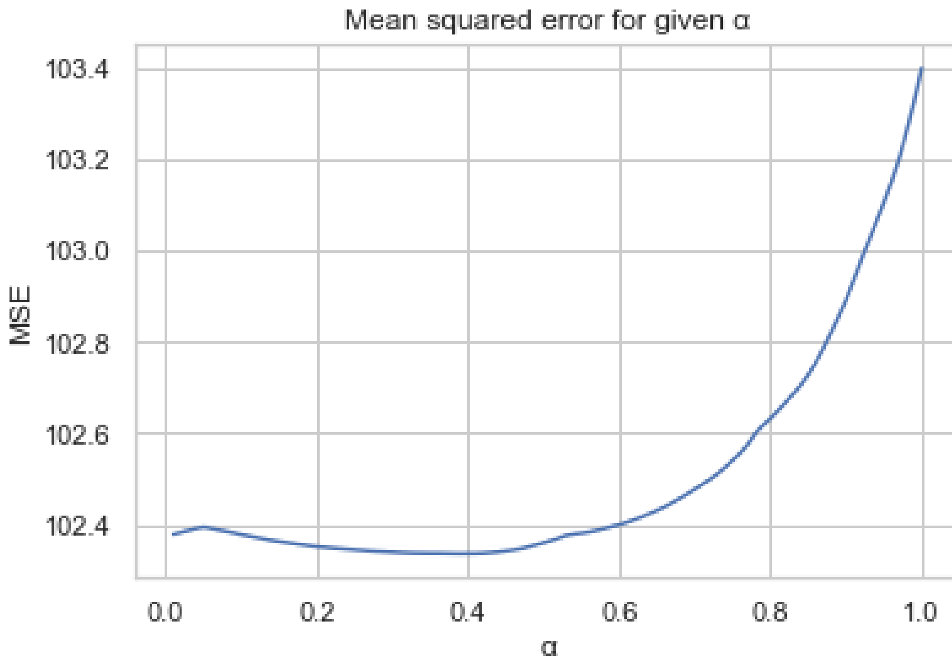

Figure 16: The MSE for all tested α ratio, using the best λ found for a given α, when tuning the hyperparameters for elastic net regression.

Figure 17 shows the MSE when tuning the penalty parameter λ, given $\alpha = 0.39$, which has the lowest MSE of all tested α.

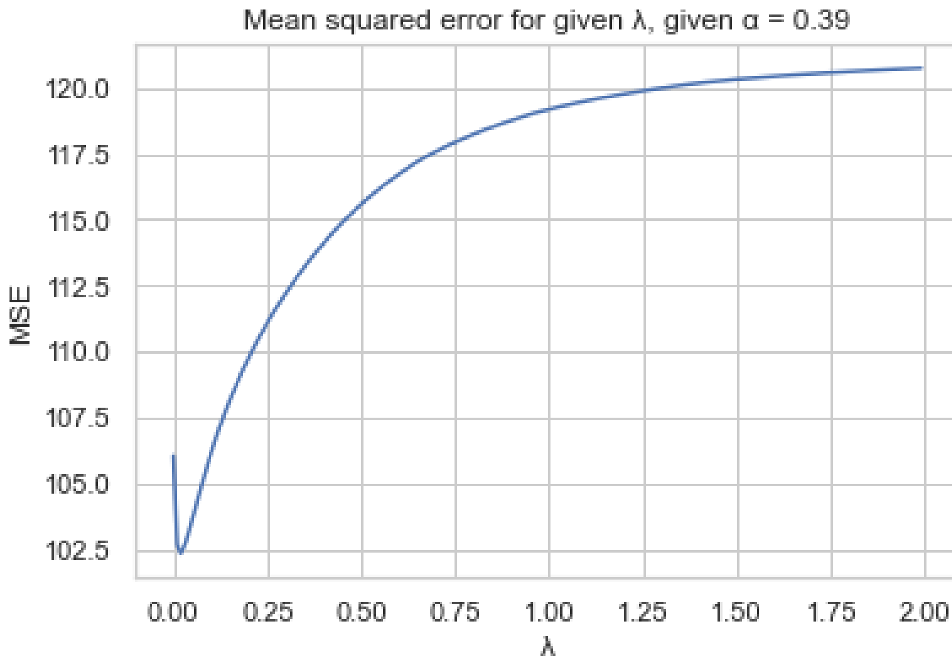

Figure 17: The MSE per tested λ for the best l1 ratio found, when tuning the hyperparameters for elastic net regression.

Random forest

We optimize the hyperparameters in the RF in 2 steps, combining random search and grid search (Probst et al., 2019). In step 1 we explore a wide range of hyperparameters through a random search. In a total of 500 iterations we explore random combinations of hyperparameters. In step 2 we optimize further by using grid search, which tests for a given range of combinations of hyperparameters which are close to the best performing combination from step 1.

Table 4 shows the ranges of hyperparameters that are tested in step 2. The ranges are based on the best performing hyperparameters from the random search in step 1. The max features parameter determines the number of variables considered at each split. The max depth, min samples leaf and min samples split are the stopping criteria for growing each tree. The max depth is the maximum depth of the tree, i.e. the number of layers in the tree. The min samples split is the minimum number of samples that a child node should have when we split the data. The min samples leaf is similar to min samples split, except a leaf node is considered (a node with no child nodes).

Table 4: Range of values used in hyperparameter tuning for the RF.

Parameter	Range	Increment	Best value
Max depth	{40,80}	10	70
Min samples split	{8,12}	1	11
Min samples leaf	{1,6}	1	5
Number of trees	{100,400}	100	100

5.2 Baseline forecast

Currently the hospital does not make explicit forecasts on daily ED patient arrivals. However, indirectly the hospital does make forecasts, for example when deciding on how much nurses to allocate each day. The hospital schedules their nurses based on historical data and experience. Logically the hospital allocates more nurses to periods where they expect more patient arrivals. We use this nurse allocation to estimate a baseline prediction, which we use to compare to our developed models.

Table 5 shows the nurse allocation during the week, along with the estimated baseline prediction. On Monday, Friday, and Saturday more nurses are allocated, meaning the ED expects more patients on these days. To come up with exact numbers we first calculate the ratio of patients to nurses by dividing the average arrivals on a weekday by the number of allocated nurses, resulting in the patients per nurse. We then take the average over all weekdays, resulting in 4.52 patients per nurse. The final baseline prediction is then determined by multiplying the average patients per nurse with the number of nurses allocated on each weekday.

Table 5: Number of nurses allocated per week day.

Weekday	Nurses	Avg arrivals	Patients per nurse	Baseline prediction
Monday	17	77.38	4.55	76.92
Tuesday	16	72.61	4.54	72.40
Wednesday	16	71.32	4.46	72.40
Thursday	16	72.33	4.52	72.40
Friday	18	81.58	4.53	81.45
Saturday	17	77.46	4.56	76.92
Sunday	16	72.28	4.52	72.40
Average		74.99	4.52	74.98

5.3 Model performance

In this section we analyse the performance and their prediction errors of the tested models.

Performance

Table 6 shows the performance of the models when predicting daily arrivals.

Table 6: Performance measures of tested prediction models on daily ED patient arrivals.

Model	MAPE	MAE	RMSE
Baseline prediction	12.29	8.17	10.18
MLR	11.67	7.88	9.76
Lasso	11.24	7.60	9.44
Ridge	11.24	7.59	9.42
Elastic net	11.25	7.59	9.43
RF	11.21	7.56	9.38

The lasso, ridge, and elastic net are very similar in performance. RF is performing slightly better, but still similar to the penalized linear models. The MLR model performs the worst in all aspects, which is expected as this is the simplest model. In addition, lasso, ridge, and elastic net are extensions to MLR, that can reduce to MLR if we set $\lambda = 0$. Therefore, we expect the penalized linear models to perform at least as good as MLR. However, the difference in performance between MLR and the rest of the models is small, meaning a simple model such as MLR can still give satisfactory results.

An attempt to apply a more sophisticated machine learning method, through the use of a RF, results in similar performance with the penalized linear models. This is in line with the results of earlier work (Jones et al., 2008; Vollmer et al., 2021; Choudhury & Urena, 2020), in which attempts to apply more complex models to the ED arrival forecasting problem often results in similar or even slightly worse results.

Given the similar performance of RF, lasso, ridge, and elastic net, we select the lasso as it has the lowest model complexity and it only uses 42 out of the 80 total variables. Furthermore, the intuitiveness is higher, because only the variables which have the best predictive power obtain a non-zero coefficient.

The lasso models shows an improvement of 1.05% in MAPE compared to the baseline prediction. In relative terms, this is an improvement of 8.54% over the baseline prediction. Bigger improvements are hard to achieve. Both due to the underlying stochastic arrival process, but also because the baseline prediction is not bad, because of the experience planners have at the hospital. Still a 8.54% relative improvement is useful in practice.

Prediction errors

For the analysis of prediction errors we focus on the lasso predictions. Figure 18 shows the predictions in comparison with the actual observations, during the test data period, using the lasso. This figure shows the model can recognize seasonal patterns quite well, such as weekdays and months. However, the figure also emphasizes the difficulty of ED forecasting, as the model fails to captures outliers. This is expected, as the underlying arrival process is stochastic.

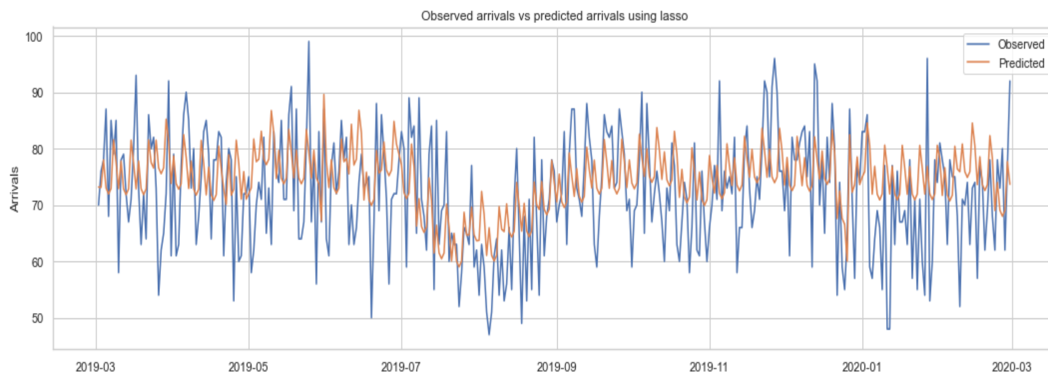

Figure 18: Observed arrivals and predicted arrivals by the lasso, on the test data period.

Figure 19 shows a histogram of the distributions the lasso predictions in comparison with the observations from the test data. Similar to the line graph in Figure 18 the histogram shows the inability of the prediction model to capture outliers. Appendix D shows the distributions of the remaining prediction models we used.

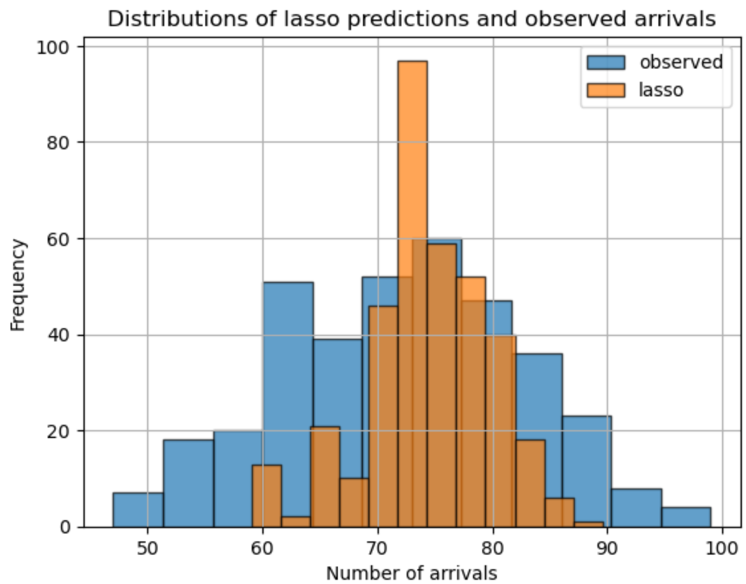

Figure 19: Observed arrivals and predicted arrivals by the lasso, on the test data period.

Figure 20 shows the distribution of the residuals (defined as predicted arrivals - observed arrivals) of the lasso predictions. We assume the lasso residuals to be normally distributed with $\mu = 2.37$ and $\sigma = 9.15$ (see Appendix C). The positive mean of the distribution of the residuals indicates the lasso predictions overestimate the number of daily arrivals, by 2.37 on average. This is caused by the negative trend in the data, which is only partly solved by the trend correction method proposed in Section 4.3.

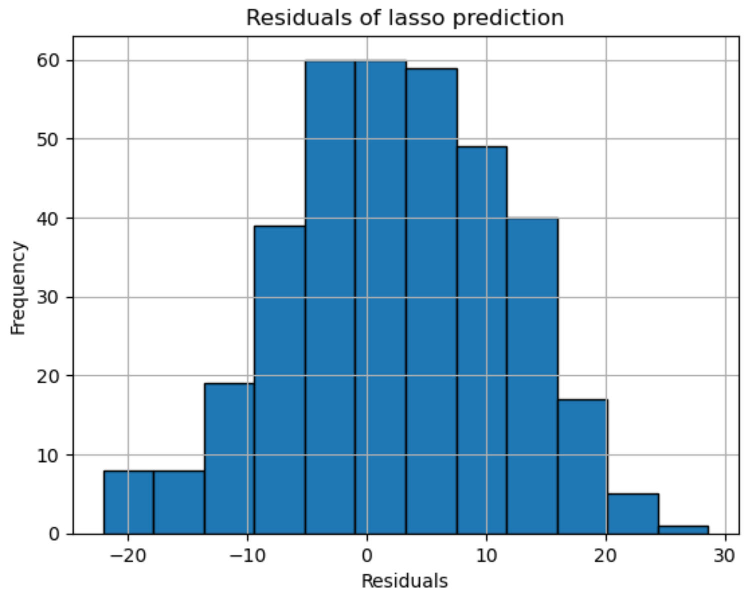

Figure 20: Distribution of residuals of predictions on the test data, using lasso regression.

5.4 Variable importance

In this section we investigate the variable importance in the linear models and in the RF model. While we prefer the use of the lasso, the variable importance of the similarly performing models can provide us with additional insights.

Linear models

The coefficients obtained by applying lasso regression are most intuitive among the applied linear models. Unimportant variables have coefficients reduced to zero, while the remaining variables obtain coefficients according to their predictive strength. The coefficients obtained by ridge regression are never zero, as explained in Section 4.2, instead every variable has a non-zero coefficient, making the total use of coefficients higher. While the performance of the models are similar, the model from ridge regression is less intuitive because of the reasons mentioned above. Similarly, the coefficients obtained by elastic net regression are less intuitive. While coefficients can be zero, most of them still have non-zero coefficients, making it harder to distinguish variable importance.

For these reasons we use the coefficients obtained by applying lasso regression to analyse variable importance. Recall that predictions in the lasso are made by summing up the intercept with variables multiplied by their coefficients, as shown in Equation (2). In case of the lasso model we developed, the intercept equals 72.24. The size of a coefficient indicates the importance of the corresponding variable. Positive coefficients indicate an increase in ED arrivals and negative coefficients indicate a decrease.

Table 7 shows the coefficients of week days. Monday, Friday and Saturday have the biggest impact on ED arrivals. This is in line with expectations, since Friday and Saturday are known to be busy days due to the weekend, where people get more active. Monday is also known to be a busier day, again due to the postponement effect which takes place on Sunday, similarly to the day after public holidays. Other weekdays have only minor impact on ED arrivals. On Sunday and Wednesday we expect slightly less arrivals. Tuesday and Thursday have zero as coefficient, indicating their impact is minor to none.

Table 7: Coefficients of weekdays obtained by applying lasso regression on the ED patient arrivals at Diakonessenhuis Utrecht, during the period 1-1-2017 to 2-29-2020.

Weekday	**Coefficient**
Monday	4.94
Tuesday	0
Wednesday	-0.54
Thursday	0
Friday	8.37
Saturday	5.20
Sunday	-0.32

Table 8 shows the importance of months. The majority of the months do not have significant impact on ED arrivals. However, July and August do show a relatively large decrease in ED arrivals, which can be explained due to Dutch people going abroad for holidays. May and June show slight increases in ED arrivals. An intuitive explanation is people are more active during these months, for example participating in outdoor activities, as these months take place during spring.

Table 8: Coefficients of months obtained by applying lasso regression on the ED patient arrivals at Diakonessenhuis Utrecht, during the period 1-1-2017 to 2-29-2020.

Month	**Coefficient**
January	-3.11
February	0
March	0
April	0
May	1.87
June	0.21
July	-4.66
August	-8.26
September	0
October	0
November	0
December	0

Table 9 shows the coefficients of week numbers. Most of the week number have few to none predictive power, as their coefficients are zero or close to zero. However, there are some interesting weeks that have higher coefficients. Week 1 has a relatively high positive coefficient, which can be explained by new years celebrations, which is known as one of the busiest days at the ED. Week 52 has one of the lowest negative coefficients, most likely because Christmas is usually during this week, which is known to be a quiet period at the ED, as the postponement effect holds for the days around Christmas. Weeks 29, 30, 31 and 32 all have relative high negative coefficients. This period is known in the Netherlands as the "bouwvak", in which employees from several job sectors have their holidays. As a result this period is known as the period in which most Dutch people go on vacation. As the population in Utrecht is temporarily lower, so are the ED arrivals.

While including the week numbers has a positive impact on the predictive power of the models, it comes with a risk, the importance of week numbers will shift over time. For example, the week numbers in which the "bouwvak" period takes place can differ per year.

Table 9: Coefficients of week numbers obtained by applying lasso regression on the ED patient arrivals at Diakonessenhuis Utrecht, during the period 1-1-2017 to 2-29-2020.

Week	**Coefficient**	**Week**	**Coefficient**	**Week**	**Coefficient**
1	5.32	19	0.80	37	0
2	0	20	0	38	0
3	0	21	0	39	1.35
4	0	22	0	40	0
5	0.29	23	0	41	0.77
6	-1.09	24	1.87	42	0
7	0	25	-2.00	43	-2.55
8	-0.10	26	0.90	44	-3.97
9	-2.85	27	5.24	45	0
10	0.78	28	0	46	0
11	-0.07	29	-3.84	47	0.78
12	0	30	-7.30	48	4.01
13	1.01	31	-2.31	49	0
14	0	32	-5.05	50	1.26
15	0	33	0	51	1.10
16	0	34	0	52	-6.08
17	-0.08	35	3.67	53	0
18	-3.22	36	-0.93		

Table 10 shows the coefficients of the remaining variables. These are all binary variables, indicating an event either did or did not happen. The importance of these variables is the difference in coefficients between the two variables. For example, when it is not a public holiday we expect 6.03 more ED arrivals over when it is a public holiday. This means the number of ED arrivals is lower on public holidays. Due to the postponement effect we see an increase in arrivals the day after. The coefficients of both weather warnings and FC Utrecht matches are close to zero, indicating low importance.

Table 10: Coefficients of public holidays, the day after public holidays, weather warnings and FC Utrecht matches obtained by applying lasso regression on the ED patient arrivals at Diakonessenhuis Utrecht, during the period 1-1-2017 to 2-29-2020.

Variable	**Coefficient**
Public holiday	0
No public holiday	6.03
Day after holiday	0
Not day after public holiday	-5.46
Weather warning	0
No weather warning	0.03
FC Utrecht match	0
No FC Utrecht match	0.13

Random forest

Recall that when building a tree in the RF, at each node m random variables out of the total p variables are chosen at random. From these m variables, the variable which splits the remaining data points such that the MSE of the remaining partitions is minimized is picked for the current split. Intuitively, the variables that split the data in such a way that the MSE is reduced the most have the best predictive power. Keeping track of the predictive power of each variable, at each split, in each tree, allows us to rank the variable importance. Figure 21 shows the relative importance of the most important variables which are found during the cross validation hyperparameter tuning. This means that the variables with the highest relative importance are picked most often, when building the RF trees. Appendix E shows the relative importance of all variables used in the RF.

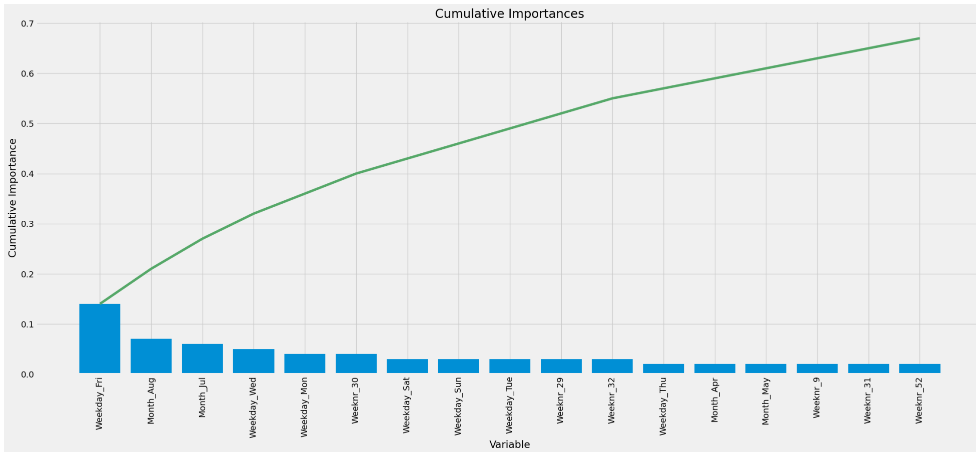

Figure 21: Variable importance plot obtained from fitting a RF on the training data, through cross validation. Only variables with relative importance of greater or equal to 0.02 are shown. The green line shows the cumulative importance.

The variable importance of the RF shows similar behavior as the coefficients of the lasso. Weekdays are the strongest predictors in emergency patient arrival at Diakonessenhuis, according to the RF. In particular, Friday is a strong predictor of ED arrivals, with a relative importance of 14%. This means that Fridays have the greatest impact on the number of ED arrivals. Among the weekdays, Tuesday is the least strong predictor. August and July are strong predictive months, which correspond to the summer months, in which most people have their holidays. This is in line with the strong drop in arrivals as seen in Figure 5, from the exploratory data analysis. The remaining months are not very strong predictors, however they do carry some predictive power.

The majority of the week numbers have low predictive power. However, week number 29, 30, and 32 are among the stronger predictors. These weeks are in the period between the end of July and the start of August. As explained earlier, this period corresponds to the bouwvak, in which a lot of Dutch people go on holidays.

According to the RF trees, the weather color codes do not have any predictive power on the arrivals. The variables that indicate FC Utrecht matches both have an importance of 1%, which indicates there is some predictive power, but very limited. In addition, due the limited number of data used in this research, this can also be a coincidence.

5.5 Conclusion

In this chapter we tuned the hyperparameters for the prediction models. Also, we developed a baseline prediction, derived from nurse scheduling at the hospital. The baseline prediction serves as a comparison for the prediction models we developed.

Among the tested prediction models the RF performs the best, with a MAPE of 12.21. The penalized linear models are very close in performance, with a MAPE of around 12.24. While the RF is slightly better, we prefer the use of the lasso model, due to its simplicity and intuitiveness. The lasso model has a 8.54% relative improvement over the baseline prediction.

The variables that contribute the most to the prediction are weekdays and months. Especially, Friday, Saturday, and Monday are busy days. In August, July, and January arrivals are lower than throughout the rest of the year. Public holidays also have predictive power. More specifically, on public holidays the arrivals are lower than usual. The day after a public holidays the arrivals are more than usual, due to the postponement effect.

In the next chapter, we provide a plan to implement the lasso model, intended for decision makers at Diakonessenhuis Utrecht.

6 Implementation

This chapter is intended for decision makers at Diakonessenhuis Utrecht to assist implementation. We answer the following research question:

How can the developed forecasting models assist decision-making in the hospital?

In Section 6.1 we discuss how to use the information from the forecasting model and to support decision making processes. In Section 6.2 we discuss which methods can be used to implement the forecasting model into practice. In Section 6.3 we provide a discussion on model maintenance.

6.1 Decision making support

The developed ED arrival prediction model predicts daily patient arrivals. These forecasts can be used to improve staff scheduling at the ED. Furthermore, the predictions can be used to predict downstream care demand. For example, at the X-ray department, surgery department, or to predict bed occupation at nursing departments.

Apart from the actual numbers coming out of the prediction model, the insights gained from the variable importance analysis in Section 5.4 can be used to improve planning activities throughout the hospital.

6.2 Implementation method

The hospital aims to add a daily forecast of ED arrivals to their existing KPI dashboard for the ED. We have provided the hospital with a forecasting tool, which takes input from an Excel file, creates predictions in Python, and write the output back to the Excel file. The most elegant option is to connect the dashboard software with the Python prediction object we provided, through an Application Programming Interface (API). This API would take as input a combination of input variables which hold for a given day, such as weekday and month etc., and outputs a prediction of daily patient arrivals at the ED. This option is the hardest to implement, but is the most robust to future changes, as changes made to the prediction model will automatically be updated.

A simpler option is to build the prediction model inside the dashboard software environment. Since the lasso prediction model is a linear combination of input variables and coefficients, implementing this into the dashboard environment is not difficult. The coefficients of the lasso model can be found in Section 3.3. Coefficients can also be extracted from the Python prediction object, in case changes are made to the model. Alternatively, predictions can be made in advance, using the prediction model in Python. Since short term variables, such as weather warnings, do not show any predictive power we can make predictions for a long period ahead. These predictions can be imported into the dashboard. However, these methods are less robust to updates to the prediction model.

6.3 Model maintenance

As seen in Section 4.1, the number of ED arrivals tend to change over time. In case of the data used in our research, a negative trend exists in ED arrivals. Trends have an impact on the performance of the models, as shown in Section 5.3. In addition, the weight of variables can change in the future. For example, people tend to postpone their hospital visits on Sundays, but this might change in the future. In addition, as discussed in Section 5.4, the importance of week numbers is likely to slightly change from year to year.

Currently our model uses a relatively simple method to deal with trends, as shown in Section 4.3. However, this method does not take into account potential changes in variable importance. More complex methods of model updating exists, such as continuous updating, where each day a new data point is collected and the model is refitted. However, continuous updating is hard to implement as it would require an automated updating procedure. The improvements of a continuously updated model may be minimal, making it potentially a waste of effort. Continuous methods are more suitable for rapidly changing environments, while the number of daily arrivals at a hospital are not likely to quickly change.

Another option is periodic updating, for example once a year. This method does not necessarily need an automated approach, but could be performed manually. Updating includes adding new data and refitting the model. Potentially new variables can be added to the model. The model will decide whether to use them, i.e. assign coefficients to the new variables.

We recommend periodic updating, because the model is sensitive to trends in ED arrivals. Additionally, model performance is likely to improve with more available data.

6.4 Conclusion

The ED forecasting model can assist decision makers at the hospital in staff scheduling. Furthermore, the forecasts can be used to predict downstream care demand. Also, the insights from variable importance can be used to better understand the underlying arrival process at the ED.

Model maintenance is important to make the model robust in the future. The model can be updated either continuously or periodical. While continuous updating is likely to outperform periodic updating, the marginal benefits may not justify the added complexity of implementation.

In the next chapter we summarize our research by providing conclusions, recommendations and a discussion.

7 Conclusion

This chapter finalizes this thesis. In Section 7.1 we conclude on our research. Section 7.2 contains recommendations, both for Diakonessenhuis as well as options for future research.

7.1 Conclusion

In this section we conclude on the most important findings in our research.

Model performance

The best performing model is the RF with a MAPE of 11.21, which is a 8.79% relative improvement over the baseline prediction. The penalized linear models lasso, ridge, and elastic net are close in performance to RF, with a MAPE of 11.24, 11.24 and 11.25, respectively. MLR scores a MAPE of 11.67.

Variable importance

The variables with the most predictive power are weekdays and months. More specifically, Mondays, Fridays and Saturdays are busier than other days of the week. Busy Mondays are explained by the postponement effect that takes places on Sunday. Friday and Saturday are the busiest days, as on these days people tend to be the most active. The summer months August and July are the strongest predictors amongst the months, with a decrease in arrivals. These periods correspond to periods in which most people go on holidays abroad. Most week numbers have little to no predictive power. However, some week numbers do show predictive power, especially the weeks that correspond to the "bouwvak" period, the most popular period for holidays in the Netherlands.

Public holidays also have an effect on ED arrivals. Similar to the postponement effect on Sundays, people tend to postpone their hospital visit on public holidays. This results in a decrease on public holidays, but an increase in the day after public holidays.

Weather warnings issued by the KNMI and FC Utrecht matches do not show any predictive power for ED arrvials.

Contribution to theory and practice

The theoretical contribution of our research is three-fold. Firstly, it adds another case study to the existing literature, further strengthening the methods used. Secondly, it addresses a gap in the literature by focusing on ED patient arrival forecasting specific to the Dutch Healthcare system. Furthermore, while most variables in this research have been studied before, the inclusion of a variable indicating extreme weather conditions is a novel contribution to the field.

The practical contribution is the developed prediction model, as well as the insights gained into variable importance. Both can be used to improve resource allocation, which counteracts crowding at the ED and therefore improves quality of care and staff satisfaction. Furthermore, improved resource allocation reduces costs, as demand better matches supply, reducing wasted resources.

7.2 Recommendations

In this section we provide recommendations on model choice, model implementation, and model maintenance.

Model choice

Even though the RF has the best performance amongst the tested models, we recommend the use of the lasso model. The lasso has a similar performance and is easier to understand, implement and maintain.

Model implementation

To implement the forecasting model we developed we recommend building an API, which takes input data from the hospital's database and outputs daily predictions. Building an API has the most flexibility and scalability, but it is also the most technologically challenging to build.

An alternative is to use the tool we developed to predict a period ahead, for example one year, which writes results to Excel. Then these results can be manually imported to the ED KPI dashboard.

Model maintenance

To keep model performance in the future, model maintenance is necessary. Model maintenance is even likely to increase the model performance. As continuous updating is technologically challenging and has limited benefits, we recommend periodic updating each year. Updating includes adding new data and refitting the model. Potentially new variables can be added. The lasso model automatically decides if the new variables should be included in the model.

7.3 Discussion

In this section we discuss limitations to our research and provide directions for further research.

Model choice

As discussed in Chapter 3, many more prediction models exist. We only tested a small subset of the available models for ED forecasting. Future research can be conducted to test different models, potentially leading to an improvement of model performance. Especially the use of state of the art machine learning models can lead to improvement, which has been the focus of recent scientific research in the field of ED forecasting. However, the downside of these models is a decrease in intuitiveness and they are harder to implement in practice.

Model performance

While a relative improvement of 8.79% over the baseline prediction is a significant and valuable improvement, further research can improve this even more. In our research the available data is limited to 3 years of ED arrivals. More available data is likely to improve the predictive power of the models. Furthermore, future research could attempt to include more variables that can explain ED arrivals. For instance, in this research we include professional soccer matches in the hospital area as a variable. Future research can attempt to include more events that could influence ED arrivals.

Related research directions

The ultimate goal of ED forecasting is to improve the utilization resources, such as nurses, doctors or technical equipment. Besides an attempt to improve our forecasting models, future research can focus on improving resource utilization related to ED forecasting. As a result of our research we know how accurate we can predict ED arrivals, including a distribution of errors. This is useful input for the analysis of related processes. For example, staff scheduling at the ED or the patient flows to auxiliary departments in the hospital, such as diagnostic or nursing departments.

References

Batal, H., Tench, J., McMillan, S., Adams, J., & Mehler, P. S. (2001). Predicting patient visits to an urgent care clinic using calendar variables [Journal Article]. *Academic Emergency Medicine*, *8*(1), 48-53. doi: 10.1111/j.1553-2712.2001.tb00550.x

Baxter, M., & King, R. G. (1999). Measuring business cycles: approximate band-pass filters for economic time series. *Review of economics and statistics*, *81*(4), 575–593.

Binns, M., & Ayub, H. M. U. (2021). Model reduction applied to empirical models for biomass gasification in downdraft gasifiers. *Sustainability*, *13*(21), 12191.

Carvalho-Silva, M., Monteiro, M. T. T., Sá-Soares, F. D., & Dória-Nóbrega, S. (2018). Assessment of forecasting models for patients arrival at emergency department [Journal Article]. *Operations Research for Health Care*, *18*, 112-118.

Chopra, S., & Meindl, P. (2016). *Supply chain management: Strategy, planning, and operation.* Pearson.

Choudhury, A., & Urena, E. (2020). Forecasting hourly emergency department arrival using time series analysis [Journal Article]. *British Journal of Health Care Management*, *26*(1), 34-43.

Christiano, L. J., & Fitzgerald, T. J. (2003). The band pass filter. *international economic review*, *44*(2), 435–465.

de Brito, F. G., Resende, E. S., Rodrigues, A. A. A., Junqueira, M. A. B., Barreto, V. R., & Filho, J. B. D. E. S. T. R. O. (2019). Demand forecast in the emergency department in minas gerais, brazil [Journal Article]. *Bioscience Journal*, *35*(5), 1640-1650.

Gul, M., & Celik, E. (2020). An exhaustive review and analysis on applications of statistical forecasting in hospital emergency departments [Journal Article]. *Health Systems*, *9*(4), 263-284. doi: 10.1080/20476965.2018.1547348

Harrou, F., Dairi, A., Kadri, F., & Sun, Y. (2020). Forecasting emergency department overcrowding: A deep learning framework [Journal Article]. *Chaos, Solitons and Fractals*, *139*.

Harrou, F., Sun, Y., Kadri, F., Chaabane, S., & Tahon, C. (2016). Early detection of abnormal patient arrivals at hospital emergency department [Conference Proceedings]. In (p. 221-227).

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.

Hu, X., Barnes, S., & Golden, B. (2018). Applying queueing theory to the study of emergency department operations: a survey and a discussion of comparable simulation studies. *International transactions in operational research*, *25*(1), 7–49.

Jilani, T., Housley, G., Figueredo, G., Tang, P. S., Hatton, J., & Shaw, D. (2019). Short and long term predictions of hospital emergency department attendances [Journal Article]. *International Journal of Medical Informatics*, *129*, 167-174.

Jones, S. S., Thomas, A., Evans, R. S., Welch, S. J., Haug, P. J., & Snow, G. L. (2008). Forecasting daily patient volumes in the emergency department [Journal Article]. *Academic Emergency Medicine*, *15*(2), 159-170.

Kadri, F., Harrou, F., & Sun, Y. (2018). A multivariate time series approach to forecasting daily attendances at hospital emergency department [Conference Proceedings]. In (Vol. 2018-January, p. 1-6).

Knmi waarschuwingen. (2022). Retrieved from `https://www.knmi.nl/kennis-en-datacentrum/uitleg/knmi-waarschuwingen`

Kudyba, S. (2018). A hybrid analytic approach for understanding patient demand for mental health services [Journal Article]. *Network Modeling and Analysis in Health Informatics and Bioinformatics*, *7*(1). doi: 10.1007/s13721-018-0164-2

Kuhn, M., Johnson, K., et al. (2013). *Applied predictive modeling* (Vol. 26). Springer.

Lin, W. D., & Chia, L. (2017). Combined forecasting of patient arrivals and doctor rostering simulation modelling for hospital emergency department [Conference Proceedings]. In (Vol. 2017-December, p. 2391-2395).

Liu, G., & Guo, X. (2022). Forecasting stock market volatility using commodity futures volatility information. *Resources Policy*, *75*, 102481.

Mataloni, F., Pinnarelli, L., Perucci, C. A., Davoli, M., & Fusco, D. (2019). Characteristics of ed crowding in the lazio region (italy) and short-term health outcomes [Journal Article]. *Internal and Emergency Medicine*, *14*(1), 109-117.

Ong, M. E. H., Ho, K. K., Tan, T. P., Koh, S. K., Almuthar, Z., Overton, J., & Lim, S. H. (2009). Using demand analysis and system status management for predicting ed attendances and rostering [Journal Article]. *American Journal of Emergency Medicine*, *27*(1), 16-22.

Probst, P., Wright, M. N., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, *9*(3), e1301.

Rema, V., & Sikdar, K. (2021). Time series modelling and forecasting of patient arrivals at an emergency department of a select hospital [Serial]. In (Vol. 1333 AISC, p. 53-65).

Rocha, C., & Rodrigues, F. (2021). Forecasting emergency department admissions [Article]. *Journal of Intelligent Information Systems*, *56*(3), 509 – 528. (Cited by: 3)

Savage, D. W., Woolford, D. G., Weaver, B., & Wood, D. (2015). Developing emergency department physician shift schedules optimized to meet patient demand [Journal Article]. *Canadian Journal of Emergency Medicine*, *17*(1), 3-12. doi: 10.2310/8000.2013.131224

Schoolvakanties. (2023). Retrieved from `https://www.anwb.nl/vakantie/schoolvakanties`

Sudarshan, V. K., Brabrand, M., Range, T. M., & Wiil, U. K. (2021). Performance evaluation of emergency department patient arrivals forecasting models by including meteorological and calendar information: A comparative study [Journal Article]. *Computers in Biology and Medicine*, *135*.

Vollmer, M. A., Glampson, B., Mellan, T., Mishra, S., Mercuri, L., Costello, C., ... Bhatt, S. (2021). A unified machine learning approach to time series forecasting applied to demand at emergency departments. *BMC Emergency Medicine*, *21*(1), 1–14.

Wargon, M., Guidet, B., Hoang, T. D., & Hejblum, G. (2009). A systematic review of models for forecasting the number of emergency department visits [Journal Article]. *Emergency Medicine Journal*, *26*(6), 395-399. doi: 10.1136/emj.2008.062380

Wester, R. A., Rubel, J., & Mayer, A. (2021). Covariate selection for estimating individual treatment effects in psychotherapy research: A simulation study and empirical example.

Wiler, J. L., Griffey, R. T., & Olsen, T. (2011). Review of modeling approaches for emergency department patient flow and crowding research [Journal Article]. *Academic Emergency Medicine*, *18*(12), 1371-1379. doi: 10.1111/j.1553-2712.2011.01135.x

Xing, L.-M., & Zhang, Y.-J. (2022). Forecasting crude oil prices with shrinkage methods: Can nonconvex penalty and huber loss help? *Energy Economics*, *110*, 106014.

Xu, K., & Chan, C. W. (2016). Using future information to reduce waiting times in the emergency department via diversion [Journal Article]. *MSom-Manufacturing Service Operations Management*, *18*(3), 314-331. doi: 10.1287/msom.2015.0573

Xu, M., Wong, T. C., & Chin, K. S. (2013). Modeling daily patient arrivals at emergency department and quantifying the relative importance of contributing variables using artificial neural network [Journal Article]. *Decision Support Systems*, *54*(3), 1488-1498.

Xu, Q., Tsui, K.-L., Jiang, W., & Guo, H. (2016). A hybrid approach for forecasting patient visits in emergency department. *Quality and Reliability Engineering International*, *32*(8), 2751–2759.

Zhang, Y., Zhang, J., Tao, M., Shu, J., & Zhu, D. (2022). Forecasting patient arrivals at emergency department using calendar and meteorological information [Journal Article]. *Applied Intelligence*.

A Systematic literature review process

In this appendix we show how we performed a SLR.

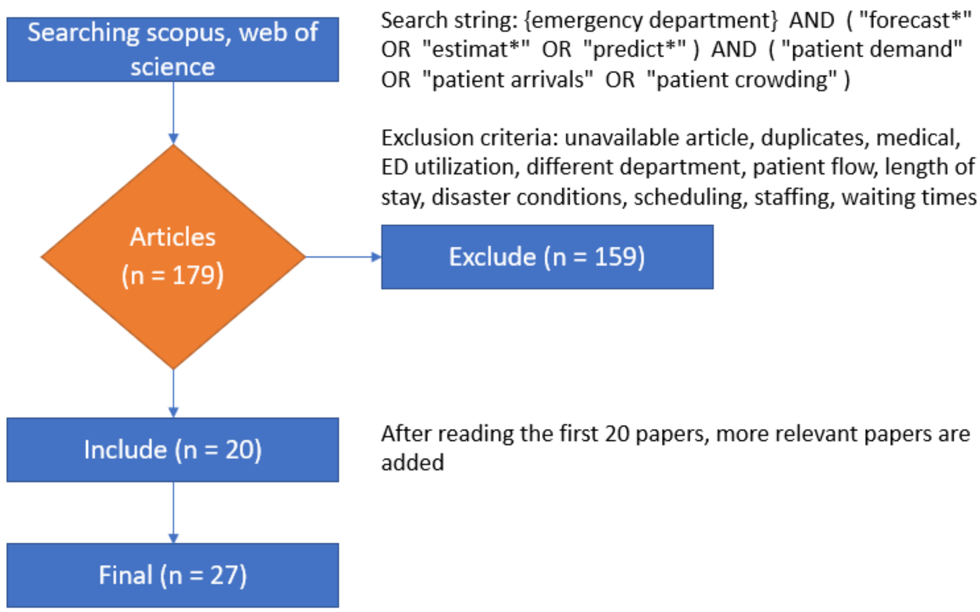

Figure 22: SLR process

B Normality check patient arrivals

In this appendix we check the daily patient arrivals for normality. We use a histogram, a Q-Q plot and a Shapiro-Wilk test to check for normality.

Figure 23 shows signs of a normal distribution, according to its bell shaped curve.

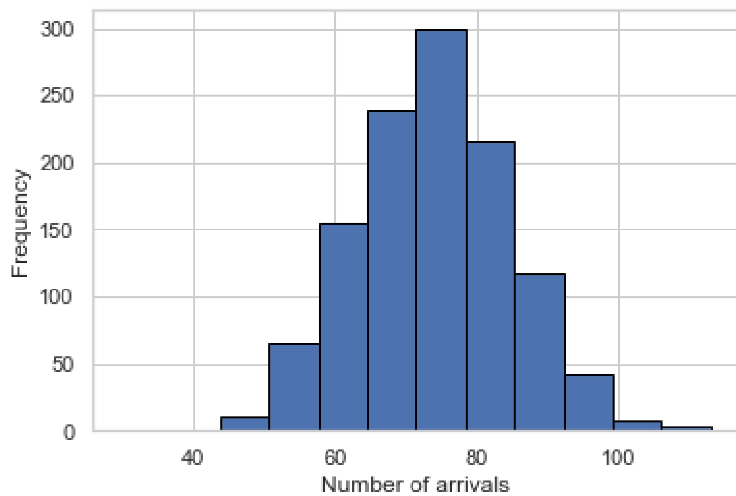

Figure 23: Histogram of daily patient arrivals (period 1-1-2017 to 2-29-2020, n = 1155, source: Diakonessenhuis Utrecht).

The Q-Q plot in Figure 24 shows evidence of a normal distribution. Most of the data is around the expected values. The only exceptions are the outer quantiles.

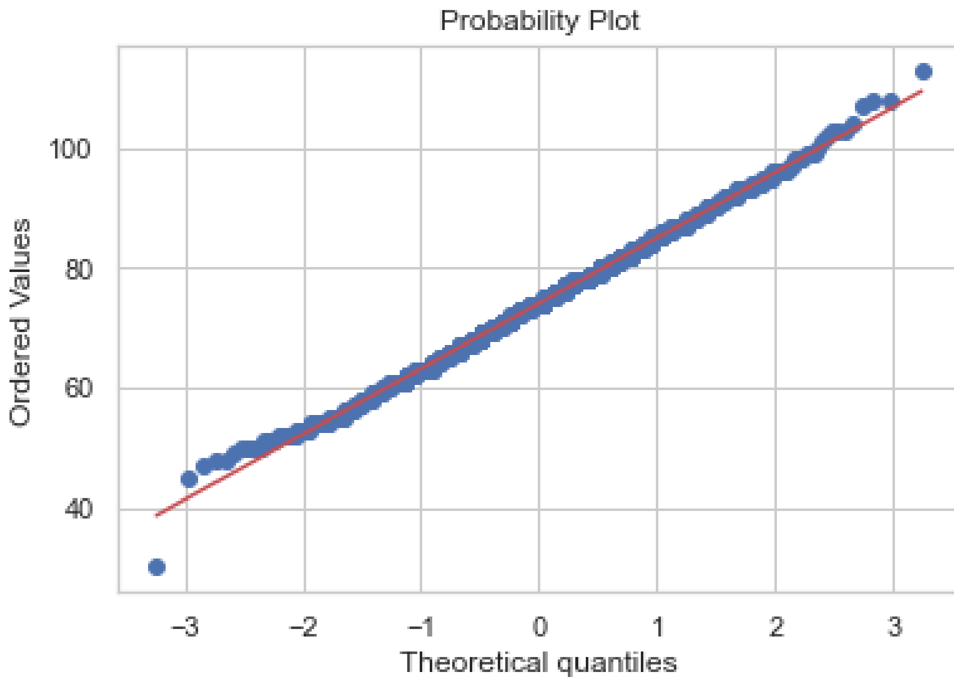

Figure 24: Q-Q plot to check normality of daily patient arrivals (period 1-1-2017 to 2-29-2020, n = 1155, source: Diakonessenhuis Utrecht).

Performing a Shapiro-Wilk test on the data with $\alpha = 0.05$ yields a test statistic of 0.997 with a corresponding p value of 0.066. Since the p value is larger than 0.05, we do not reject the null hypothesis, which states the data comes from a normal distribution. Therefore, based on the Shapiro-Wilk test, we have reason to believe the data is normally distributed.

Based on the histogram, the Q-Q plot and the Shapiro-Wilk test we have statistical evidence to assume the daily patient arrivals are normally distributed with $\mu = 74.02$ and $\sigma = 10.92$.

C Normality check lasso residuals distribution

In this appendix we check the residuals of the lasso predictions for normality. We use a histogram, a Q-Q plot and a Shapiro-Wilk test to check for normality.

Figure 25 shows signs of a normal distribution, according to its bell shaped curve.

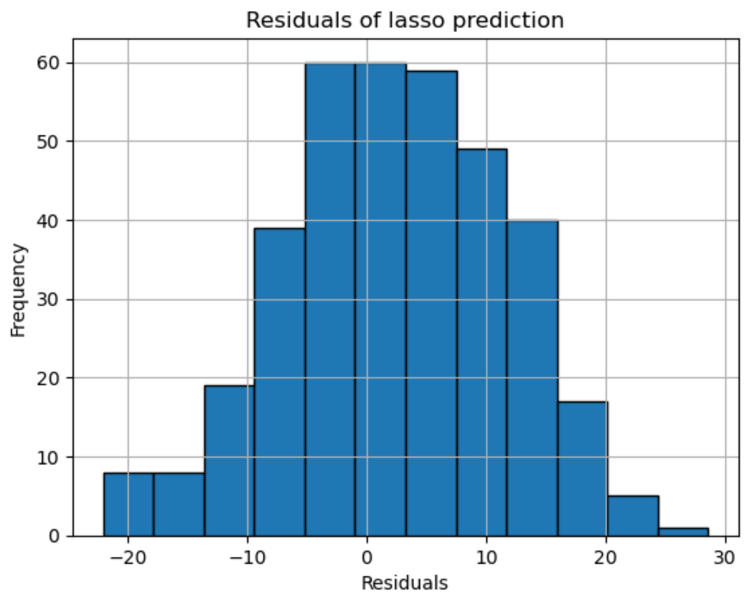

Figure 25: Distribution of residuals of predictions on the test data, using lasso regression.

The Q-Q plot in Figure 26 shows evidence of a normal distribution. Most of the data is around the expected values. The only exceptions are the outer quantiles, which is in line with the original arrival distribution, as seen in Appendix B.

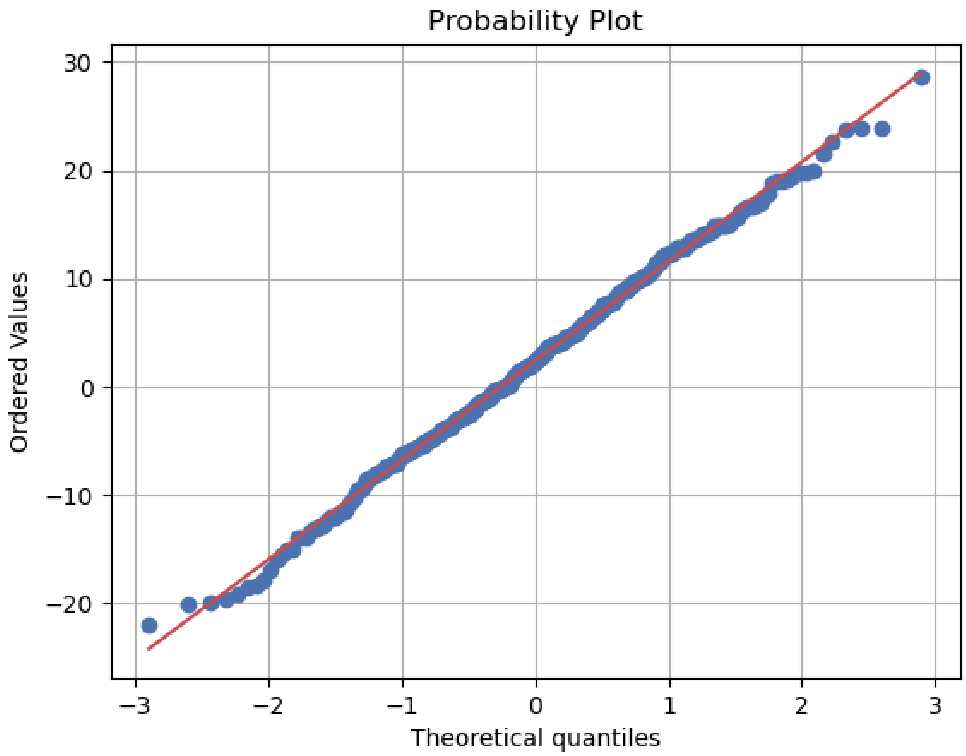

Figure 26: Q-Q plot to check normality of residuals of lasso predictions.

Performing a Shapiro-Wilk test on the data with $\alpha = 0.05$ yields a test statistic of 0.997 with a corresponding p value of 0.754. Since the p value is larger than 0.05, we do not reject the null hypothesis, which states the data comes from a normal distribution. Therefore, based on the Shapiro-Wilk test, we have reason to believe the data is normally distributed.

Based on the histogram, the Q-Q plot and the Shapiro-Wilk test we have statistical evidence to assume the residuals of lasso predictions are normally distributed with $\mu = 2.37$ and $\sigma = 9.15$.

D Distributions of predictions

In this appendix we show the distributions of the predictions on the test data for the ridge, elastic net, RF and the baseline prediction.

Note that the baseline prediction in Figure 30 only predicts three different values, and therefore has a unusual distribution.

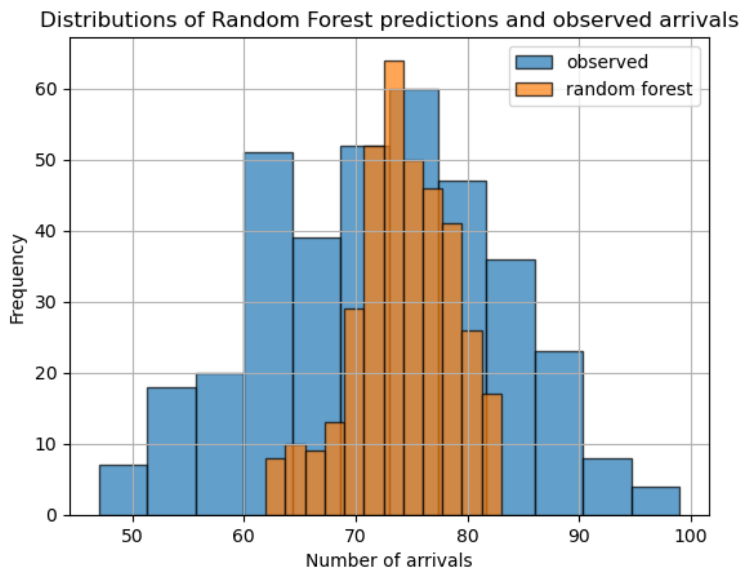

Figure 27: Distribution of residuals of predictions on the test data, using RF.

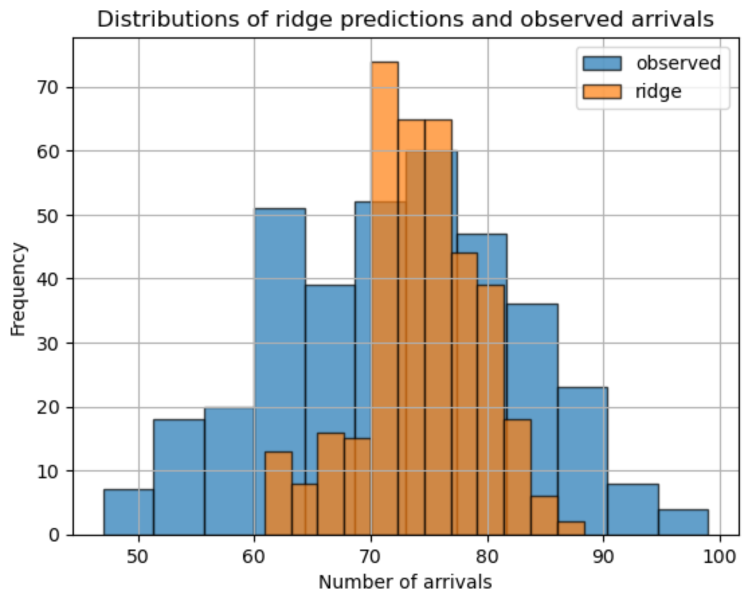

Figure 28: Distribution of residuals of predictions on the test data, using ridge regression.

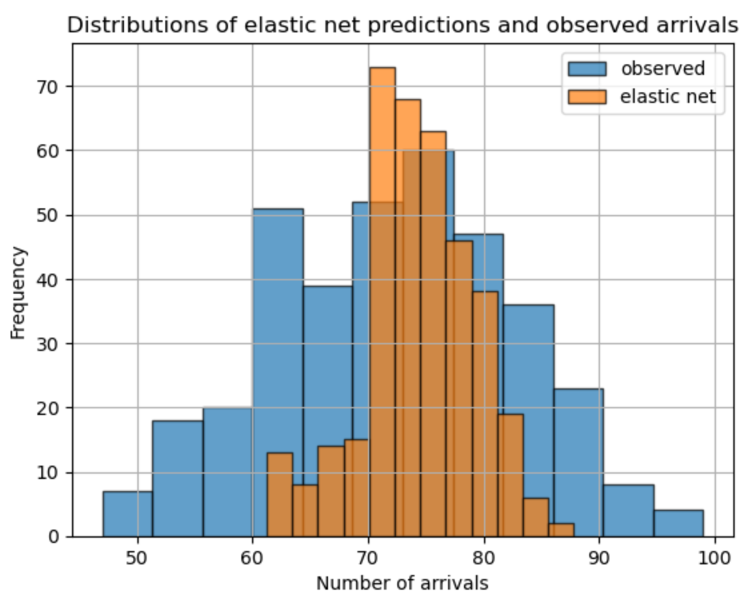

Figure 29: Distribution of residuals of predictions on the test data, using elastic net regression.

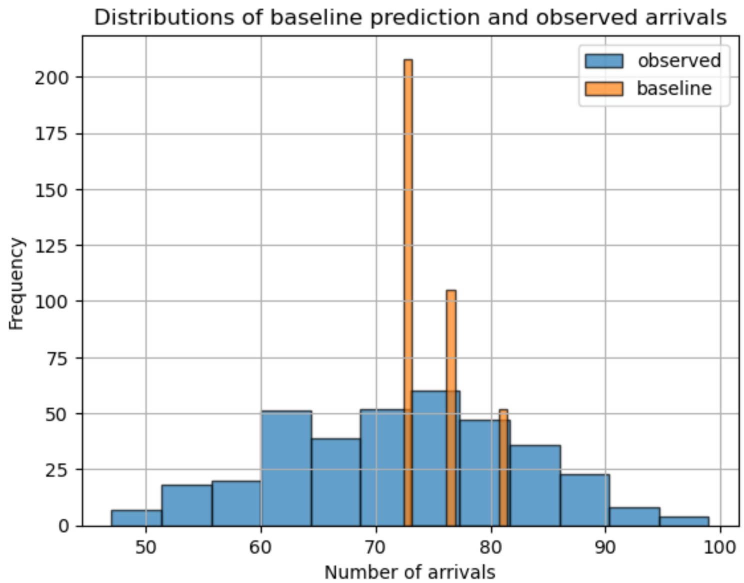

Figure 30: Distribution of residuals of predictions on the test data, using the baseline method.

E RF variable importance

In this appendix we provide the relative performance of variables when fitting a RF model to the ED arrival data.

Table 11: Relative importance of variables obtained by applying a RF on the ED patient arrivals at Diakonessenhuis Utrecht, during the period 1-1-2017 to 2-29-2020.

Var	**Imp.**	**Var**	**Imp.**	**Var**	**Imp.**	**Var**	**Imp.**
Fri	0.14	Jun	0.01	Week 24	0.01	Week 20	0
Aug	0.07	Mar	0.01	Week 25	0.01	Week 22	0
Jul	0.06	Nov	0.01	Week 26	0.01	Week 23	0
Wed	0.05	Okt	0.01	Week 27	0.01	Week 28	0
Mon	0.04	Sep	0.01	Week 33	0.01	Week 34	0
Week 30	0.04	Holiday F	0.01	Week 42	0.01	Week 35	0
Sat	0.03	Holiday T	0.01	Week 43	0.01	Week 36	0
Sun	0.03	After holiday F	0.01	Week 51	0	Week 37	0
Tue	0.03	After holiday T	0.01	Warning F	0	Week 38	0
Week 29	0.03	FC Utrecht F	0.01	Warning T	0	Week 39	0
Week 32	0.03	FC Utrecht T	0.01	Week 2	0	Week 40	0
Thu	0.02	Week 1	0.01	Week 5	0	Week 41	0
Apr	0.02	Week 3	0.01	Week 6	0	Week 44	0
May	0.02	Week 4	0.01	Week 8	0	Week 45	0
Week 9	0.02	Week 7	0.01	Week 11	0	Week 46	0
Week 31	0.02	Week 10	0.01	Week 12	0	Week 47	0
Week 52	0.02	Week 13	0.01	Week 15	0	Week 48	0
Dec	0.01	Week 14	0.01	Week 16	0	Week 49	0
Feb	0.01	Week 19	0.01	Week 17	0	Week 50	0
Jan	0.01	Week 21	0.01	Week 18	0	Week 53	0