

Scientific Paper for the  
Interdisciplinary Project (IDP)  
at Technische Universität München

## Titel

Betreuer: M. Sc. Alexander Döge  
M. Sc. Christian Ruf  
Lehrstuhl für Operations Management  
Technische Universität München

Studiengang: Informatics Master

Eingereicht von: Johannes Neutze  
Haager Strasse 74  
85435 Erding  
Tel.: +49 157 78948410  
Matrikelnummer: 03611960

Eingereicht am: 14.07.2015

The food delivery market has grown over the last few years but is now almost saturated. In order to survive, delivery services have to improve their way of working. Volo has done this step by introducing an improved travelling salesman algorithm to save resources and time for the drivers. But the path finding algorithm can still be improved by tuning the time component which cannot be calculated by the algorithm itself. This interdisciplinary project has the goal to create a reliable forecast for the preparation times of food for restaurants so the driver can arrive right on time.

# Table of Contents

<b>List of Figures</b>	<b>iii</b>
<b>List of Tables</b>	<b>iv</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Review of Literature and Research</b>	<b>2</b>
2.1 Research . . . . .	2
2.2 Approach and Basics . . . . .	2
2.2.1 Time Series . . . . .	3
2.2.2 Algorithms for Forecasting . . . . .	3
2.2.2.1 Moving Average . . . . .	3
2.2.2.2 Weighted Moving Average . . . . .	4
2.2.2.3 Simple Exponential Smoothing . . . . .	4
<b>3 Methodology</b>	<b>5</b>
3.1 Problem Definition . . . . .	5
3.2 Gathering Information . . . . .	5
3.3 Preliminary Analysis . . . . .	6
3.3.1 Raw Data Cleanup . . . . .	6
3.3.2 Restaurant Wise Proceeding . . . . .	9
3.4 Choosing and Fitting Models . . . . .	10
3.4.1 Categorizing of Orders . . . . .	10
3.4.2 Combination of Categorizations . . . . .	11
<b>4 Results</b>	<b>13</b>
<b>5 Conclusion</b>	<b>14</b>
<b>Bibliography</b>	<b>15</b>
<b>Appendix</b>	<b>17</b>
1 Appendix One . . . . .	17
2 Appendix Two . . . . .	18

## List of Figures

3.1 An example figure . . . . .	7
3.2 An example figure . . . . .	7
3.3 An example figure . . . . .	8
3.4 An example figure . . . . .	8
3.5 An example figure . . . . .	9
3.6 An example figure . . . . .	9
3.7 An example figure . . . . .	10
3.8 An example figure . . . . .	10
3.9 An example figure . . . . .	12

## List of Tables

4.1 An example table . . . . .	13
--------------------------------	----

# 1 Introduction

The food delivery market is a fast growing market with an giant volume. Rocket internet predicts the market to have a value of 90 billion Euro by 2019. CITE!! This money attracts many companies which fight for the supremacy in the market. In order to achieve this goal they have to differentiate from their competitors. Some do this by being the cheapest, others have the biggest variety of restaurants they offer and still others try to optimize the process from ground up.

An example is is the Munich based startup called VOLO. The idea of VOLO is to provide a great and effective experience in food delivery. Starting with premium restaurants with an appealing shop, followed by an process optimized driver fleet for the restaurants and fast deliveries. The optimized and fast deliveries are backed by an underlying algorithm. The algorithm is based on the travelling salesman algorithm and calculates the best solution for a number of drivers who have to fulfill a number of deliveries. It assigns each driver a route and deliveries he has to fulfill including the route the driver should take and the time needed to complete the task. This enables drivers have less empty drives and to minimize idle times from which the company and the driver profits. The driver is able to earn more money from loaded times and tips and the company has to pay few idle times where the driver is not used.

This algorithm is perfect from the point in time the driver instantly picks up the food, leaves the restaurant and starts driving. The only problem is that the food is almost never ready at the moment the algorithm sends the driver to the restaurant. It is crucial for a food delivery to pick the meal up at the exactly right time. On the one hand being late is bad, because if the customer gets his food cold or too late he is likely not to order again from the company. On the other hand when the driver is too early at the restaurant he has to wait for the order to be prepared. In this case the time gain is lost and with it the advantage over the other competitors. This is why VOLO has the need to create a forecast which tells the algorithm the forecasted preparation time for the food so the driver can be send to the restaurant just in time. The following chapters will explain the proceeding to solve the problem. It will start with the resources used to generate knowledge, then focus on the methodology used to forecast and it will finish with evaluating the result and a conclusion.

## **2 Review of Literature and Research**

### **2.1 Research**

In order to get a good overview over the problem, a search in Google Scholar was done. The goal was to find similar problems and approaches to it. Different words were chosen for the search, like "preparation time", "meal", "restaurant" and "forecast". The results did not give the right output for this problem. Most of the forecasting in restaurant was about the amount of staff needed, the amount of meals which will be sold over a period or the size a buffet has to be. These topics did not fit the problem given because the problem stated is pretty unique. On the one hand, most of the delivery services have their own fleet of drivers and are not that time critical as volo is. On the other hand has not yet a company combined a food delivery fleet with a routing algorithm. This is why a custom approach had to be made.

A plan with different steps was created. As a first step, a solid base about forecasting had to be made by searching for basics of forecasting. In the second step, this knowledge should be applied to the case before evaluating it in a third step.

### **2.2 Approach and Basics**

After searching for fitting materials, the book "Forecasting: Principles and Practice" by R. Hyndman and A. Athanasopoulos was chosen as a starting point, since it covers the topic pretty good for forecasting beginners. After reading the book it was determined to follow the steps the book suggests to create the forecast. The book had five basic steps for forecasting.

First of all, you have to define the problem. This is simple but often the most time consuming part of the whole task. The forecaster has to talk to everybody who uses the forecast as well as the people who support the data to generate it. Questions regarding the way it will be used, by who it is required and how it fits into the process of the ones who required it have to be asked. Also the people who support the forecast with collecting data, maintaining the data and use the forecast for future planning have to be figured out. After the problem is defined, in a second step, all information available is gathered. This requires to get all statistical data available, but often there will be not enough historical data to create a decent statistical model. The available data is combined with the expertise of the people who collected this

data and use the forecasts. In the next step, preliminary analysis is done. The data is put into a graph to get a first rough overview over the behavior. This is a simple way to see patterns, trends, seasonalities or business cycles. Outliers can be spotted and can be questioned early. Now the real forecasting can begin. In the next step the model is chosen and fitted to the data. The models are created from the historical data and all other knowledge which can support the forecast. Different potential models are used and later compared. This is done in the last step, when the model is used and evaluated. This is done by comparing the outcomes of the different models with the actual events which are forecasted.

Now that there is a basic approach, more information about the different models is needed. This information is taken from PowerPoint slides which are provided by the Department of Logistics. The slides are from lectures with forecasting as a topic and are used to get a rough overview over different simple algorithms often used for forecasting.

### 2.2.1 Time Series

Since the events are happening on a time line, time series decomposition was the choice. Time series can have different patterns that is why putting the data in a graph first is very useful. The book and the slides likewise addressed common patterns and methods which can be used on time series and improve forecasts. The time series can then be split into different parts to support the corresponding pattern. There are three types of time series patterns. The first pattern is the trend. The trend is an decrease or increase over a long period of time. The second is the season pattern. It is influenced by seasonal factors and has a known and fixed time period and. The third one is the cyclic pattern. Cyclic pattern influence usually has fluctuations of at least 2 years and the period is not fixed. The available data should be transformed to a time series and divided into suitable components, in case patterns are present.

### 2.2.2 Algorithms for Forecasting

Now that the time series is investigated for patterns, it is time to forecast. The book and slides offer different algorithms to fulfill this task. Three algorithms are chosen to be presented more closely.

#### 2.2.2.1 Moving Average

A classical method is the moving average. It is used to calculate the next forecast value of a time series iteratively. For the time series it is done by taking all prior events in account for the next value. The calculated values are independent. The formula for the average at time  $t$  is:  $m(t) = 1/n \sum_{i=0}^{n-1} x(t-i)$



### 2.2.2.2 Weighted Moving Average

The window which is taken into the calculation of the next value can also be weighted. In contrast to the normal Moving Average, a window for the amount of values included in the creation of the average value. The window moves with the time series. When calculating the next value, the first value is removed from the window. This way only a specific amount of values stays in the calculation.

### 2.2.2.3 Simple Exponential Smoothing

Cite papers with `\cite{BibTeX key}`

A comprehensive introduction to Operations Research: Winston (2007)

## 3 Methodology

The following part explains the single steps of the forecasting process for this Interdisciplinary Project in detail. The forecast will be presented from the very beginning. First of all, the problem will be defined properly. Then the information needed will be specified and gathered. This data is then analyzed in a preliminary step and after this models for the data will be examined. In the end the result of the models will be evaluated and further proceeding will be explained in the final conclusion.

### 3.1 Problem Definition

Before the forecasting can start, some general questions have to be solved. Most important is the question of who is requesting the forecast and what will the forecast be used for. In order to acquire this information, the COO, Emanuel Pallua, who is requesting this forecast, is interviewed. The aim is to forecast the preparation time of a restaurant. This is needed to be able to send the driver to the restaurant just in time when the meal is prepared. This improves the algorithm and reduces driver idle times. The forecast should be done once to be evaluated and later the most accurate method should be integrated into the algorithm itself for automatization. In addition to this, Sebastian Sondheimer gave some input regarding the forecast parameters. The forecast should include different periods, like the current time, day and a limit time of the past. This will be explained in more detail later.

After having the requirements from business side, the tech side is examined. Sergej Krauze, CTO and responsible for the algorithm, explained that the algorithm takes all restaurants with open orders, the customers and the drivers to calculate the most efficient route to pickup and delivery the food. The algorithm is written in Java so it would be welcome to develop the forecasting in Java as well. Finally Stefan Rothlehner, CTO and responsible for the backend, had to be questioned, what data is available and how it can be extracted. Every order is stored in a PostgreSQL database on heroku, on which it can be accessed and downloaded.

### 3.2 Gathering Information

The database is dumped at two dates in time. The training data set was downloaded at the 26th of March 2015 to create a forecast which later can be compared to a test

data set. This test data set was extracted from the server on the 29th of April. Speaking in number of orders, the csv file of the first extraction has 3034 entries while the second one has 4973 entries. In addition to the historic data, the operations team was interviewed for their personal experience with the preparation times. They said that estimating around 15 minutes is more or less accurate except for some restaurants which are known to take much longer.

This information is valuable since the historical data has some weaknesses. The biggest problem is the size of the gathered data. Forecasts need big amounts of information to generate a meaningful result so the data.

In order to use the datasets from the csv files, it has to be parsed into objects in the Java program. For this purpose a csv parser library is used. The library reads the csv file and matches the orders from the database to the OrderModel.class of the code. Not all attributes of the database are used, only the one related to the forecast are picked from the information. Since there is no tracking in the restaurant for the preparation time, it was decided to take the time interval from the point in time at which the restaurant knows from the order until the driver leaves the restaurant. In the process of volo, the printer in the restaurant prints the recipe at the same time the driver accepts the delivery, which is saved in the database as `accepted_at`. The timestamp of the driver leaving the restaurant is saved in `"delivery_started_at"`. *Since it is not the task to figure out the exact the preparation time but the time*

### 3.3 Preliminary Analysis

The weather or season can also affect the preparation time. There can be seasonal cycles or trends caused by rising brand awareness. All these different situations have to be taken into account. There can also occur other factors, like marketing campaigns, but this has not happened in the recorded time frame. In order to discover these patterns and get a feeling for the data, a preliminary analysis is done.

#### 3.3.1 Raw Data Cleanup

First of all the average preparation time of the orders is calculated. This is done to get a feeling in which how long the restaurants usually take. It can give a rough estimation of what can be expected and compared to the forecast later to find out whether the result is total unrealistic or pretty close to what is possible. In addition to the average time, a visualization of the data is done. The best way to do this is to use a time series plot. In the graph it is very easy to see anomalies and patterns.

The average time and the visualized raw data is shown and explained below.

The analysis of the diagram reveals the first problems of the data. The orders from the trainings dataset still contain unfinished orders, bugged orders, which were finished at a wrong point in time, and orders which have corrupted time stamps. Un-

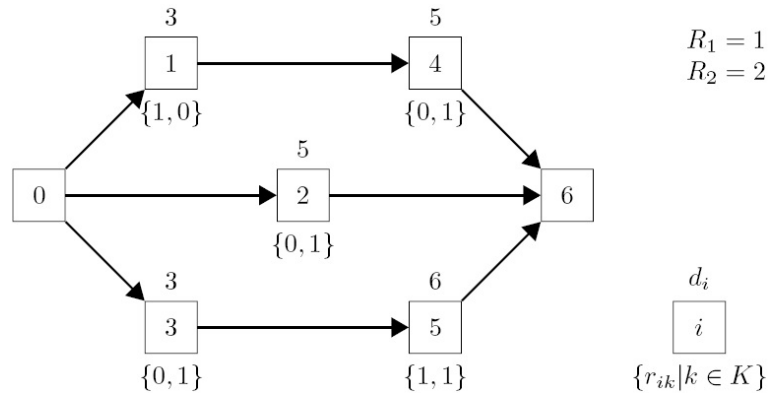


Figure 3.1: An example figure

finished orders have missing timestamps and like unreadable timestamps they get as "total"-time attribute -1 to be detectable. This can be observed in many cases in the diagram above. Also an order which was finished the next day can be seen around the 100 mark. This flaws in the data lead to a result of 12 minutes of average preparation time. This cannot be taken as a serious results since the -1 results lower the average drastically below of what the real average would be like. These values have no value to the forecast and have to be removed from the pool. This is done by removing all -1 from the used data. The result can be seen in [] below. This increases the average preparation time to 16 minutes. This can be explained because of the missing -1 values.

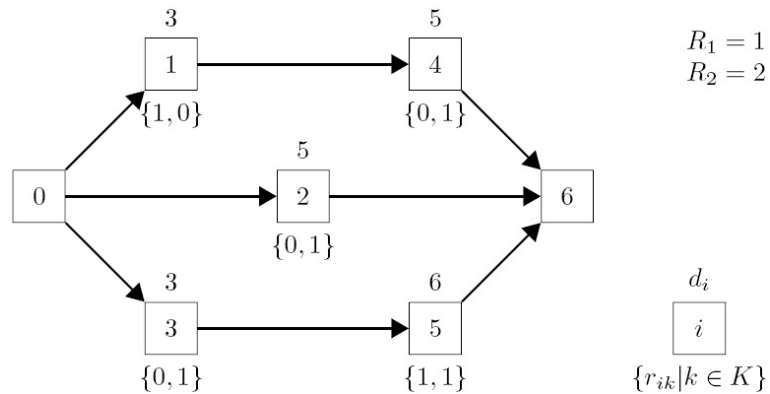


Figure 3.2: An example figure

After removing the obviously unusable orders, the ones which were finished long time after completing the delivery, mostly because of bugged software, have to be removed. For this purpose the Grubbs training for outliers is applied to the dataset and all values seem not to come from a normally distributed population are removed. This results in an average preparation time of 15 minutes which is the same result as the operations team suggest to use as a basis.

Now the usable training dataset is extracted from the input data, we can analyze it for patterns for patterns or similarities.

0000<sub>g</sub>rubbedisveryfineandsortedbyorder.Itwascreatedtogetanoverviewfocusedonorders

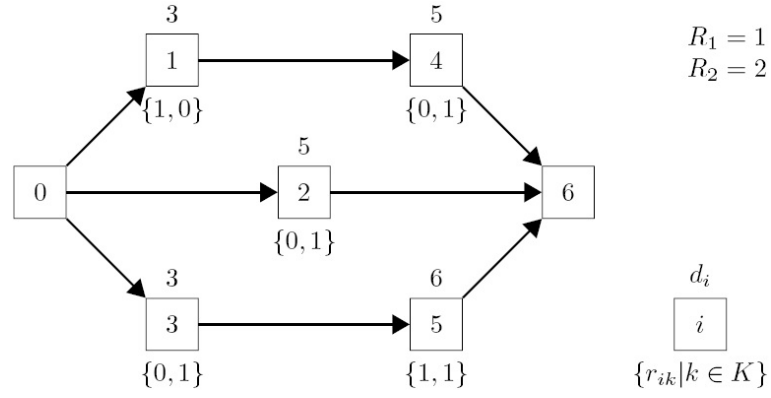


Figure 3.3: An example figure

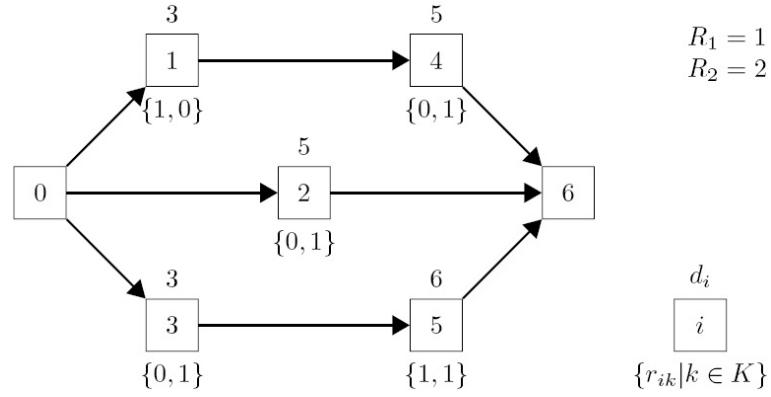


Figure 3.4: An example figure

Since no clear trend or pattern can be observed in a day by day analysis, a box and whisker diagram is created from the data. It shows the median and average preparation time as line and point inside the box. The box represents the upper and lower quartile of preparation times in which 50 % of the data is while the red whiskers visualize the area without outliers, which are red circles and a red triangle in case they are out of scale. The purpose of this graph is to give an overview of the set of values of the preparation time and in which boundaries most of the orders are located. The data was divided into different categories. The first diagram [] is on week basis. It was created to see if there is a rough trend over time as the company expertise grows. As the diagram shows there cannot be made any conclusions for this categorization. After fluctuating preparation in the first weeks it gets more constant towards the end. Since this is not helping a box whisker diagram for each slot was created []. There are three slots and each is a part of the delivery time of the day. The day is divided into the big meals lunch and dinner as well as the time between these, the afternoon, which should be not as busy as the meal times. The diagram shows no real difference in preparation times between the slots. This has to be inspected in a different categorization of the slots so the idea is to have a look at slots on different days. The choice is to identify special behaviour according to the weekday since weekdays can have strong varying load

for the restaurant. For example on a sunday evening many people like to go to a restaurant and do not order while in the week offices sometimes order big deliveries for lunch. But the values in the diagram do not vary that much which can be due to a low training dataset. This leads to the last overall diagram []. It shows the preparation time for each weekday, split by slot. This is done since it contains two different key information. Slots and weekdays alone do not have as much information as the two combined. The weakness of slots and weekdays alone is e.g. at the weekend ordering for lunch is not as popular as ordering lunch at work. Also restaurants will have more workload on the weekdays when everyone is coming for lunch than on the weekend. It could be separated between weekend and weekdays but there is not enough data to do this and gain additional information. In order to maximize information gain a diagram containing slots per weekday is introduced. The difference between slots on weekdays as well as between weekdays is more significant than all other diagrams before and should be considered when creating the model.

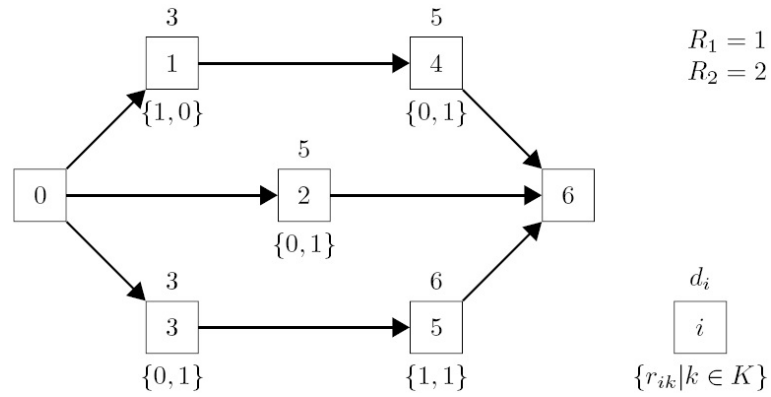


Figure 3.5: An example figure

### 3.3.2 Restaurant Wise Proceeding

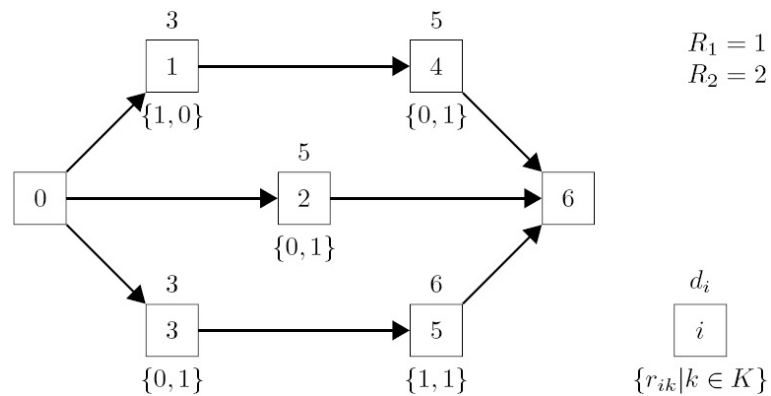


Figure 3.6: An example figure

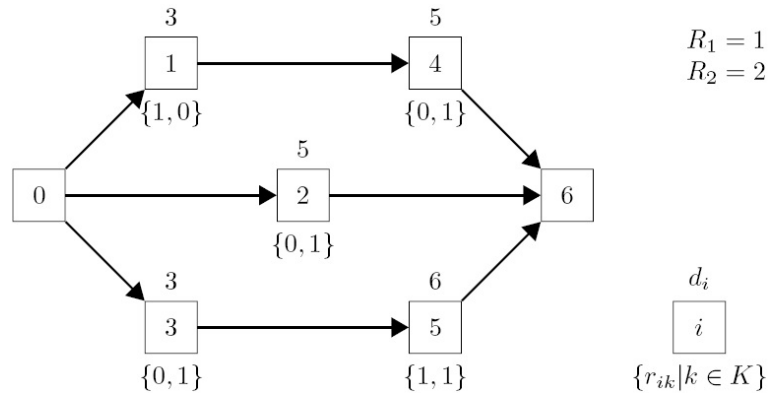


Figure 3.7: An example figure

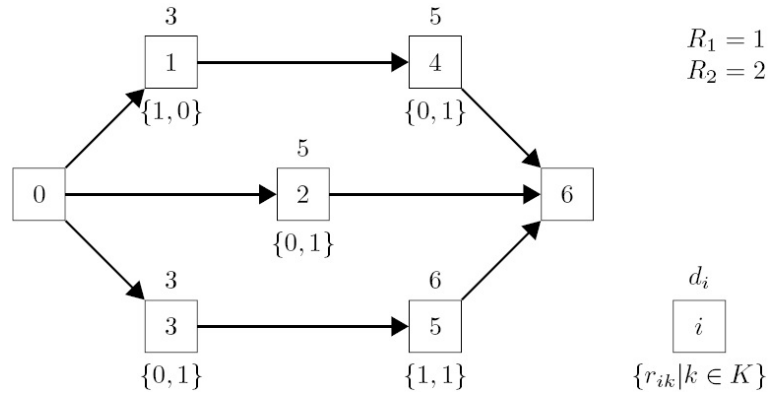


Figure 3.8: An example figure

## 3.4 Choosing and Fitting Models

### 3.4.1 Categorizing of Orders

**No time categorization** First of all, no time categorization is done. The forecast is done as all predecesing orders are relevant to the current one. This gives a rough estimation over the forecasting possibilities. It is not very accurate since the more orders are already included the closer the forecast is to the average. The next forecast is done with the same assumption but for each slot. Each order in the slot is forecast by all orders in that slot before it. This can eliminate the difference in case preparation times vary between the different slots. In the end the mean of the three mean square errors is taken and the root is calculated as overall root mean square error.

**No time but slot categorization** **Day categorization** After doing uncategorized forecasts, a time component is introduced. Since preparation times can vary from day to day, e.g. weekdays or holidays, they should be independent. For this reason, orders are separated by day and forecast according to this time constraint. Forecast is then done for each day as well as each slot for each day. In the end the errors of each forecast model are summed up and the average is returned. In case of the slots of each day, the average is first calculated for each slot day by day and in the end the mean errors of each slot are summed up.

**Day and slot categorization**

**Week categorization** The next time categorization is on weekly basis. Like in the daily calculation, the error for each order by taking all predecesing orders of the week. This is also done for orders in a slot of each week similar to the day basis.

**Week and slot categorization Weekday categorization** The same procedure as for day and week categorization is done for weekday as well. **Weekday and slot**

**categorization Single slot categorization** The final simple categorization is done by separating orders into their day and slot. Each slot is then treated separately (the same way as in the day only calculation). In the end the mean square error of each slot is added up and the average is rooted to get the overall root mean square error.

Before the calculation can be done, the edge case of not having predecesing orders for the current order has to be resolved. Since the operations teams suggested a 15 minute basic time for preparation, this time is always taken when no forecast value can be generated for the current order.

### 3.4.2 Combination of Categorizations

For the next step, Sebastian Sondheimer was interviewed. He is in the business department and joined later which is the reason he was not part of the problem definition. His focus for the forecast lies on two pillars. One is the best and most useful data to increase accuracy and the other one is to keep factors in the calculation to a minimum in order to improve speed. These two factors can work against each other and thus have to be carefully adjusted. The result is the combination of six factors which weights have to be evaluated. The first factor is the forecast for the current slot in which the order is. The second value is the forecast calculated from all orders of this day. This is combined with the forecast for all orders in the last 7 days as well as the forecast for the current slot of the last day. The last two factors are the forecast for this slot of

Refer to symbols or abbreviations with

```
\gls{symbolname} \glstext{symbolname} \glsfirst{symbolname}:
first N additional N first N
```

Bachelorarbeit (BA) Diplomarbeit (DA) Masterarbeit (MA)

Refer to other sections with `\ref{labelname}`:

A reference to this subsection: ??

Include figures with

```
\begin{figure}
```

```
...
```

```
\caption{An example figure}\label{fig:example}\end{figure}
```

Include only PostScript images (.eps) if you want to create a PostScript document using dvips and only .pdf, .png, .jpeg and .gif images if your goal is a PDF document using pdflatex.



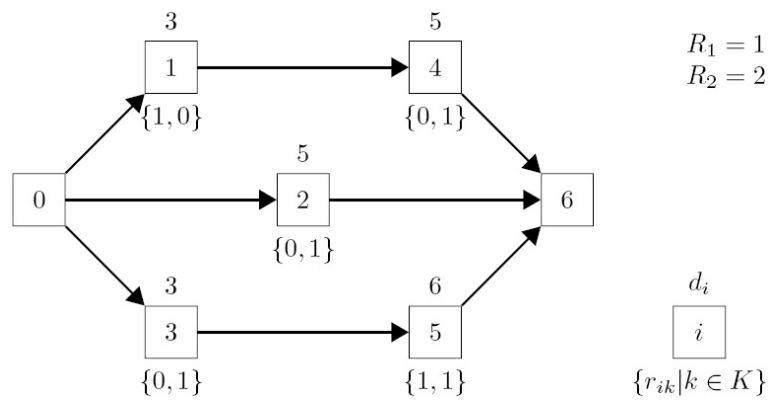


Figure 3.9: An example figure

## 4 Results

Include tables with

```
\begin{table}
\caption{An example table}\label{tab:example}
\begin{tabular}{ccc}
...
\end{tabular}\end{table}
```

Table 4.1: An example table

Parameter	Levels				
<i>A</i>	1	2	3	4	
<i>B</i>	1	2	3	4	
<i>C</i>	1	2	3	4	5
<i>D</i>	1	2	3	4	
<i>E</i>	1	2	3	4	

## 5 Conclusion

## Bibliography

Winston, Wayne L. 2007. *Operations research: Applications and algorithms*. 4th ed. Brooks/Cole Thomson Learning, Belmont 2007.

# Appendix

# **1 Appendix One**

## **2 Appendix Two**

## **Ehrenwörtliche Erklärung**

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit selbständig angefertigt habe. Die aus fremden Quellen direkt und indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Ich weiß, dass die Arbeit in digitalisierter Form daraufhin überprüft werden kann, ob unerlaubte Hilfsmittel verwendet wurden und ob es sich - insgesamt oder in Teilen - um ein Plagiat handelt. Zum Vergleich meiner Arbeit mit existierenden Quellen darf sie in eine Datenbank eingestellt werden und nach der Überprüfung zum Vergleich mit künftig eingehenden Arbeiten dort verbleiben. Weitere Vervielfältigungs- und Verwertungsrechte werden dadurch nicht eingeräumt. Die Arbeit wurde weder einer anderen Prüfungsbehörde vorgelegt noch veröffentlicht.

Ort, Datum

Unterschrift