

Demand forecast models for online supermarkets

Evers, J.M.; Tavasszy, Lorant; van Duin, Ron; Schott, Dingena; Gorte, Frank

Publication date

2018

Document Version

Final published version

Published in

E-groceries, digitalization and sustainability: Which governance, planning and regulation mix do our cities need?

Citation (APA)

Evers, J. M., Tavasszy, L., van Duin, R., Schott, D., & Gorte, F. (2018). Demand forecast models for online supermarkets. In *E-groceries, digitalization and sustainability: Which governance, planning and regulation mix do our cities need?* (pp. 1). Molde University.

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.



Demand forecast models for online supermarkets

J.M. Evers^{12*}, L.A. Tavasszy¹, J.H.R. van Duin¹³, D.L. Schott⁴ & F.B. Gorte²

¹*Delft University of Technology, Faculty of Technology, Policy & Management, Transport & Logistics, Jaffalaan 5 1, 2628 BX, Delft, the Netherlands*

²*Picnic Supermarkets B.V., van Marwijk Kooistraat 15, 1096 BR, Amsterdam, the Netherlands*

³*Knowledge Centre Sustainable Port Cities, Rotterdam University of Applied Sciences, Heijplaatstraat 23, 3089 JB, Rotterdam, the Netherlands*

⁴*Delft University of Technology, Faculty Mechanical Maritime and Materials Engineering, Transport engineering and logistics, Mekelweg 2, 2628 CD, Delft, the Netherlands*

Abstract

Food waste and incomplete orders at online supermarkets mainly occur due to inaccurate demand forecasting which leads to incorrect ordering of products. The objective of this study is to develop an accurate demand forecast model at product-level based on historical customer order data, give recommendations on implementation and describe impact on logistical planning at an online supermarket in the Netherlands. The product of research in this case study was bread, because of the habitual order behaviour of customers.

It is found that, using historical customer behaviour, model accuracy can be increased by forecasting the bread order probability for every customer individually compared to a total order regression. Decision regression trees and random forest regression models are implemented to forecast product sales on short term and show high accuracy. The forecast accuracy of predicting the number of breads per day is about 99.9 percent, given the number of customers that is going to order. This implies that it is feasible to order bread directly at supplier with a significant level of reliability such that waste due to overprediction and incomplete orders due to underprediction can be reduced to acceptable levels.

The main advantages of using tree models for demand forecasting of products for an online supermarkets is the fast run time, accurate forecasts and easy interpretation. Within a few seconds, thousands of customers are analysed and conclusions can be drawn on the future demand of bread based on historical demand of each customer. Average breads per order in historical orders and number of orders are main predictors for future demand. Decision trees and random forest regression outperform linear regression in this case study.

Keywords (3): demand forecast models; random forest regression; online supermarkets
JEL classification: *code*.

* Corresponding author. Email: jenneke.evers@teampicnic.com.

Key remarks:

- **Background & context:** online customer behaviour is significantly different from offline shopping behaviour. The availability of rich customer-specific historical order data broadens the range of useable methodologies for demand forecasting from linear regression and time-series methods to machine learning techniques.
 - **Literature review:** Tree models are most suitable for forecasting at online supermarkets. The full exploitation of data and easy interpretation have been the leading arguments for this decision.
 - **Data specification and methodology:** Bread forecasts are generated per customer. The outcome of the models is the expected number of breads that each customer is going to order, given that he/she is going to order.
 - **Results:** It can be concluded that adding customer specific data adds significant value to the forecast of bread sales. The features reveal non-linear relationships in the data.
 - **Model comparison:** Tree models outperform linear regression models in this case study.
 - **Conclusion:** Not only the accuracy improves, it also gives new insights in customer behaviour at Picnic and dominant variables within forecasting on product level. Just-in-time delivery of products is feasible with reliable demand forecasting.
-

1. Background & context

Online supermarkets are supermarkets where people order their groceries via an app or website. In 2016, almost 1.6 million customers in the Netherlands ordered their groceries online at least once and the market share of online groceries is expected to grow up to 5% in 2017 (Ecommerce News, 2016a). This research is executed at Picnic, an online-only supermarket. All the data used in this research is provided by Picnic.

The strategy of online-only supermarket Picnic is based on the belief that physical shops are unnecessary to serve customers. The company combines groceries per customer at their fulfilment center and forwards complete orders with trucks to hubs. Orders are delivered from these hubs to the customers with electric vehicles. The electric vehicles drive more or less fixed routes according to the bus model. Routes are optimized and updated daily using shortest-path methodology. As a growing business, more products are ordered directly at suppliers, instead of being storepicked at the distribution center of another supermarket that supported Picnic in their uplift. This collaboration allowed Picnic to expand fast, but also limited their expansion drift and logistical planning. More reliable demand forecasting is a prerequisite for independent operation and the trigger for this research.

Demand forecasting for online supermarkets is susceptible to forecast margins. Too much inventory leads to waste of products, as the perishability date of bread expires after one day. Products must be thrown away. Too little inventory leads to incomplete orders and/or lost sales. In this case customers do not receive their pre-ordered product or are unable to order their desired product. Demand forecast models aim to minimize the forecast error of products.

Demand forecasting is key for efficient logistics and is used within the traditional supermarket business for a long time. However, online customer behaviour is significantly different from offline shopping behaviour. Firstly, online supermarkets experience other peak days compared to traditional supermarkets. Peak days of traditional supermarkets are Fridays and Saturdays,

when consumers tend to have time for grocery shopping. Peak days of online supermarkets are Monday and Friday. This is due to the different opening hours of the supermarkets. Secondly, customers that shop via mobile phones, purchase more frequently compared to customers who shop offline. As customers shop online, they tend to purchase habitual products (Wang, Malthouse, & Krishnamurthi, 2015). This implies that historical orders have predictive value for future orders. Online supermarkets can analyse this habitual behaviour based on precise historical orders. Thirdly, ordering via apps, or websites, enables online retailers to collect detailed data on historical orders. The availability of rich customer-specific historical order data broadens the range of useable methodologies for demand forecasting from linear regression and time-series methods to machine learning techniques.

The objective of this study is to develop an accurate demand forecast model at product-level based on historical customer order data, give recommendations on implementation and describe impact on logistical efficiency at an online supermarket in the Netherlands.

This research has been split up in four sections. The first section reviews literature on forecasting methods and sets requirements. The second section elaborates on the performance of tree models for demand forecasts at online supermarkets. Data specification and methodology are explained first, where after the main results on forecast performance are presented. The third section presents two benchmark models and the fourth section presents the conclusions and policy suggestions.

2. Literature review on forecasting methods

This section provides a background in frequently used and novel methods in demand forecasting for the food & retail industry that use customer data. Novel methods often employ machine learning techniques to learn and leverage features for prediction. The methods are assessed according to requirements set for Picnic.

Multilinear regression

Linear regression is the most common algorithm for regression tasks. Regression of the past grocery sales allows for a linear relationship (Ali *et al.*, 2009). The main advantages of linear regression is the simplicity of the model which makes it straightforward to understand and explain. The weakness is that linear regression performs poorly if non-linear relationships are involved. When sophisticated input data is available, non-linear regression models become interesting.

Several researchers show that autoregressive integrated moving average (ARIMA) models outperform simple linear regression, because seasonal and cyclical trends are taken into account. Taylor (2007) proposed interval forecasts from quantile predictions using exponentially weighted quantile regression to forecast daily supermarket sales. Lasek, Cercone and Saunders (2016) compared several forecasting techniques in their chapter on 'Smart Restaurants' and found that ARIMA models are suitable if data is nonstationary and mean and variance change over time. This is mostly the case with peak- and nonpeak days. Arunraj and Ahrens (2015) proposed a hybrid seasonal ARIMA and quantile regression for forecasting daily food sales of bananas from a retail store in Germany. The same researchers found that the combination of the ARIMA model with seasonal and external factors improved forecasting daily sales of perishables (Arunraj, Ahrens and Fernandes, 2016). Aburto and Weber (2007) use a combination of SARIMA and neural networks trained with the residuals of time series model

to forecast grocery sales. Ramos, Santos and Rebelo (2015) showed that ARIMA models produce comparable results as an exponential smoothing technique to forecast future retail sales of shoes.

Machine learning techniques allow learning non-linear relationships (Ali *et al.*, 2009). This is especially interesting if other information besides time series are involved. There are two variations of machine learning techniques, supervised and unsupervised. Supervised means that the algorithm is trained based on a dataset that includes the right outcome. Unsupervised means that the algorithm is trained to find underlying structure or distribution in data without requiring a data label (Brownlee, 2016). Forecasting demand using historical data is a supervised problem because a label is often available. In this case, the label is the bread demand per day. In literature, three methods are often employed to represent a forecast model learned by labelled data: Neural Networks, Tree methods and Support Vector Machines

Neural networks

Artificial neural networks (ANNs) allow complex nonlinear relationships between the response variable and its predictors. Alon, Qi and Sadowski (2001) compared ANNs and traditional models, like linear regression, ARIMA and exponential smoothing, and concluded that generally ANNs are more successful in forecasting demand for restaurants. Thiesing and Vornberger (1997) trained neural networks to forecast future demand per item in a German supermarket. They included influences like price changes, advertising campaigns and holiday season. The experimental results showed that neural networks outperformed naïve and statistical approaches. The main advantages of neural networks is that they are able to handle large datasets and generally perform quite well compared to linear methods. The main disadvantage of neural networks is that the outcome of the model cannot be easily traced back. In other words, the outcome is black box.

Tree models

Tree models, like decision trees and random forest regression, are named after their structure of splitting data into several branches. An example in demand forecasting is (Agarwal, 2017) who trained a dataset of over 3 million online grocery orders from over 200.000 customers using a random forest regressor model to predict number of future orders. The model performed very well on predicting this value (0.93 on a scale of 0 to 1). The main advantage of decision tree methods is the easiness of interpretation and their fast computation times (Louppe, 2015). These methods support heterogeneous data input, which makes preparation time faster. Decision or regression trees have some limitations, one the biggest being that they tend to overfit the data in an attempt to account for outliers in the data.

Support Vector Machines

Support Vector Machines (SVM) has been applied in time-series modelling problems, like stock price prediction (Madge, 2015) and arrival forecasts of vessels (Parolas, 2016), but there are not much examples in literature of forecasting demand in the food and retail sector. SVM is algorithmically complex and therefore not practical to use in demand forecasting (Suykens *et al.*, 2003). It also requires extensive memory and training and testing takes relatively long.

Table 1: Model requirements for demand forecasting of bread

	Data-related requirements			Practical requirements		
	<i>Full exploitation of data</i>	<i>Ability for multi-label input</i>	<i>Suitable for time-series data</i>	<i>Short computation time</i>	<i>Easy interpretation</i>	<i>Flexibility and adaptiveness</i>
Multiple regression	✗	✗	✓	✓	✓	✓
ARIMA	✗	✗	✓	✓	✓	✓
Neural network	✓	✗	~	~	✗	✓
Decision Tree	✓	✓	~	✓	✓	✓
Random Forest	✓	✓	~	~	✓	✓
Support Vector Machine	✓	✓	✗	✗	✗	✗

Model requirements

There are three data-driven and three practical requirements for demand forecast models at Picnic. Data-related requirements involve the input data set of the model. These are that the model should be able to exploit the full potential of the data. Non-linear relationships should be revealed as well. The second data-related requirement is that input data should cope with multi-label variables. This shortens the preparation time. The third is that the model should be suitable for time-series data.

The practical requirements for demand forecast models involve the applicability within daily operation of an online supermarket. The run time should be less than 1 hour, because updates to the supplier are required every hour. The model should be easily interpretable, because purchase order operators should understand the reasons for the output as well. The last practical requirement is that the demand forecast model should be flexible for adjustments in demand. The models are compared to the requirements in Table 1. Based on the requirements, tree models are most suitable for forecasting at online supermarkets. The full exploitation of data and easy interpretation have been the leading arguments for this decision. Multiple regression is chosen as well, because of its embedded position in the field of demand forecasting.

3. Data specification and methodology

The available data for this research consists of historical order- and customer data between January 2016 and October 2017. The total data set includes more than 70.000 unique customers. The order count per customer is known, as well as how long ago the customer placed its first order. There are almost 800.000 unique orders that can be linked to one of the customers. Each order has its own features. Other relevant information per order is weekday, number of products, bread types and amount spent.

Methodology

Two tree models are designed in this research. This first is a Decision Regression Tree (DT) and the second is a Random Forest Regression (RF). Random forest models combine a lot of parallel constructed decision trees with different subsets of data and features and averages the result. The goal of random forest model is to avoid overfitting of one decision tree model. Forcing to exclude a random subset of data, individual decision trees will not have a strong correlation and makes average predictions more reliable. It eliminates bias and reduces variance. Predictions are made by averaging the outcome of each tree.

According to Tabachnick and Fidell (2013) cross-validation with a second sample is highly recommended. Therefore the entire data set is split up in two parts, the training and test data. Training data is used to define the model and test data is used to validate the model and measure its performance. The most common way to split the original data set is to randomly select 30% as test set. The remaining 70% serves as input for training the decision tree. The splitting criterion used is the mean squared error. The final nodes are called leaves of the tree. Each leaf node stores the final test result or answer.

Three aspects of tree models are investigated in this research. The first is if tree models are suitable for demand forecasting of bread at an online supermarket for next-day delivery. The second is to find out if features of customer data differ in their predictive value and the third is if selecting data input and number of leaf nodes influence performance.

Model set-up

Bread forecasts are generated per customer. The outcome of the models is the expected number of breads that each customer is going to order, given that he/she is going to order. All these individual forecasts of the customers are then added together to make the total forecasted bread sales per day. The time span of forecasting is one day, because bread is delivered daily to the distribution centre. Tree models require an input that is made up of rows and columns that represent features. There are 24 features, or predictors, in the dataset. There are three main parameter settings: maximum leaf node, the test size and the random state.

- Maximum leaf nodes = {10, 100, 1000}
- Test size = 30%
- Random state = 200

The maximum leaf node is the number of leaf nodes of one tree. More leaf nodes give more possible output values, but also increases the chance of data overfitting. The test size of 30 percent is standard in decision trees. The random state guarantees that the split of the test and train set will always be the same. It does not matter what number is chosen. It is useful for reproduction of results. The model is trained once and after that the model is run 14 times for all test data. Note that the model is not trained with information of the test data.

Evaluation of forecast performance

Evaluation metrics for an online supermarket are different from traditional supermarkets or businesses because of the importance that demand is not underestimated. Picnic aims to deliver 99% of orders complete, meaning correct articles and quantity. The maximum level of total negative forecast error per day is 1% at moment of ordering at suppliers. All models are evaluated based on the forecast accuracy. A confusion matrix, or error matrix, makes it possible to visualise the accuracy of machine learning models. It is known for evaluating classification errors by comparing the type of misclassifications. The confusion matrix can be used for real numbers as well as for normalized numbers.

Most evaluation metrics described in literature value negative and positive errors equally. The case study in this research demands a different approach. Negative errors (underprediction of bread) is valued as worse than positive errors (overprediction). Weighing negative errors more than positive errors is a common phenomenon in social sciences, where for example things of negative nature have greater effect on psychological state of an individual than positive things.

This is called a “negativity bias”. Parallel can be drawn to customers of online shopping, where the bad feeling of missing an item in their order is stronger than the happy feeling of an extra (unexpected) item in their order. It is interesting and useful to quantify the negative impact on customers in case of missing items because of underprediction in more detail, but this is out of scope for this research.

4. Forecast performance

Tree models fit the purpose of demand forecasting very good and contribute to forecast on product level. The designed models have shown that forecasting on customer level is better compared to the baseline model that uses non-granular historical order data.

Forecast performance

The forecast performance of tree models is extremely accurate, almost 100 percent. Forecasting the number of breads per customer, given customer-specific features, is hard to excel. It can be concluded that adding customer specific data really adds value to the forecast of bread sales. The features reveal non-linear relationships in the data. The decision tree model takes about 4 seconds to analyse over 500.000 data points. The random forest regression takes about one hour. Figure 1 shows the percentual deviation between forecast and actual number of breads. Decision tree seems to underestimate the number of breads per day. Random forest on the other hand, seems to overestimate the number of breads per day. Random forest regression is therefore more suitable to use for online supermarkets, because of the negativity bias.

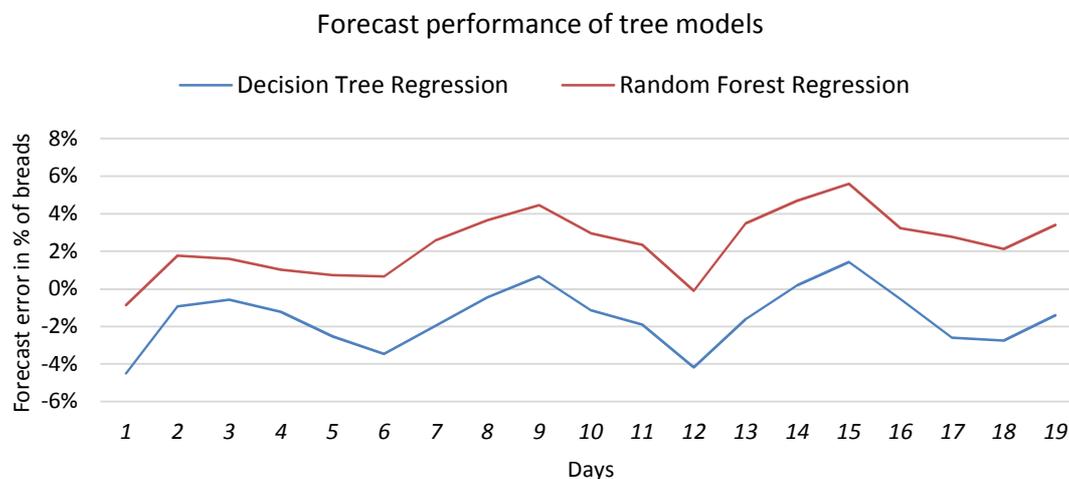


Figure 1: Over- and underprediction of tree models

Feature interpretation

One of the advantages of tree models compared to other machine learning methods, is the possibility of interpretation of features. This is useful for model improvement and explain model behaviour. The number of orders and the cumulative bread count for each customer on Monday, Friday and Saturday have predictive value as well. The peak days include more customers, and thus data points. These days are therefore more reliable and make them appear in the tree interpretation and graph with feature importance. The decision tree model is run three times with different number of features to test if this impacts the performance.

The model including 24 the features performs better compared models with 23 or 22 features. The process time is almost equally short and therefore not a reason for eliminating variables.

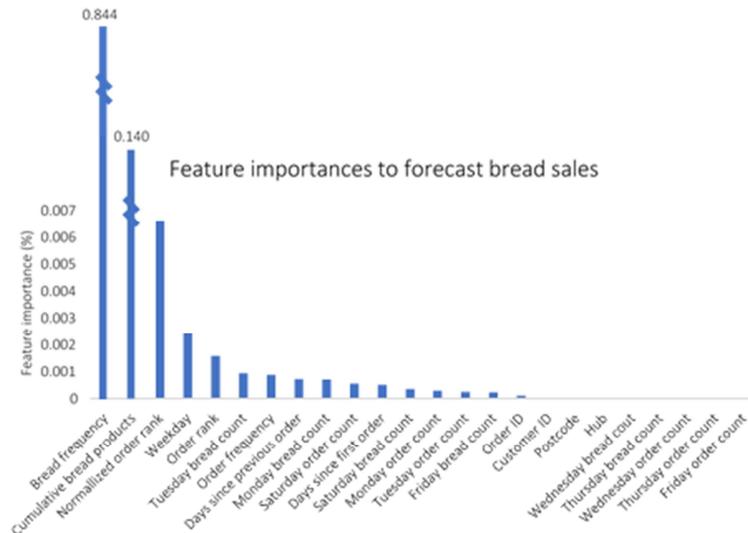


Figure 2: Feature importance of trained Tree Models with 24 features

Figure 2 shows the feature importance of the trained decision tree with 24 features. The feature 'bread frequency' appears to have large influence on the number of breads that will be ordered by that specific customer. These contribute for 84% and 14% to the forecast of the bread sales. The other 2 percent is made up of 14 features and 8 features do not contribute at all to the outcome.

Maximum leaf node impacts performance

The decision tree is trained with maximum number of leaf nodes of 10, 100 and 1000. Confusion matrices allow to visually compare the forecast performance of the decision tree model. Figure 3 shows the normalised confusion matrix of each test. The normalized confusion matrix shows the ratio of actual number of breads per order compared to the forecasted number of breads per order. The x-axis represents the number of forecasted breads per customer and the y-axis represents the actual number of breads per customer. The darker the blocks, the more customers in that block. If all customers were predicted correctly, all the blocks would be located on the diagonal from top-left to bottom-right of the confusion matrices.

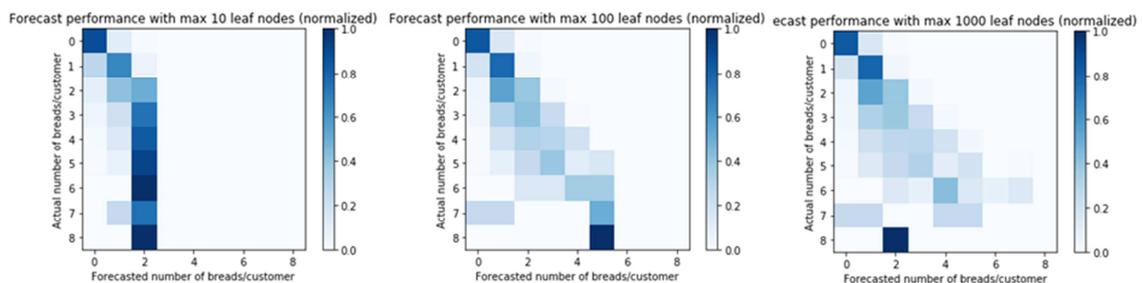


Figure 3: Normalized confusion matrixes with maximum number of leaf nodes {10, 100, 1000}

Three conclusions can be drawn from these graphs. The first is that 100 leaf nodes gives the best results in this comparison. If the model is trained for 1000 leaf nodes, it starts to show effects of overprediction. The second is that there is a skewness in predictions towards underpredictions, because there are darker areas below the diagonal in the normalized graphs. The third is that the maximum number of breads that is predicted is 5, while the data set does include up to 8 breads per order. These are apparently outliers and are therefore never can be predicted.

As end nodes increase, the number of maximum predicted breads per customer increases as well, because there are more output possibilities. Selecting data input influences forecast performance. More input data increases the process time, but increases performance as well. The random forest model is trained three times with input data of one year, six months and six weeks. One year of input data shows better results than six months or six weeks of input data.

5. Model comparison

Two other demand forecast methods have been evaluated in this research, which are the baseline model of Picnic and a linear regression. All models were tasked to forecast the future demand of bread at Picnic.

Baseline Model

The baseline model at Picnic linearly extrapolates the average number of breads per customer on a specific time of day, until the end of that day. It can be classified as a sort of smoothing model. There are no external effects taken into account. Forecasts are generated per hour for next-day delivery. The baseline model is different from the tree models, because forecasts are generated per breadtype, while the tree models do not distinguish per breadtype.

The performance of the current demand forecast model is evaluated by comparing the forecasts and the actual sales of bread at the end of the day. The average forecast error in number of breads one hour before order deadline is about 8 percent. The average forecast errors at earlier times are higher, because of more uncertainty on number of customer orders. The baseline model is highly dependent on the total customer orders.

Linear Regression Model

The linear regression model estimates the coefficient of five independent variables to forecast the dependent variable, bread demand per day. No distinction is made between the breadtypes. The independent variables are average customer orders per day of last 3 weeks, forecasted customer orders per day, historical article order rate and weekday. The accuracy of the estimated linear regression model with forecasted customer orders and historical AOR is 94 percent. Adding weekday dummy variables increase the model performance. Tuesday, Wednesday and Thursday have significant negative impact on the total bread orders. The total customer orders on those non-peak days are lower, resulting in the need for other independent variables to explain the dependent variable. The negative coefficient is straightforward, but Monday, Friday and Saturday are expected to have positive coefficients. One of the reasons that coefficient of the peak days is not significant is that the number of total customer orders is higher, meaning that these are more reliable. This model is not recommended to use at Picnic, because forecasts tend to lean towards more negative errors.

Table 2: Average accuracy per day of bread forecasts (during three-week period)

Model	Input	Output	Accuracy
Baseline Simulation	Order data (3 weeks)	Forecast per breadtype & per hour	92% ^a
Linear Regression Model	Order data (3 weeks)	Total bread forecast	94%
Decision Regression Tree	Customer data (1 year)	Total bread forecast	99.9% ^b
Random Forest Regression	Customer data (1 year)	Total bread forecast	99.9% ^b

^a The baseline model updated hourly with new information. Accuracies per hour differ. This % is at 21:00. Order deadline is 22:00.

^b This model forecasts the number of breads given the (number of) customers that are going to order next day.

Comparison

Tree models outperform linear regression models in this case study. Table 2 compares the four models. The output of the baseline model has additional levels of detail and cannot directly be compared. However, the accuracy of the model with one hour of unknown demand is much lower compared to the other three, which gives the conclusions that forecasting on customer level increases forecast performance compared to the baseline model.

6. Conclusion

This research has demonstrated that tree models are applicable in demand forecasting and outperform linear regression. The accuracy of the models is more than 99 percent. The decision tree shows skewness towards underprediction and the random forest shows more positive forecast errors. The features 'bread frequency', 'cumulative bread orders' and 'number of orders' have most predictive value. 100 leaf nodes gives the best performance. Run time is less than 1 hour. Tree models have shown that forecasting on customer level is better compared to forecasting currently employed at Picnic. Not only the accuracy improves, it also gives new insights in customer behaviour at Picnic and dominant variables within forecasting on product level.

Several new insights have been presented regarding the set-up of the model. The maximum leaf nodes are set at 100 after experimenting with the set-up. 10 maximum leaf nodes resulted in lower accuracy. 1000 leaf nodes resulted in overprediction. For these type of models it is best to use one year of input data. One year of input data shows better results than 6 months or 6 weeks of input data.

Reliable demand forecasting on product level has positive impact on the logistical planning of e-grocers. If online supermarkets implement machine learning methods instead of smoothing models, waste due to overprediction can be reduced and incomplete orders due to underprediction can be eliminated. Just-in-time supply chain for all fresh products is possible.

Acknowledgements

I could not have completed this research without the help of many others. I am grateful for the support for the research and its presentation from Picnic Supermarkets and the VREF Center of Excellence for Sustainable Urban Freight Systems network partner of Delft University of Technology. I would like to thank co-authors Lori Tavasszy, Ron van Duin and Dingena Schott from Delft University of Technology, both offline and online. Special thanks to Frank Gorte and Joris-Jan Kraak from Picnic Supermarkets.

References

Article in a journal:

- Aburto, L., & Weber, R. (2007). Improved supply chain management based on hybrid demand forecasts. *Applied Soft Computing Journal*, 7(1), 136–144.
<https://doi.org/10.1016/j.asoc.2005.06.001>
- Ali, Ö. G., Sayin, S., van Woensel, T., & Fransoo, J. (2009). SKU demand forecasting in the presence of promotions. *Expert Systems with Applications*, 36(10), 12340–12348.
<https://doi.org/10.1016/j.eswa.2009.04.052>
- Alon, I., Qi, M., & Sadowski, R. (2001). Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods. *Journal of Retailing and Consumer Services*, 8(3), 147–156.
- Arunraj, N. S., & Ahrens, D. (2015). A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting. *International Journal of Production Economics*, 170, 321–335. <https://doi.org/10.1016/j.ijpe.2015.09.039>
- Arunraj, N. S., Ahrens, D., & Fernandes, M. (2016). Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry. *International Journal of Operations Research and Information Systems*, 7(2), 1–21. <https://doi.org/10.4018/IJORIS.2016040101>
- Ramos, P., Santos, N., & Rebelo, R. (2015). Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and Computer-Integrated Manufacturing*, 34, 151–163. <https://doi.org/10.1016/j.rcim.2014.12.015>
- Taylor, J. W. (2007). Forecasting daily supermarket sales using exponentially weighted quantile regression. *European Journal of Operational Research*, 178(1), 154–167.
<https://doi.org/10.1016/j.ejor.2006.02.006>
- Wang, R. J. H., Malthouse, E. C., & Krishnamurthi, L. (2015). On the Go: How Mobile Shopping Affects Customer Purchase Behavior. *Journal of Retailing*, 91(2), 217–234.
<https://doi.org/10.1016/j.jretai.2015.01.002>

Chapter in a book:

- Lasek, A., Cercone, N., & Saunders, J. (2016). Smart restaurants: Survey on customer demand and sales forecasting. In *Smart Cities and Homes* (pp. 361–386). Elsevier Inc.
<https://doi.org/10.1016/B978-0-12-803454-5.00017-1>
- Suykens, J., Horváth, G., Basu, S., Micchelli, C., & Vandewalle, J. (2003). *Advances in Learning Theory: Methods, Models and Applications* (Series III). Amsterdam, Berlin, Oxford, Tokyo, Washington D.C.: IOS Press - NATO Science Series.
- Tabachnick, B. and Fidell, L. (2013). *Using multivariate statistics*. Pearson, 6th edition
- Thiesing, F. M., & Vornberger, O. (1997). Forecasting Sales Using Neural Networks. *Computational Intelligence*, 321–328.

Working paper:

- Madge, S. (2015). *Predicting Stock Price Direction using Support Vector Machines*. Independent Report. Princeton University. Princeton, NJ.
- Parolas, I. (2016). *ETA prediction for containerships at the Port of Rotterdam using Machine Learning Techniques*. Master thesis. Delft University of Technology.

Online resource:

Agarwal, S. (2017). Analysing shopping behaviour of users using Machine Learning. Retrieved October 16, 2017, from <https://medium.com/towards-data-science/analysing-shopping-behaviour-of-users-using-machine-learning-part-1-518483e72a2c>

Brownlee, J. (2016). Supervised and Unsupervised Machine Learning Algorithms. Retrieved October 23, 2017, from <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>

Ecommerce News. (2016a). 1 in 6 Dutch will order groceries online in 2017. Retrieved July 10, 2017, from <https://ecommercenews.eu/1-6-dutch-will-order-groceries-online-2017/>

Loupe, G. (2015). Tree models with scikit-learn. In *PyData*. Paris. Retrieved September 20, 2017 from <https://www.youtube.com/watch?v=GuRfLbAU9Ho&t=682s>