DEGREE PROJECT IN COMPUTER ENGINEERING,
FIRST CYCLE, 15 CREDITS
STOCKHOLM, SWEDEN 2017

Predicting sales in a food store department using machine learning

ROBERT SIWERZ

CHRISTOPHER DAHLÉN

KTH ROYAL INSTITUTE OF TECHNOLOGY
SCHOOL OF COMPUTER SCIENCE AND COMMUNICATION

Predicting sales in a food store department using machine learning

ROBERT ÅBERG
CHRISTOPHER DAHLÉN

Degree project in Computer Science, DD142X
Date: June 12, 2017
Supervisor: Pawel Herman
Examiner: Örjan Ekeberg
Swedish title: Försäljningsprediktion i en matvarubutik med hjälp av maskininlärning
KTH, School of Computer Science and Communication

Abstract

Prediction of sales is an important field in the food industry and due to new technologies it has recently gained alot of attention in order to improve business operations and profitability. However, historically the industry has relied on traditional statistical models but in recent years more advanced machine learning methods has gained traction.

This study aims to compare three machine learning methods for sales prediction in the food industry: Multilayer Perceptron (MLP), Support Vector Machine (SVM) and Radial Basis Function Network (RBFN). The methods were compared in terms of their prediction accuracy on daily sales in a food store department. The performance of the models was determined using the performance measures: Mean Average Percentage Error (MAPE) and Root Mean Squared Error (RMSE).

The results show that the SVM performed lower error measures than the other two methods. The repeated measure analysis of variance (rANOVA) was used in order to determine if there was a difference between the methods. The test indicated a statistical significant difference between the afformentioned methods.

Sammanfattning

Försäljningsprediktion är ett viktigt område inom livsmedelsindustrin och tack vare nya teknologier har området nyligen fått stor uppmärksamhet i syfte att förbättra affärsverksamheten och lönsamheten i en mataffär. Historiskt sett har livsmedelsindustrin dock använt sig av traditionella statistiska modeller men under senare år har mer avancerade maskininlärningsmetoder vunnit mark.

Denna studie ämnar att jämföra tre maskininlärningsmetoder för försäljningsprognostisering inom livsmedelsindustrin: Multilayer Perceptron med hjälp av backpropagation (MLP), Support Vector Machine (SVM) och Radial Basis Function Network. Metoderna jämfördes med avseende på deras prediktionsträffsäkerhet på daglig försäljning i en butiksavdelning och mättes med hjälp av mätningsvärktygen: medelprocentfelet (MAPE) och rotmedelfelet (RMSE).

Resultaten visar att SVM presterade lägre prediktionsfel än de andra två metoderna. Upprepad variansanalys (rANOVA) användes för att avgöra om det förelåg någon skillnad mellan metoderna. Testet indikerade en statistiskt signifikant skillnad mellan metoderna.

Contents

Chapter 1

Introduction

The amount of sales data has increased steadily in recent years and the importance of taking advantage of this data is continuously increasing. One of the most valuable assets a company has today is the data generated by the customers and it has become popular to try and win business benefits from analyzing this data [8, 22]. Within the data there might be patterns that could be used to guide a company in how to take decisions regarding marketing, organization and sales [2]. The food industry is currently undergoing a digital transformation where new technology has created opportunities to better understand the underlying factors which play part in the daily sales [24].

The use of existing resources such as sales data has become a major focus in order to efficiently improve the business operations. One use case is prediction of sales, which is an important and complex factor in running a grocery store [23]. There are a wide range of applications where sales prediction can have an impact in the food industry and although prediction tools are used in aiding the ordering of supplies in a grocery store, there are still many areas to explore [24]. There exists several techniques to predict sales and historically the food industry has relied on traditional statistical models [5]. However, machine learning has in the last decade grown to become a field with broad application potential and has gradually gained ground in the food industry, among other areas [5, 13, 18].

To predict outcomes of a future event a machine learning model is exposed to data from which it learns patterns that are used to predict the outcome. There are several methods for this purpose and studies have shown promising results for the use of, Multilayer Perceptrons

(MLPs) and various network architectural models such as Radial Basis Function Networks (RBFNs) and Support Vector Machines (SVMs).

MLP is a concept that is extensively used for sales prediction because of its capability of learning non-linear trends [3, 9, 20]. A commonly used MLP is the Backpropagation Neural Network which has shown to outperform conventional statistical methods in salesprediction[3, 20]. Moreover, RBFNs shares the capability of learning non-linear trends and has also been applied to various problems in the food industry. The method has also shown promising results as a tool for sales prediction [25, 13]. In addition, SVMs is another popular machine learning algorithm. It is used to solve both classification- and regression problems and, like the network architecture models mentioned above, it has also been used as a method to predict sales in the food industry [25, 18, 20]. An accurate forecasting model can increase

the sales revenue and is of high importance in a company in order to improve the ordering of supplies and to better understand how to scheduele personnel [24, 5].

1.1 Problem statement

Studies have shown that machine learning models could be useful in the retail industry to gain knowledge of a business. Hence, customer segmententations and product correlation analysis are already widely used to better understand the market [2, 26, 11]. Moreover, predictive analysis is an important area in the food industry that are under constant development. There are numerous methods described in the literature and three of the most commonly used are SVM, MLP and RBFN [13, 29, 18]. These methods are usually compared against traditional statistical methods [9] and limited research has been done on comparing them against each other in food sales prediction. Having this in mind and by using sales data provided by one of the largest food companies in Sweden, this report aims to answer the following: *Which of the; MLP, RBFN and SVM achieves the highest accuracy when predicting the number of sold products in a food store department.*

Due to a non-disclosure agreement the company will not be mentioned by name.

Chapter 2

Background

2.1 Algorithms

2.1.1 Fundamentals of Multilayer Perceptrons

An MLP consists of computational units arranged in a layer-structure where the units are connected to eachother via weighted links [15]. The conceptual architecture is inspired by a biological neuron and how it is connected to a network of other neurons. The computational unit in an MLP represents the biological neuron and the weighted links represent the neurons axons as can be seen in 2.1.

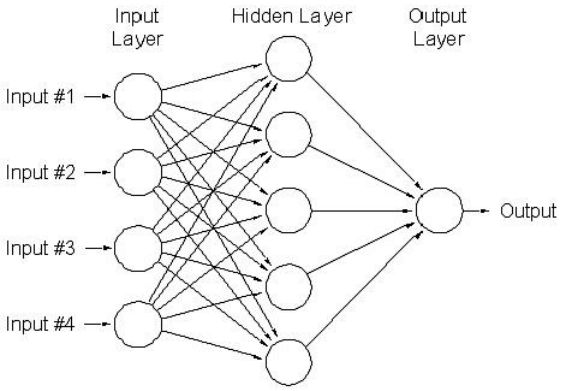


Figure 2.1: The visualization of a MLP. Consisting of 4 input nodes, 5 hidden nodes and 1 output node. source (May 10, 2017): $http$: $//www.xlpert.com/nn\_Solve.htm$

Each unit has an activation function by which the output of the

neuron is calculated as the weighted sum of the units in the previous layer. The output of the units in one layer is then multiplied by a corresponding weight as it is fed to the next layer, and each unit is, just like a biological neuron, activated when it reaches a certain threshold [15]. There are several choices for an activation function and a commonly used one in an MLP is the sigmoid function:

$$f(x) = \frac{1}{1 + e^{-x}}$$

which have the characteristic S-curve as shown in fig 2.2 [15]. This

characteristic means that the value on each neuron's S-curve decides how much of a contribution is made to the next layer [15].

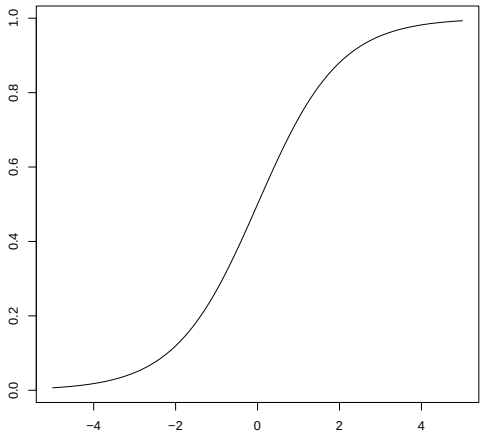


Figure 2.2: Sigmoid function

An MLP bases its prediction not only on the desired output but also on the input parameters. The flexible structure in which all neurons in a layer have a weighted connection to the neurons in the next layer allows the model to learn complex patterns without prior knowledge about the input data [21]. In an MLP the weights can be trained to fit the output of the training data by using a cost function which describes the difference between the desired output and the predicted value [24]. The cost function can then be minimized by using the backpropagation algorithm, which means that the difference between the desired output and the predicted value is minimized with respect to the weights. This is done iteratively in each layer, stepping backwards through the network starting from the output layer [9]. An MLP that uses the backpropagation learning algorithm is called a Backpropagation Neural Network.

The Backpropagation algorithm has its drawbacks, in terms of slow convergence and runs the risk of being trapped in a local minima of the cost function [14, 21]. However, it has been extensively used as a tool for learning since it was first presented in the 1960s. Modifications of the method over the years has resulted in that it is still one of the most used methods for training an MLP [24, 31].

2.1.2 Fundamentals of Radial Basis Function Networks

Another network architecture is the RBFN and there are a number of ways in how the network is trained. The training procedure can be divided into one or several phases [28, 16]. In a three phase approach, the initial phase usually consists of choosing centers which can be done by unsupervised or supervised learning. Another way to generate the centers is by randomly selecting them from the training data [28]. In the second phase the weights between the hidden and output layers can be solved for through the pseudo-inverse. The last phase then minimizes the error between the calculated output and the real output [28].

The architecture of RBFNs are similar to MLPs but uses a different activation function. The RBFN-activation function has the property that each units function is affected only by the radial distance, typically Euclidean, from the centers c initialized in the first phase[6]. A commonly used activation function is the Gaussian function:

$$G(x, c) = e^{-\gamma||x-c||^2} \quad , \quad \gamma = \frac{1}{2\sigma^2}$$

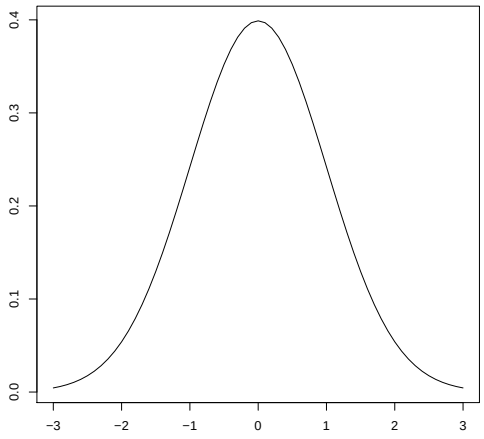


Figure 2.3: Gaussian function

where σ is the standard deviation. The function is characterized by being symmetric around the center and the further distance away from the center you travel the smaller the Gaussian functions value gets, as

shown in fig 2.3 [19]. This is a key component for the RBFN, hence, a point in the data set will affect the nearby points more than it will affect the far away points [19]. The spread, γ, in each units activation function is indicating how fast or steep the decrease is per distance unit from the center.

In order to minimize the output-error the network parameters are adjusted. This can be done in multiple ways, for example, the Expectation Maximization algorithm can be used to adjust γ and the weights. Another approach is by using backpropagation [28, 27].

There are differences between RBFNs and MLPs, for example, the typical structure of a RBFN is a single-hidden-layer design whereas an MLP might have several hidden layers between the input and output layer [16]. The activation function in the output layer of a RBFN differs from the hidden layer in that the output layer is simply a linear combination of the units in the hidden layer, while the hidden units is given by the non-linear gaussian function [16].

2.1.3 Fundamentals of Support Vector Machines

The SVM is a supervised learning algorithm which was originally used for classification problems but is now widely used for regression problems [25]. Since the non-linear version of SVM was presented in the early 90's, it has been a widely used method for solving complex problems [12, 25]. The method has been proposed for solving regression problems in the food industry, such as sales forecasting and customer demand [25, 20]. The SVM has several advantages compared to traditional prediction methods as it does not need any prior knowledge about the relationship between the input and the output. Also the method does not face the problem of ending up in a local minima of the error function, as might be the case with neural networks architectures [9, 20].

The goal of a SVM is to minimize the standard deviation and some measure of complexity. The complexity of a function depends on a vector W, which is perpendicular to the plane spanned by the input, where small values in the vector W renders in a flat and less complex approximated function.

In the case that the input is not linearly separable, a Gaussian kernel function can be used to extract as much meaningful information as possible from the training data [25]. The SVM then tries to approxi-

mate a function that describes the relationship between the input parameters and the output [18, 9].

2.2 Related Work

An accurate forecasting model can increase the sales revenue and is of high importance in a food company [24]. Studies have shown that businesses that use a data-driven approach to make decisions are, on average, 6% more profitable than those that do not [7]. In a report conducted by several food companies it was shown that there is room for improvement in terms of sales prediction practice but also in how to enjoy the benefits of gathering vast amounts of sales data [1]. Sales prediction has also been suggested as a tool for reducing inventory management and ordering which is of high relevance in the food industry [7, 17].

Moreover, a large part of the products sold in a food store is seasonally dependent due to holidays, cultural habits or weather [23]. Thereby a study by Thiesing et al. [30], highlighted that the sales of some products in a food store are more difficult to predict than others. Therefore it has been suggested that products could be classified as predictable or random before a prediction model is applied [32].

Even though there is no clear standard on what features to consider when predicting sales in the food industry, reports claim that the use of determenistic features such as calendar events makes a model more robust in that the model does not have to consider stochastic features such as outdoor temperature or rainfall [20, 30].

The use of SVM as a prediction tool in the food industry has been proposed by Levis et al. [20] who predicted the monthly sales of a single product 12 months ahead. The results show that the SVM scored a Mean Average Percentage Error (MAPE) of 7% and the authors suggests that a SVM could be an alternative to using neural networks. This conclusion is also supported by Pillo et al. [25], who found that the SVM outperforms RBFN on daily sales predictions on a single type of product. Although a larger MAPE was obtained for all three models the SVM scored the lowest average MAPE (61%) which was achieved by training on one year of data. SVM has also been successfully used for predicting sales on aggregate level such as total sales of a store as shown by Krause-Traudes et al. [18].

Besides the use of SVM, multiple reports show that neural networks in general have emerged as a technology with a great promise for modeling sales patterns in the food industry. In particular, MLP and RBFN are two commonly used methods for this purpose. These methods have been assessed for a variety of predictions including sales of several single products [19] and sales of product categories [30]. This was also shown by Chen et al. (2010) [10] who used a backpropagation-MLP to predict the daily sales of fresh food. Furthermore, Hasin et al. [17] proposed that a single-hidden-layer backpropagation-MLP can be used as a prediction tool for cumulative monthly sales in a food store. In another study, Thiesing et al. [30] concluded that an MLP using backpropagation outperforms conventional statistical methods in predicting sales of a product group. They also suggested further research could be done on the selection of inputs according to reduce the complexity of the models. Furthermore, Slimani et al. [29] also emphasize the strengths of a backpropagation-MLP in handling the nonlinearities in daily sales of single products [29].

Doganis et al. [13] predicted the sales of short-shelf life food products using RBFN and stated that the method can provide a good fit to the actual sales. This was also stated by Pillo et al. [25] who used RBFN as a method to predict sales. Moreover, A study from Kuo et al. [19] showed the use of an enhanced learning algorithm in conjunction with a RBFN to predict the sales of a food product. The result showed that the RBFN was significantly better than conventional statistical methods.

Chapter 3

Methods

The models compared in this study (MLP, RBFN, SVM) have been used for numerous problems in the food industry. They are chosen considering their recent popularity in research on sales prediction [19, 25, 17, 29]. In addition, the data in this study is provided by a Swedish food company and the methods are quantified by various statistical methods.

3.1 The Data

An important aspect to consider when validating a model is how the dataset is partitioned into a training set and a validation set. The approach taken can seriuosly influence the performance of the model and although there is no clear standard for this in the food industry, there are several different partitionings used [25, 17]. However, since this study has a limited amount of data available a 10-fold cross-validation [4], is used to evaluate the models. The available data is provided by a provided by a Swedish food company and consists of daily sales from one department in one store from year 2012 to year 2016. Each validation set then consists of 180 daily sales in a department.

Furthermore, the selection of features which are shown in table 3.1, are based on findings in the literature which partly emphasize the advantage of non-stochastic features [25, 23].

Table 3.1: Features used in the study and descriptions of them

Features	Description
Input:	
1. Day of year	A number between 1 and 366.
2. Day of month	A number between 1 and 31.
3. Day of week	A number between 1 and 7.
4. Holiday	A boolean value.
5. Department price	The sum of the price of the set of products in the department. A measure of how expensive the department is.
Output:	
6. Amount of sales	Number of sold products in the department.

3.1.1 Multi Layer Perceptron

The network architecure is derived from Hasin et al. [17] and is set to a single-hidden-layer architecture. As proposed by Pillo et al. [25], the number of neurons are chosen by successively increasing the numbers of neurons until the mean of the Mean Absolute Percentage Error (MAPE) converges. Hence, 6 neurons is used in the hidden layer. Furthermore the sigmoid activation function is utilized.In addition the initial weights are randomly generated [30] and the number of iterations made by the backropagation algorithm is set to 100.

3.1.2 Radial Basis Function Network

The RBFN uses a three phase learning process as described above with random initialization of the centers, psuedo-inverse and backpropagation, The network consists of 9 neurons in the hidden layer, where the amount of neurons are chosen based on the same approach used for chosing the neurons in the MLP. Furthermore, the Gaussian function is used as a activation function and the number of iterations is set to 50, as is described by Schwenker et al. [28]

3.1.3 Support Vector Machine

There are multiple choices for a kernel function, however this report uses a Gaussian kernel function. While further research is currently in progress, there are mainly two approaches to choose the initial parameters for an SVM; a heuristic approach and an exhaustive grid search [20]. The latter is used in this report, and although it is a time-consuming approach as data grows larger it could result in more optimal parameter values [20].

3.2 Prediction Measuring

To measure the accuracy of a prediction model, a comparison between the prediction and the actual value is conducted. There are a number of known error measures and in this study we will use the MAPE and Root Mean Squared Error (RMSE). A low MAPE implies a small absolute deviation between the actual and predicted value [20]. The RMSE amplifies and severely punishes large errors, hence a large RMSE does not nessesarely need to indicate a poor model but rather that some errors are very large.

3.2.1 Mean Average Percentage Error

The MAPE value is the average percentage errors of the predicted and real value.

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{A_n - F_n}{A_n} \right|$$

Where A_n, F_n is the actual value and predicted value respectively and n is the number of training examples.

3.2.2 Root Mean Squared Error

The RMSE is the average squared error of the predicted and real value

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (F_n - A_n)^2}$$

where the variables are the same as in the MAPE equation.

3.3 Analysis of Variance

As this report aims to compare different prediction models, in terms of accuracy, we have to clarify if there is a statistically significant difference. The data was divided into 10 different but dependent data sets and therefore we made use of the Repeated Measure Analysis of Variance (rANOVA), using an F-test [21]. The purpose of an F-test is to tell if there is a difference between the models and the F-value value is calculated by:

$$F = \frac{\text{Variance between the models}}{\text{Variance within the models}}$$

In conjuction with the F-value a p-value determines if the result is statistically significant. Informally, a low F-value means that there is a small difference between the models. In addition, if the p-value is less than a certain significance level α, chosen as 0.05 in this study, the null hypothesis can be rejected and vice versa. The null hypothesis to be investigated in this study is that there are no differences between the means of the methods, under the assumptions of rANOVA holds.

Chapter 4

Results

In this section the results of the three models will be presented. The results were obtained by applying the three models; SVM, RBFN and MLP on a 10-fold crossvalidation data set. The rANVOA test was then used to check whether there were any statistical significant differences between the models.

4.1 Multi Layer Perceptron

Figure 4.1 shows the run where the MLP scored the lowest RMSE and MAPE value. However, the low error measures are a deviation from the rest of the MLP predictions as can be seen in table 4.2 and 4.1.

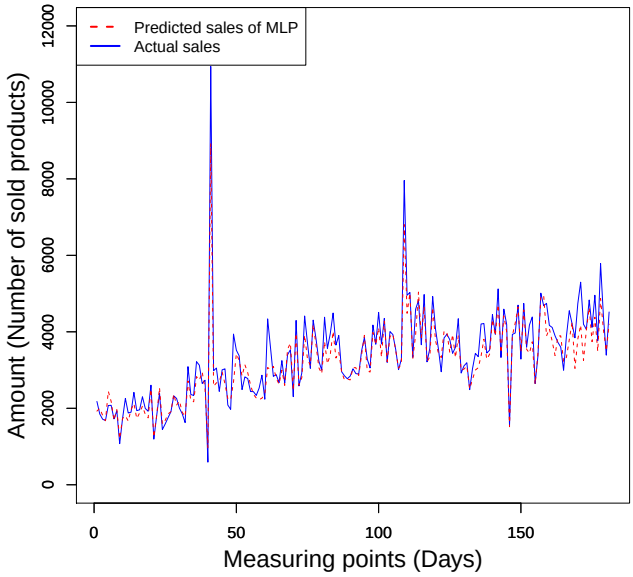


Figure 4.1: The figure shows the run in which the lowest MAPE of the MLP-predictions was achieved.

4.2 Radial Basis Function Network

Figure 4.2 shows the predicted number of sales when the lowest value of MAPE was achieved. The RBFN does match the curve well and overall provides a good fit to the curve as observed in table 4.1. More interestingly, it finds the drops in the actual sales. But since some spikes are not met the RMSE is resulting in a higher value, as is shown in table 4.2.

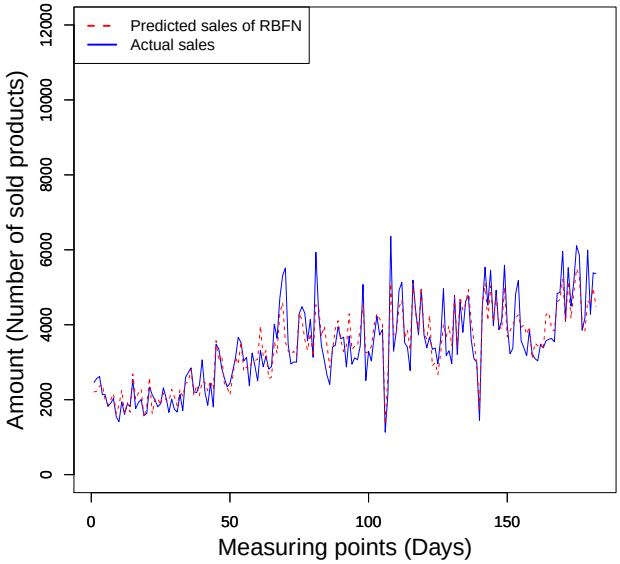
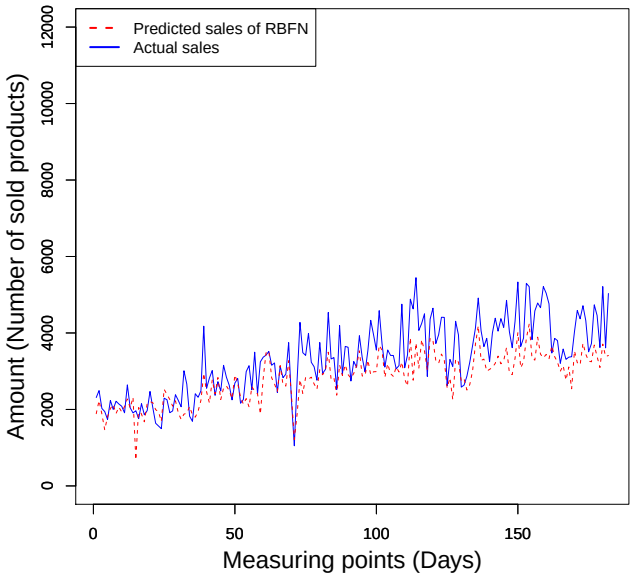


Figure 4.2: The figure shows the run in which the lowest MAPE of the RBFN-predictions was achieved.

The prediction which resulted in the lowest RMSE is shown in 4.3. The figure shows a relatively flat prediction curve compared to the actual sales. However, the RMSE value is the lowest of the RBFN. Also noteworthy is the MAPE, at table 4.1, which is lower than the average.



-

Figure 4.3: The figure shows the run in which the lowest RMSE of the RBFN-predictions was achieved.

4.3 Support Vector Machine

The figure 4.4 shows the SVM prediction where the lowest MAPE was scored. It is noteworthy that the value of RMSE is the highest indicating that there are one or more large errors in the prediction. Further stating the fact the RMSE punishes large errors.

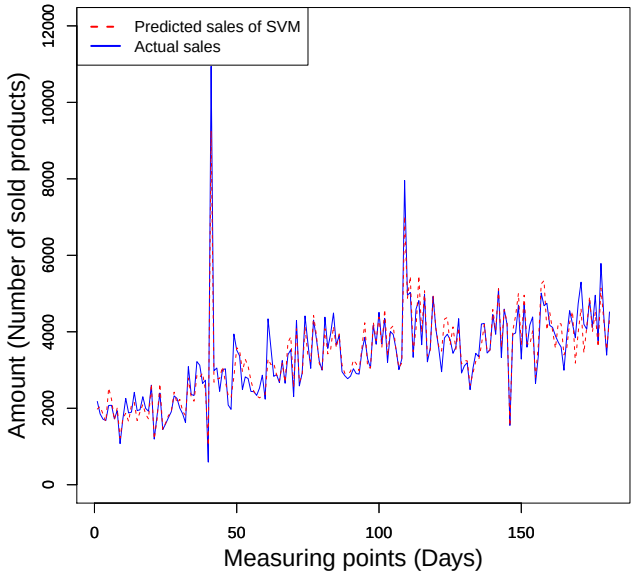


Figure 4.4: The figure shows the run in which the lowest MAPE of the SVM-predictions was achieved.

The SVM prediction with the lowest RMSE is shown in figure 4.5. Looking at table 4.1, also the MAPE value is below average.

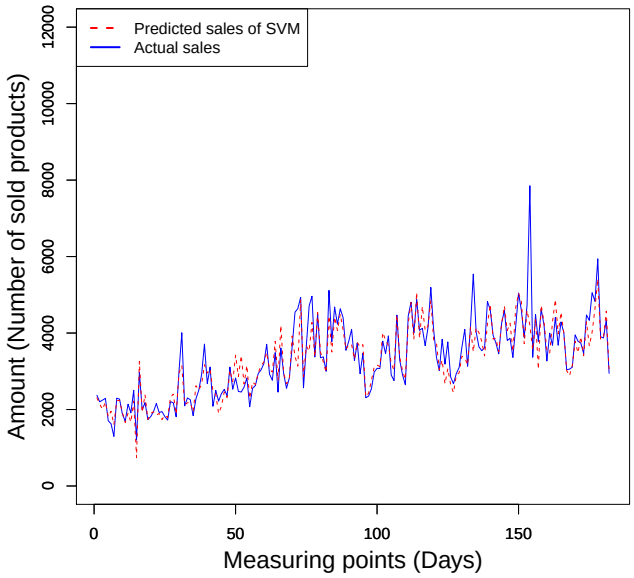


Figure 4.5: The figure shows the run in which the lowest RMSE of the SVM-predictions was achieved.

4.4 Error measures & and standard deviation

In table 4.1, 4.2 and 4.3 the MAPE, RMSE and standard deviations (σ) are shown respectively. In most of the runs the RMSEs of the MLP is considerably larger than the RMSE of the other models. This indicates that the MLP consistently fails to match large deviations in the actual sales. The MLP also shows a larger average MAPE than the other models and a high σ of the MAPE also informs us that the model provides deviating predictions. The same can be stated for the σ of the RMSE.

Table 4.1 shows that the SVM provides the lowest average value of the MAPEs. This in conjunction with the low σ of the MAPE indicates that there is a lower deviation between the predicted and the acutal sales than the other models. Moreover the predictions by the SVM also has the lowest σ of the RMSE which further emphasizes a more consistent performance than the other models.

Table 4.1: Collection of MAPEs obtained by each run

Method	1	2	3	4	5	6	7	8	9	10	**Average**
SVM_{MAPE}	8.1%	8.2%	8.6%	8.1%	7.6%	7.7%	7.6%	8.5%	7.3%	8.5%	8.0%
MLP_{MAPE}	42.8%	45.7%	33.4%	19.1%	59.3%	15.7%	14.6%	38.3%	7.7%	13.5%	29.0%
$RBFN_{MAPE}$	10.2%	15.7%	14.0%	16.2%	36.2%	11.5%	12.8%	15.7%	18.4%	15.3%	16.6%

Table 4.2: Collection of RMSEs obtained by each run

Method	1	2	3	4	5	6	7	8	9	10	**Average**
SVM_{RMSE}	1007.1	943.0	958.0	1001.5	903.9	987.0	1092.0	942.4	1103.0	1058.0	999.6
MLP_{RMSE}	1903.8	1767.2	1514.1	1537.0	1705.6	1511.0	1475.0	2043.2	1010.1	1305.6	1577.2
$RBFN_{RMSE}$	975.8	623.7	1182.0	1195.1	1030.0	816.9	858.8	1184.4	973.5	853.7	969.3

Table 4.3: Standard deviations of the error measures

Standard deviation (σ)	MLP	RBFN	SVM
σ_{MAPE}	0.17	0.07	0.004
σ_{RMSE}	296.7	186.9	66.9

4.5 Analysis of Variance

The values obtained by rANOVA, which can be seen in table 4.4, shows that we can reject the null hypothesis of no difference between the models. As both p-values are lower than the α of 0.05 there is a statistically significant difference between the models.

Table 4.4: Result of rANOVA

rANOVA	MAPE	RMSE
F	17.14	27.63
p	$1.56 \cdot 10^{-5}$	$2.94 \cdot 10^{-}7$

Chapter 5

Discussion

5.1 Discussion of results

In order to verify our problem statement we chose to use the MAPE and RMSE as error measures. As the SVM scored a lower average MAPE and RMSE than the MLP and the RBFN, our result is similar to the one obtained by Pillo et al. [25], who showed that the SVM outperformed a RBFN. However, our MAPE is considerably lower than Pillo et. al [25]. A reason for that might be that this study used data from a food store department instead of a single product. Moreover the models in this study was trained on a larger data set which also might be a contributing factor to the generally lower error measurements. The performance of the SVM was not unexpected since Levis et al. [20] used a SVM to predict the monthly sales of a single product and showed a similar MAPE compared to this study. Our result is also in line with Krause-Traudes et al. [18] who concluded that SVMs can be used to predict aggregated sales in food a store.

We did not expect the poor performance of the MLP, since multiple studies have shown its usage in sales prediction [10, 30, 29, 17]. The MLP shows surprisingly high standard deviations in both MAPE and RMSE compared to the other models. However our result could be a consequence of the parameter settings used since there are multiple ways to choose initial weights and the amount of neurons in the hidden layer.

Another noteworthy finding is the RBFNs ability to produce a low average RMSE compared to the other models which was also shown by Kuo et al. [19]. Producing a low RMSE can be considered important

in the food industry since large prediction errors can cause trouble in maintaining the inventory when using a model for automated ordering of supplies.

As was stated by Pechenizkiy et al. [23], large deviations in actual sales might occur in the food industry as a consequence of promotional days in which the majority of the department is on sale. This report did not contain campaign indicators which could be further looked into, in order to achieve more accurate predictions. There is also a possibility that certain features used in this study fit one model better than the others. However, finding the best possible match of features to the models was not in the scope of this study.

Performance measures used in this study might not be the only interesting quantifiers when comparing models. Another useful metric would be the amount of time it takes to train the model. If the prediction of sales is to be performed daily, over a large number of stores, speed is of essence. Further research could therefore be done on the tradeoff between accuracy and speed for sales prediction models in the food industry.

5.2 Limitations

Due to the time constraint of this study, finding the optimal parameter settings for each model was downprioritized. This can be seen as one of the stronger limitations in this study, which is due to a lack of deeper knowledge in how to find the optimal parameter settings. A more detailed information about these settings in [10, 30, 29, 17] could have been helpful to better compare our results to existing literature. Hence a deeper understanding of what features that affect the sales would potentially result in a more accurate prediction. In addition, the result of a statistical comparison between machine learning methods might not be very informative without sufficient amounts of data. Therefore a loss of generality has to be considered. In addition, the lack of time and available data limited us to only one store. Relinquishing on any of these two factors might have led us to more generalizable result.

5.3 Further Reaserch

As there are vast amount of sales data in the food industry it would be of interest to evaluate what features that are of higher importance for predicting sales. Also, it would be of value to investigate how small amounts of data and how few parameters is sufficient for the purpose of sales prediction. To verify which of the models that are statically different from the others additional statistical tests such as a post hoc analysis could have been used.

Chapter 6

Conclusion

There is a statistically significant difference between the SVM, MLP and RFBN when predicting the sales in a food store department. The SVM performed lower error measures than the other two methods. Since this study used limted data, thus, one could hardly draw the conclusion that the SVM is always the most accurate method to use for sales prediction in a food store department. However, the result of this study can indicate what methods to look at when implementing machine learning methods to predict sales in the food industry.

Bibliography

[1] Mann Adebanjo. "Identifying problems in forecasting consumer demand in the fast moving consumer goods sector". In: *Benchmarking: An International Journal* 7 (2000), pp. 223–230.

[2] Syed Riaz Ahmed. "Applications of Data Mining in Retail Business". In: *ITCC '04 Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04)* 2 (2016), p. 455.

[3] Ilan Alon, Min Qi, and Robert J. Sadowski. "Forecasting aggregate retail sales a comparison of artificial neural networks and traditional methods". In: *Journal of Retailing and Consumer Services* 8 (2001), pp. 147–156.

[4] Sylvain Arlot. "A survey of cross-validation procedures for model selection". In: *Statistics Surveys* 4 (2010), pp. 40–79. ISSN: 1935-7516. DOI: 10.1214/09-SS054.

[5] Nari Sivanandam Arunraj and Diane Ahrens. "A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting". In: *Internation Journal Production Economics* 170 (2015), pp. 321–335.

[6] Christopher M. Bishop. *Pattern Recogniton and Machine Learning*. Springer, 2007, pp. 2–301. ISBN: 978-0387310732.

[7] Marko Bohanec, Mirjana Kljajic Borštnar, and Marko Robnik-Šikonjac. "Explaining machine learning models in sales predictions". In: *Expert Systems With Applications* 17 (2017), pp. 416–428.

[8] Kim Brynjolfsson Hitt. "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance". In: (2011). URL: http://ebusiness.mit.edu/research/papers/

2011.12_Brynjolfsson_Hitt_Kim_Strength%5C%20in%5C%20Numbers%5C_302.pdf.

[9] Real Carbonneau, Kevin Laframboise, and Rustam Vahidov. "Application of machine learning techniques for supply chain demand forecasting". In: *European Journal of Operational Research* 184 (2008), pp. 1140–1154.

[10] Chen-Yuan Chen et al. "The study of a forecasting sales models for fresh food". In: *Expert Systems with Applications* 37 (2010), pp. 7696–7702.

[11] Ching-Wu Chu and Guoqiang Peter Zhang. "A comparative study of linear and nonlinear models for aggregate retail sales forecasting". In: *International Journal of Production Economics* 86 (January 2003), pp. 217–231.

[12] Orinna Cortes and Vladimir Vapnik. "Support-vector networks". In: *Machine Learning* 20(3) (1995), pp. 273–297.

[13] Philip Doganis et al. "Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing". In: *Journal of Food Engineering* 75 (2006), pp. 196–20.

[14] Estrella Funes et al. "A Review: Artificial Neural Networks as Tool for Control Food Industry Process". In: *Journal of Sensor Technology* 5 (2015), pp. 28–43.

[15] Martin T. Hagan et al. *Neural network design, 2 edition*. Martin Hagan, 2014, pp. 2.2–2.12. ISBN: 978-0971732117.

[16] C. Harpham and C.W. Dawson. "The effect of different basis functions on a radial basis function network for time series prediction: A comparative study". In: *Neurocomputing* 69 (2006), pp. 2161–2170.

[17] M. Ahsan Akhtar Hasin, Shuvo Ghosh, and Mahmud A. Shareef. "An ANN Approach to Demand Forecasting in Retail Trade in Bangladesh". In: *International Journal of Trade, Economics and Finance* 2(2) (2011), pp. 154–160.

[18] Maike Krause-Traudes et al. *Spatial data mining for retail sales forecasting*. Tech. rep. Fraunhofer-Institut Intelligente Analyse- und Informationssysteme (IAIS), 2008.

[19] R.J. Kuo, Tung-Lai Hu, and Zhen-Yao Chen. "Application of Radial Basis Function Neural Network for Sales Forecasting". In: *Journal of Industrial Engineering International* 12 (2016), pp. 389–400.

[20] A. A. Levis and L. G. Papageorgiou. "Customer Demand Forecasting via Support Vector Regression Analysis". In: *Chemical Engineering Research and Design* 8 (2005), pp. 1009–1018.

[21] F. Lollia et al. "Single-hidden layer neural networks for forecasting intermittent demand". In: *International Journal of Production Economics* 183 (2017), pp. 116–128.

[22] Brynjolfsson McAfee. "Big data the management revolution". In: *Harvard Business Review* (October 2012), pp. 59–68.

[23] Mykola Pechenizkiy and Patrick Meulsteea. "Food Sales Prediction: "If Only It Knew What We Know"". In: *IEEE International Conference on Data Mining Workshops* 128 (2008), pp. 128–143.

[24] Dilek Penpece and Orhan Emre Elma. "Predicting Sales Revenue by Using Artificial Neural Network in Grocery Retailing Industry: A Case Study in Turkey". In: *International Journal of Trade, Economics and Finance* 5(2) (2014), pp. 435–440.

[25] Gianni Di Pillo et al. *An application of learning machines to sales forecasting under promotions*. Tech. rep. Sapienza, Universiteta di Roma, 2013.

[26] Kumar Rajaram and Christopher S. Tang. "The impact of product substitution on retail merchandising". In: *European Journal of Operational Research* 135 (2001), pp. 582–601.

[27] Marcelino Lázaro Ignacio Santamaría and Carlos Pantaleòn. "A new EM-based training algorithm for RBF networks". In: *Neural Networks* 16 (2003), pp. 69–77.

[28] Friedhelm Schwenker, Hans A. Kestler, and GuÈnther Palm. "Three learning phases for radial-basis-function networks". In: *Neural Networks* 14 (2001), pp. 439–458.

[29] Ilham Slimani, Ilhame El Farissi, and Said Achchab. "Artificial neural networks for demand forecasting: Application using Moroccan supermarket data". In: *Intelligent Systems Design and Applications (ISDA)* (2015).

[30] Frank M. Thiesing and Oliver Vornberger. *Forecasting Sales Using Neural Networks*. Tech. rep. Department of Mathematics and Computer Science, 1997.

[31] Li Y. et al. "The Improved Training Algorithm of Back Propagation Neural Network with Self-adaptive Learning Rate". In: *International Conference on Computational Intelligence and Natural Computing* 1 (2009), pp. 73–76.

[32] Indre Zliobaité, Jorn Bakker, and Mykola Pechenizkiy. "Beating the baseline prediction in food sales: How intelligent an intelligent predictor is?" In: *Expert Systems* 39(1) (2012), pp. 806–815.

TRITA CSC, KTH, 2017-06-12