Forecasting of Product Sales Prices Team 6

David Jung (13646168), Pierre Blanchet (13919172), Mannes Mokkenstorm (11922222), Ella Cordus (11027703)
Applied Forecasting in Complex Systems 2021
University of Amsterdam

ABSTRACT

Making accurate forecasts of future sales and thus creating an optimal stock is one of the key problems in the retail business. In this research, various forecasting methods will be used to predict the next 28 days of sales for the food department of one Walmart store in Texas. The methods used are naïve forecasting models, the Croston method, a Poisson regression and ARIMA methods. The various methods are evaluated based on accuracy and simplicity to define the best performing model. The ARIMA model has the best performance on accuracy. However, this is a more computationally heavy model. Therefore, we suggest the use of the Croston model, which has the best trade-off between accuracy and simplicity. One of the limitations of this study was the computational power at hand and we suggest the exploration of Neural Networks in forecasting this data in future research.

1 INTRODUCTION

A fundamental challenge in the retail business is the forecasting of product sales. Inaccurate predictions or estimations will possibly leave the retailer with a stock shortage, resulting in lower sales. Additionally, when the retailer has too much in stock, more storage space is needed which is limited and expensive, with the possibility of products going to waste. Therefore, it is economically desirable to optimise stocking. This can be done by predicting the future sales of products and adapting the stocking and products orders accordingly. Especially in the food department, where products can only be stored for a limited amount of time, stocking optimisation is very important. The aim of this research is to contribute to solving the stocking problem by forecasting the sales of food products. This is done for a Walmart store, namely TX3, in the state of Texas. The past number of past sales per product, past prices, and calendar information such as weekdays and holidays will be used to make the most accurate forecasting. This will be done by comparing different forecasting methods that predict the next 28 days of sales. To determine the most viable model, several factors will be taken into consideration, such as accuracy, complexity and running time. To predict the sales of the next 28 days for the Walmart TX3 food department, it is important that the model used needs to be robust and account for the fact that the sales do not start at the beginning of the time series for each product. The use of exponential smoothing is suggested for this cause [9]. In exponential smoothing, over time exponentially decreasing weights are assigned. This means that more recent data have larger weights than older data. Moreover, J. D. Croston developed an exponential smoothing method for sales forecasting with intermittent demands [4]. That is why we would expect this method to perform well compared to simpler forecasting methods. Therefore, the Croston's forecasting method is applied in this research to forecast the next 28 days of sales. Additionally, the seasonal ARIMA method with various parameters will be used to produce forecasts. In earlier research, it is stated that sales prediction is more of a regression problem then a time series problem [7]. Hence, we will also forecast the data using a Poisson regression analysis. In summary, four forecasting methods are used: Naïve methods as a benchmark, the Croston method, a Poisson regression and Autoregressive integrated moving average (ARIMA) models.

Multiple studies have been conducted comparing sales forecasting methods [3][8], comparing linear and non-linear models. A recent trend can be seen in the deployment of

Recurrent Neural Networks for time series forecasting, e.g. in the NN5 competition, where the most accurate model was found to be a Recurrent Neural Network, while still closely followed by an ETS and ARIMA model.

Based on relevant research, we hypothesise that the ARIMA model will perform best on accuracy. However, since this is a computational heavy model, the Croston model will have the best trade-off between accuracy and simplicity"

2 EXPLORATORY DATA ANALYSIS

The data set consists of daily sales data from 5.4 years. In total, 1943 days and 823 products from the food department of the Walmart store TX3 in Texas are included in the data set, as well as the daily sell prices of each product. This data set is split up into 1914 days of training data and 28 days of validation data in order to be able to compare different forecasting performances. This is represented in Figure 1. The calendar data includes information on the day of the week and events occurring on those days, such as sporting, religious, cultural or national events. The event types are further specified by event names, such as the Superbowl, Mother's day or Memorial Day to name a few examples. Since the training, validation and calendar data set shared the variable d with the days stored in it, a natural inner join was performed to merge the needed information into one data frame respectively.

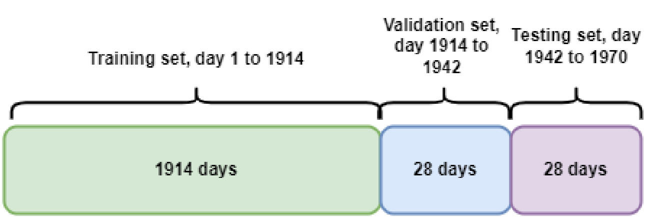

Figure 1: Training, validation and test sets

Firstly, the data was grouped and plotted to examine the overall sales of the TX3 store. This can be seen in Figure 2. From figure 2 it can be concluded that there is a strong cyclic behaviour with a slight upward trend and a steep drop off at the end of the year. Secondly, it can be observed that there is one day each year with zero sales for the whole food department. This indicates a full closure of the store or the food department of the store, on Christmas Day: December 25th. Moreover, two positive spikes stand out. The first one appears on August 24th, 2014 and the second one on July 15th, 2015. Remarkable events on the 24th of August include a large pop concert of the band One Direction in the AT&T stadium, Texas [1], and an earthquake in the state of California [11]. No noteworthy events happened on the 15th of July. Since it is unknown what the exact location is of the Walmart store of this research, we cannot conclude that the spikes in the data correspond to the events named. Based on the EDA performed, there is no indication that the data has to be transformed. The decision for no transformation is made because there is no sign of an increasing variance over time, in combination with having a lot of different products. Since there are so many products, a certain transformation can be needed for a certain product, but not necessary for the entire dataset. This is indicated by figure 5, 6 and 7. There it can be seen that the distribution of sales highly differs between products, not allowing for a one size fits all transformation.

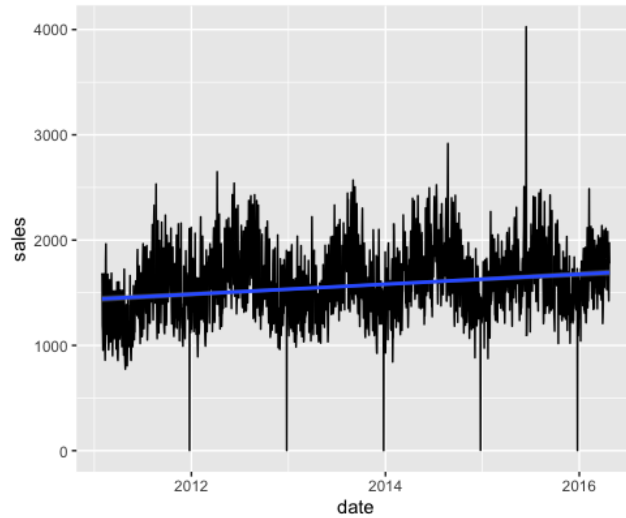

Figure 2: Overview of all sales

Not all product time series are equal in length, since some products were added to the assortment later, examples of which can be seen in Figures 6 and 7. This also implicates the possibility that the overall upward trend, seen in Figure 2, can

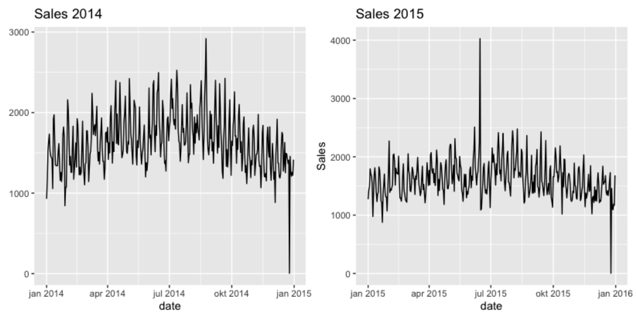

Figure 3: Sales in 2014 and 2015

be (partially) attributed to the addition of new products over the years. Figures 5, 6 and 7 show time series from randomly selected products in the data set. This shows the variety per product in the data set. Some products have periods of zero sales, such as product 1 which can be seen in Figure 5 where nothing is being sold during the summer. Additionally, outliers are apparent as well which can be seen in Figure 4 with product 412. One of the challenges in forecasting such data, is the diversity in the data. In forecasting, we will need a robust model that can follow the seasonality of each products and that performs well considering that some products sales do not always start at the beginning of the data frame.

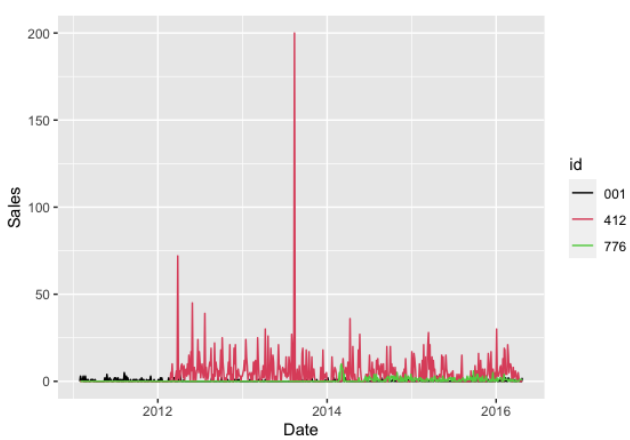

Figure 4: Sales of three products

In addition to the wide range of underlying patterns between the products, the data has another feature which is remarkable, namely that the time series consist of counts.

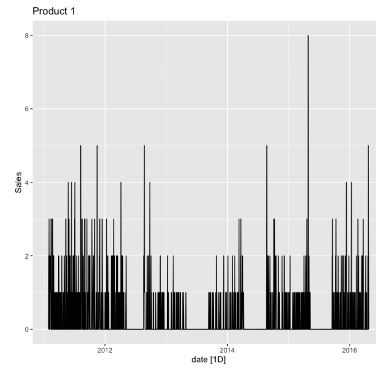

Figure 5: Time Series of Product 1 Sales

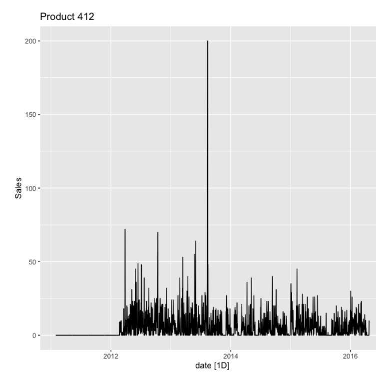

Figure 6: Time Series of Product 412 Sales

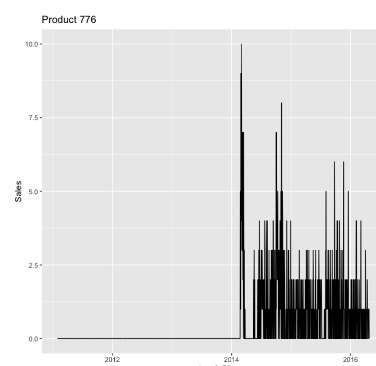

Figure 7: Time Series of Product 776 Sales

This means that one seeks to have a forecasting method which is more appropriate for a sample space of non-negative integers. Before taking a closer look at the forecasting methods, which will be done in the next section, one should take a closer look at the distribution of the prediction variable `sales`. First, the summary of `sales` can be seen in Table 1. Second, the histogram of `sales` can be seen in Figure 8.

Table 1: Sales

Metrics	Values
Min.	0.000
1st Qu.	0.000
Median	0.000
Mean	1.904
3rd Qu.	1.000
Max.	385.000

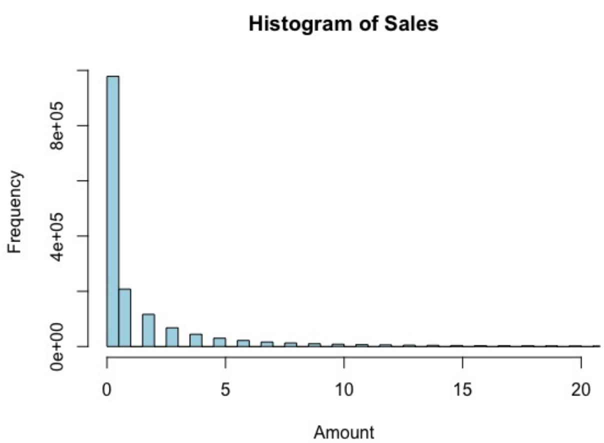

Figure 8: Histogram of Sales

It is clearly visible that a strong right-skewness is apparent with a large range between 0 and 385. Please note that the X-axis has been truncated in Figure 5 to make the distribution more visible. The frequency of sales with 0 amount of product sold is very high for several reasons. First, not all products were present at the beginning of the time frame and therefore they have a high numbers of leading zeros in their time series. Second, as we can see in Figure 5, 6 and 7, the pattern of sales varies a lot among products. Some have high variability and no seasonality whereas other have a strong seasonal pattern. The following Section takes a closer look at the different forecasting methods which are being used in this paper.

3 FORECASTING METHODS

As described in the previous section, this research aims to produce accurate forecasts for the full data set, meaning predicting the next 28 days of sales for each individual products. In this section, various forecasting methods will be explained and the implementations of these models in this research will be discussed. All forecasts are carried out using R. For the Poisson Regression, Python is used. The Naïve methods will be used as a benchmark. Furthermore, the data is not transformed for this research. As can be seen in Figure 2, there is no obvious changing variance in the data that transforming the data would solve.

The accuracy of each forecasting method will be computed by the Root Mean Square Error (RMSE) as this is a more robust error measure [9]. The formula for the RMSE can be seen in Equation 1.

$$RMSE = \frac{1}{n}\sum_{k=0}^{n}\sqrt{\frac{1}{28}(y_t - \hat{y}_t)^2} \tag{1}$$

NAIVE Methods

Even though it is tempting to use more advanced and sophisticated methods right away, one should first compute the naïve forecast methods since they are simple to implement and produce fairly accurate forecasts and therefore represent a good benchmark. This observation was also noted by Hyndman, R.J, & Athanasopoulos [6] "This method works remarkably well for many economic and financial time series". As a brief refresher:

- **Naive:** Simply takes the last known value as a forecast
- **Drift:** Identifies the current drift and uses it as a forecast
- **Mean:** Takes the mean as a forecast
- **Seasonal Naive:** Identifies the seasonality and uses the last seasonal pattern as a forecast

Croston's Method

When Croston's Method is being used, two new time series are being constructed from the original one by noting which time periods contain zero values and which periods contain non-zero values. q_i is defined to be the ith non-zero quantity, and a_i is the time between q_{i-1} and q_i. Two separate exponential smoothing forecasts are being conducted on the two new time series a and q. Since Croston's method is often applied to time series of demand for items, which is also the case in this case study, q is often called the "demand" and a the "inter-arrival time" [6].

Equation 1 & 2 are given by Croston's method if $\hat{q}_{i+1|i}$ and $\hat{a}_{i+1|i}$ represent the one-step forecasts of the $(i+1)_{th}$ demand and inter-arrival time respectively based on the data to demand i.

$$\hat{q}_{i+1|i} = (1 - \alpha_q)\hat{q}_{i|i-1} + \alpha_q p_i \quad (2)$$

$$\hat{a}_{i+1|i} = (1 - \alpha_q)\hat{a}_{i|i-1} + \alpha_q a_i \quad (3)$$

The two used smoothing parameters α_a and α_q can take values between 0 and 1. The time since the last observed positive observation is being denoted by j. Consequently, the h-step ahead forecast for the demand at time T + h is given by the ration in Equation 4 [6].

$$\hat{y}_{T+h|T} = \frac{\hat{q}_{j+1|j}}{\hat{a}_{j+1|j}} \quad (4)$$

In R, the two smoothing parameters α_a and α_q are being automatically estimated from the data.

Poisson Regression

A Poisson regression was also carried out which is often being used to model count data and contingency tables. Within a Poisson regression, it is assumed that the response variable Y (sales in this case) has a Poisson distribution and thus the logarithm of its expected value can be modelled by a linear combination of unknown parameters. The probability mass function of a Poisson distribution can be seen in Equation 5 where $P_X(k)$ is the probability of seeing k events in time t, λ is the event rate (number of events per unit time), t is time and k is the number of events.

$$P_X(k) = \frac{e^{-(\lambda t)} * (\lambda t)^k}{k!} = Poisson(\lambda t) \quad (5)$$

The expected value for a Poisson distribution is λ. Therefore, one can expect λt events during a time interval t. In this case study, the following formula was initially used for the Poisson regression which represents the base model for the Poisson regression:

sales ~ DAY + MONTH

Please note that for the Poisson regression, the programming language Python was being used for performance reasons, instead of R. The model itself was fitted with the `statsmodels` package.

ARIMA Method

ARIMA models are built to describe the autocorrelations of the data. These models can capture complex relationships as they take observations of lagged terms and error terms. Therefore, past points of time series can impact current and future value points. A weight is applied to each of the past term and these can be adjusted depending on how recent or not the data points are.

Non-seasonal ARIMA models have 3 properties. **p** is the order of the autoregressive part. **d** is the degree of first differencing involved and **q** us is the order of the moving average part. Given the apparent seasonality visible in Figure 1, we however wanted to take the seasonality into account. We therefore use the the following ARIMA model:

ARIMA(p,d,q)(P,D,Q)m

Where m is the seasonal period and P,D,Q are the seasonal part of the model. They are terms similar to the non-seasonal components of the model but includes backshifts of the seasonal periods.

In this research, two approaches for the ARIMA method were used. In the first method, for each food item an ARIMA model was fitted automatically by R and used for their respective forecasts. In the second method, it was specified in R that the ARIMA model should search for the best hyperparameter. That was done by precising *stepwise = False* and *approximation = False*.

4 RESULTS

Table 2 shows the results of the forecasting methods used for this research. The data set was split up into a training and a validation set. The training set contains all the data up

Table 2: Forecasting Methods

Forecasting Methods	RMSE
Mean	2.113
Drift	2.349
Naïve	2.340
Seasonal Naïve	2.322
Croston	1.899
Croston0.15	1.899
Croston0.001	1.889
Pois. Reg. 1	2.107
Pois. Reg. 2	2.08
ARIMA (approximation = False)	1.827

until 28.03.2016 and the validation set contains the following 28 days, namely 29.03.2016 - 24.04.2016 in order to perform a 28 days forecast which can be evaluated against real world data. Please note that with this forecast the RMSE values were calculated. Therefore, these values do not represent the ones from the kaggle competition.

Naïve Methods

As discussed in Section 2, the amount of most food items that are being sold during a day remain 0 and the data is in general right skewed. As a result, the naïve methods perform relatively well in forecasting such data. In particular the mean method stands out. Additionally, the running times for the naive methods remain low even though the whole data set was being taken into consideration for the calculations.

Croston's Method

Since one is dealing with a time series of counts, the next logical step is to try Croston's method and indeed, this method is able to beat the score of the naive method. The running time for this method is slightly longer, but remains under a minute which is a surplus. The default value for the smoothing parameter α is 0.1 in R. Therefore, this value has been changed to 0.15 and 0.001 in order to obtain potential improvements. However, the results remained the same.

Poisson Regression

Table 3 shows the metrics of the base model and it is already clearly visible that not all variables are significant. Figure 8 shows a Poisson forecast for food item 1 for 28 days with the base model. Only item 1 is being represented in this paper but the Poisson regression was performed on each of the 827 products individually. One can see that it resembles the mean method which can be explained by the large occurrence of zeros in the time series. However, since it is a regression the values do change over time, as opposed to the mean method.

Table 3: Poisson Regression Fit 1 Item 1

variable	coef	std err	P>|z|
Intercept	-0.9845	0.108	0.000
DAY	-0.0028	0.005	0.534
MONTH	-0.0138	0.012	0.232

Since more information can easily be included in a Poisson regression, the variable `weekday` was included as a next approach. The new regression equation looks as follows.

$$\text{sales} \sim \text{DAY} + \text{MONTH} + \text{weekday}$$

This lead to the following metrics for fit 2 which can be seen in Table 4. Again, most variables are not significant. However, the performance got better with an RMSE of 2.08.

Table 4: Poisson Regression Fit 2 Item 1

variable	coef	std err	P>|z|
Intercept	-1.0381	0.147	0.000
Monday	-0.0473	0.155	0.761
Saturday	-0.0763	0.156	0.626
Sunday	0.2724	0.144	0.058
Thursday	-0.0234	0.154	0.880
Tuesday	0.4921	0.138	0.000
Wednesday	-0.5711	0.181	0.002
DAY	-0.0027	0.005	0.555
MONTH	-0.0139	0.012	0.229

Compared to the fit 1 it is clearly visible that more information is being taken into consideration and the fit does not resemble the mean method as much as fit 1. This time, the Poisson regression tries to include some peaks which roughly occur whenever an item is being sold.

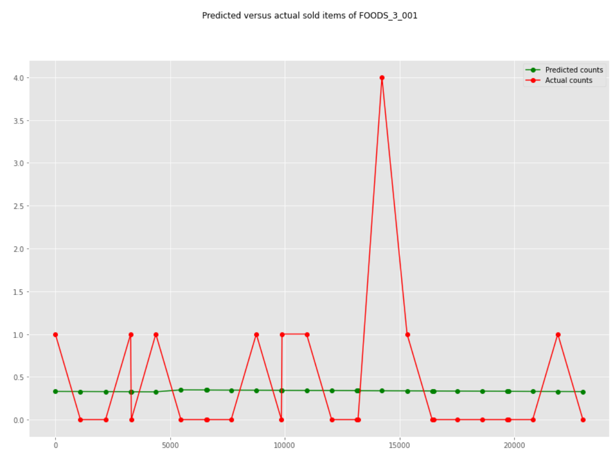

Figure 9: Poisson Regression of Item 1 Fit 1

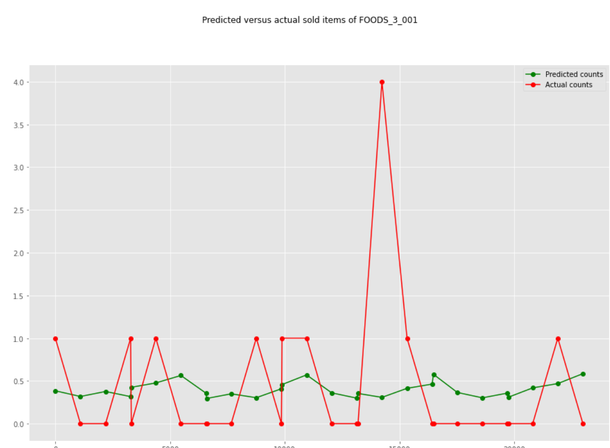

Figure 10: Poisson Regression of Item 1 Fit 2

ARIMA Method

An approach with the ARIMA method was also carried out in this work. Without specifying any hyper parameter, the ARIMA forecast was able to undercut Croston's method by a small margin. Unfortunately, fitting the model according to the second approach was not possible in a reasonable time frame on our machines, as these are very heavy calculations. Nonetheless, with the ARIMA (sales, approximation = FALSE) model some interesting results were obtained that took into account the variability of sales for each product. Each product was therefore fitted with the appropriate ARIMA model. A few examples can be found in table 5.

Table 5: Sample of products & respective ARIMA models

Product	ARIMA model
`001`	ARIMA(2,1,2)(2,0,0)7
`008`	ARIMA(1,1,4)(0,0,1)7
`400`	ARIMA(1,1,1)
`600`	ARIMA(0,1,2)
`008`	ARIMA(3,1,1)(2,0,0)7

5 CONCLUSION

This paper explored different forecasting methods on the sales data of the Walmart store TX3 in Texas. It has been observed that the worst performing method is the Drift models. The two naïve models and the mean method comes next with surprisingly accurate performance. The Poisson method follows the naïve models with an accurracy slightly higher. However, the Croston method and the ARIMA models produce the best forecasts for this dataset. The accuracy of the Croston method can be explained by the presence the intermittent sales data in the dataset. This method is developed for handling such data. ARIMA works well because it can capture complex relationship as well as seasonal and non-seasonal patterns.

The aim of this research was to find the best model to forecast the next 28 days of sales. In addition to the accuracy, the complexity and run time are important factors in defining the best model. The naïve models seem to have a good trade-off between accuracy and complexity, however, when this model's outcomes would be used in a real life setting, the store would have too little or too much stock almost every day. The same goes for the Poisson regression. On the upside, this regression does take the calendar data into consideration. The ARIMA model performs best in terms of accuracy, however the complexity and the runtime of this model are higher, it being the slowest of the 3 best methods tested in this study. The Croston method is very close to the ARIMA model in terms of accurracy. Also, the runtime of this model is lower and the model is less complex. Therefore, we conclude that the Croston model is the best performing model in this research, with an optimal trade-off between accuracy, runtime and complexity. Especially since only a subset of the data is given in this challenge. With a store as big as Walmart with thousands of store locations as well as products scalability is something of high importance when it comes to forecasting.

6 DISCUSSION

The biggest limitation of this research was the computational power at hand. For the ARIMA model, this meant that we were unable to perform the second approach of this model as described in sections 3 and 4. This was also reflected in the inability to use Neural Networks (NN), which is suggested by various authors for time series forecasting [5], [2], [7], [10], [5]. In earlier research the outcomes produced by NNs came out as very accurate forecasts. At the same time, it is suggested using NNs when there is a small amount of historical data available [7]. Neural networks were used to compare forecasting accuracy on individual products. Unfortunately the heaviness of the calculation renders any forecast on the entire dataset impossible with the equipment of the authors. Therefore, we suggest to use this method in future research. Another potential improvement would be to have variable time series length. Indeed some products time series start with a many zeros which do not represent zero sales but simply show the unavailability of the product at the time. Therefore they skew the forecast. The methods using exponential smoothing manage to overcome this limitation to an extent by diminishing the weights attributed to old observation, but this is still an area that could be explored further. A third aspect where improvements could be made is using the historical sales prices and the calendar data in the forecast. It would be interesting to examine to what extent the past sales prices influence the forecast. It can be hypothesised that high prices decrease the amount of sales. The inclusion of calendar data in the forecasting models would be interesting, since sales can spike around certain events or weekdays. The aggregated data did not show such trends except those mentioned in section 2. However, a closer look into the specific food items or clusters of items could reflect calendar-related trends. This could be investigated further in future research, to provide more specific forecasts for food products.

REFERENCES

[1] [n.d.]. One Direction Concert Setlist 2014. https://www.setlist.fm/setlist/one-direction/2014/atandt-stadium-arlington-tx-53ce0bb5.html. Accessed: 2021-12-09.

[2] Angela P Ansuj, ME Camargo, R Radharamanan, and DG Petry. 1996. Sales forecasting using time series and neural networks. *Computers & Industrial Engineering* 31, 1-2 (1996), 421–424.

[3] Ching-Wu Chu and Guoqiang Peter Zhang. 2003. A comparative study of linear and nonlinear models for aggregate retail sales forecasting. *International Journal of Production Economics* 86, 3 (2003), 217–231. https://doi.org/10.1016/S0925-5273(03)00068-9

[4] John D Croston. 1972. Forecasting and stock control for intermittent demands. *Journal of the Operational Research Society* 23, 3 (1972), 289–303.

[5] Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara. 2021. Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting* 37, 1 (2021), 388–427.

[6] Athanasopoulos G. Hyndman, R.J. 2021. Forecasting: principles and practice, 3rd edition. (2021).

[7] Bohdan M Pavlyshenko. 2019. Machine-learning models for sales time series forecasting. *Data* 4, 1 (2019), 15.

[8] Sanjay-kumar Rao. 1985. An empirical comparison of sales forecasting models. *The Journal of product innovation management* 2, 4 (1985), 232–242.

[9] James W Taylor. 2011. Multi-item sales forecasting with total and split exponential smoothing. *Journal of the Operational Research Society* 62, 3 (2011), 555–563.

[10] Frank M Thiesing and Oliver Vornberger. 1997. Sales forecasting using neural networks. In *Proceedings of International Conference on Neural Networks (ICNN'97)*, Vol. 4. IEEE, 2125–2128.

[11] Author Unknown. [n.d.]. California earthquake strikes San Francisco Bay area. *BBC News* ([n. d.]). https://www.bbc.com/news/world-us-canada-28918600

Glossary

List of Figures

1	Training, validation and test sets	2
2	Overview of all sales	2
3	Sales in 2014 and 2015	3
4	Sales of three products	3
5	Time Series of Product 1 Sales	3
6	Time Series of Product 412 Sales	3
7	Time Series of Product 776 Sales	3
8	Histogram of Sales	4
9	Poisson Regression of Item 1 Fit 1	7
10	Poisson Regression of Item 1 Fit 2	7

List of Tables

1	Sales	4
2	Forecasting Methods	6
3	Poisson Regression Fit 1 Item 1	6
4	Poisson Regression Fit 2 Item 1	6
5	Sample of products & respective ARIMA models	7