Individual Work – Sales Forecasting on Time Series Data

Name: Brij Popatbhai Patel
Student ID: 1146509

Sales forecasting in the fashion retail sector is a challenging subject that needs cutting-edge and effective techniques, particularly when projections are made for novel goods that lack sales historical data [1]. To estimate sales in the fashion sector and forecast the sales of new individual goods in upcoming seasons, this study investigates the use of a deep learning technique. The dataset being analyzed is a genuine dataset that was given by the fashion retail corporation whose purchasing procedures are the subject of this investigation.

This study aims to investigate the possibility of a deep learning strategy in this context while trying to forecast the sales of new fashion goods using data mining techniques. Based on previous sales data, the traits of both old and new products, and an examination of their commonalities, it is possible to anticipate future product sales. This paper offers to investigate several regression data mining approaches to create a model to predict the sales of new items. The parameters in this model that define the products, those about the company's internal organizational structure and logistics, and those that incorporate judgments made by subject-matter specialists are used as independent factors to forecast a product's sales (dependent varying) for a specific season. For the model's learning process, historical sales information for the goods that are similar Use of time.

This paper offers to investigate several regression data mining approaches to create a model to predict the sales of new items. The data mining techniques used for the research and have been compared are Decision tree, Random Forest, Support Vector Regression, Artificial Neural Network, and Deep Neural Network. The study also enables concluding that no one strategy can be regarded as the best because none of the strategies under consideration displays better performance across all parameters examined.

The limitation drawn out of the study is the lack of historical data by the companies might result in prediction. The dataset contains data on product sales from prior collections and predictive characteristics that describe current and upcoming items. Secondly, it tried algorithms like ARIMA and XGBoost that might have been taken into consideration for the prediction of the product which has shown quite a good result for the forecasting models.

One of the less well-known abilities in the analytics field is time series [2]. Due to the non-linear trends and noise present in the series, seasonality, trends, and cycles do exist in data, but they are challenging to recognize and predict with accuracy. Their strategy is based on a dilated convolutional neural network that has been modified for time series forecasting. A significant quantity of time series data may finally be analyzed with respectable speed and accuracy by using this approach iteratively to batches of n samples. Test data is divided into portions and utilized for the accurate evaluation of models on both public and private leaderboards, whilst train data comprises data for model training. 125,497,040 train observations and 3,370,464 test observations make up Corporacion Favorita. Datasets included information on sales by date, shop, item, and promotion. Perishable items are given a weight of 1.25 whereas all other items are given a weight of 1.00.

Regardless of how well a forecast is made, it may still be completely off. Using historical data is an accurate approach to predicting if market circumstances stay mostly constant. They have discussed a strategy for employing CNN WaveNet, a sequence-to-sequence architecture, for sales forecasting, which proved to be a very successful way to deal with time series prediction issues. For increasingly challenging problems, CNN architectures with more layers should be considered. Another promising future goal is to adapt the pipeline for other data kinds and domains. The use of deep learning techniques for time series forecasting jobs is briefly reviewed by the authors.

Considering the research, CNN architectures with more layers should be considered. To train a more sophisticated architecture, a large volume of data is necessary. Another promising future goal is to adapt the pipeline for other data kinds and domains. Investigating other ensemble strategies may also result in a respectable accuracy improvement. Additionally, the proposed methodology has not been compared to algorithms like ARIMA and XGBOOST which have shown comparative results in time series forecasting.

One of the most significant commercial sectors for data mining and data science applications is retail [3]. This study suggests a method for forecasting demand and sales for retail at each location. Because efficiency is the future of most sectors today, the authors want to broaden their solution to assist retailers in boosting production and income by utilizing Data Analysis.

Extreme Gradient Boosting, sometimes known as XGBoost, is a library created and enhanced for boosted algorithms. The library's goal is to give large-scale tree boosting a scalable, adaptable, and precise architecture. It is an advancement over the current Gradient Boosting method. They have used several data mining techniques to achieve sales forecasting for retailers. Predicting sales for every store on any given day was part of the challenge. Authors have reviewed prior work in the field, including Time Series Algorithm and a Spatial method, to get conversant with the problem. The data underwent extensive examination to find trends and outliers that may help or hurt the prediction system. Results from the implementation of data mining techniques including XGBoost, Random Forest Regression, and Linear Regression were compared. The technique known as XGBoost, an enhanced gradient boosting algorithm, was shown to have the greatest prediction performance.

Since efficiency is the future of most businesses today, increasing revenue should be the goal. Additionally, the model has the ability to outperform the traditional time series model such as ARIMA, and if the model was compared to the Deep Learning algorithm, would have a broader perspective on factor selection. Secondly, classifying products according to the category in a particular store would help the store to generate revenue.

The Study proposes [4], an organized pre-processing architecture to address the issues in the e-commerce industry Additionally, they present several product grouping techniques to support LSTM learning schemes in cases when a product portfolio's sales trends are dissimilar. Based on data from a real-world online marketplace provided by Walmart.com, the authors experimentally evaluate the suggested forecasting paradigm. The experimental set-up utilized to empirically assess the suggested forecasting framework was provided by the authors. The datasets, error measurements, hyper-parameter selection approach, benchmark techniques, and LSTM variations utilized to conduct the experiments are all included, as are the outcomes. Based on two datasets obtained from Walmart.com, they evaluated our forecasting methodology. First, a selection of 1724 products from the product home category which has 15 distinct subcategories—was used to assess the framework. Then, extract a collection from a single super-department, which comprises 16 separate categories, to scale up the number of goods to 18254. The authors have grouped the product into categories, A product group with a high sales ranking and a low zero sales density is represented by the first group (G1). Group 2 (G2), in contrast, stands for the product category with the highest zero sales density and the lowest sales rating. The remaining items are represented by Group 3 (G3). Improving the accuracy of sales forecasts in G1 is crucial because, from an e-commerce standpoint, products in G1 are the "head items" that contribute the most to the company.

The LSTM cluster and LSTM are G1 components with the greatest potential for commercial impact. In terms of the Mean and Median mMAPE, group variations exceed the rest of the benchmarks. The Nave forecast and the LSTM variations, on average, produce the greatest results in G2 and G3, when product sales are less plentiful than in G1. The authors have trained a global model across the products offered in a product assortment hierarchy using the suggested technique to estimate the sales demand.

Their advancements provide our basic model with various systematic grouping techniques that are particularly helpful when there are few product sales. The ConvLSTM is developed for reading two-dimensional spatial-temporal data but can be adapted for use with univariate time series forecasting, this might be useful as there are clusters of the product that might end up in a different result.

Forecasting sales based on the influence:

1) Forecasting for Sales Cycle Length

This forecasting technique makes use of information on how long it generally takes a lead to become a paying customer. For instance, if your sales rep has been working with a prospect for two months and an average sales cycle lasts four months, there is a 50%

probability that your rep will close the deal. Because the length of the sales cycle is fully objective, forecasting using this method is excellent because your prediction is not reliant on a subjective aspect like your reps' gut feeling that they'll close or their feelings of optimism regarding this prospect. It is entirely impartial. You can use it on other lead sources to get a more precise forecast. Depending on the customer you are selling to, the sales cycle may change. For SMBs, it will be shorter; for enterprises, it will be longer. Data should be diligently tracked by your sales team. Your projections could be drastically off-base with only a minor mistake. Your sales projection can be significantly affected by changes to your product. These adjustments could involve adding a new feature in response to high demand, eliminating a pointless feature, or fixing faults. These product tweaks can help salespeople clinch more deals and reduce their sales cycle.

2) Intuitive Forecasting:

The Intuitive Forecasting approach is predicated on your confidence in your sales representatives. You begin by inquiring about their level of assurance regarding the timing and completion of their sale. It considers the feedback of the sales representatives who are near the prospects because they are most likely to be aware of how things are going. You can easily understand the drawbacks of this approach. It is arbitrary. Since sales representatives tend to be upbeat, their estimations will probably be generous. Additionally, there is no other way to confirm this evaluation than to review every encounter between your sales representative and the client, which is just more labor. It's a terrific technique for freshly established businesses to make sales estimates without using historical sales data. It is quite individualized. Reps frequently have more faith in the success of their opinions. Similar products created by different companies can be found in any industry. The market share of the industry might change and be impacted by factors like new technology, design, competitors running promotional efforts, or new firms joining the market, all of which will affect your sales estimates.

Work Progress:

Considering the Literature review individually, we have discussed our future goals and steps to perform individually. We have decided to make certain models and compare them with each other so that we can identify which one is the perfect fit for our dataset and can be used by other business users for future profit.

As an individual work, I have decided to make the LSTM model (Long Short Term Memory) which is A unique variety of recurrent neural networks called an LSTM that can recognize long-term dependencies in data. This is made possible by the model's recurring module, which consists of four levels that interact with one another. A cell state, three gates, and an LSTM module provide them the ability to selectively learn, unlearn, or retain information from each of the units. By allowing only a small number of linear interactions, the cell state in LSTM aids in the uninterrupted flow of information across the units. Each component contains an input, an output, and a forget gate that can add or remove data from the cell state. The forget gate utilizes a sigmoid function to determine which information from the previous cell state should be ignored. Using point-wise multiplication of "sigmoid" and "tan h," the input gate regulates the information flow to the active cell state. The output gate ultimately determines which information should be transferred to the following hidden state [5]. For the sales forecasting of time series store data, I have to follow some steps on data to make it compatible with the evaluation of the model. In the beginning, I have come across 2 different approaches of the LSTM model to select features; 1st is by selecting families of items as features and 2nd is by considering date and trend as features.

1. **Considering Family as a Feature:** For the family feature, first I have considered the training table of the dataset which consists of item id, date, store number, the family of the item, sales, and promotion. To train and evaluate the model, I have split the data into training and testing sets of 250 thousand training data and 50 thousand testing data. The family feature has been encountered by the "Ordinal Encoder" which fits and transforms the training and testing data as per the family of items.

   Along with that, a pivot on the training data has been performed with an index of "date", values as "Sales", and columns "Store Number" and "Family". This process has narrowed down the training data from 250 thousand to 1403 days of sales data. After that, I used the MinMaxScaler which reduces the data inside the specified range, often between 0 and 1. By scaling features to a predetermined range, it changes data. It scales the values to a particular value range while preserving the original distribution's shape [6].

   LSTM works on Time Steps that are used to predict upcoming future sales. For the proposed method, I have created a "split series" function that creates a new dataset with a window size of 5 which is used to predict near-future sales. After creating a new dataset for LSTM, I made an LSTM model with a unit size of 256, and input shape = (5, no of features) and also added the Dropout and Batch Normalization layer. The Mean Square Error value for LSTM with the family feature is as below.

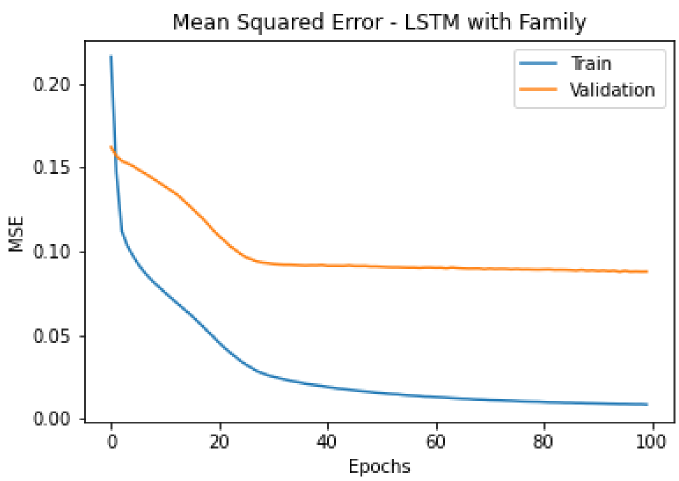

Fig 1. MSE values for LSTM with Family Feature

2. **Considering the Date and Trend as Feature:** Along with individual work, all team members discussed our progress and the similarity of feature selection. At the end of our discussion, we decided to work on the date and trend feature because it is feasible for all of our models so it would be a more appropriate feature to compare our models. Among all the given datasets provided on Kaggle [7], there are 4 main datasets that are important to us namely “Training Set”, “Holiday Events”, “Transactions”, and “Oil Price” as it contains features that affect the sales predictions. In the beginning, I grouped data of transaction, oil, and training table on “date” and then merge those tables with the index of date.

   As per the previous method, LSTM works on the time step so I have also created the new dataset but this time with a time step of 10 days. So here function used the training and testing dataset and generate the new dataset with a slot of 10 data to predict the 11th data. In LSTM, I have only considered the date and sales values because LSTM predicts future value by looking at previous sales values and trends. I calculated scalar values of sales data. After that, I made the LSTM model with the activation Function “Relu”, and input shape of (10, 1). For the compilation, I applied the “Adam” optimizer and a learning rate of 0.001. To compare two different feature selection methods, I have also generated an MSE graph as below.

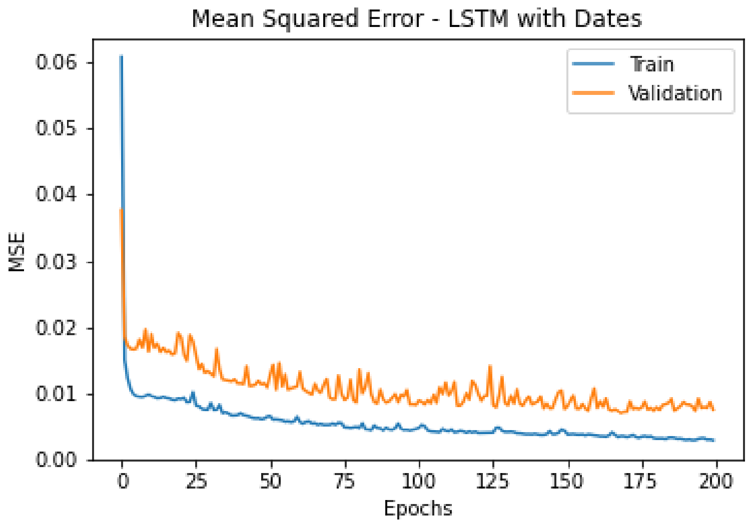

Fig 2. MSE values for LSTM with Date and Trend Feature

By looking at the values of both MSE graphs, I can say that for the given dataset LSTM with the date feature has performed well compared to LSTM with a family feature so for further implementation and prediction, I will consider the Date feature and other groups members also find date feature is better for their models.

For the comparison of the model’s performance, I have inversed the predicted scaled value to its original sales value and plotted those values on the graph. I have also calculated RMSE (Root Mean Square Error) and Normalized RMSE values.

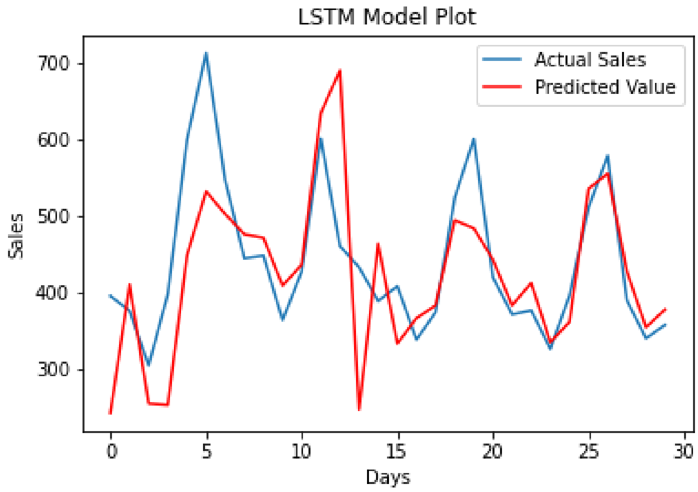

Fig 3. Actual and Predicted values for the LSTM model

```
rmse_LSTM = sqrt(mean_squared_error(new_data_lstm_test[11:], test_pred))
print('Test RMSE: %.3f' % rmse_LSTM)
```

```
Test RMSE: 116.232
```

```
NormalizedRMSE_LSTM = rmse_LSTM / (max(new_data_lstm_test) - min(new_data_lstm_test))
print('Test Normalized RMSE: %.3f' % NormalizedRMSE_LSTM)
```

```
Test Normalized RMSE: 0.143
```

Fig 4. RMSE and nRMSE values for LSTM Model

References:

[1] Loureiro AL, Miguéis VL, da Silva LF. Exploring the use of deep neural networks for sales forecasting in fashion retail. Decision Support Systems. 2018 Oct 1;114:81-93.

[2] Kechyn G, Yu L, Zang Y, Kechyn S. Sales forecasting using WaveNet within the frameworkof the Kaggle competition. arXiv preprint arXiv:1803.04037. 2018 Mar 11.

[3] Jain A, Menon MN, Chandra S. Sales forecasting for retail chains. San Diego, California: UC San Diego Jacobs School of Engineering. 2015.

[4] Bandara K, Shi P, Bergmeir C, Hewamalage H, Tran Q, Seaman B. Sales demand forecast ine-commerce using a long short-term memory neural network methodology. InInternational conference on neural information processing 2019 Dec 12 (pp. 462-474). Springer, Cham.

[5] (2020). Time series - LSTM. TutorialsPoint. https://www.tutorialspoint.com/time_series/time_series_lstm_model.html

[6] Data Pre-Processing with Sklearn using Standard and Minmax scaler - GeeksforGeeks. (2021, February 21). Retrieved December 12, 2022, from https://www.geeksforgeeks.org/data-pre-processing-wit-sklearn-using- standard-and-minmax-scaler/

[7] Store Sales - Time Series Forecasting | Kaggle. (n.d.). Retrieved December 12, 2022, from https://www.kaggle.com/competitions/store-sales-time-series-forecasting