Time Series Forecast For Non-Linear Sales Trend

Yousef Humsi, Dr. Waleed Al-Sitt

Abstract: Sales forecast is an essential tool for any company seeking success, as it gives insight into how a company should manage its workforce, income and resources. Additionally, sales forecast help companies to allocate their internal assets viably. In this paper will report two different approaches for forecasting non-linear sales trends using Long Short-Term Memory (LTSM) and Facebook developed model (Prophet) by implementing it on multiple sources of historical sales data.

General Terms: Forecast, Machine learning, Prophet, Seasonality, Outliers detection. Hampel filter.

Keywords: LTSM, Normalizing, ML, One-hot encode, Outliers

1. INTRODUCTION

The goal of forecasting is not to predict the future more to put to your notice what you need to know to take proper, near to flawless actions in the present. According to Oxford dictionary, Forecasting is defined as: “A calculation or estimate of future events, especially coming weather or a financial trend. 'Past generations of economists were able to make forecasts based on trends in industrial activity”. We tend to forecast in order to minimize risk as primary goal, as well as manage our resources at the upmost effectiveness and efficiency to be able to maximize on outcome, forecasting help in getting all resources work on a higher synergy level with lower rates of risk to have better outcome. With the increased level of competition in today's market among products, and globalization breaking all boarders of trading and resourcing, forecasting importance has never been emphasized on more, no company's KPI is free of forecasting accuracy, in fact forecasting now plays major role in deciding a success of a company/product or a failure. While some companies use forecast to minimize loss, others use it to find and create future opportunity, following a combined approach is a key of company’s success formula. Forecasting is one of the oldest methods used by humans to predict future outcomes, going back to the Primary Era, humans were able to predict the weather by simply monitoring the movement of birds. this was called the observation of nature patterns, in late 19th century prediction was developed using modern statistics as Karl Pearson introduced today’s widely used Pearson product-moment correlation coefficient and John Galton developed key concepts such as standard deviation, correlation and even regression analysis. In this paper, our aim is to forecast non-linear sales trends by analyzing and spotting hidden insights which will provide us with additional features that can support and enhance our model prediction for the future and will enable us to set clear targets for our future sales to optimize our outcomes and reduce the risk of over supplying the market and increase the opportunities of creating new sales.

2. LITERATURE REVIEW

2.1 Related Work:

Recently, several studies for time series forecast proposed on using different approaches such as supervised machine learning. Other studies prove that sales prediction is rather a regression problem than a time series problem. Practice shows that the use of regression approaches can often give us better results compared to time series methods[3]. Machine-learning algorithms make it possible to find patterns in the time series., in general the approaches can be divided into two main approaches, one of them by implementing supervised learning techniques on machine learning classifiers that we will explain in this paper, other techniques covered unsupervised learning by creating new features such as trendline, yearly and weekly seasonality.

2.2 Background:

2.2.1 Hampel Filter:

The Hampel Filter used in this project is a moving-window implementation of the Hampel identifier described by Davies and Gather [1]. the Hampel Filter is used to detect the outliers by implementing a sliding window that compute the median of surrounding samples along with the standard deviation in each sample, if the samples differ from the absolute median by three standard deviations, we replace it with the rolling median of the same window.

2.2.2 LTSM:

Long Short-Term Memory networks, usually just called “LSTMs”, are a special Recurrent Neural Networks (RNNs) that are suitable for learning long-term dependencies[2]. The model uses deep learning to select useful features to predict the targeted time series variables, LTSM requires categorical data to be transformed using one-hot encoding before its implemented in the model.

2.2.3 Prophet:

Introduced by Facebook in 2018, Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, Prophet works with decomposable time series with three main components: trend, seasonality, and holidays[5]. The model provides deep insights such as changepoints that highlights the changes that encounter our time series trends as shown in figure 1.

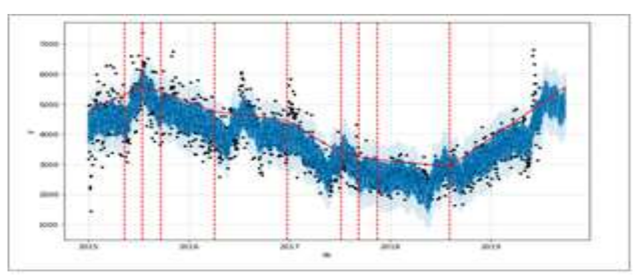

Fig. 1: *Prophet Change Points in Time series*

3. METHODOLOGY

3.1 Data Preprocessing:

Historical data was imported from multiple sources in which sales data was imported from firm internal system for the sales occurred between 2015 and 2019, weather data was imported from third part website in which it provides the weather conditions for the same period in the country we are performing the forecast in, the third dataset includes the holidays and special occasions that occurred in the same period. Preprocessing consist of the following steps: dimensional reduction, replacing missing records, normalizing outliers, One-hot encoding and categorical classification as shown in figure.2.

Fig. 2: *Data Preprocessing Phases*

A. Dimensionality Reduction

Models for time series forecast requires a single record for each date, and in order to fit the data to our model, we need to perform vertical dimensional reduction by aggregating each date as single record. Vertical dimensional reduction compresses the data from its current state which will reduce the numbers of total records in our data.

Table. 1*: Record before dimensional reduction*

Date	Country	Sales	Temp
01-01-15	Coun_1	754	10.94
01-01-15	Coun_1	702	10.94
01-01-15	Coun_1	422	10.94
01-01-15	Coun_1	386	10.94

Table. 2*: Records after dimensional reduction*

Date	Country	Sales	Temp
01-01-15	Coun_1	2264	10.94

B. Replacing Missing Records

Imported data had missing records such as weather conditions for multiple dates, missing values were replaced with the average condition of the same month within the same year same approach was used for the sales data and holidays dataset.

C. Normalizing Outliers

Outliers were found within the records and effected the model accuracy in which they had the influence on our prediction therefor to eliminate their influence we had to identify the outliers and normalize them, traditional outliers detection using histogram was inefficient to detect time series outliers therefore we implemented Hampel Filter technique in which it use sliding windows while moving through our time series, Hampel filter observe the surrounding records for each record and its classified as outliers based on the surrounding median

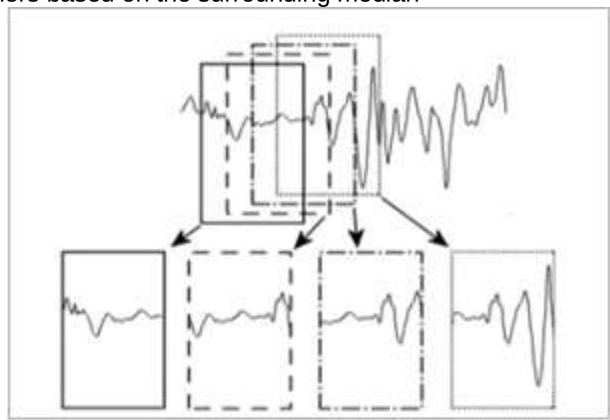

Fig. 3*: Hampel Filter Outliers detection*

absolute deviation, once the outliers is identified we replace it with the rolling median of the same window surrounding the record.

Status
Overcast
Partially cloudy
Clear

Overcast	Rainy	Clear
1	0	0
0	1	0
0	0	1

Fig. 4*: Categorical One-Hot encoding*

D. One-Hot Encoder

In order to fit the data to our machine learning model we had to implement one-hot encoder to our categorical data in which it converts each variable to a new binary variable by expanding the number of features.

E. Categories Classification

Our aim in this model is to predict not only the total of sales but also to predict each product separately, therefore each product had to be divided into its own dataset and will be fed to the model separately.

3.2 Features selection:

Feature selection, as a dimensionality reduction technique, aims to choosing a small subset of the relevant features from the original features by removing irrelevant, redundant or noisy features [4], in this project features shows a weak to moderate correlation with our targeted variables, Features correlation is demonstrated in figure 5.

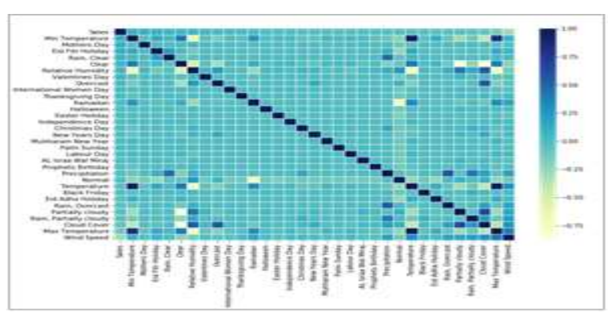

Fig. 5*: Features correlation heatmap*

Based on the above demonstration of features correlation we can spot the features that has impact on our targeted values in which it will save us from importing unrelated features to our model which will enhance our model prediction and reduce the computational time for our machine.

3.3 Training Models:
As mentioned earlier we have two different approaches to forecast our non-linear time series data in the first model using supervised learning with (LTSM) we prepare our data with the selected features we split the data into two sets, one for training which possess 90% historical sales data of the total dataset while the other 10% is used to test.

```
-------------------------
splitting data into train and test . . .
train: Counter({1560, 9})
test: Counter({174, 9})
-------------------------
```

Fig. 6: *Splitting data set to test and train*

Afterwards, the training set will be fed to our classifiers for learning and the model will iterate through each line in our test set to predict the targeted attributes of each observation. For our second model using Prophet data is fed entirely to the model with no features in which the model only depends on tuned dataset for holiday and seasonality and the historical records of the sales, after preparing both dataset we tune the model for the requested prediction period by adding dates with empty observations for the sales to predict the desired period. During the training of Prophet, the model compute additional features for sales trend, weekly and yearly seasonality

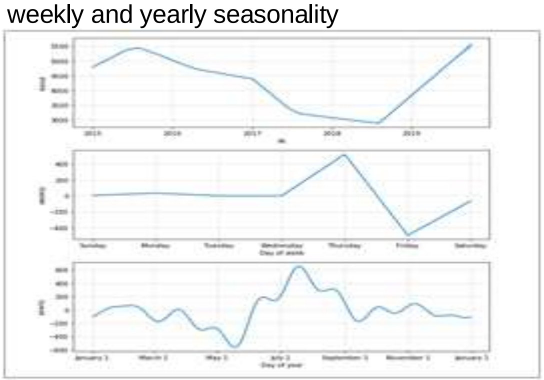

Fig. 7: *Seasonality Features*

After creating the additional features, Prophet forecast the sales for the desired period.

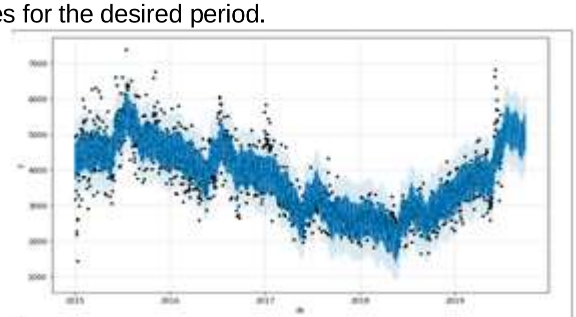

Fig. 8: *Prophet Sales Forecast*

4. RESULTS AND DISCUSSIONS
Generally, the model approach for time series prediction is distinct from traditional evaluation techniques using the four indexes: Accuracy, Precision, Recall and F1-score which is calculated using confusion matrix results for each classifier. Our model predicts sales values which classify it as regression model, therefore the accuracy is evaluated based on the residuals between the predicted values and the actual results:

A. LTSM model evaluation:
After training our data and fitting it to our model we can iterate through all records in the testing set to predict the actual values, afterwards the model iterates the predicted values with the actual values in the testing set to evaluate the accuracy of the model.

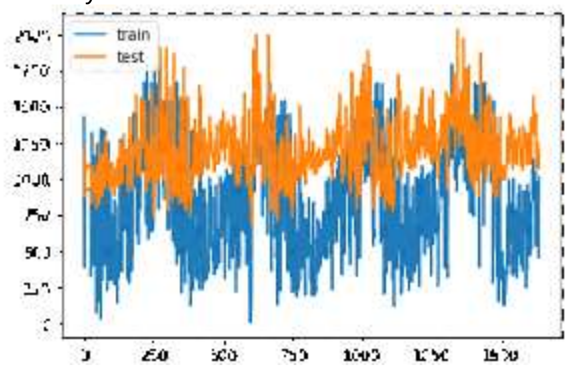

Fig. 9: *LTSM Predicted Sales*

B. Prophet model evaluation:
By computing the differences between the actual sales and the predicted sales we can evaluate our model following to two steps: the first by plotting the actual sales and the predicted sales we calculate the differences between the predicted values with the actual values in our dataset.

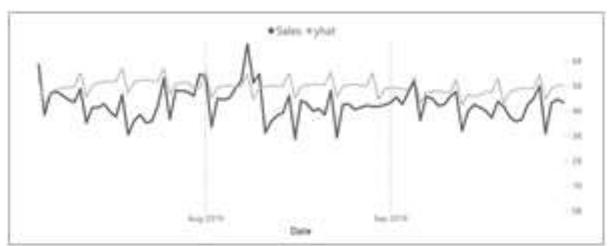

Fig. 10: *Actual Sales and Predicted value trendline*

For the second step we plot the difference between both datasets in two methods, one is by calculating the accuracy on daily level, and the second by plotting the accumulated accuracy for the predicted period.

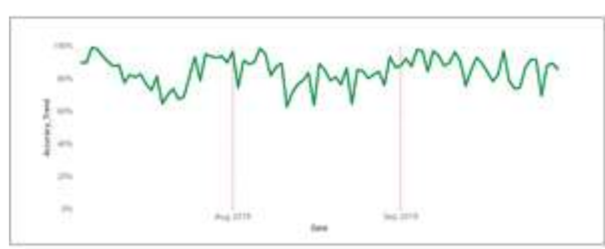

Fig. 11: *Daily Accuracy*

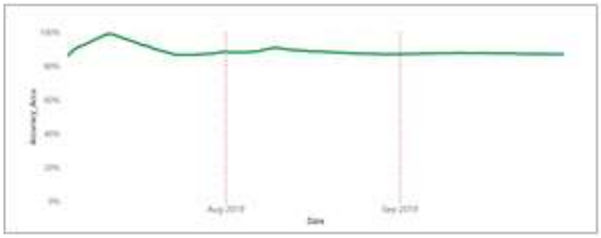

Fig. 12: *Accumulated Accuracy*

After plotting the accuracy, we noticed some fluctuation in the daily accuracy due to abnormal sales that occurred in those days with minimum accuracy of 64% and maximum of 99% per each predicted value and actual values, therefore the accumulated accuracy will provide better insights in evaluating our model, below figure represent the Prophet Plot for the predicted value in respect to the actual historical sales occurred in the model.

5. CONCLUSION

In conclusion, we can observe that our two models had two different approaches, LTSM approaches the data based on the features provided by the user, while Prophet compute its own features. However, after conducting the experiment LTSM model showed a low accuracy rate as its dependent on the features fed to the model, and due to low correlation between the features and the targeted variable the model was not able to provide accurate results, Prophet showed reliable results in respect to the actual values as market seasonality trends showed a consistency among the years which enhanced the model predictivity. Going deeper into the model we noticed the data has high fluctuation in the daily forecast due to abnormal sales in selected dates, the model provided higher accuracy for wider windows sales such as weekly and monthly forecast which is more efficient in predicting the sales to provide proper supply specially for firms that are dependent on factories out side the country of residency.

6. FUTURE WORK

Enhancing our model by providing other sources of data such as sell-in data, sell-thru and firm on hand stocks which has direct influence to the sales, this will provide wider space for our model to practice the prediction of the sales by overviewing the entire supply chain process.

7. REFERENCES

[1] Navratil, Miroslav & Kolková, Andrea. (2019). Decomposition and Forecasting Time Series in Business Economy Using Prophet Forecasting Model. Central European Business Review. 8. 26-39. 10.18267/j.cebr.221.

[2] Hua, Yuxiu & Zhifeng, Zhao & Li, Rongpeng & Chen, Xianfu & Liu, Zhiming & Zhang, Honggang. (2019). Deep Learning with Long Short-Term Memory for Time Series Prediction. IEEE Communications Magazine. PP. 1-6. 10.1109/MCOM.2019.1800155.

[3] Pavlyshenko, Bohdan. (2019). Machine-Learning Models for Sales Time Series Forecasting. Data. 4. 15. 10.3390/data4010015.

[4] Miao, Jianyu & Niu, Lingfeng. (2016). A Survey on Feature Selection. Procedia Computer Science. 91. 919-926. 10.1016/j.procs.2016.07.111.

[5] Pearson, Ronald & Neuvo, Y. & Astola, J. & Gabbouj, Moncef. (2015). The class of generalized hampel filters. 2501-2505. 10.1109/EUSIPCO.2015.7362835.

[6] Alon, Ilan & Qi, Min & Sadowski, Robert. (2001). Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods. Journal of Retailing and Consumer Services. 8. 147-156. 10.1016/S0969-6989(00)00011-4.