A Study of Sales Forecasting and Prediction

Submitted in partial fulfilment of the requirements of the

Research Project

By

Paras Jain
Birla Institute of Technology & Science, Pilani
Rajasthan, INDIA

Under the supervision of

Dr Dmitry M. Klionskiy
PhD, Associate Professor
Department of Information Systems
Department of Software Engineering and Computer Applications

Saint Petersburg Electrotechnical University "LETI"
Russian Federation, Saint Petersburg
February, 2021

“Study hard what interests you the most
in the most undisciplined,
irreverent and original manner possible”
RICHARD FEYNMAN

Acknowledgement

I would like to express my deepest appreciation to all those who provided me the possibility to complete this report. I wish to express my sincere gratitude to my supervisor Dr Dmitry M. Klionskiy, for his patience, interest, insightful comments, practical advice and unceasing ideas that have always helped me enormously in my research and writing of this project report. Without his support and guidance, this project would not have been possible. I could not have imagined having a better supervisor in my study.

I would also like to thank the International Office at the Saint Petersburg Electrotechnical University for giving me this opportunity to pursue this research project and for their constant support throughout.

Abstract

Sales forecasting has long been a crucial aspect for firms and companies which allows these institutions to pursue better planning and management thereby making informed decisions and increase the profit margins. Currently, various methods exist to forecast daily, weekly and monthly sales for various commodities and items. However, each of these techniques suffers from their drawbacks. In this study, we apply various statistical and Machine Learning techniques to predict car sales in the European Union. Precisely, we use several algorithms including HWES, AR, MA, ARMA, ARIMA and SARIMAX. We also used other machine learning algorithms including Bayesian Ridge, XGBoost, LightGBM, Random Forest, KNN, SVM, artificial neural networks and LSTM based networks. We also tested ensemble-based methods to carry out superior forecasting and prediction on our dataset. Our proposed method gives a MAPE of 0.1491 hence paving way for further research.

1 Introduction

Sales prediction and forecasting is an indispensable part of business processes throughout the world with many businesses opting for formal forecasting of their products before or after the launch of their products. Accurate sales forecasting allows managers to take informed decision and devise strategies and solutions for long and short term basis. The onus to provide accurate forecasts lies on Business Analysts who consider various features like the complex market environment, demand, competitiveness and viability before rolling out mathematical models and hypothesis to describe the sales process.

Retail sales in India is one of the fastest growing markets in the whole world which is expected to reach \$1.75 trillion by 2026 and currently accounts for 10% of the GDP. The E-commerce industry stands at \$30 billion in 2019 and is expected to grow at 30% CAGR for gross merchandise value set at 2026. India's retail market is ranked at No. 2 in Global Retail Development Index (GRDI) in 2019 and is responsible for the generation of 8% of the total employment. The share of the organized retail is expected to grow by 20-21% by 2022 with a potential to reach an upwards of \$140-160 billion. In these circumstances, it is inevitable for the enterprises to leverage the growing economic conditions and make plans that conform to market rules.

Sales demand for various products observe an erratic trend. In case of high demand, retailers have to maintain adequate inventory and efficient supply-chain management systems to satisfy the needs of the customers. In case of low demand, maintaining huge stocks is detrimental to the retailer and adds to the overhead costs and losses for the retailer. Demand is harder to predict in case of pre-launch products where little to no data is available but forecasting is even much crucial to maintain customer satisfaction. Therefore efficient forecasting is vital to managers of supply chain and increasing profit margins.

Sales forecasting impacts the marketing plans of the enterprises. It is important to consider various macroeconomic factors and economic environment in the study to generate robust results. The current rate of inflation, consumer price index, monetary policies and the state of financial markets add to the growing list of indicators a researcher must look at while developing forecasts aimed at increasing customer retention rate and gross margin returns on investments.

With the advent of Big Data and analytics starting to revolutionize most fields, retail marketing and sales is one such domain where results are on a positive growth. Big data is helping managers grow their sales to an all-time high and improving prospects considerably over the past years. In marketing too, big data is providing insights into customer satisfaction and helping enterprises invest in a more informed manner thus helping them attain greater customer responsiveness. Artificial intelligence with open source libraries and resources are helping Business analytics tap resources and gain actionable intelligence into the key drivers for Businesses.

At present, various methods for forecasting can be divided into four dominant domains: Qualitative analysis (Delphi Method), Quantitative Time Series analysis (LSTM modelling, Bass Model, ARIMA), Quantitative Machine Learning analysis (Feedforward networks), and combined methods. The Delphi method originated in a series of studies that the RAND Corporation conducted in the 1950s. The objective was to develop a technique to obtain the most reliable consensus of a group of experts[1]. The Bass model, proposed in 1969, is one of the most frequently cited studies that captures forecasting of sales using Time-series analysis methods. It considers the diffusion of a particular product using two parameters, i.e. word of mouth or imitation and external publicity or innovation. It is able to forecast the number of people buying a particular product and can also be extended to the sales of durable items.

Various studies have been undertaken that have attempted to modify the original Bass model to encompass other factors as well. Long short-term memory networks are a special kind of Recurrent Neural Networks that are able to learn long-term dependencies. It has been shown that these networks perform better than the traditional methods such as ARMA, SARIMA and ARMAX which have been used for decades. These networks can also be applied to a wide-array of time series problems since conformity to stationary data is not needed while making a model. There also exist studies that use hybrid models using traditional and newer methods which have been shown to perform better.

Modern neuro-fuzzy networks are represented as special multilayer feedforward neural networks which is based on a fuzzy system trained by a learning algorithm derived from neural network theory. It can be viewed as a 3-layer feedforward network where the first layer represents input variables, the middle layer represents fuzzy rules and the output layer is the last layer. An interesting aspect of Neuro-Fuzzy approach is that the obtained model can be represented in terms of if...then rules on linguistic form. In particular, if the rules are expressed in the so-called Takagi-Sugeno form, i.e. with the consequent part expressed as a linear combination of the input mapping.

The rest of the report is divided as follows: the next section discusses the motivation behind this project. Section 3 discusses the previous works while Section 4 encapsulates our analysis onto the dataset and model interpretation. This is followed by a discussion on the result and general limitations with the conclusion.

2 Motivation

Many people are acquainted with machines such as Alphazero which prides itself in being able to play chess with users at high rates of proficiency. Success stories like these are a part of a growing global phenomenon where machines are taking leverage of the myriad of data available to compute complex models beyond the reach of an average human mind. In a nutshell, the powers of deep neural networks are being used in many fields of science and technology to perform intricate tasks. While various models have been developed using them, there still needs a lot to be done in the application of such models.

Modern Businesses churn out a lot of data. However, not all businesses have the capacity to use this data in a structured and coherent way for their own advantage. While there are several advantages and disadvantages in the present computational models proposed by various studies, there exists a lot of grey area in selection and further optimization of such models. Also, some of these studies use a wide array of assumptions which might not reflect the actual scenario where various macroeconomic factors come into the frame. Thus, the purpose of this study is to devise a more optimized working model which takes into account the major factors to forecast sales processes efficiently and accurately. The aim is to help concerning businesses to increase their profit margins and make better use of their data.

3 Related Works

Various studies have been undertaken in the field of sales forecasting. This section provides a brief summary to the related work. In the past decade, many of the research work has been centered on refining the existing models of prediction, adding subsequent layers on top. These methods exist with their own set of advantages and disadvantages.

Autoregressive Integrated Moving average (ARIMA) is a form of regression analysis that gauges the strength of one dependent variable relative to other changing variables. Many statistical methods like ARIMA and ARMA have been used for creating sales forecasting models. A 10 year data based on House sales in Turkey was trained using various ARIMA models differing in parameters by [2]. They compared these models with each other based on MSE and selected the one that gave the minimum error. Bowen et al. [3] combine BP neural network and ARIMA to forecast cargo sales for the next five day showing that the two models lack complementarity. Wen Hu et al. [4] in their study forecast commodity sales and find that the residual error obtained by ARIMA model has autocorrelation and thus propose an improvised adaptive neural network to decompose the residual time series into subsequences and input them into the network for optimization.

The Bass Model has been used by numerous researchers to forecast demand of future products based on data collected from the sales of other products. In [5], the authors extend the Bass model to predict the pattern of services like E-mail using diffusion pattern of similar communication based services. In this study, they show that the level of innovation of a certain product can be applied to other technologies as well. In a study by Lee et al. [6], the bass model was modified for the sale of Hybrid cars and Industrial robots. This study also collected web search traffic data and patent data in order to modify the Bass model itself and is shown to work better than the original model. In [7], a sentiment analysis was performed to through User-generated reviews and then converted into product Word-of-mouth index. The innovation part is borrowed from the Baidu index and finally macroeconomic indicators are combined to perform robust forecasting. The authors of the study in [8] use the sales data of 87 products to create various product and diffusion attributes. Various regression algorithms are then used to predict the parameters of the concerning Bass model. In [9], a new demand forecasting model is proposed to reflect special conditions owing to new technology shifts.

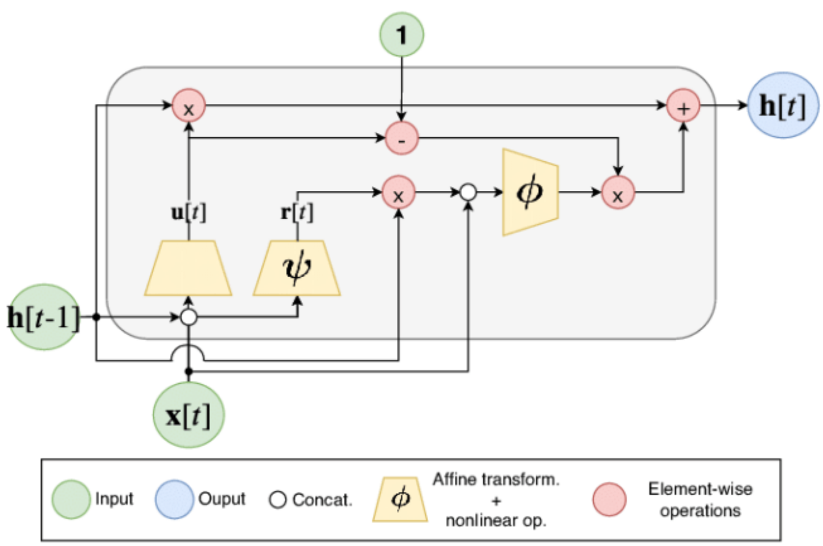

Figure 1: Diagram of a recurrent neural network [10]

The ability of Long Short-term Memory (LSTM) networks, a special type of RNNs, to forecast has been shown by various researches since more variables can be easily fed into the network to obtain more robust forecasts and can even learn the seasonality within the data. In [11], LSTM network has been proposed to obtain forecasts for the next 24 hours, 48 hours, 7 days, and 28 days. In [2], an LSTM network has also been proposed to forecast house sales data in Turkey over 10 years and the model is then combined with an ARIMA model to give better forecast results. Livieris et al. [12] use a combined forecasting method encompassing LSTM layers with convolution layers to forecast prices of Gold over a period. Convolutional and pooling layers are able to develop features of the input data while the LSTM layers exploit the general features.

Neuro-fuzzy analysis in forecasting in used frequently in several studies to obtain logical results with relative ease[13]. The monthly sales retail data by the US census bureau from 1992 to 2016 was used for prediction services using de-trended time series by Neuro-Fuzzy and Feed forward neural network by [14]. An adaptive model comprising of neuro-fuzzy inference system and Genetic algorithm was proposed by [15] to predict price of copper over varying time periods and the performance was compared with other statistical techniques such as generalized autoregressive conditional heteroscedasticity (GARCH) and Support Vector Machines (SVM). The daily price change of cryptocurrencies like Bitcoin was predicted using a hybrid Neuro-fuzzy controller, named PATSOS. This model outperforms other computational intelligence models and thus paves way for further research. Signals obtained through ultrasound devices was used to predict electrical insulator faults based on adaptive neuro-fuzzy inference system (ANFIS) by [16]. Further, three inference system were evaluated: grid partition, fuzzy c-means clustering, and subtractive clustering.

Studies like [17] show that sales forecasting can be modelled as a regression based problem instead of a time-series problem which can allow us to obtain more robust results since the data is not immune to outliers and missing information. We also need to have long historical data which may not be available in the initial phase of product launch. Studies that focus on such supervised techniques include random forests and gradient boosting machines. In [18], the data is firstly divided into clusters using K-means clustering and for each subsequent cluster, an extreme learning machine method is utilized to predict sales data. This approach has the advantage that clustering takes care of outliers in the model thus improving the performance of the model. A similar approach is used by [19] where a clustering based random forest method was proposed to forecast sales on big promotion days. This study also shows that clustering increase the performance of the model with higher accuracy. In [20], a complete framework for sales forecasting of short-shelf life food products was carried out combining the radial basis function (RBF) neural network architecture and genetic algorithm.

In some studies, several machine learning algorithms are implemented on sales data and then a conclusion is drawn. In [21], retail sales data was forecasted using several algorithms and boosting algorithms like AdaBoost. It was found that boosting algorithms perform better than traditional regression algorithms. In a similar study [22], C-XGBoost model was developed to forecast e-commerce data and comparisons were drawn amongst other boosting algorithm as well[23].

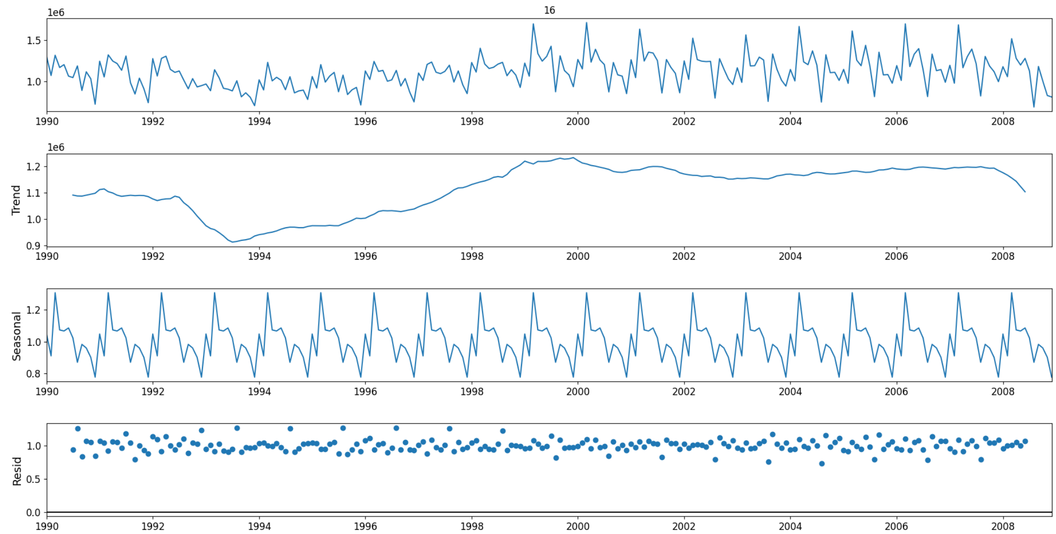

Figure 2: Seasonal-Trend decomposition of the data

4 Methods

4.1 The Data

The data used for this study is collected from the European Automobile Manufacturers Association (ACEA) [24], which publishes two press releases reporting the sales of new passenger and Commercial vehicle registration in European Union (EU) every month. The analysis is carried out on the consolidated data by country. The data contains sales of passenger vehicles sold in EU-15 countries from 1990-2008. Monthly sales of every month were considered and a time series was constructed. Individual sales of every country were also considered while making multivariate models comprising of multi-level perceptrons/inputs. The data is subsequently split into sizes of a window for the efficient training process. Various sizes of the window are experimented with leading to the best outcome.

Our initial analysis comprises time series decomposition as an abstract method for better understanding of the problem. A given time series is thought to consist of three systematic components including level, trend, seasonality, and one non-systematic component called noise. The multiplicative model is used which suggests $y(t) = Level * Trend * Seasonality * Noise$. Time series decomposition also has an advantage of modelling the trend and subtracting it from your data, or implicitly by providing enough history for an algorithm to model a trend if it may exist. Subsequent analysis shows that while there is an overall increasing trend from 1994 to 2000, the overall trend seems to be random with spikes of seasonality combined with White Gaussian Noise. In Particular, there is a spike at the starting of April which is possibly due to the start of the financial year. A high level of white Gaussian noise suggests that there will always be factor uncertainty attached to statistical analysis.

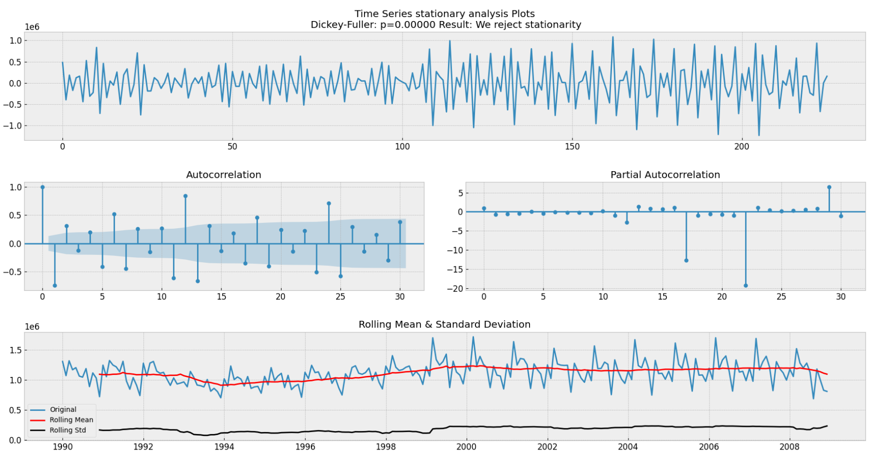

Figure 3: Auto-correlation plots of the data

Further, auto-correlation was computed to calculate the degree of similarity between the given series and a lagged version of itself over successive time series. The resulting output can range from 1 to -1, in line with the traditional correlation statistic. It is essentially useful to compute how much of an effect the past values have on the current values. The analysis shows that the series is somewhat correlated to the previous values with a coefficient > 0.25 in some lags. A sustained amount of growth in sales is thus inherent from such analysis and auto-manufacturers can make use of the traction in the market during such circumstances. The series was also made to be stationary using differencing.

To compare the performance of various models, a few major performance metrics were used. The mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) are the major metrics used to assess the prediction and forecasting abilities of the model. A lower value of all these metrics implies a better accuracy of the model. These metrics are defined as:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \tag{1}$$

$$MAE = \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{n} \tag{2}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y_i} - y_i|}{|\hat{y_i}|} \tag{3}$$

where y_i denotes the predicted value for the i^{th} value and y denotes the original value.

4.2 Model Implementations

Some models are more robust towards outliers but have worse performance than the more sensible models. Selection of the model is thus critical and depends on the use by cases basis. Our initial models were implemented using statistical packages and software and used univariate time series modelling. In the first model, Holt-Winters Exponential Smoothing (HWES) is used for forecasting time series data that exhibits both a trend and a seasonal variation. The Holt-Winters technique is made up of the following four forecasting techniques stacked one over the other: Weighted average, exponential smoothing, Holt ES and Holt-Winters ES. The data was split into a set of 60% for training and 40% for testing. The level of the time series seems to be increasing linearly, so we set the trend as an additive.

We subsequently implement Auto-Regressive models which forecast the variable of interest using a linear combination of past values of the variable. The term autoregression indicates that it is a regression of the variable against itself. The model can be written as

$$\hat{y_t} = a_1 y_{t-1} + a_2 y_{t-2} + ... + a_p y_{t-p} + c + \epsilon_t \tag{4}$$

where ϵ_t denotes the white gaussian noise. p is the parameter associated with the auto-regressive aspect of the model, which incorporates past values i.e lags of dependent variable. Autoregressive models are remarkably flexible at handling a wide range of different time series patterns. In our model, training and prediction is done 1 day ahead for all the test data. In due course, we implemented Moving Average method which relies on using past forecast errors in a regression-like model. The equation of such a model is given by:

$$\hat{y_t} = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \epsilon_{t-q} + c \tag{5}$$

where ϵ is the one used in the previous equation of autoregression. This model is referred to as a moving average model of q order. This model is different from the exponential smoothing we have implemented previously. The parameter q was found with the help of statistical packages which gives the best output in terms of q.

We implemented the ARMA model or Autoregressive Moving Average model, which is to describe weakly stationary stochastic time series in terms of two polynomials. The first of these polynomials is for autoregression, the second for the moving average. This model is referred to as $ARMA(p, q)$ model where p is

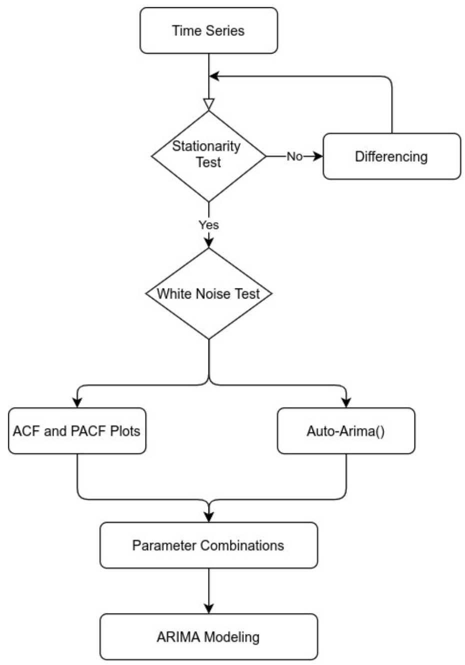

Figure 4: Parameter determination of ARIMA Model

the order of the autoregressive part, while q is the order of the moving average polynomial. This model can be seen as an ensemble of the previous two models described. Parameters selection is similar to the previous methods which are done through PACF.

$$\hat{y}_t = \sum_{i=1}^{p} \psi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t + c \tag{6}$$

Similar to the ARMA model, there is another model called ARIMA, which stands for Autoregressive Moving Integrated Average model. It involves another parameter d, a measure of how many non-seasonal differences are needed to achieve stationarity. A problem with ARIMA is that it does not support seasonal data. That is a time series with a repeating cycle. ARIMA expects data that is either not seasonal or has the seasonal component removed, e.g. seasonally adjusted via methods such as seasonal differencing.

The various shortfalls of ARIMA model lead us to eventually implement the SARIMA model, an acronym for Seasonal Autoregressive Integrated Moving Average. This model is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component. It adds three new hyperparameters to specify the autoregression, differencing and moving average for the seasonal component of the series, as well as an additional parameter for the period of the seasonality. The modelling procedure is almost the same as for non-seasonal data, except that we need to select seasonal AR and MA terms as well as the non-seasonal components of the model. AUTO-SARIMA tests various hyperparameters on its own and gives the best parameter to be used. In our case, we have used $(0, 0, 3)$ as parameters which give us the optimum performance.

We now shift our attention to Multivariate models which use the sales of individual countries to predict cumulative car sales in the whole European Union. Bayesian Ridge regression is very effective when the size of the dataset is small. In the Bayesian approach instead of maximizing the likelihood function alone, we would assume prior distributions for the parameters and use Bayes theorem. The Bayesian approach is a tried and tested approach and is very robust, mathematically. So, one can use this without having any extra prior knowledge about the dataset. We use grid search to look for the best hyperparameters of the model such as $\alpha_1, \alpha_2, \lambda_1$ and λ_2. The inference of the model can be time-consuming

We then implement the Random Forest Algorithm which is a flexible and easy to implement machine learning algorithm which doesn't need much emphasis on parameter-tuning. It is one of the most widely implemented algorithms used in several pieces of research [25]. Random forest is a supervised learning algorithm. It builds an ensemble of decision trees and can be used for both, classification and regression purposes. It is also helpful in measuring the relative importance of a certain feature and adds randomness to the model for superior prediction services to avoid overfitting of the model. As per our problem, the random forest algorithm works well during the initial testing phase but fails to capture the trend in the second half.

Taking inspiration from such tree-based model, we subsequently implemented LightGBM (Gradient Boosting Method) which is a framework based on decision trees to increases the efficiency of the model and reduces memory usage [26]. It uses two inherent and novel algorithms: Gradient-based One Side Sampling and Exclusive Feature Bundling (EFB) which fulfils the limitations of the histogram-based algorithm that is primarily used in all GBDT (Gradient Boosting Decision Tree) frameworks. LightGBM splits the tree leaf-wise as opposed to other boosting algorithms that grow tree level-wise. It chooses the leaf with maximum delta loss to grow. The LightGBM algorithm gives us a good performance on our dataset.

Another decision-tree based ensemble algorithm that is similar to LightGBM is called XGBoost which also uses a gradient boosting framework, which was developed as a part of the research project at the University of Washington. XGBoost can be applied to various problem statements including regression, classification, and other user-defined problems. It uses various system-level optimization including parallelization and tree pruning as it uses a greedy framework for the stopping criterion and uses optimum utilization of hardware available.

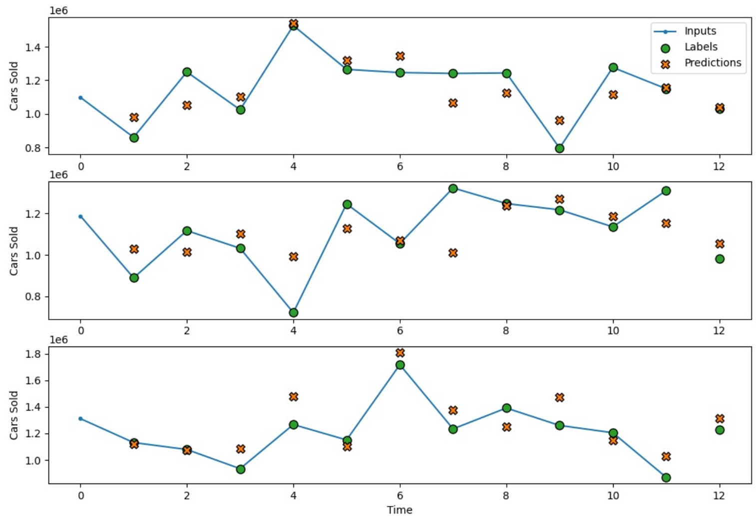

Figure 5: Output of Neural-network based model

Various studies have shown that XGBoost outperforms other state-of-the-art methods including Gradient Boosting, Random forest and Bayesian Regression. In our study too, this algorithm can outperform other methods and its emphasis on lesser use of hyperparameters make it a more suitable algorithm for prediction services.

Turning our attention to regression-based algorithms, we subsequently implement Support Vector Machine Algorithm, a set of supervised learning methods used for classification, regression and outliers detection. There are three different implementations of Support Vector Regression: SVR, NuSVR and LinearSVR. LinearSVR provides a faster implementation than SVR but only considers the linear kernel, while NuSVR implements a slightly different formulation than SVR and LinearSVR. We implement the SVR version of support vector machines with radial basis function as the kernel. We set the values of γ as 0.1 and $\epsilon = 1$.

Before implementing a Long-Short term memory-based network, we tested an artificial neural network-based network that is useful in providing robust solutions to non-linear problems and has been applied in various domains of classification and regression types of problems. We implemented a Neural network with 3 layers and used ReLU and sigmoid activation functions. We used hyperparameter tuning to select the various parameter used for training and used 500 epochs for training with a batch size of 32. The weights of the neural network were trained using backpropagation and dropout is used to avoid over-fitting by the network. As the neural network is devoid of any memory, it fails to carry out any causation-based model dependencies but is still able to predict the outcome given a time window which is 12 in our case. Finally, we used three optimation based techniques, ADAM, RMSPROP, and adaGRAD and found that ADAM performs the best in terms of the output.

Since an LSTM network can calculate long term based dependencies, it is an obvious model of choice in various machine learning-based solutions. We implemented two LSTM-based models, one that predicts 1-month sales, while the other model predicts 6-month sales. We used the sequence length of 12 and used *tanh* activation function for both the networks. A dropout value was set to 0.2 in the first model, while it was set to 0.3 in the second model. We used 150 epochs to train the model and used Adam optimization technique, similar to the implementation of our neural network. All of these hyperparameters were found using grid search which gave us the least error on the training model.

In our last implementation, we experimented with ensemble methods, combining various models to give better outcomes. This is done to get the best characteristics of every model. We used an ensemble method testing of LightGBM and XGBoost. We calculate the correlation of these methods and get a value of 0.78 which shows that these two models are quite similar and it would thus make sense to make an ensemble method combining them. We use a weighted average mean of these two methods and initially tested using a 0.4 : 0.6 ratio since the XGBoost method gave us partially better results. We, however, found that using the mean of predictions gave us much better outcomes.

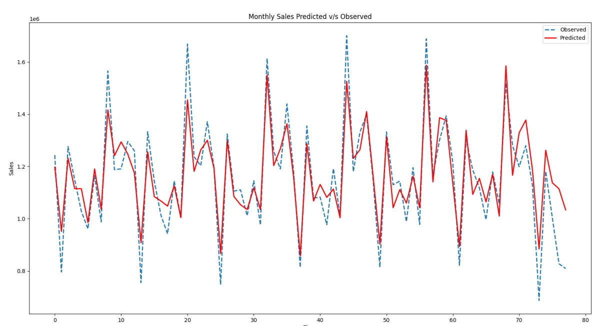

Figure 6: Using AutoARIMA on our test set gave us accurate predictions

5 Results and Discussion

In terms of the statistical methods, SARIMAX and auto aroma gave us the best outcome in terms of error. We obtained a Mean Absolute Percentage Error

(MAPE) of 0.3925 during testing the autoARIMA model. This can also be attributed to fine parameter tuning. The model is especially useful to understand how statistical methods can give good outcomes although being much simpler than machine learning algorithms.

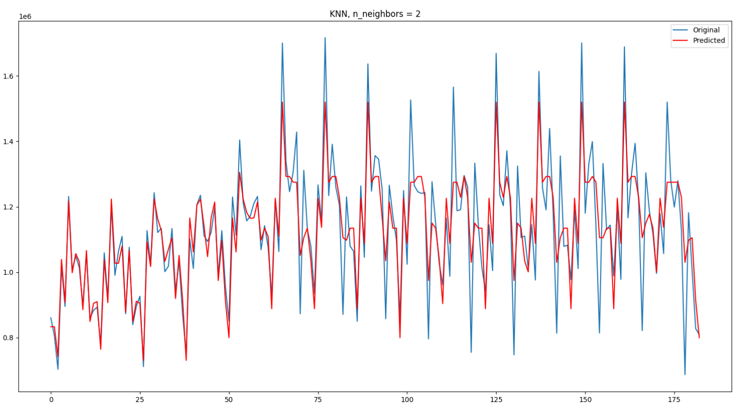

Figure 7: K-Nearest Neighbor model prediction outcome

K-Nearest neighbour algorithm is one of the most influential, yet simple algorithms in machine learning which can give us robust results with little overheads. KNN has been used in statistical estimation and pattern recognition already since the beginning of the 1970s as a non-parametric technique. The results obtained from using this algorithm suggest that it gives us a good outcome compared to its relative simplicity. A mean absolute percentage error of 0.6510 was obtained by training the algorithm. We had also used hyperparameter optimization and restricted the value of n to only 2 to get the most optimum results.

The artificial neural network-based model gave us accurate predictions for a one-time step ahead and was crucial to our understanding and use of LSTM networks. We had implemented two LSTM models, one which gave predictions for 1-month sales while the other predicted sales for the next 6-month time gap. The first LSTM network comprised of a MAPE of 0.2909 which shows its superiority over statistical methods. It also is much more computationally expensive and more robust giving better predictions.

The ensemble-based method comprising of LightGBM and XGBoost performed exceptionally well and was easier to train and deploy as well. The ensemble method gave us a Mean absolute percentage error of 0.1492 which signifies the importance and characteristics of using such ensemble-based methods. This ensemble-based method is also robust to outliers and can predict the trend during the testing phase. It also performed better than the LSTM based networks which predicted for 1 month and 6 months in advance.

6 Limitations and Relevance

The prediction services that are offered through any algorithm or software application are directly proportional to the type, quantity and quality of the dataset on which analysis is carried out. In our case, we restricted our analysis on monthly car sales of European Union from 1990 to 2008. While we carried out extensive testing and method testing on our dataset, more granularity of the dataset would have prompted better and more credible research outcomes in this direction. Individual sales across cities and regions would have spearheaded lesser error in the various multivariate models that we built. Additionally, weekly sales data would have provided more insight into the data and would allow car manufacturers to focus on advertising around a specific time thus leading to better margins and profits.

We tested and carried out analysis using various algorithms that have been used for various machine learning tasks. However, we were unable to test and implement some other methods which involve qualitative analysis to show how machine learning and mathematical analysis perform concerning these methods.

We also performed hyperparameter tuning during the testing of all algorithms where it could have been used. We used already available machine learning packages to perform this task. We were able to improve the performance of various algorithms by as much as 30% in some cases. However, extensive hyperparameter tuning would have certainly improved results which require a high level of computation power. Using pretrained models and transfer learning is an optimum way of training in computer vision models. However, we trained all our models on our processor locally and therefore, training time was time-consuming.

7 Conclusion

Business decisions such as production, sales management, planning, and inventory highly depend on the sales forecast which can highly improve any company's business prospect and propel it further. The automobile industry is a major part of the economy of many countries and faces uneven demand throughout the year.

We studied and implemented various statistical and machine learning algorithms to improve our current forecasting abilities. We used high-end methods and used sophisticated hyperparameter optimization to train our models on the dataset. We also showed that sales forecasting can be vastly improved by using good parameter optimization techniques. In some cases, the performance increase is manifold due to various optimizations in the running algorithm. We found that machine learning algorithms can provide much better and robust prediction outcomes as they were able to calculate long-term dependencies. Empirical results also show that the ensemble-based method comprising of LightGBM and XGBoost gave us the model with the minimum error. Other machine learning models were too complex and overfitting could not be avoided in such case partially due to lack of much data. Further scope of the research includes better feature extraction models and suitable deployment of machine learning and other algorithms for better prediction services.

References

[1] Chitu Okoli and Suzanne D Pawlowski. The delphi method as a research tool: an example, design considerations and applications. *Information & management*, 42(1):15–29, 2004.

[2] Ayşe Soy Temür, Melek Akgün, and Günay Temür. Predicting housing sales in turkey using arima, lstm and hybrid models. *Journal of Business Economics and Management*, 20(5):920–938, 2019.

[3] Tan Bowen, Zhang Zhe, and Zhang Yulin. Forecasting method of e-commerce cargo sales based on arima-bp model. In *2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, pages 133–136. IEEE, 2020.

[4] Wen Hu and Xinchao Zhang. Commodity sales forecast based on arima model residual optimization. In *2020 5th International Conference on Communication, Image and Signal Processing (CCISP)*, pages 229–233. IEEE, 2020.

[5] David Wong, Kenneth B Yap, Bradley Turner, and Nexhmi Rexha. Predicting the diffusion pattern of internet-based communication applications using bass model parameter estimates for email. *Journal of Internet Business*, (9):27–50, 2011.

[6] Won Sang Lee, Hyo Shin Choi, and So Young Sohn. Forecasting new product diffusion using both patent citation and web search traffic. *PloS one*, 13(4):e0194723, 2018.

[7] Chuan Zhang, Yu-Xin Tian, and Ling-Wei Fan. Improving the bass model's predictive power through online reviews, search traffic and macroeconomic data. *Annals of Operations Research*, 295(2):881–922, 2020.

[8] Hakyeon Lee, Sang Gook Kim, Hyun-woo Park, and Pilsung Kang. Prelaunch new product demand forecasting using the bass model: A statistical

and machine learning-based approach. *Technological Forecasting and Social Change*, 86:49–64, 2014.

[9] Yoon Seong Kim, Eun Jin Han, and So Young Sohn. Demand forecasting for heavy-duty diesel engines considering emission regulations. *Sustainability*, 9(2):166, 2017.

[10] Recurrent Network Tutorial, https://petamind.com/recurrent-neural-network-time-series-data-part-1/.

[11] Shahzad Muzaffar and Afshin Afshari. Short-term load forecasts using lstm networks. *Energy Procedia*, 158:2922–2927, 2019.

[12] Ioannis E Livieris, Emmanuel Pintelas, and Panagiotis Pintelas. A cnn–lstm model for gold price time-series forecasting. *Neural computing and applications*, 32(23):17351–17360, 2020.

[13] George S Atsalakis, Ioanna G Atsalaki, Fotios Pasiouras, and Constantin Zopounidis. Bitcoin price forecasting with neuro-fuzzy techniques. *European Journal of Operational Research*, 276(2):770–780, 2019.

[14] Giuseppe Nunnari and Valeria Nunnari. Forecasting monthly sales retail time series: A case study. In *2017 IEEE 19th Conference on Business Informatics (CBI)*, volume 1, pages 1–6. IEEE, 2017.

[15] Zakaria Alameer, Mohamed Abd Elaziz, Ahmed A Ewees, Haiwang Ye, and Zhang Jianhua. Forecasting copper prices using hybrid adaptive neuro-fuzzy inference system and genetic algorithms. *Natural Resources Research*, 28(4):1385–1401, 2019.

[16] Stéfano Frizzo Stefenon, Roberto Zanetti Freire, Leandro dos Santos Coelho, Luiz Henrique Meyer, Rafael Bartnik Grebogi, William Gouvêa Buratto, and Ademir Nied. Electrical insulator fault forecasting based on a wavelet neuro-fuzzy system. *Energies*, 13(2):484, 2020.

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

[18] Chi-Jie Lu and Ling-Jing Kao. A clustering-based sales forecasting scheme by using extreme learning machine and ensembling linkage methods with applications to computer server. *Engineering Applications of Artificial Intelligence*, 55:231–238, 2016.

[19] Hongyan Dai, Haoyang Yu, Qing Xiao, and Weihua Zhou. A clustering-based sales forecast method for big promotion days in o2o on-demand retailing. In *2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pages 1079–1083. IEEE, 2019.

[20] Philip Doganis, Alex Alexandridis, Panagiotis Patrinos, and Haralambos Sarimveis. Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. *Journal of Food Engineering*, 75(2):196–204, 2006.

[21] Akshay Krishna, V Akhilesh, Animikh Aich, and Chetana Hegde. Sales-forecasting of retail stores using machine learning techniques. In *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, pages 160–166. IEEE, 2018.

[22] Shouwen Ji, Xiaojing Wang, Wenpeng Zhao, and Dong Guo. An application of a three-stage xgboost-based model to sales forecasting of a cross-border e-commerce enterprise. *Mathematical Problems in Engineering*, 2019, 2019.

[23] FL Chen and TY Ou. Sales forecasting system based on gray extreme learning machine with taguchi method in retail industry. *Expert Systems with Applications*, 38(3):1336–1345, 2011.

[24] ACEA, https://www.acea.be/statistics/tag/category/consolidated-figures.

[25] Tim F Cootes, Mircea C Ionita, Claudia Lindner, and Patrick Sauer. Robust and accurate shape model fitting using random forest regression voting. In *European Conference on Computer Vision*, pages 278–291. Springer, 2012.

[26] Cheng Chen, Qingmei Zhang, Qin Ma, and Bin Yu. Lightgbm-ppi: Predicting protein-protein interactions through lightgbm with multi-information fusion. *Chemometrics and Intelligent Laboratory Systems*, 191:54–64, 2019.