Predictive Analysis of Retail Sales Forecasting using Machine Learning Techniques

Muhammad Sajawal[1], Sardar Usman[2] , Hamed Sanad Alshaikh[3], Asad Hayat[4] and M. Usman Ashraf[5]

[1] Department of Computer Science & IT Lahore Leads University, Lahore, Pakistan,
[2] Department of Computer science, Grand Asian University Sialkot, Pakistan.
[3] College of Telecommunications and Electronic Jeddah Saudi Arabia
[4] Department of Computer Science Leads university Lahore, Pakistan.
[5] Department of Computer Science, GC Women University, Sialkot, Pakistan.

Abstract

In a retail industry, sales forecasting is an important part related to supply chain management and operations between the retailer and manufacturers. The abundant growth of the digital data has minimized the traditional system and approaches to do a specific task. Sales forecasting is the most challenging task for the inventory management, marketing, customer service and Business financial planning for the retail industry. In this paper we performed predictive analysis of retail sales of Citadel POS dataset, using different machine learning techniques. We implemented different regression (Linear regression, Random Forest Regression, Gradient Boosting Regression) and time series models (ARIMA LSTM), models for sale forecasting, and provided detailed predictive analysis and evaluation. The dataset used in this research work is obtained from Citadel POS (Point Of Sale) from 2013 to 2018 that is a cloud base application and facilitates retail store to carryout transactions, manage inventories, customers, vendors, view reports, manage sales, and tender data locally. The results show that Xgboost outperformed time series and other regression models and achieved best performance with MAE of 0.516 and RMSE of 0.63.

Keywords:

Machine learning, Time Series, Sales Forecasting, Regression, Gradient Boosting, LSTM, ARIMA, Random Forest.

1. INTRODUCTION

Sales forecasting is the most challenging task for the inventory management, marketing, customer service and Business financial planning for the information technology chain store. To develop sales forecasting accurate model, it is a very difficult task due to multiple reasons like over-forecasting model that increases operation cost and generates unnecessary products and under forecasting model lose customer satisfaction and its sales opportunities (Ofoegbu, 2021). Accurate and robust sales forecasting results can lead to customer satisfaction, enhanced channel relationships, and significant monetary savings.

There are different Back Propagation Neural Network (BPN) techniques for sales forecasting due to their ability to capture functional relation among the empirical data but there is difficult to control large parameter and has risk of model over-fitting. Support vector regression (SVR) algorithm has been used for solving the nonlinear regression estimation problem. So, prediction result of SVR is better than the BPN due to capability in obtaining a unique solution among the empirical data. SVR has been mostly used for the time series prediction such as traffic flow prediction, financial time series forecasting and wind speed prediction. But SVR cannot show accurate results when many potential independent variables are considered.

To overcome the problem, Multivariate Adaptive regression splines (MARS) is a suitable methodology for modeling complex nonlinear and non-parameter regression problem. MARS has much power for building model that has huge dataset like electricity price forecasting, credit scoring and network intrusion detection (Deo et al., 2017).

Sales forecasting is important for enterprises to make business plans and gain competitive advantages. There are different time series methods contributing to field of sales forecasting but they only deal for the traditional linear data and ignoring nonlinear data(Álvarez-Díaz et al., 2018). So, to overcome this traditional method many researchers are using soft computing skills for solving the non-linear data problem like fuzzy neural network, fuzzy logic, neural network, and evolutionary algorithm etc. for robust sales forecasting.

Different types of sales forecasting algorithm and statistical models have been generated to solve to problems like, ARIMA model that do forecast within few seconds based on hundreds of historical data points(Shumway and Stoffer, 2017). But these models are unable to process when the complex data patterns are given to that model for sales forecasting. Although ANN based algorithms can solve this problem but when we consider improvement in predication accuracy then these models take time in completing simple sales forecasting. ELM has ability to learn much faster with a higher performance than the traditional gradient-based learning algorithms, but it also reduces many difficulties faced by gradient-based learning methods such as learning rate, stopping criteria, the over-tuned problems, learning epochs, and local minima as well as the ELM model minimize learning time of ANN quickly and ELM is being used in real-time applications such as real-time controlling system(Lu and Kao, 2016).

Sales forecasting is an important part related to supply chain management and operations between the retailer and manufacturers. Manufacturer needs to predict the actual future demand to inform production planning. Similarly, retailers need to predict sales for purchasing decision and minimize the capital costs. So, it depends upon the end users.

Therefore, depending upon the nature of the business, sales forecasting can be done through human planning and statistical model or by combining both methods. In this paper authors used the Partial Recurrent Neural Networks (PRNN) statistical model for sales forecasting. The proposed methodology can extract the pattern from the past sales and facilitates future sales forecasting.

The aim of this research work is to investigate the various sales forecasting methods executed in financial area and evaluate the performance of the chosen machine learning algorithms to find out the best suitable and efficient model for the chosen data set. We have used machine learning based regression models (Linear regression, random forest and Xtreme Gradient boosting) and time series models (LSTM, ARIMA) for sale forecasting using Citadel POS data set. Results showed that Xtreme Gradient boosting out performed both time series models and other regression techniques.

2. Literature Review

2.1. Background

Supply chain contains different business parties that share physical good and customer services related to goods and money. Supply chain to be developed in two different areas: Supply chain execution and supply chain planning.

Forecasting Concept

Forecasts are nothing but predictions about future. Maybe forecasts of sunrise and sunset can be predictable without any mistake but it is not the scenario in business. Business equations changes as time goes and hence prediction may give error. (Mentzer and Moon, 2004)describes sales forecast as a projection into future of expected demand, given a started set of environmental conditions. We should not confuse the planning process and forecasting process. Planning is nothing but managerial actions which should be taken to meet or exceed the sales forecast. The aim of right forecast is to predict demand perfectly. Forecasting have been used in all kinds of companies, service sectors, government organizations, andused as input to the planning project or set of activities.

(Hofmann, 2013)summarizes the characteristics of sales forecast as follows:

- Forecasts are always wrong and hence one should always expect evaluation of errors in it.
- Long term forecast is normally less accurate than short time forecasts. This is because larger standard deviation of error relative to mean than short term forecasts.
- Aggregate forecasts are normally more accurate than disaggregate forecasts. Aggregate forecast contains smaller standard deviation of error than disaggregate forecasts.
- The Greater the distortions of information in supply chain the higher are the errors in sales forecast.

Sales forecasting need in Planning

Manufacturing industries work on principle to satisfy customer demand by appropriate supply. According to(Mentzer and Moon, 2004), companies consider the sales forecasting as integral part of this process. End customers create demand, and it can be increased by activities like promotions. Hence marketing focus on end customers for creating demand. Sales department ease the same by different strategies such as servicing other parties in this streamline like wholesaler and retailers. Supply should be enough to meet demand. Different management functions like manufacturing, purchasing and logistics work together to maintain the supply.

Forecasting Methods and Techniques

Several standardized methods for forecasting are available. They differ in terms of the relative performance in forecasting over the level of quantitative sophistication used, and the logic base (historical data, expert opinion, or surveys) from which the forecast is derived. Those methods could be categorized into three different groups: historical projection, qualitative, and casual(Ballon, 2004).

(Ballon, 2004) states that "when a reasonable amount of historical data is available and the trend and seasonal variations in the time series are stable and well defined, projecting these data into the future can be an effective way of forecasting for the short term". He also mentions that the quantitative nature of the time series supports the use of mathematical and statistical models as primary forecasting tool. By using such tools accuracy can be reached for forecasted periods. These methods are most appropriate when the environmental situation is stable, and the basic demand pattern does not vary significantly from year to year.

According to (Mentzer and Moon, 2004), it is not possible to forecast every product with the same time series technique and that is why we need different time series technique for eachproduct. He also points out that there are many techniques available in the general category of time series analysis. Time series techniques have common characteristics and endogenous techniques. It means that time series technique looks at the patterns of the history of actual sales. These patterns can be identified and projected to derive forecast. Time series techniques looks only on patterns that are the parts of the actual history. Despite which time series technique used they all are examined by four basic time series patterns: level, trend, seasonality, and noise.

Machine Learning Techniques

There are three main machine learning algorithms i.e., Supervised, Unsupervised, and Reinforcement Learning.

In supervised learning, we are given a labeled data set (labeled training data) and desired outcome is already known, where every pair of training data has relationship. Supervised learning is where you have input variables (x) and an output variable (Y), and you use an algorithm to learn the mapping function from the input to the output.Random forest, linear regression, and long short-termmemoryare supervised machinelearning techniques(Pavlyshenko, 2019).

In the unsupervised machine learning approach, the model is trained by using unlabeled or non-classified data objects. The unsupervised learning approach is difficult fromsupervised learning because in this technique neither trained a model or machine by using training dataset. Two main types of unsupervised machine learningare Association Rule Mining and Clustering (Hussain et al., 2018).

Association Rule Mining

In this unsupervised technique, Association rule mining is a technique to identify underlying relations between different items. Take an example of a Super Market where customers can buy variety of items. For instance, mothers with babies buy baby products such as milk and diapers. In short, transactions involve a pattern(Kaur and Kang, 2016).

Clustering

Clustering is the task of dividing the population or data points into several groups such that data points in the same groups are more like other data points in the same group than those in other groups. Simply, clustering is to segregate groups with similar traits(Sinaga and Yang, 2020).

2.2. Related Work

(Catal et al., 2019) predicted the actual sales accurately by using different machine learning algorithm like linear regression, Random Forest Regression, and time series techniques like ARIMA, Seasonal Arima, Non-Seasonal Arima as well as Seasonal ETS. They used the Walmart's public online sales data to predict the sales by using different regression algorithms in Azure Machine Learning (ML) Studio as well as several time series analysis methods were implemented using R packages through R programming language manually. They selected the best model to predict the sales and made web service of that model and deployed on the Azure Cloud platform. Azure sent the output in the form of JSON format. Author after the experimental results, identified the best method which was that the regression techniques provide better performance as compared to the time series analysis approaches. To overcome the problem Multivariate Adaptive regression splines (MARS) is a suitable methodology for modelling complex nonlinear and non-parameter regression problem. MARS has much power for building model which has huge dataset like electricity price forecasting, credit scoring and network intrusion detection.

(Lu, 2014) proposed the hybrid two stage model using MARS and SVR by focusing on above mentioned drawbacks and for the sales prediction accurately. To evaluate the performance of the proposed hybrid sales forecasting procedure, three IT product sales data, i.e.,notebook (NB), LCD monitor and mother board (MB), collected from an IT chain store in Taiwan are used as the illustrative examples.

(Omar and Liu, 2012) proposed a model on based Back Propagation Neural Network (BPNN) to improve sales forecasting by using popularity information in magazines through Google Search engine. According to Author view, popular content in magazine can boost the sales. In the proposed model he used popular celebrity words as keywords to interact user for sales forecasting. They used some tools to estimate the popularity of words like Digg allows user to submit links in news. They used the nonlinear historical data to evaluate the forecasting performance that our proposed model can improve the sales forecasting. They used the Chinese publication magazines data.

Recently, (Feng et al., 2009) proposed a new learning algorithm called Extreme Learning Machine (ELM) for single-hidden-layer feed forward neural networks (SLFNs) which randomly chooses hidden nodes and analytically determines the output weights of SLFN. Author predicted the book sales using the ELM by combining with the statistical method for the popular e-commerce company in China.

(Müller-Navarra et al., 2015)used the statistical model which is called Partial Recurrent Neural Networks (PRNN) for sales forecasting. They proposed the methodologies that can extract the pattern from the past sales and facilities future sales forecasting. They used it as a tool for business planning and after that they performed an empirical benchmark that was the prevailing approach in forecasting. Real-world sales series show non-linear pattern due to different types of reasons like trend, seasonality, or introduction of new product model. That's why PRNN has the capability to handle nonlinear and well-suited model to solve sales forecasting problems.

(Holt, 2004)used the Exponentially Weighted Moving Averages (EWMA) model to measure the seasonal impact on sales trend. They combined two feature cluster related queries algorithm and seasonal time series sales behavior. They compared four models query feature only, seasonal feature without EWMA model, seasonal feature with EWMA model, proposed model seasonal feature with EWMA combined with query feature and showed best performance by developing proposed model.

3. Methodology

First to find the results we learned current existing relevant research. These literature review results are being used as input to our analysis of retails sales using machine learning techniques. Our main goal in this research work is to evaluate the performance of machine learning models like linear regression, Random Forest regression, and Xtreme Boosting Regression on the sales data from point of sale. The figure 4.1 shows the complete methodology of the proposed solution.

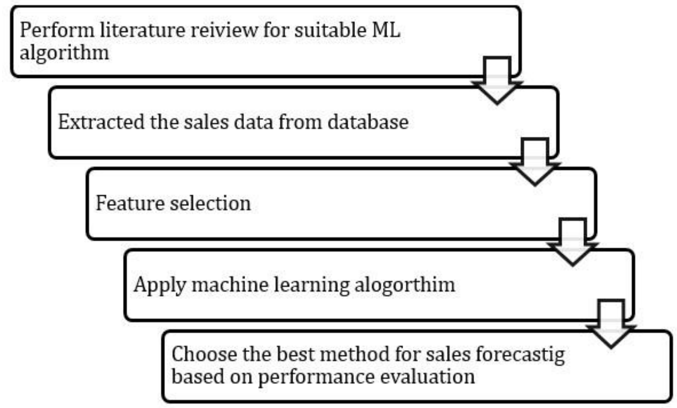

Figure 4.1: Proposed Methodology for Sales Forecasting

This research work is performed using python programming language and multiple libraries like pandas, Numpy, matplotlib, seaborn, and sklearn.

3.1. Dataset description

In this paper, we presented a methodology which is implemented for a retail Point of Sale system in a test set of |S| =32 locations in early 2007.We have citadel point of sales system that has all the records related to sales. We collected the data from the different table using SQL queries. There are multiple stores that contain the different items for sales. Each store has five stations. We took the one customer history data. This customer has 228 invoices. Each invoice contains average five items. We collected the data from 2013 to 2018 data and performed the testing on 2020 data. Train data contains item id, store number, total sales items, and total sales of each item. The training data set contains total 87847 rows.

3.2. Data Pre-processing

Several standardized methods for forecasting are available. They differ in terms of the relative performance in forecasting over the level of quantitative sophistication used, and the logic base from which the forecast is derived. We converted the data into days, week, year then we check the outlier, all missing or null values and removed all the outlier and missing values. We refined the dataset to perform testing. Those methods could be categorized into three different groups: historical projection, qualitative, and casual (Ballon, 2004).

Augmented Dickey-Fuller Test

This is basically the statistical test which is used to test the null hypothesis that a unit root is present in an autoregressive model.

The null hypothesis of the test is that the time series can be represented by a unit root, that it is not stationary (has some time-dependent structure). The alternate hypothesis is that the time series is Stationary Glynn et al. (2007).

Null Hypothesis (H0): If failed to be rejected, it suggests the time series has a unit root, meaning it is non-stationary. It has some time-dependent structure.

Alternate Hypothesis (H1): The null hypothesis is rejected; it suggests the time series does not have a unit root, meaning it is stationary. It does not have a time-dependent structure.

The result can be explained as follows:

P-value > 0.05: Fail to reject the null hypothesis (H0), the data has a unit root and is nonstationary.

P-value <= 0.05: Reject the null hypothesis (H0), the data does not have a unit root and is stationary.

3.3. Feature Selection

As there are many factors that play important role for machine learning success. Feature selection is very important factor which have huge influence on machine learning model performance. It helps from over fitting by removing data redundancy, reduces the training time and improves the accuracy of the model. There are different approaches which we used to overcome these problems like correlation method. Feature set having negative co relations with target variables have been removed during feature selection process.

3.4. Implementation

We implemented the following model to see our results and compared the performance of these models:

- Linear Regression Model
- ARIMA
- Random Forest Regression
- LSTM model
- Gradient Boosting Regression

Two metrics mean absolute error and root mean square error are evaluated for each machine learning regression model.

Mean Absolute Error is a standard measure of forecast error in the time series analysis.MAE is one of the many metrics for evaluating the performance of the machine learning model. The mean absolute is a quantity used to measure how close forecasts or prediction are to the eventual outcomes. As the name suggests, the mean absolute error is an average of the absolute errors (Chai and Draxler, 2014). Lower the error implies greater the accuracy of the model.

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

Where yi represents actual values and ˆ*yi* represents the forecasted values.

Root Mean Square Error (RMSE) is the square root of the mean square error. It is the root of the average of squared differences between prediction and observation. Lower the error implies greater the accuracy of the model(Chai and Draxler, 2014).

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

Where yi represents actual values and ˆyi represents the forecasted values.

4. Results and Discussion

To find the result, we must have to check stationary and non-stationary time series. Non stationary data is not predictable and cannot be modeled or forecasted due to change in mean, variance, and co-relation. So, we must convert it into stationary time series for reliable results.

4.1. Citadel POS Dataset

The Citadel POS is basically a point of sales system which is working in US. There are 32 different locations which have different items for sale each item has different prices. There are two types of customers who visit in these stores i.e. loyalty customers and non-loyalty customers. Loyalty customers are regular customers with frequent shopping, but non-loyalty are not regular customers with occasional visits. Fig 5.1 shows abstract store wise data containing total invoice, sales tax, Grand total including sales and sales without including tax.

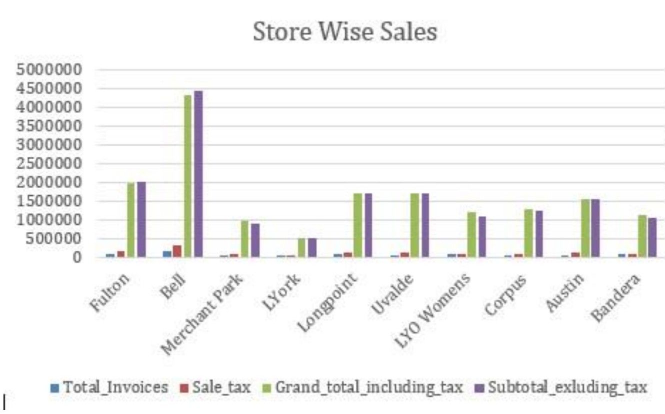

Figure 5.1: Store Wise Sales

Fig 5.1 represents the overall 2019 store wise sales data. Blue lines show the total invoices of each store and red line represents the sales tax on each invoice, green line show the total sales on each store including tax and dark blue line represents the sales without including tax. Figure shows that Bell store has highest sales for the year 2019.

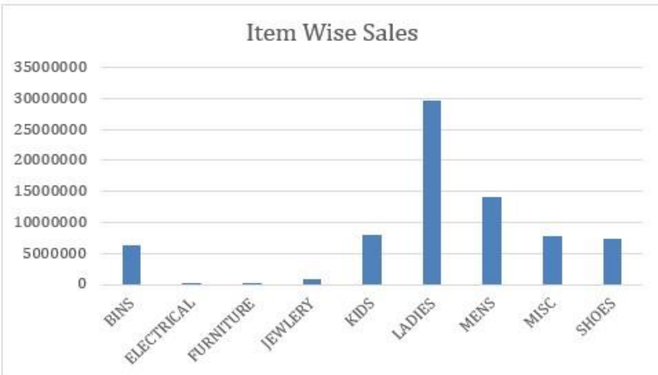

Figure 5.2 Item Wise Sales

Figure 5.2 represents item wise sales of different stores. We took sales data from 2013 to 2018 containing Ladies, Men, Shoes, Misc., Kids, Jewelry, Furniture, Electrical and Bins. Dataset contains 27343 rows having attributes item id, item name, total number of items sales, and total sales from each item.

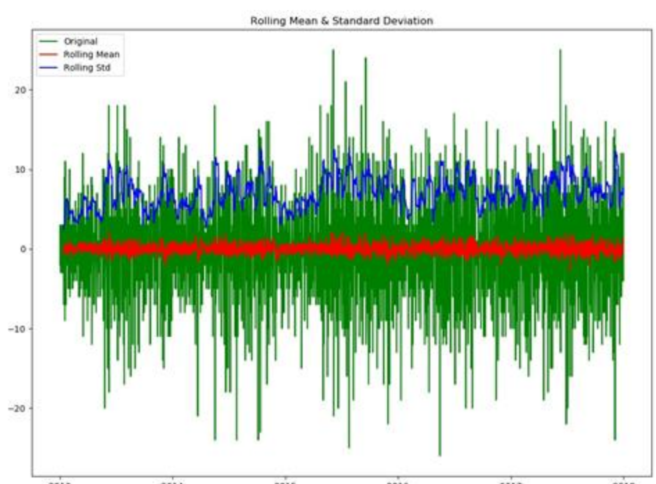

Figure 5.3 Stationary Data Checkpoint Using Rolling Mean and Rolling STD

Fig 5.3 shows the rolling and standard deviations of sales data. Dickey fuller test also used to check stationary data. By visualizing the data, we can check that data is stationary or not. Stationary data means in which mean value continuously increase with time. If p value is less than the significance level which is 5%, or if test static value is greater than the critical value then our data would be stationary, here our p value is 5.70503.

4.2. Predictive Analysis

Linear Regression

Linear regression is basically machine learning algorithm based on supervised learning techniques. It is used to perform regression task; it is used to predict the dependent variable (y) based on the independent value (x).

y=m*x +c

First, we got the dataset of retail sales from 2013 to 2018 dataset to perform prediction task. We trained our model with sales data. We performed some preprocessing task on dataset. After training the dataset, we made sure that using small data set its working fine then we performed task on large dataset.

Index	**Score**
RMSE	0.96849
MAE	0.82136

Table 5.1: Model Performance Using Linear Regression

Table 5.1 represents the Root Mean Squared Error and Mean Absolute Error acquired by Linear Regression on validation test. In this table RMSE is the standard deviation of the prediction error, which means measure of how far data points from the regression line which value is **0.96849** and MAE measures the average magnitude of the error without considering the directions between the actual and prediction observation which value is**0.82136.**

Figure 5.4: Linear Regression Sales Forecasting

Figure 5.4represents the actual and forecasted sales of target variant obtained using the linear regression model where blue line represents the actual sales value and red line represents forecasted sales of the targeted variant.

4.3. ARIMA Model

This model is used to forecast the sales basically it's the statistical method for time series sales. There are following ARIMA model parameters:

P: Trend auto regression order. D: Trend difference order.

Q: Trend moving average order

There is other four differential seasonal elements which is not the part of ARIMA model it can be handled using SARIMA model like SARIMA (p, d, q) (P, D, Q) m.

Index	**Score**
RMSE	1.04959
MAE	1.01265

Table 5.2: ARIMA Model Result

Table 6.2 represents the Root Mean Squared Error and Mean Absolute Error acquired by ARIMA model on validation test. In this table RMSE is the standard deviation of the prediction error, which means measure of how far data points from the regression line which value is **1.04959** and MAE measures the average magnitude of the error without considering the directions between the actual and prediction observation which value is **1.01265**

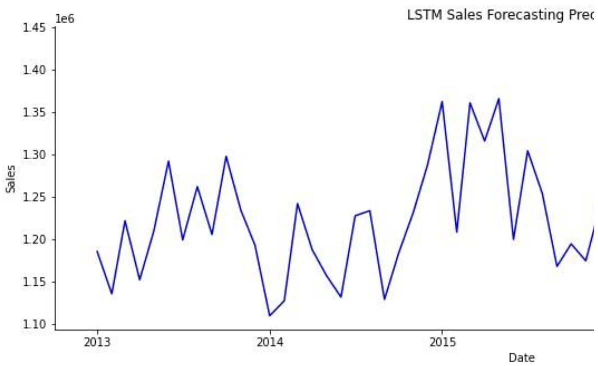

Figure 5.5: Sales Forecasting Using ARIMA model

Figure 5.5 represents the actual and forecasted sales of target variant obtained using the ARIMA regression model where blue line represents the actual sales value and red line represents forecasted sales of the targeted variant.

4.4. LSTM Model

There are multiple sequence predictions problems have been for a long time, due to these types of problems it's very difficult to solve the time series problem. For example, predicting sales to find pattern in stock market's data. To handle the sequence problem in dataset LSTM model has been applied to predict the sales. It is used to predict the sales based on previous history dataset of the retail sales.

Index	**Score**
RMSE	0.99964
MAE	0.81910

Table 5.3: LSTM Model Performance Results

Table 5.3 represents the Root Mean Squared Error and Mean Absolute Error acquired by Long Short-Term Memory (LSTM) Regression on validation test. In this table RMSE is the standard deviation of the prediction error, which means measure of how far data points from the regression line which value is **0.99964** and MAE measures the average magnitude of the error without considering the directions between the actual and prediction observation which value is **0.81910.**

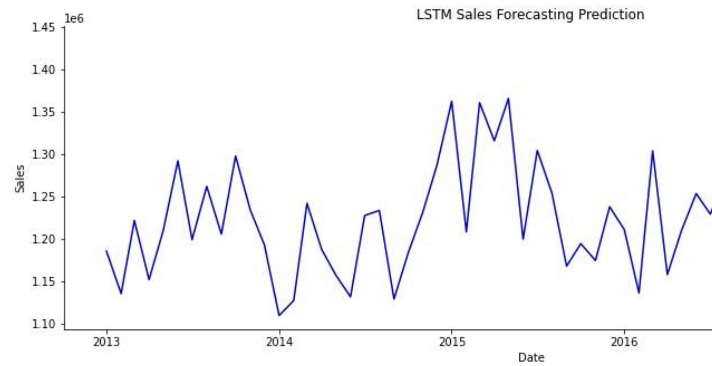

Figure 5.6: LSTM Forecasting Performance

Figure 5.6 represents the actual and forecasted sales of target variant obtained using theLSTM regression model where blue line represents the actual sales value and red line represents forecasted sales of the targeted variant.

4.5. Random Forest Regression

We used to Random Forest regression model to improve our results. It is used to improve the computation power. Random forest is a supervised machine-learning technique, in which a decision-tree mechanism used when training the model. Build multiple models (decision trees) using random training dataset with replacement and then compute the accuracy of each model. And increase the weight of the model that has maximum accuracy.

Index	Score
RMSE	0.69460
MAE	0.59121

Table 5.4: Random Forest Performance Results

Table 5.4 represents the Root Mean Squared Error and Mean Absolute Error acquired by Random Forest Regression on validation test. In this table RMSE is the standard deviation of the prediction error, which means measure of how far data points from the regression line which value is **0.69460** and MAE measures the average magnitude of the error without considering the directions between the actual and prediction observation which value is **0.59121**

Figure 5.7: Random Forest Regression Sales Forecasting

Figure 5.7 represents the actual and forecasted sales of target variant obtained using the Random Forest regression model where blue line represents the actual sales value and red line represents forecasted sales of the targeted variant.

4.6. Extreme Gradient Boosting Regression

Three concepts are involved in Xgboost algorithm: extreme, gradient, and boosting. Starting from basics boosting is one of the systematic ensemble methods aims to converting weak learners (regression trees in this case as this is a tree based Xgboost model; there is also a linear type) into stronger learners to obtain more accurate predictions. SMAPE error score is **10.14 %**

Index	Score
RMSE	0.63010
MAE	0.51599

Table 5.5: Xgboost Model Performance Results

Table 5.5 represents the Root Mean Squared Error and Mean Absolute Error acquired by Gradient Boosting Regression on validation test. In this table RMSE is the standard deviation of the prediction error, which means measure of how far data points from the regression line which value is **0.63010** and MAE measures the average magnitude of the error without considering the directions between the actual and prediction observation which value is **0.51599**

Figure 5.8:Xgboost Model Sales Forecasting

Figure 5.8 represents the actual and forecasted sales of target variant obtained using the Gradient Boosting regression model where blue line represents the actual sales value and red line represents forecasted sales of the targeted variant.

4.7. Performance evaluation and comparison results

We applied different machine learning algorithm. We have implemented different machine learning algorithm on dataset of retail sales. We implemented two evaluation Root Mean Squared Error and Mean Absolute Error to check performance of our different machine learning model. When we compared all different model then we concluded that Xgboost is the best suitable model for our dataset of retail sales based on performance evaluation of all the models.

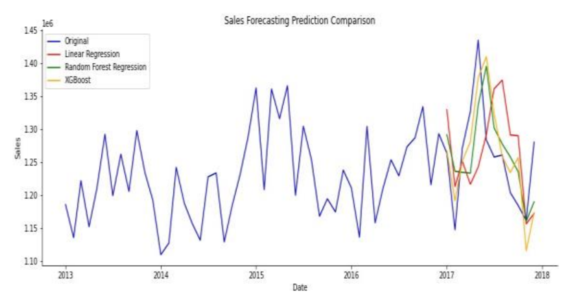

Figure 5.9: Model Prediction Comparison

Figure 5.9 represents the comparison of different model prediction. We implemented different model on sales dataset more than 87746 rows. Here blue line shows the original value, red line represents the linear regression result, green line represents the Random Forest regression, and orange line represents the Xgboost results. Xgboost is the best suitable model to predict the future sales and shows the nearest prediction values as compared to other model like LSTM, Linear regression, and Random Forest Regression.

	Index	RMSE	MAE
0	**Random Forest**	0.69460	0.59121
1	**Linear Regression**	0.96849	0.82136
2	**ARIMA**	1.04959	1.01265
3	**LSTM**	0.99964	0.81910
4	**Xgboost**	0.63010	0.51599

Table 5.6: Regression Model Error Comparison

Table 5.6 shows the results and performance of the models with their basic configuration and default parameter. From this table it's clear that Gradient Boosting regression and Random Forest performed well with both Matrices RMSE and MAE, Xgboost has least error in sales forecasting when compared to the linear regression, ARIMA, LSTM. ARIMA model showed the worst performance with the higher error in both matrices.

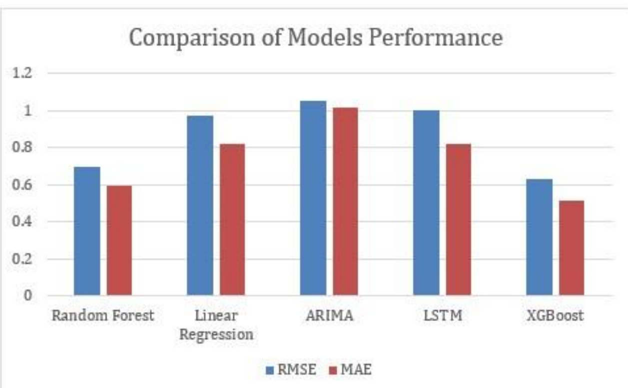

Figure 5.10: Comparison Machine Learning Model Error Results

Figure 5.10 represents the Mean Absolute error and Root Mean Squared error from the results of the produces by the random forest regression, linear regression, ARIMA, LSTM, random forest and Xgboost. From the figure Xgboost has the least RMSE and MAE and is best-performed algorithm for point of sales retail sales data.

5. Conclusion

In this paper, we concluded that sales forecasting is the most challenging task for the inventory management, marketing, customer service and Business financial planning for the information technology chain store. Sales forecasting is an important part related to supply chain management and operations between the retailer and manufacturers. Manufacturer needs to predict the actual future demand to inform production planning. Similarly, retailers need to predict sales for purchasing decision and minimize the capital costs. Therefore, depending upon the nature of the business, sales forecasting can be done through human planning and statistical model or by combining both methods. To develop sales forecasting, accurate model is a very difficult task due to different reason like over-forecasting and under-forecasting. Therefore, accurate and robust sales forecasting results can lead to customer satisfaction, enhanced channel relationships, and significant monetary savings. We applied both time series model like LSTM and ARIMA model to predict the sales as well as machine learning regression algorithm like Linear Regression model, Random Forest model and Xgboost model. We found the Xgboost is the most suitable model for Citadel POS dataset.

In future, deep learning approach can be used for sales forecasting by increasing the size of dataset. Similarly, accuracy can be increased on large dataset of retail sales using deep learning models.

References

[1]. ÁLVAREZ-DÍAZ, M., GONZÁLEZ-GÓMEZ, M. & OTERO-GIRÁLDEZ, M. S. 2018. Forecasting international tourism demand using a non-linear autoregressive neural network and genetic programming. Forecasting, 1, 7.

[2]. BALLON, R. 2004. Business logistics/supply chain management. Planning, organizing and controlling the supply chain.

[3]. CATAL, C., KAAN, E., ARSLAN, B. & AKBULUT, A. 2019. Benchmarking of regression algorithms and time series analysis techniques for sales forecasting. Balkan Journal of Electrical and Computer Engineering, 7, 20-26.

[4]. CHAI, T. & DRAXLER, R. R. 2014. Root mean square error (RMSE) or mean absolute error (MAE). Geoscientific Model Development Discussions, 7, 1525-1534.

[5]. DEO, R. C., KISI, O. & SINGH, V. P. 2017. Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. Atmospheric Research, 184, 149-175.

[6]. FENG, G., HUANG, G.-B., LIN, Q. & GAY, R. 2009. Error minimized extreme learning machine with growth of hidden

nodes and incremental learning. IEEE Transactions on Neural Networks, 20, 1352-1357.

[7]. GLYNN, J., PERERA, N. & VERMA, R. 2007. Unit root tests and structural breaks: A survey with applications.

[8]. HOFMANN, E. 2013. Supply Chain Management: Strategy, Planning and Operation, S. Chopra, P. Meindl. Elsevier Science.

[9]. HOLT, C. C. 2004. Forecasting seasonals and trends by exponentially weighted moving averages. International journal of forecasting, 20, 5-10.

[10]. HUSSAIN, S., ATALLAH, R., KAMSIN, A. & HAZARIKA, J. Classification, clustering and association rule mining in educational datasets using data mining tools: A case study. Computer Science On-line Conference, 2018. Springer, 196-211.

[11]. KAUR, M. & KANG, S. 2016. Market Basket Analysis: Identify the changing trends of market data using association rule mining. Procedia computer science, 85, 78-85.

[12]. LU, C.-J. 2014. Sales forecasting of computer products based on variable selection scheme and support vector regression. Neurocomputing, 128, 491-499.

[13]. LU, C.-J. & KAO, L.-J. 2016. 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.

[14]. MENTZER, J. T. & MOON, M. A. 2004. Sales forecasting management: a demand management approach, Sage Publications.

[15]. MÜLLER-NAVARRA, M., LESSMANN, S. & VOß, S. Sales forecasting with partial recurrent neural networks: Empirical insights and benchmarking results. 2015 48th Hawaii International Conference on System Sciences, 2015. IEEE, 1108-1116.

[16]. OFOEGBU, K. 2021. A comparison of four machine learning algorithms to predict product sales in a retail store. Dublin Business School.

[17]. OMAR, H. A. & LIU, D.-R. Enhancing sales forecasting by using neuro networks and the popularity of magazine article titles. 2012 Sixth International Conference on Genetic and Evolutionary Computing, 2012. IEEE, 577-580.

[18]. PAVLYSHENKO, B. M. 2019. Machine-learning models for sales time series forecasting. Data, 4, 15.

[19]. SHUMWAY, R. H. & STOFFER, D. S. 2017. ARIMA models. Time series analysis and its applications. Springer.

[20]. SINAGA, K. P. & YANG, M.-S. 2020. Unsupervised K-means clustering algorithm. IEEE access, 8, 80716-80727.

[21].

[22]. Bukhsh, Madiha, et al. "An Interpretation of Long Short-Term Memory Recurrent Neural Network for Approximating Roots of Polynomials." IEEE Access 10 (2022): 28194-28205.

[23]. Tufail, Hina, M. Usman Ashraf, Khalid Alsubhi, and Hani Moaiteq Aljahdali. "The Effect of Fake Reviews on e-Commerce During and After Covid-19 Pandemic: SKL-Based Fake Reviews Detection." IEEE Access 10 (2022): 25555-25564.

[24]. Mumtaz, Mamoona, Naveed Ahmad, M. Usman Ashraf, Ahmed Alshaflut, Abdullah Alourani, and Hafiz Junaid Anjum. "Modeling Iteration's Perspectives in Software Engineering." IEEE Access 10 (2022): 19333-19347.

[25]. Asif, Muhammad, et al. "A Novel Image Encryption Technique Based on Cyclic Codes over Galois Field." Computational Intelligence and Neuroscience 2022 (2022).

[26]. Mehak, Shakra, et al. "Automated Grading of Breast Cancer Histopathology Images Using Multilayered Autoencoder." CMC-COMPUTERS MATERIALS & CONTINUA 71.2 (2022): 3407-3423.

[27]. Naqvi MR, Iqbal MW, Ashraf MU, Ahmad S, Soliman AT, Khurram S, Shafiq M, Choi JG. Ontology Driven Testing Strategies for IoT Applications. CMC-Computers, Materials & Continua. 2022 Jan 1;70(3):5855-69.

[28]. S. Tariq, N. Ahmad, M. U. Ashraf, A. M. Alghamdi, and A. S. Alfakeeh, "Measuring the Impact of Scope Changes on Project Plan Using EVM," vol. 8, 2020.

[29]. Asif M, Mairaj S, Saeed Z, Ashraf MU, Jambi K, Zulqarnain RM. A Novel Image Encryption Technique Based on Mobius Transformation. Computational Intelligence and Neuroscience. 2021 Dec 17;2021.

[30]. Ashraf, Muhammad Usman. "A Survey on Data Security in Cloud Computing Using Blockchain: Challenges, Existing-State-Of-The-Art Methods, And Future Directions." Lahore Garrison University Research Journal of Computer Science and Information Technology 5, no. 3 (2021): 15-30.

[31]. Ashraf MU, Rehman M, Zahid Q, Naqvi MH, Ilyas I. A Survey on Emotion Detection from Text in Social Media Platforms. Lahore Garrison University Research Journal of Computer Science and Information Technology. 2021 Jun 21;5(2):48-61.

[32]. Shinan, Khlood, et al. "Machine learning-based botnet detection in software-defined network: a systematic review." Symmetry 13.5 (2021): 866.

[33]. Hannan, Abdul, et al. "A decentralized hybrid computing consumer authentication framework for a reliable drone delivery as a service." Plos one 16.4 (2021): e0250737.

[34]. Fayyaz, Saqib, et al. "Solution of combined economic emission dispatch problem using improved and chaotic population-based polar bear optimization algorithm." IEEE Access 9 (2021): 56152-56167.

[35]. Hirra I, Ahmad M, Hussain A, Ashraf MU, Saeed IA, Qadri SF, Alghamdi AM, Alfakeeh AS. Breast cancer classification from histopathological images using patch-based deep learning modeling. IEEE Access. 2021 Feb 2;9:24273-87.

[36]. Ashraf MU, Eassa FA, Osterweil LJ, Albeshri AA, Algarni A, Ilyas I. AAP4All: An Adaptive Auto Parallelization of Serial Code for HPC Systems. INTELLIGENT AUTOMATION AND SOFT COMPUTING. 2021 Jan 1;30(2):615-39.

[37]. Hafeez T, Umar Saeed SM, Arsalan A, Anwar SM, Ashraf MU, Alsubhi K. EEG in game user analysis: A framework for

expertise classification during gameplay. Plos one. 2021 Jun 18;16(6):e0246913.

[38]. Siddiqui N, Yousaf F, Murtaza F, Ehatisham-ul-Haq M, Ashraf MU, Alghamdi AM, Alfakeeh AS. A highly nonlinear substitution-box (S-box) design using action of modular group on a projective line over a finite field. Plos one. 2020 Nov 12;15(11):e0241890.

[39]. Ashraf, Muhammad Usman, et al. "Detection and tracking contagion using IoT-edge technologies: Confronting COVID-19 pandemic." 2020 international conference on electrical, communication, and computer engineering (ICECCE). IEEE, 2020.

[40]. Alsubhi, Khalid, et al. "MEACC: an energy-efficient framework for smart devices using cloud computing systems." Frontiers of Information Technology & Electronic Engineering 21.6 (2020): 917-930.

[41]. Riaz S, Ashraf MU, Siddiq A. A Comparative Study of Big Data Tools and Deployment PIatforms. In2020 International Conference on Engineering and Emerging Technologies (ICEET) 2020 Feb 22 (pp. 1-6). IEEE.

[42]. Ashraf MU, Eassa FA, Ahmad A, Algarni A. Empirical investigation: performance and power-consumption based dual-level model for exascale computing systems. IET Software. 2020 Jul 27;14(4):319-27.

[43]. Ashraf, Muhammad Usman, et al. "IDP: A Privacy Provisioning Framework for TIP Attributes in Trusted Third Party-based Location-based Services Systems." , International Journal of Advanced Computer Science and Applications (IJACSA) 11.7 (2020): 604-617.

[44]. Manzoor, Anam, et al. "Inferring Emotion Tags from Object Images Using Convolutional Neural Network." Applied Sciences 10.15 (2020): 5333.

[45]. Alsubhi, Khalid, et al. "A Tool for Translating sequential source code to parallel code written in C++ and OpenACC." 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA). IEEE, 2019.

[46]. Ashraf MU, Naeem M, Javed A, Ilyas I. H2E: A Privacy Provisioning Framework for Collaborative Filtering Recommender System. International Journal of Modern Education and Computer Science. 2019 Sep 1;11(9):1.

[47]. Ashraf MU, Ilyas I, Younas F. A Roadmap: Towards Security Challenges, Prevention Mechanisms for Fog Computing. In2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE) 2019 Jul 24 (pp. 1-9). IEEE.

[48]. Ashraf MU, Qayyum R, Ejaz H. “State-of-the-art Challenges: Privacy Provisioning in TPP Location Based Services Systems.”, International Journal of Advanced Research in Computer Science (IJARCS). 2019 Apr 20;10(2):68-75.

[49]. Ashraf MU, Arshad A, Aslam R. Improving Performance In Hpc System Under Power Consumptions Limitations. International Journal of Advanced Research in Computer Science. 2019 Mar;10(2).

[50]. Javed, Rushba, et al. "Prediction and monitoring agents using weblogs for improved disaster recovery in cloud." Int. J. Inf. Technol. Comput. Sci.(IJITCS) 11.4 (2019): 9-17.

[51]. Ali, Muhammad, et al. "Prediction of Churning Behavior of Customers in Telecom Sector Using Supervised Learning Techniques." 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE). IEEE, 2018.

[52]. Ashraf MU, Eassa FA, Albeshri AA, Algarni A. Performance and power efficient massive parallel computational model for HPC heterogeneous exascale systems. IEEE Access. 2018 Apr 9;6:23095-107.

[53]. Ashraf MU, Eassa FA, Albeshri AA, Algarni A. Toward exascale computing systems: An energy efficient massive parallel computational model. International Journal of Advanced Computer Science and Applications. 2018 Jan;9(2).

[54]. Ashraf MU, Arif S, Basit A, Khan MS. Provisioning quality of service for multimedia applications in cloud computing. Int. J. Inf. Technol. Comput. Sci.(IJITCS). 2018;10(5):40-7.

[55]. Ashraf MU, Eassa FA, Albeshri AA. Efficient Execution of Smart City’s Assets Through a Massive Parallel Computational Model. InInternational Conference on Smart Cities, Infrastructure, Technologies and Applications 2017 Nov 27 (pp. 44-51). Springer, Cham.

[56]. Alrahhal, Mohamad Shady, et al. "AES-route server model for location based services in road networks." International Journal Of Advanced Computer Science And Applications 8.8 (2017): 361-368

[57]. Ashraf MU, Eassa FA, Albeshri AA. High performance 2-D Laplace equation solver through massive hybrid parallelism. In2017 8th International Conference on Information Technology (ICIT) 2017 May 17 (pp. 594-598). IEEE.