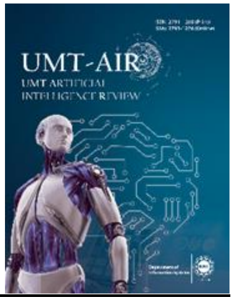
UMT Artificial Intelligence Review (UMT-AIR)
Volume 5 Issue 2, Fall 2025
$ISSN_{(P)}$: 2791-1276, $ISSN_{(E)}$: 2791-1268
Homepage: https://journals.umt.edu.pk/index.php/UMT-AIR



Check for updates

Title:	**Store Sales Prediction Using Machine Learning, Statistical Modeling, and Deep Learning Frameworks: A Comparative Analysis**
Author (s):	Asif Ahsan[1], Ayesha Awan[2], and Syed Mohammad Irteza[3]
Affiliation (s):	[1]Department of Computer Science & IT, The University of Lahore, Pakistan [2]Department of Intelligent Systems, The University of Lahore, Pakistan [3]Department of Computer Science, National Institute of Technology, Lahore, Pakistan
DOI:	https://doi.org/10.32350.umt-air.52.05
History:	Received: July 22, 2025, Revised: October 18 2025, Accepted: November 23, 2025, Published: December 16, 2025
Citation:	A. Ahsan, A. Awan, S. M. Irteza "Store Sales Prediction Using Machine Learning, Statistical Modeling, and Deep Learning Frameworks: A Comparative Analysis," *UMT Artif. Intell. Rev.*, vol. 5, no. 2, pp. 74–88, Dec. 2025, doi: https://doi.org/10.32350.umt-air.52.05
Copyright:	© The Authors
Licensing:	This article is open access and is distributed under the terms of Creative Commons Attribution 4.0 International License
Conflict of Interest:	Author(s) declared no conflict of interest

A publication of
Department of Information System, Dr. Hasan Murad School of Management
University of Management and Technology, Lahore, Pakistan

Store Sales Prediction Using Machine Learning, Statistical Modeling, and Deep Learning Frameworks: A Comparative Analysis

Asif Ahsan[1], Ayesha Awan[2], and Syed Mohammad Irteza[3]

[1]Department of Computer Science & IT, The University of Lahore, Pakistan

[2]Department of Intelligent Systems, The University of Lahore, Pakistan

[3]Department of Computer Science, National Institute of Technology, Lahore, Pakistan

ABSTRACT Sales forecasting or prediction are essential tools for businesses to make informed decisions about resource allocation, budgeting, and marketing strategies. Sales prediction involves predicting future sales of a product or service. There are several methods that businesses can use to forecast sales, including statistical analysis, market research, and expert opinion. Key factors that may impact sales predictions include market conditions, economic trends, competition, and customer behavior. However, forecasting sales can be a challenging task, as it requires businesses to accurately forecast future demands and predict future sales in the face of changing market conditions and unpredictable customer behavior. Furthermore, there are several uncertainties that businesses may face when forecasting or predicting sales, such as economic conditions, competition, customer behavior, product life cycles, and external events. The current study focused on sales prediction and provided a comparative analysis of three widely-used techniques in the field of forecasting, namely: statistical modeling, Machine Learning (ML), and Deep Learning (DL). In the analysis, the data given by the Rossmann supply chain was used. This is the second largest drugstore chain in Europe, and to the best of researchers' knowledge, no one has ever used the DL approach to predict sales of these stores. Therefore, the current study attempted to forecast sales and improve the results of the prediction.

INDEX TERMS comparative analysis, deep learning, machine learning, sales forecasting, statistical modeling

I. INTRODUCTION

Data science and data mining are essential in the retail industry due to the volume of data and the complexity of optimization. It is possible to optimize prices, discounts, suggestions, and stock levels using data analytic techniques. Response modelling, recommendation algorithms, demand fore-casting, pricing discrimination [1], sales event planning, and category management [2] are all typical data mining issues. Now organizations must contend with a highly competitive environment in the modern global economy and use a variety of techniques to achieve an advantage. Mass customization (MC), which blends mass manufacturing and personalized output, is one such tactic [3]. MC is a phenomenon that may be seen in a wide range of digital and physical goods and services that are offered for sale. Demand planning is essential for managing business operations [4]. Despite their complexity and execution, forecasting processes in various industries have the same objective. This objective includes to estimate the demand of the goods or any services that users need and to estimate it with high accuracy by using the historical data of the goods and other influential factors, such as current political and socioeconomic situations in order to strategize the business and planning.

Ordered data values are found in time series data. The estimation of future demand must be done quickly. A considerable amount of data is gathered on a regular basis. Time series data includes things, such as market prices, business data, sales income, and weather observations. Time series analysis and forecasting are well-liked academic subjects [5].

The objectives of time series analysis include finding correlations between time series variables and drawing conclusions. Using the past data, time series prediction predicts future data points. Time series analysis can be used to predict future electricity consumption. The number of Machine Learning (ML) algorithms for the problems of time series prediction has increased over the past ten years as a result of Artificial Intelligence (AI), becoming more and more relevant. ML techniques do not make any assumptions about the structure of the data, in contrast to conventional time series prediction methods. Using supervised ML, it might be possible to recognize patterns in the sales process. The ML algorithms, that is, Random Forest (RF) [6] and Gradient Boosting are based on trees. Regression algorithms work under the assumption that past patterns would reappear in new data. High-performance computers have recently posed a threat to the widespread adoption of DL models despite the 1940s origins of neural networks. The performance of computer hardware was so poor that it was completely useless. Then, ML models took their place to address those issues [7].

DL models have improved the performance of pattern recognition, including speech and image recognition. Artificial Neural Networks (ANNs) are flexible, error-tolerant, and self-learning systems. Complex nonlinear models are precisely fitted to multidimensional data. Neural networks have been used in several studies [8] for classification and regression. In the past, Long-Short Term Memory (LSTM), Gated Recurrent Units (GRU), Multi-Layer Perceptron (MLP), and Convo-lutional Neural Networks (CNN) were utilized. In complicated data, DL models may identify both linear and nonlinear trends [9]. This enables algorithms to dependably anticipate innovative, uncertain, noisy, and impartial data. This document aimed to forecast the sales of Europe's second-largest corporation, that is, Rossmann. The Rossmann Company provided the data for 1115 out of 3000 stores, thus sales forecasts must be generated for each store. The study intended to establish which components are most important for sales forecasting based on demand. Moreover, it compared the performance of advanced ML, statistical, and DL models in time series prediction. In time series prediction, ML, statistical models, and DL perform admirably. Furthermore, the study also utilized the Prophet Model for sales forecasting. This is because better results could not be obtained from the dataset with base paper. Therefore, the study incorporated the best approach after analyzing the literature and then constructed the most accurate model possible to surpass that accuracy. The following sections of this manuscript can be classified as follows:

In Section 2, the problem statement and research objectives are defined. Section 3 investigates the prior research on demand forecasting methods, algorithms, and associated publications. Section 4 describes the dataset; the study provided strategies for data preparation and how the given dataset was used to get into a desired shape to feed into a model. Section 5 describes the ML methodologies utilized for this research. The final Section 6 discusses the conclusion of this research.

II. LITERATURE REVIEW

Sales forecasting has various applications across all lines of businesses, including inventory management, replenishment of goods, supply chain management, monitoring fraudulent activities, increased sales, making policies to protect the environment, predicting market situations, smart city initiatives, sales prediction, intelligent business decisions, and more. To conduct the current research, the main focus was on the Rossmann store literature. However, many other time series forecasting papers and techniques were also reviewed. For instance, what advancements people have made in the past few years and what are the better options to deal with the sales data. When dealing with industrial and rea-time problems, there might not be an advantage of using the datasets. This is the reason that most of the datasets are private. Some might be freely available on public websites, and some are available upon request. In the study [10], to make sales time series projections for the Rossman store, the author looked at both the FDR and Support Vector Regression (SVR) models. Due to the magnitude of the data variables, SVR was able to outperform FDR. The current investigation showed that the most effective SVR technique makes use of a polynomial kernel that is subject to regularization. The study [11] offered a method for DL that employs temporal trend correction to anticipate non-stationary time series. Using a subnetwork block, the neural network model calculates the sales value based on a temporal trend term, that is, Block model. The predicted weight value and normalized time value are parts of the time trend term. The results showed that DL model's trend correction block can enhance forecasts for non-stationary sales with time patterns. In the research [12], the authors studied how bad weather affected goods' sales. The authors used a daily transactional record from 45 Walmart stores between the time period 2012 and 2014 for the analysis. From adjacent weather stations, these localities obtained their weather data. Weather-related datasets were used. The sets were put to test after the train had been disassembled and its components engineered. Sales unit forecasting and training were permitted by ANN. Mean Squared Error (MSE) started to grow after increasing the number of layers from three to four (MLP-L3-N70 and MLP-L3-N90 for different datasets). ANN had the lowest MSE on both the training set and test set when compared to Recurrent Neural Networks (RNNs), bagging, and time lag. While requiring substantially more time than training neurons, bagging typically yielded ten outcomes.

The study [13] focused on deep model learning applications in sales predictive analytics. This study aimed to show how important it is to use primary methodologies and in-depth case studies for projecting sales. DL's generalizability was considered. When there is historical sales data available, such as when a new product or merchant is introduced, this might be used to estimate sales. One illustration is the debut of a new product. Researchers have stacked the single-outfit regression ensembles. The goal of this study was to create a data mining use case to forecast retail sales and demand. A prediction model for the retail locations of a renowned European pharmacy business was created using Extreme Gradient Boosting (XGBoost). This model anticipates sales. Using historical data on the sales, business promotions, school holidays, accessibility, retail competition, store location, state holidays, and the season of the year, future sales were forecasted. The model was developed using logic, data analysis, and conclusive findings. In comparison to linear regression and RF regression, XGBoost's performance was evaluated. This study explored the use of data mining to forecast retail sales. It highlighted the retail need. For the stores of a major

European pharmaceutical retailer, XGBoost was utilized to precisely predict retail sales. Future sales projections are based on historical sales information, store promotions, retail competition, state and federal holidays, store location and accessibility, as well as the present season. Common sense and analytical abilities developed through data analysis served as the process' guiding principles, and the outcomes were persuasive. Comparisons were made between XGBoost's accuracy and that of linear and RF regression. In the study [14], a specific loss function for an LSTM model was created via hyperparameter search. The study used the Kaggle public Rossmann sales dataset to show their effectiveness. The experiment showed that the suggested method performed better in sparse data prediction than Au-toML (Auto Machine Learning) models. This was proven through the experiment. In the study [15], ML algorithms for predictive sales analytics were examined. In this study, sales forecasting in the real world has been discussed along with ML techniques. It has been thought about as to whether ML could be generalized. When there is minimal past sales data available, such as when was the new product introduced and the store's entry too, that influence might be used for the estimation of sales. In addition, at the time when the new shop or product was introduced, the idea of stacking individual models to create regression ensembles was considered. The study found that stacking techniques enhanced sales time series forecasting predictive models, an issue that many furniture manufacturers' face is covered with [16].

Manufacturers and wholesalers must lower stockouts and maintain enough inventory levels for a successful supply chain and optimal inventory. In order to produce the required items and deliver them on time, furniture producers need an accurate prediction of their inventory. Using feedforward and backpropagation network architectures as well as ANN, the model received training assistance from a private furniture business. Approximately, 38 months were needed to collect the data (July 2007–August 2010). Between the eight input qualities, there is a hidden layer with logistic activation functions. ANN was contrasted with a linear model built on BDS. The Mean Absolute Precision Error (MAPE) of the BDS model was 44.84. In the study [17], the author introduced a straightforward implementation of a tree-based ML method that substantially outperforms traditional forecasting methods. Based on 4,523 products of a leading Belgian retailer and a variety of external variables (e.g., promotions and national events), it was determined that the proposed framework improved the forecast accuracy of commonly used statistical methods by 9.34% up to 20.52% while being computationally efficient. Using re-fine values and slightly adjusted tree-based models, the study empirically analyzed the reason for this outperformance and identified the necessary requirements to make investment in this ML forecasting method worthwhile. It was found that global tree-based model benefits more by having access to a variety of external variables and can learn non-linear time series patterns. Extensive numerical experimentation finally shows how the forecast superiority of the proposed framework translates to higher service levels, lower inventory costs, and improvements in the bullwhip of orders and inventory. In the study [13], heavy-duty vehicles, such as intercity buses were the subject of the study. It is preferable to use highways since they produce less wind and noise. Monitoring gasoline use may boost fuel efficiency and reveal dishonest activity but the result is a wealthier company.

Three ML implementations were compared in this analysis, that is, RAF, Gradient Boosting (GB), and neural networks (NN). The two most crucial variables in calculating fuel consumption included the distance traveled and the average speed. For a single bus to make the entire trip, the route is also split into inbound and outgoing trips. The RAF starts the trial. By changing the number of trees and randomly selecting the variables, the model is adjusted. The split criterion is 2 and tree count (250, 500, 750). The GB package of mboost is used by R. The model is trained using three neurons in the first hidden layer and two in the second using the neural net R program. The database is categorized, and the effectiveness of each model is evaluated using the Nash-Sutcliffe efficiency coefficient. Results are analyzed with bias, MAE, and RMSE. RAF performed better than other models on this dataset.

In the study [18], the author showed how ML can be used to forecast retail sales in the experiment that follows. Store managers can create the best personnel schedules with the help of these forecasts. Modeling, analysis, prediction, and visualization were all done in Python. NumPy, scikit-learn, pandas, and mat-plotlib were used. Regression, ensemble learning, and XGBoost regression techniques were applied for time-series forecasting. Forecast performance was evaluated using the Root Mean Square Error (RMSE), which served as the primary metric for assessing prediction accuracy. The Autoregressive Integrated Moving Average (ARIMA) method and classification-based modeling are two popular forecasting methods. These are forecasting and modeling techniques. In the study [19], academics and business-related people intending to forecast sales both focused on the analysis of sales patterns. Trends in sales analysis are of importance to both parties. It is difficult to create a framework that can consistently anticipate future item sales due to the nonlinear nature of sales data. Recommender systems may help an organization or merchant make decisions by identifying sales patterns and producing precise projections of future revenue. The current study used neural network algorithms to estimate sales data, producing accurate sales predictions. The data was divided into different categories based on assortment, promotions, school, and state holidays. Outliers and unnecessary data were also eliminated. a.csv file in the data bundle provides data specific to each retailer. To estimate future sales and offer support for that version, data is fed into a trained classifier. Resultantly, exceptional qualities are developed.

III. METHODOLOGY

A. DATA ACQUISITION AND PREPARATION

The current study utilized Rossmann's dataset. It contained 2.5 years' worth of sales information collected from 1,115 European Rossmann stores. The dataset was hosted by Kaggle [10]. Rossmann's test set sales were predicted using this dataset. The model was developed thoroughly using time series techniques. The behaviors and patterns of the data were interpreted using several approaches. To comprehend Rossmann sales, it is necessary to investigate seasonality, patterns, and other elements. It may be vital to comprehend why salespeople behave in this manner. Exploratory data analysis, or EDA, would make use of Pandas and other graph libraries. These libraries are typically used for graphic data display. Time series data analysis is made simpler by these formats. The elements that have the greatest impact on sales are identified through a key component analysis.

B. DATASET OVERVIEW

The dataset obtained from Kaggle has three different files, namely: •Train.csv: It includes information about sales, customers, school holidays, state holidays, days of the week and details, such as when the store opens and closes, as well as a breakdown of sales for each day of the week. •Store.csv: It contains the data of stores and related information, such as store types, promotion details, competitor information, and assortment. •Test.csv: This file anticipates sales for the data being provided, that is, data for a period of six weeks. The data properties and their descriptions are listed below.

C. EXPLORATORY DATA ANALYSIS

The study performed EDA on the dataset. The Rossmann store dataset contained 10M+ records in it. It had data of almost 2.5 years for the time period January 2013 to July 2015. Table 1 shows the train data attributes. It had 10M+ records according to the daily sales and 8 unique columns. In the same way, the study performed EDA on the store dataset as well. It had 1115 rows that were store records, and there are 10 columns against those stores. The details of the store dataset are given below in Table 2.

D. STORE DETAILS OVER THE DAYS OF THE WEEK

The study checked the counts of the stores as to when the store was closed and opened. From the chart, it can be concluded that the store was mainly closed on Sunday. This means that people have a holiday on Sunday and it is a resting day for everyone. Figure 1 shows the detailed graph.

However, still, it can be seen that some stores were open on Sunday as well. The details of the store IDs that were open on Sunday were 85, 122, 209, 259, 262, 274, 299, 310, 335, 353, 423, 433, 453, 494, 512, and 524.

E. SALE OVERVIEW OF EACH STORE TYPE

It can be seen that sale of the store type = b was far more than the rest of the three store types.

F. EFFECT OF PROMO OVER SALES

The effect of promotion on the sale was also checked. It was concluded that the sale due to promotion was very high as compared to the other sale when the store did not have any promo. The effect of promo can be seen in the Fig 3.

G. COMPLETE SALES OVERVIEW

Figure 4 summarizes the complete sales details over the 2.5 years of the record provided by the Rossmann store. Some trends can also be seen in the sales record, that is, after almost 2.5 months sales increased. This trend kept repeating till the end of the year. This is the effect of promo stores running on the sale, increasing the sale of the stores drastically.

H. DATA CLEANSING AND PRE-PROCESSING

After performing EDA and visualizing the data for better understanding, it can be seen that there were null values in the store.csv file. To perform prediction, there is no possibility to have data with null values. Null valued columns were 'CompetitionDistance', 'Competi tionOpenSinceMonth', 'CompetitionOpenSinceYear', 'Promo tion2Week', 'Promotion2Year', and 'PromoInterval'. These are the columns that have missing values so, these cannot be simply removed. If null values are removed, then the model's accuracy may be affected. It is better to deal with null values rather than remove them from the data. So, the meaning of the data is simply taken and the missing data is filled with their calculated means. After that, there was a column of 'PromoInterval'. Its data was in string format so, there was a need to take another approach to it.

I. FEATURE ENGINEERING

When there was clean data without null values, 'day', 'month', and 'year' were extracted as new features from the data to perform the prediction. The 'date' column cannot be simply used as an object type for training, therefore, data should be given in one format. Hence, features were extracted from the data for daily prediction of 6 continuous weeks which was the objective of this research.

J. FEATURE ENCODING

As discussed earlier, model training cannot be performed with different data formats so, data must be given to the model in one format. Therefore, dummies were used, which is Panda's library to encode the string data type attributes. It replaces the data with 0-1 values, that is, 1 means data is present and 0 means it is not present. Finally, when the dataset was there in the desired form, model training was started based on the data to achieve the desired results.

K. MODEL BUILDING

After dealing with the data's unordered format and shape, the study moved to its final phase of model building for the implementation of the base paper. For the base paper implementation, the study used 6 models namely, Lasso Regression, Linear Regression, Random Forest Regressor, Ridge Regression, Decision Tree Regressor, and Xtreme Gradient Boosting Regressor. Some of the models were being used for classification purposes too. However, as this was a regression problem so, regression modules were used. The study provided some basic idea about each of the models and then moved towards the implementation and the results of the models.

1) LINEAR REGRESSION

An LR model is used for the prediction of one variable based on the values of the other variables. This analysis estimates the coefficients of a linear equation by using one or more independent variables to predict the dependent variable with the highest possible accuracy. Linear regression minimizes the discrepancies between expected and actual output values by fitting a straight line or surface. When given a set of paired data, simple linear regression calculators using the "least squares" method can determine which line fits the data the best. The value of X (independent variable) is then estimated using Y (dependent variable). The values are calculated using the following equation:

$$Yi = f (Xi,) + ei$$

Where Yi represents the dependent variable (predicted value), Xi is the independent variable(s), Beta is the unknown parameter, and ei is the error term.

2) RIDGE REGRESSION

It is a linear regression's regularized form that adds a penalty term, known as the L2 penalty, to the model. This adjustment is useful when the data exhibits multicollinearity. Unlike least squares, which are unbiased but can produce large variances in the predicted values due to high variance in the data, ridge regression helps to reduce the variance and produce more reliable predictions.

TABLE I
TRAIN DATASET TABLE WITH ITS ATTRIBUTES

Date	Store	Day of Week	Sales	Customers	Open	Promo	State Holiday	School Holiday
2025-07-31 00:00:00	1	5	5263	555	1	1	0	1
2025-07-31 00:00:00	2	5	6064	625	1	1	0	1
2025-07-31 00:00:00	3	5	8314	821	1	1	0	1
2025-07-31 00:00:00	4	5	13995	1498	1	1	0	1
2025-07-31 00:00:00	5	5	4822	559	1	1	0	1

TABLE II
STORE DATASET OVERVIEW WITH ATTRIBUTES

Index	Store	Store Type	Competition Distance	Competition Open Since Month	Competition Open Since Year	Promo2	Promo2Since Week	Promo2Since Year	Promo Interval
0	1	c	1270	9	2008	0	NaN	NaN	NaN
1	2	a	570	11	2007	1	13	2010	Jan, Apr,Jul, Oct
2	3	a	14130	12	2006	1	14	2011	Jan, Apr,Jul, Oct
3	4	c	620	9	2009	0	NaN	NaN	NaN
4	5	a	29910	4	2015	0	NaN	NaN	NaN

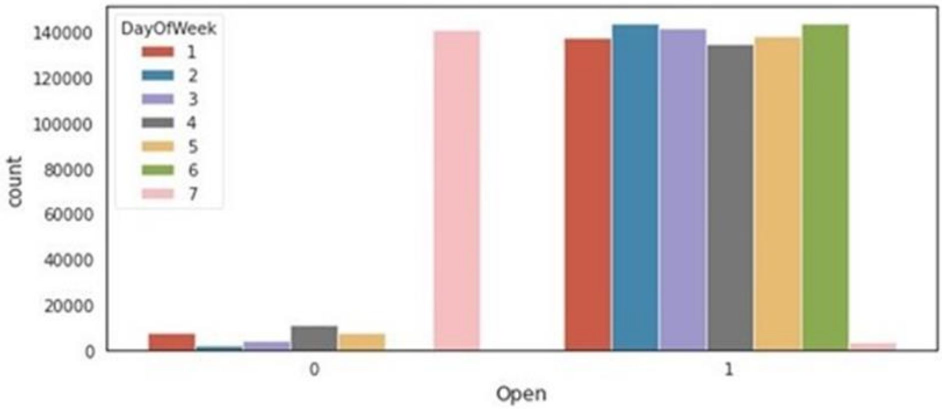


FIGURE 1. Store’s behavior on different days of the week

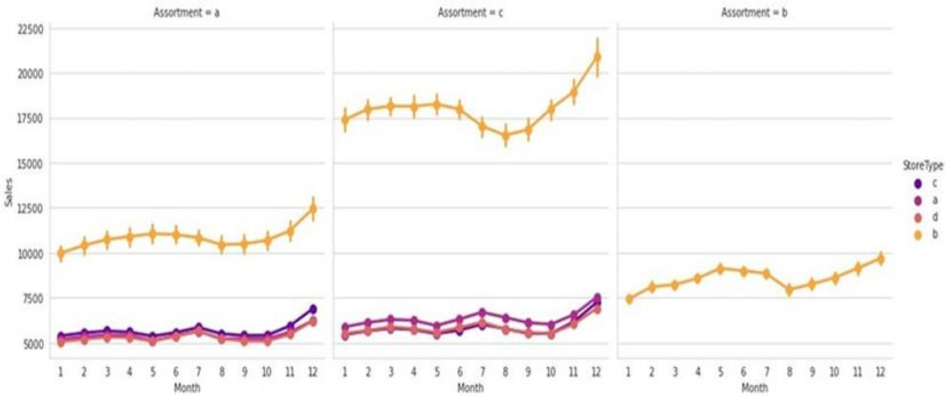


FIGURE 2. Sale of each store type over the month's linear regression (source: java point, online)

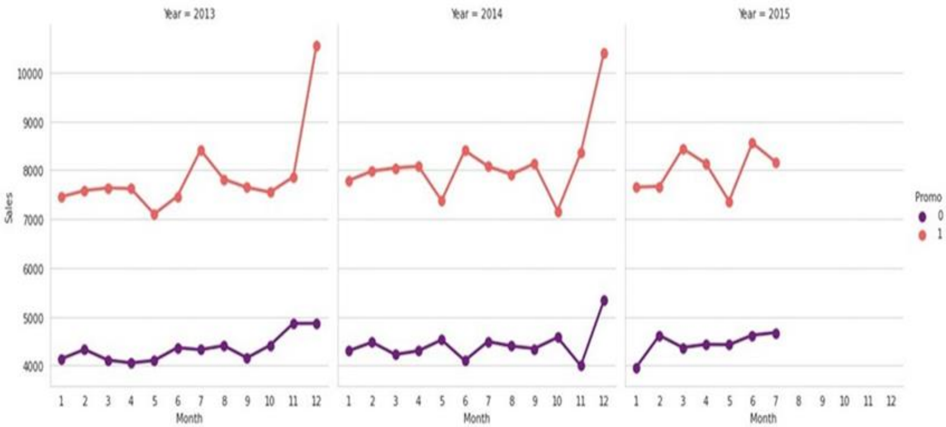


FIGURE 3. Sale of each store type over the months due to promo

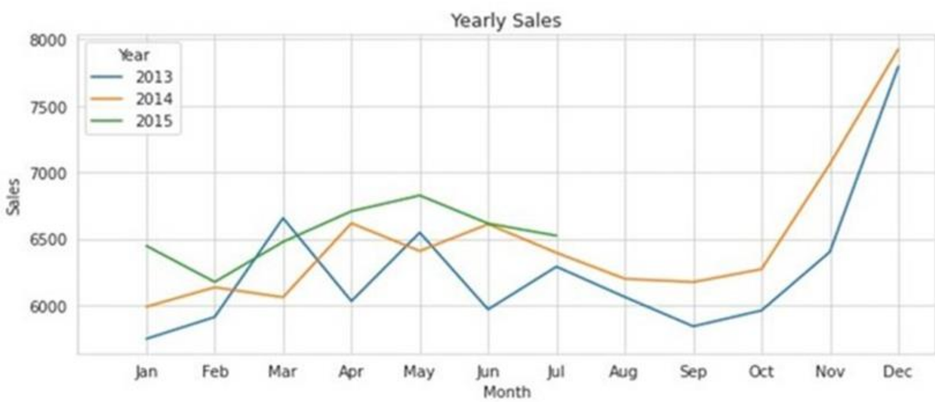


FIGURE 4. Sales trend over the 2.5 years

To get the appropriate value of lambda, cross-validation needs to be conducted first. To recalculate the error, the least value is chosen as the parameter once the cross-validation process is finished, and then that parameter is put into the formula.

3) LASSO REGRESSION

It is also a technique to regularize the Linear Regression model, as Ridge Regression introduces the L2 penalty, it adds the L1 penalty which shrinks the coefficients of determinations towards zero. Linear Regression calculates the coefficients and Lasso Regression regularizes or shrinks those coefficients to avoid the problem of overfitting. It is also used when the dataset encounters the problem of multicollinearity

, and the value of the lambda is chosen by performing the cross-validation technique which has been discussed in Ridge Regression before. It just penalizes the less important features and eliminates them, which gives the ability to select only important features and model with high accuracy without a problem of overfitting.

4) DECISION TREE (DT) AND RANDOM FOREST (RF)

Decision trees (DTs) are formed via a series of inter-connected nodes that branch out into further child nodes. The algorithm begins by starting at the first node parent (called a ’root’ node) and answers a conditional question that has been assigned to it by the model during the ’fitting’ process. The algorithm uses the features of the input sample and determines which sample response matches its answer. The algorithm then proceeds to follow the path connected to the matching answer. This path leads to the next node in the tree with its conditional question. Now, this new node would become the parent node, and the whole process would be repeated until the algorithm reaches a node with no further child nodes (that is, the leaf node). The answer on the final child node would be utilized to produce the output of the model. RF [4] models are composed of a very large number of DTs organized using an ensemble methodology. Consequently, many trees are formed, each with its unique combination of conditional questions and several nodes. The output of the RF would be the average of its separate DTs.

.5) XGBOOST REGRESSOR

There is a library dedicated to tree-boosting techniques called as XGBoost. Friedman et al. gave the initial suggestion for the gradient boosting trees model. XGBoost is a powerful algorithm that builds upon the principles of Gradient Boosting Machines (GBM) and is commonly used in supervised learning tasks where the goal is to predict the target variable using a collection of training data with many features. It is a variation of the traditional GBM approach. Although more efficient, it is similar to the gradient-boosting architecture. This means that it creates weak learning models and sequentially combines their predicted values to improve the overall performance of the model at the final stage. This process can be seen in Fig below. It has a tree learning technique as well as a linear model solver. This increases XGBoost’s efficiency over existing gradient boosting implementations by a factor of at least ten. It offers objective features including ranking, categorization, and regression. XGBoost is perfect for several contests because it is very slow to execute but has very high predictive power. The objective of the XGBoost model is: Where, the first component of the loss term is the predictor of the model’s performance, while the second portion is the regularization that prevents the model from overfitting the data. For the error evaluation, the matric used was RMSE.

IV.RESULTS AND DISCUSSION

The study started experimentations with basic statistical models and then proceeded to ML and DL models.

A. STATISTICAL MODELING

In time series, the data changes with time. With an increasing or decreasing trend, most time series have some form of seasonal trends, that is, variations specific to a particular time frame. For instance, for Christmas holidays. Stationarity of Time Series. First, we check the stationarity through Dicky-Fulller and Rolling window technique: We check the trend of sales for 1 store from each Store Type and plot the data on Weekly basis, we see that the trend is Stationary. i.e., Constant mean Constant Variance, Covariance changes with time Rolling: A rolling analysis of a time series model is often used to assess the model's stability over time. When analyzing financial time series data using a statistical model, a key assumption is that the parameters of the model are constant over time. The window is rolled on weekly basis, in which the average is taken on weekly basis rolling statistics is a visualization test, where we can compare the original data with the rolled data and check if the data is stationary or not. After analyzing the nature of the data, the next step in time series analysis is to review Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots.

From the Figure 5, the values of the P, Q were taken for statistical modeling. Afterwards, grid search was conducted and the AIC was tested for each model to select the best p, d, q parameters. The model having lowest AIC determines an accurate performance. The AIC measures how well a model fits the data while considering the overall complexity of the model. A model that fits the data well while using considerable features would be assigned a larger AIC score than a model that uses fewer features to achieve the same goodness-of-fit. Therefore, the study was aimed at finding a model that yields the lowest AIC value. When the desired values of the (p, d, q) were obtained, the data was fed into stats models and predicted the sales. The MSE, RMSE, and RMSPE were calculated. After performing many different combinations of the (p, d, q) values, the RMSPE error could not be achieved below 1. So, the researchers moved towards ML techniques which are more advanced and better in terms of statistical techniques. The results of the statistical techniques are shown in Table 3.

It can be seen that SARIMAX outperformed the older version of statistical techniques. This is because it can handle the external variables as well while incorporating the already given features. However, still, the error below 1 could not be achieved. Squares. We experimented with LASSO and Ridge regres sions, testing low learning rates (alpha = 10ˆ-4) as initial checks favored these. LASSO slightly outperformed Ridge but was still inferior to Ordinary Least Squares (OLS). Next, ensemble techniques were applied including Random Forest Regression, which yielded better results than Ridge and LASSO. Without parameter tuning, the DT Regressor provided the best scores among the models tested. Finally, the XG Boost Regressor was used, which surpassed the baseline paper's state-of-the-art score. After five rounds of cross-validation, XGBoost consistently delivered superior results. Feature creation was crucial, involving numerous experiments with different features and transformations. Log transformation was applied to the skewed sales data and outliers were removed. By running XGBoost with up to 25,000 trees, an RMSE of 0.08461 and RMSPE of 0.09323 was achieved on the testing dataset, which took several hours to compute. Table 4 summarizes the model scores. Many experiments were performed but out of them best 5 have been shown in the table. The experiment results have been written in ascending order. From the table above, it can be seen that experiment 1 had the best score RMSPE of 0.09323. This was achieved by using log transformation, creation of extra features, and handling outlier values.

TABLE III

STATISTICAL MODEL RESULTS

Statistical Model	RMSE	MSE	RMSPE
ARIMA	249.61	62305.17	1.01
SARIMA	41.85	1751.46	1.0
SARIMAX	41.85	1751.46	1.0

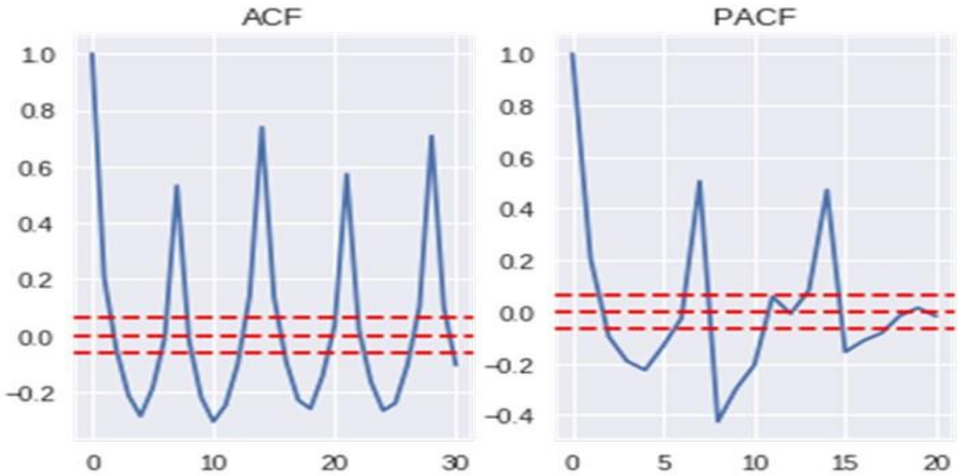


FIGURE 5. Statistical modeling results

TABLE IV

EXPERIMENTAL RESULTS WITH DIFFERENT CONFIGURATIONS

Experiment #	Including Factors	RMSE	RMSPE
Experiment 1	With log, extra features and no outliers	0.08461	0.09323
Experiment 2	Without outliers handling	0.08532	0.09608
Experiment 3	Using Log but No extra features	0.08615	0.09939
Experiment 4	Without log transformation	594.40971	0.10210
Experiment 5	Without log and outlier handling	595.66987	0.10238

V. CONCLUSION

This research aimed to enhance sales forecasting accuracy. The findings demonstrated that the XGBoost model significantly outperformed traditional statistical models, achieving an RMSPE of 0.09323 on the test dataset. The XGBoost model effectively captured complex patterns in the Rossmann stores' sales data, leveraging DTs and gradient boosting to extract valuable insights and intricate non-linear relationships. The superior performance of the XG-Boost model underscores the substantial advantages of ML over traditional statistical methods in sales prediction. This highlights the practical applicability of advanced ML techniques in improving sales forecasting accuracy for retail businesses.

A. STUDY LIMITATIONS

Despite these promising results, this study had limitations, including the exclusion of DL techniques and the reliance on a specific dataset from Rossmann stores. Future research should explore additional ML algorithms and feature engineering techniques to further enhance sales predictions. Additionally, incorporating DL approaches could provide further improvements in predictive accuracy. In conclusion, this study contributed to the existing literature by demonstrating the efficacy of the XGBoost model in sales forecasting. The low RMSPE error achieved signifies the potential of machine learning to address complex forecasting challenges in the retail industry. Future research should continue to explore innovative methods to refine and improve sales prediction models.

Author Contribution

Asif Ahsan: conceptualization, methodology, software, formal analysis, investigation, data curation, validation, visualization, writing – original draft. **Ayesha Awan:** methodology, validation, writing – review & editing. **Syed Mohammad Irteza:** supervision, formal analysis, writing – review & editing.

Conflict of Interest

The authors of the manuscript have no financial or non-financial conflict of interest in the subject matter or materials discussed in this manuscript.

Data Availability Statement

Data supporting the findings of this study will be made available by the corresponding author upon request.

Funding Details

No funding has been received for this research.

Generative AI Disclosure Statement

The authors did not use any type of generative artificial intelligence software for this research.

REFERENCES

[1] A. Ali, N. Khan, M. Abu-Tair, J. Noppen, S. McClean, and I. McChesney, "Discriminating features-based cost-sensitive approach for software defect prediction," *Autom. Softw. Eng.*, vol. 28, no. 2, Art. no. 11, June 2021, doi: https://doi.org/10.1007/s10515-021-00289-8.

[2] S. Coleman and A. Ahlemeyer-Stubbe, *A Practical Guide to Data Mining for Business and Industry*. Chichester, U.K.: John Wiley & Sons, 2014, doi: https://doi.org/10.1002/9781118763704.

[3] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*, vol. 103. New York, NY, USA: Springer, 2013.

[4] S. Palamalai, "Demand and operational planning: Optimizing efficiency and meeting customer needs," *SAARJ J. Banking Insurance Res.*, vol. 11, no. 1, pp. 94–103, 2022.

[5] C. Bandt, "Small order patterns in big time series: A practical guide," *Entropy*, vol. 21, no. 6, Art. no. 613, June 2019, doi: https://doi.org/10.3390/e21060613.

[6] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: https://doi.org/10.1023/A:1017934522171.

[7] M. R. Minar and J. Naher, "Recent advances in deep learning: An overview," *arXiv* preprint arXiv:1807.08169, Jul. 2018, doi: https://doi.org/10.48550/arXiv.1807.08169.

[8] J. Malik, A. Akhunzada, I. Bibi, M. Imran, A. Musaddiq, and S. W. Kim, "Hybrid deep learning: An efficient reconnaissance and surveillance detection mechanism in SDN," *IEEE Access*, vol. 8, pp. 134695–134706, 2020, doi: https://doi.org/10.1109/ACCESS.2020.3009849.

[9] Y. Du, J. Wu, S. Yang, and L. Zhou, "Predicting vehicle fuel consumption patterns using floating vehicle data," *J. Environ. Sci.*, vol. 59, pp. 24–29, Sep. 2017, doi: https://doi.org/10.1016/j.jes.2017.03.008.

[10] S. Lin, E. Yu, and X. Guo, "Forecasting Rossmann store sales leading 6-month sales." [Online]. Available: https://cs229.stanford.edu/proj2015/192_report.pdf.

[11] B. M. Pavlyshenko, "Forecasting of non-stationary sales time series using deep learning," *arXiv* preprint arXiv:2205.11636, May 2022, doi: https://doi.org/10.48550/arXiv.2205.11636.

[12] E. Taghizadeh, "Utilizing artificial neural networks to predict demand for weather-sensitive products at retail stores," *arXiv* preprint arXiv:1711.08325, Nov. 2017, doi: https://doi.org/10.48550/arXiv.1711.08325.

[13] S. Wickramanayake and H. D. Bandara, "Fuel consumption prediction of fleet vehicles using machine learning: A comparative study," in *Proc. Moratuwa Eng. Res. Conf. (MERCon)*, Apr. 2016, pp. 90–95, doi: https://doi.org/10.1109/MERCon.2016.7480121.

[14] Y. Dai and J. Huang, "A sales prediction method based on LSTM with hyper-parameter search," in *J. Phys. Conf. Ser.*, vol. 1756, no. 1, Art. no. 012015, Jan. 2021, doi: https://doi.org/10.1088/1742-6596/1756/1/012015.

[15] S. K. Shetty and R. Buktar, "A comparative study of automobile sales forecasting with ARIMA, SARIMA and deep learning LSTM model," *Int. J. Adv. Oper. Manag.*, vol. 14, no. 4, pp. 366–387, 2022, doi: https://doi.org/10.1504/IJAOM.2022.10052792.

[16] N. Mahbub, S. K. Paul, and A. Azeem, "A neural approach to product demand forecasting," *Int. J. Ind. Syst. Eng.*, vol. 15, no. 1, pp. 1–18, 2013, doi: https://doi.org/10.1504/IJISE.2013.055508.

[17] A. Wellens, R. Boute, and M. Udenio, "A tree-based framework to democratize large-scale retail sales forecasting with big data," *SRN J.*, 2022, doi: https://doi.org/10.2139/ssrn.4213618

[18] M. Al Ali, "Retail demand forecasting," Master's thesis, Roch. Inst. Technol., Dubai, 2021. [Online]. Available: https://repository.rit.edu/theses/11093.

[19] E. C. Alexopoulos, "Introduction to multivariate regression analysis," *Hippokratia*, vol. 14, Suppl. 1, pp. 23–28, 2010.