[1]Ravikant Suryawanshi

[2]Sandeep Musale

[3]Dr. Snehal Bhosale

Comparative Analysis of use of Machine Learning Algorithm for Prediction of Sales

Journal of Electrical Systems

Abstract: - Predicting sales holds immense importance for businesses, particularly for large retailers. It significantly influences decision-making processes and their subsequent execution. For big retail stores engaged in marketing and sales, accurate sales prediction is crucial for effectively allocating resources, estimating sales revenue, and devising future strategies. It empowers these stores to grasp market trends, make informed decisions, and adapt to evolving conditions, ensuring long-term viability and prosperity. Given the dynamic nature of trends and market conditions, traditional forecasting methods face significant challenges in keeping pace with rapid changes. This is where the application of machine learning algorithms becomes indispensable, especially when dealing with large and complex datasets. This paper delves into the exploration of various machine learning algorithms and models to predict sales, conducting comparative analyses to assess their performance.

The overarching aim is to identify the most suitable algorithm or combination of algorithms through comprehensive analysis. Furthermore, the paper scrutinizes how the characteristics of the dataset impact the performance of these algorithms, offering valuable insights for selecting the most appropriate machine learning algorithm based on performance analysis. By leveraging these insights, businesses can enhance their sales prediction capabilities and make informed decisions, thereby maximizing their success in the marketplace.

Keywords: **Machine learning, Sales forecasting, Data mining, ARIMA, LSTM, XGBoost, Hybrid model,**

I. INTRODUCTION

In the past, there were small local grocery stores instead of big supermarkets or department stores. Back then, store owners could easily know their customers and understand what they liked and didn't like. But now, we have huge franchise businesses like Dmart, Big Bazaar, and Walmart with many stores, making it hard to know each customer well. These big businesses generate a massive amount of data every day, and this data keeps growing a lot.

Many businesses focus on using and managing this data to make smart decisions for the future. One important thing they do is sales forecasting, which means predicting future sales based on past sales data. This helps them avoid having too much of products that don't sell well and not enough of popular ones, making more profit. Accurate forecasts also help businesses use their money wisely, plan for future growth, and avoid risks.

The main goal of these businesses is to make money and keep customers happy to build a good reputation. These big stores have so much data that it's too hard to handle manually. Thanks to technology and machine learning, they

use special algorithms to process all this data. These advanced programs help them understand important information from the big data, make accurate predictions, and guide how they

make and sell products.

The study uses a dataset from a Kaggle competition, covering store sales from 2013 to 2018. The goal is to predict sales for the next month. The dataset undergoes analysis with four machine learning techniques including Linear Regression, Decision Trees, Random Forest, and XGBoost, alongside two models, ARIMA and LSTM. To enhance outcomes, the paper utilizes data preprocessing and visualization methods. Evaluation of performance relies on metrics such as Root Mean Square Error (RMSE), R-square, and Mean Absolute Error (MAE).RMSE measures the average difference between predicted and actual values, R-squared gauges how well predictions match real data, and MAE evaluates accuracy for continuous variables without considering direction. The models' performances are compared, and the best one is chosen for sales prediction. Hybrid models combining the two best-performing

[1]Assistant Professor, MKSSS's Cummins College of Engineering for Women, Pune, Maharashtra, India

ravikant.suryawanshi@cumminscollege.in

[2]Professor, MKSSS's Cummins College of Engineering for Women, Pune, Maharashtra, India

sandeep.musale@cumminscollege.in

[3]Department of Electronics and Telecommunication Engg. Simbiosis Institute of Technology (SIT), Symbiosis International (Deemed University) (SIU), Lavale, Pune, Maharashtra, India

snehal.bhosale@sitpune.edu.in, http://orcid.org/0000-0002-0275-3554

algorithms are also tested, with XGBoost identified as the top-performing model. The paper is divided into six parts.

In the beginning, the first part introduces the study. The second part looks at previous research on the topic. The third part explains how the research was done. The fourth part talks about the actual experiments conducted. The fifth part goes into the analysis of the results obtained from the research. Finally, the sixth part wraps up the study by presenting the conclusions drawn from the findings.

II. LITERATURE REVIEW

Regression models find application in various fields like predicting crime rates, assessing cardiovascular risk, health sector analyses, house price predictions, and sales forecasting. Specifically, XGBoost is utilized for cardiovascular risk prediction. Sales prediction, on the other hand, focuses on estimating the sales of products across different stores. The predictability of sales increases with the growing volume of products. Python serves as the programming language for this task, and Colab is the chosen tool. Machine learning features such as supervised learning and regression functions are employed in this context.

The process involves key steps like data processing, feature engineering, model design, and testing. The regression function utilizes multiple algorithms to forecast product sales. This includes tasks such as data deletion, cleaning, and transformation. The company's profit is directly tied to the accuracy of the sales forecast.

When it comes to predicting housing prices using regression techniques, people are careful when buying a new home. They take into account their budget and consider market strategies. The prediction process evaluates housing prices by considering the financial conditions and preferences of potential buyers.

In a developing country like India, where hearing about crime regularly is not uncommon, crime prediction through KNN (K-Nearest Neighbors) can be valuable. This data can offer insights into the criminal record of an area, aiding criminal investigations.

XGBoost, pioneered by Friedman, stands out among a plethora of machine learning models including Linear Regression, Support Vector Machines, and Naive Bayes, garnering substantial attention from both researchers and engineers. Renowned for its consistent delivery of exceptional results across diverse problem domains, XGBoost has become a staple tool in various studies. For instance, Gumus et al. harnessed XGBoost to delve into the factors influencing crude oil prices and predict future trends in oil pricing. Furthermore, in an innovative application of machine learning techniques, a hybrid approach combining Random Forest and XGBoost is adopted to devise a data-centric framework for detecting faults in wind turbines. Leveraging Random Forest's feature-ranking capabilities, this approach identifies crucial features, which are then utilized by XGBoost to construct ensemble classifiers tailored to specific fault types. Gurnani et al. devised a hybrid approach blending both linear and nonlinear models to forecast drug sales, utilizing decomposition techniques. Here, a linear model anticipates the linear component, whereas a nonlinear model forecasts the nonlinear component. Furthermore, Zhong et al. introduced a prediction framework reliant on XGBoost for discerning crucial proteins. They employed a sub-expand-shrink strategy to craft composite features, complemented by a model fusion technique to bolster prediction accuracy. Moreover, XGBoost finds application in predicting PM2.5 concentration by scrutinizing factors affecting PM2.5 levels. Through the elimination of insignificant features, the model achieves enhanced estimation performance.

In their work, the authors implemented a hybrid model. Initially, two separate models were built based on different frameworks. Then, weights were assigned to integrate the models based on their prediction results. This integrated model combines the characteristics of both models simultaneously and demonstrates improved predictive ability. Specifically, in crime prediction using the K-Nearest Neighbor algorithm, two machine learning algorithms, XGBoost and Random Forest, were integrated into the model for enhanced performance.

III. METHODOLOGY

The research focused on predicting store sales using various data mining methods. After applying these techniques, the researchers carefully analyzed and evaluated them to create more dependable models. They also used feature engineering, a method to modify input data and make it suitable for machine learning algorithms. The model's accuracy was enhanced by getting rid of unimportant features, making the performance estimates more effective.

A. Exploratory Data Analysis

Before training, it's crucial to examine the data to understand which attributes it contains. This initial exploration helps determine if the dataset is suitable for training and guides the next steps. We use different graphs and plots to

visualize the dataset, extracting valuable information. By doing so, we can observe trends and obtain specific results. Conducting an exploratory analysis is essential to gain a clear understanding of the data's nature.

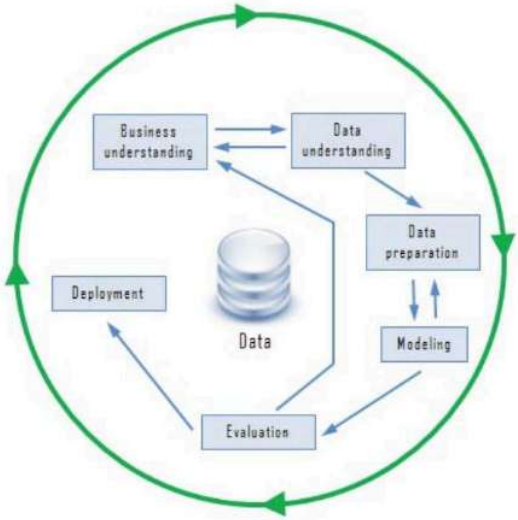


Figure 1. Data Mining Process

Data Understanding: The dataset utilized in this research is sourced from a Kaggle competition, focusing on item-level sales data across various store locations. To forecast store sales, historical sales records spanning five consecutive years (2013 to 2018) are taken into account. The dataset comprises four columns: date, store, item, and sales, totaling 913,000 rows. The date column indicates the sale date, excluding holidays or store closures. Analyzing the statistical aspects of the data, we created a plot illustrating the total monthly sales trend over time (see Fig 2). Notably, the figure indicates an upward trend in average monthly sales, revealing that our data is non-stationary.

Figure 2: Distribution of Sales per month

Data Exploration:

To make the data stationary, we take the difference between the sales of each month and add it as a new column in our data frame. First, we determine the time interval of the dataset in terms of days, years, and months. We then calculate the total sales per store and visualize it using a bar plot (refer to Fig 3). We also compute the mean monthly sales and the average monthly sales of the previous year, which serves as the forecasted sales for analysis.

Next, we analyze the time series to identify trends and seasonality, aiming to remove them and obtain a stationary series. Statistical forecasting techniques are then applied to this series. The forecasted values are converted back to the original scale by reintroducing trend and seasonality constraints. We plot the Monthly Sales before Differential Transformation in Fig 4. In Fig 5, we display the plot of Monthly Sales after Differential Transformation. Differencing involves calculating the difference between consecutive terms in the series, typically used to eliminate varying means. In this case, we compute the month-over-month difference in sales.

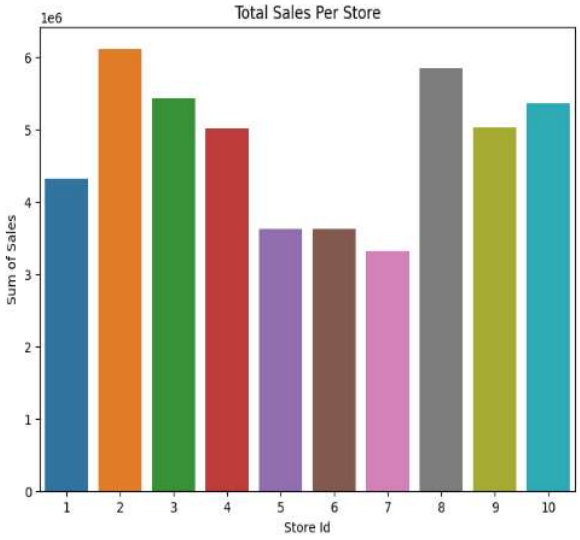


Figure 3. Total Sales per store

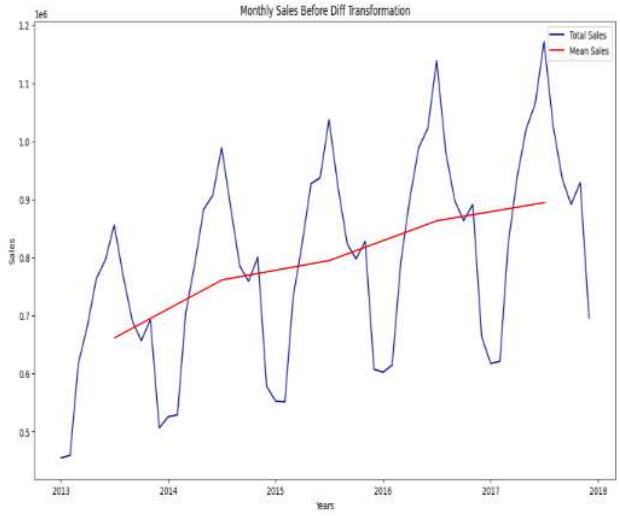


Figure 4. Monthly Sales Before Differential Transformation

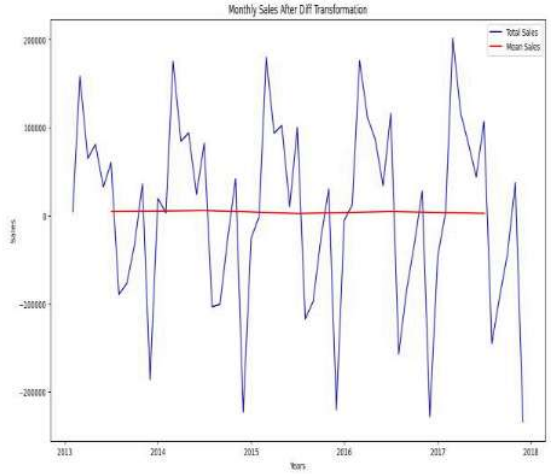


Figure 5. Monthly Sales After Differential Transformation.

Modeling:

Now that our data has been transformed to ensure stationarity and is organized into monthly sales, we begin preparing it for various model types. We establish two distinct structures for this purpose:

a. ARIMA Modeling Structure: To facilitate ARIMA modeling, we construct a data frame with a datetime index and columns representing the dependent variable (sales difference).

b. Other Models Structure: For all other models, we create a new data frame wherein each feature corresponds to the sales of the previous month.

Determining the appropriate number of months to include in our feature set involves examining autocorrelation and partial autocorrelation plots (refer to Fig 6). Following our observations, we opt for a look-back period of 12 months. Accordingly, we generate a data frame containing 13 columns: twelve for each of the preceding months and one for the dependent variable, i.e., the difference in sales (see Fig 7).

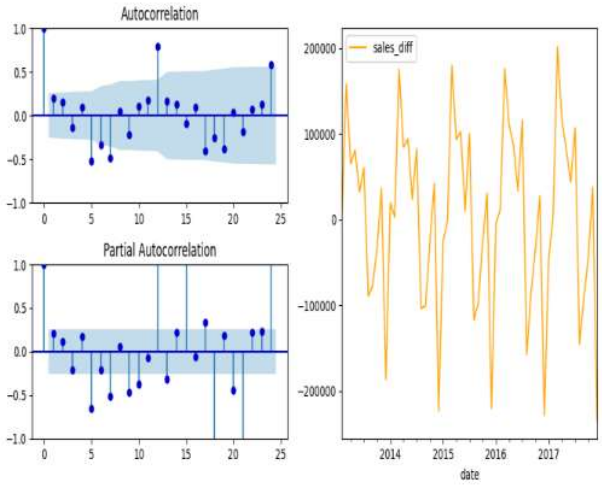


Figure 6. Plot of Auto-correlation and Partial Correlation

	date	sales	sales_diff	lag_1	lag_2	lag_3	lag_4	lag_5	lag_6	lag_7	lag_8	lag_9	lag_10	lag_11	lag_12
0	2014-02-01	528117	3130.0	19380.0	-186036.0	36656.0	-33320.0	-76854.0	-89161.0	60325.0	32355.0	80968.0	64982.0	157965.0	4513.0
1	2014-03-01	704301	175184.0	3130.0	19380.0	-186036.0	36656.0	-33320.0	-76854.0	-89161.0	60325.0	32355.0	80968.0	64982.0	157965.0
2	2014-04-01	788914	84613.0	175184.0	3130.0	19380.0	-186036.0	36656.0	-33320.0	-76854.0	-89161.0	60325.0	32355.0	80968.0	64982.0
3	2014-05-01	882877	93963.0	84613.0	175184.0	3130.0	19380.0	-186036.0	36656.0	-33320.0	-76854.0	-89161.0	60325.0	32355.0	80968.0
4	2014-06-01	906842	23965.0	93963.0	84613.0	175184.0	3130.0	19380.0	-186036.0	36656.0	-33320.0	-76854.0	-89161.0	60325.0	32355.0
5	2014-07-01	989010	82168.0	23965.0	93963.0	84613.0	175184.0	3130.0	19380.0	-186036.0	36656.0	-33320.0	-76854.0	-89161.0	60325.0
6	2014-08-01	885596	-103414.0	82168.0	23965.0	93963.0	84613.0	175184.0	3130.0	19380.0	-186036.0	36656.0	-33320.0	-76854.0	-89161.0
7	2014-09-01	785124	-100472.0	-103414.0	82168.0	23965.0	93963.0	84613.0	175184.0	3130.0	19380.0	-186036.0	36656.0	-33320.0	-76854.0
8	2014-10-01	758883	-26241.0	-100472.0	-103414.0	82168.0	23965.0	93963.0	84613.0	175184.0	3130.0	19380.0	-186036.0	36656.0	-33320.0
9	2014-11-01	800783	41900.0	-26241.0	-100472.0	-103414.0	82168.0	23965.0	93963.0	84613.0	175184.0	3130.0	19380.0	-186036.0	36656.0
10	2014-12-01	578048	-222735.0	41900.0	-26241.0	-100472.0	-103414.0	82168.0	23965.0	93963.0	84613.0	175184.0	3130.0	19380.0	-186036.0
11	2015-01-01	552513	-25535.0	-222735.0	41900.0	-26241.0	-100472.0	-103414.0	82168.0	23965.0	93963.0	84613.0	175184.0	3130.0	19380.0
12	2015-02-01	561317	8804.0	-25535.0	-222735.0	41900.0	-26241.0	-100472.0	-103414.0	82168.0	23965.0	93963.0	84613.0	175184.0	3130.0

Figure 7. Supervised Structure which includes Lags

B. Train-Test Split data

To assess the performance of machine learning algorithms, we implemented a train-test split with a ratio of 80:20. This means that 80% of the data is assigned to the training set, where the model learns from known outputs. The goal is for the model to generalize well to new data. In this case, the model was trained using sales data from 2013 to 2017. The remaining 20% of the data forms the test set, allowing us to evaluate the model's predictions. For testing, we used sales data specifically from the year 2018.

C. Algorithms/Models used

Having completed the preceding phases, the dataset is now prepared for building models. In this paper, we introduce a set of techniques of machine learning, including Linear Regression, Decision Tree, Random Forest, XGBoost, and two deep learning models, namely ARIMA and LSTM. These models will be applied to predict and analyze sales based on the transformed and structured data.

1. Linear Regression

Linear Regression is a statistical method commonly employed for predictive analysis. It highlights the linear relationship between variables, demonstrating how the dependent variable changes concerning the values of the independent variable. The formula for Linear Regression is often represented as:

$$Y=aX+b$$

Here:

Y is the target variable.

X is the predictor variable.

a is the coefficient of Linear Regression.

b is the intercept of the line.

In this context, *X* and *Y* represent the datasets used in the Linear Regression model. The model aims to find the best-fitting line that describes the relationship between these variables.

2. Decision Trees

A Decision Tree functions as a supervised learning method, operating as a tree-shaped classifier wherein internal nodes represent dataset features, branches encapsulate decision rules, and each leaf node signifies a particular outcome. Within this tree structure, decision nodes facilitate decisions with multiple branches, whereas leaf nodes serve as the ultimate output of those decisions, leading to no further branches. Decision Trees offer significant utility in structuring and illustrating decision-making processes systematically and hierarchically.

3. Random Forest

The Random Forest concept is rooted in ensemble learning, a methodology that amalgamates multiple classifiers to tackle intricate problems. Specifically, Random Forest constitutes a classifier comprising numerous decision trees, each trained on distinct subsets of the provided dataset. Subsequently, the ensemble aggregates the predictions from these trees to bolster the predictive accuracy of the dataset. Typically, a greater number of trees within the forest correlates with heightened overall accuracy. This technique of consolidating multiple decision trees serves to enhance the model's resilience and capacity for generalization.

4. XGBoost

XGBoost, which stands for Extreme Gradient Boosting, represents an implementation of Gradient Boosted decision trees. Within the gradient boosting framework, each predictor endeavors to rectify the errors of its predecessor. This is achieved by training each predictor using the residual errors of the preceding one as labels, leading to sequential creation of decision trees.

In XGBoost, the weights assigned to independent variables hold considerable significance. These weights are allocated to all independent variables and fed into the decision tree for result prediction. Notably, the weights of variables that the tree incorrectly predicts are augmented. Subsequently, these mispredicted variables are employed as input for the subsequent decision tree in the sequence. Ultimately, these individual classifiers or predictors ensemble together, culminating in a robust and highly accurate model.The iterative nature of XGBoost, with its emphasis on rectifying errors, contributes significantly to its effectiveness in predictive modeling.

5. ARIMA

ARIMA, short for Autoregressive Integrated Moving Average, represents a statistical analysis model tailored for time series data, utilized to glean insights from a dataset or forecast future trends. This model leverages autocorrelations and moving averages across residual errors within the data to predict forthcoming values.

a. *Auto Regressive (AR):* The AR component of ARIMA hinges on the autocorrelation principle. In this regression model, the dependent variable is predicated on its past values, whereby the current value is forecasted based on its own previous values.

b. *Integrated:* The integrated aspect of ARIMA endeavors to transform the non-stationary nature of time-series data into a relatively more stationary form. This involves differentiating the data to render it more amenable to analysis.

c. *Moving Average (MA):* The MA element of ARIMA aims to mitigate noise in the time series data by employing an aggregation operation on past observations, accounting for the residual error ε. This aids in smoothing out fluctuations within the data.

By amalgamating these three components, ARIMA furnishes a holistic approach to scrutinizing and predicting time series data.

6. LSTM

Long Short-Term Memory (LSTM) represents an artificial recurrent neural network engineered to excel at learning long-term dependencies, particularly in sequence prediction tasks. The hallmark of an LSTM model lies in its memory cell, known as a 'cell state,' which preserves its state across time, enabling the model to retain information from past inputs over an extended period.

LSTM comprises three pivotal gates:

Forget gate: Responsible for determining what information from the cell state to discard or retain, the forget gate considers both the previous output and the current input.

Input gate: The input gate governs the update to the cell state by determining which new information to incorporate. It employs the sigmoid activation function to decide which values warrant updates.

Output gate: Tasked with determining the subsequent hidden state and generating the model's output, the output gate employs the tanh activation function to control the values for output.

By integrating these three gates, LSTM networks are adept at capturing and leveraging long-term dependencies, making them particularly effective in sequence prediction challenges. The combination of these gates, along with the sigmoid and tanh activation functions, enables LSTMs to effectively manage and utilize information over prolonged sequences, making them particularly adept at addressing long-term dependencies in data.

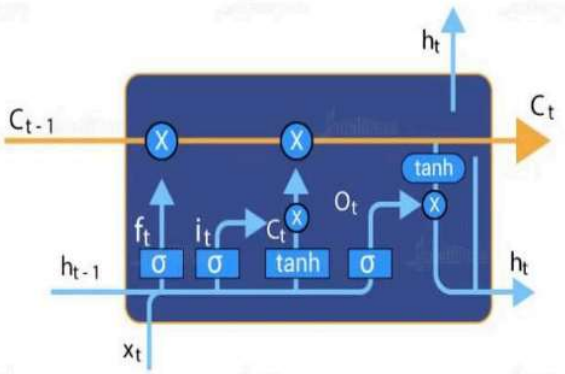


Figure 8. LSTM model

IV. EXPERIMENTATION

In addition to our main dataset, we also explored two additional datasets: the Black Friday Sales dataset and the Big Mart Store dataset. The Black Friday Sales dataset focuses on predicting company sales during the shopping event following Thanksgiving. With 8523 observations and 12 features including Product ID, User ID, age, gender, occupation, years stayed in the current city, marital status, and product categories, the goal is to forecast sales based on the Black Friday seasonality.

We applied various machine learning algorithms such as Linear Regression, Decision Tree, Random Forest, and XGBoost to this dataset. Among these, the Random Forest algorithm demonstrated the highest accuracy, achieving around 81%. Random Forest excels in handling tabular data with both categorical and numerical features, capturing nonlinear relationships between features and target variables. By combining multiple decision trees, it provides

more reliable predictions compared to individual trees. Additionally, the Random Forest model incorporates outlier detection, contributing to its superior performance.

Our research suggests that when predicting the product, a customer is likely to purchase based on factors like gender, age, and occupation, this model yields accurate results. This highlights the importance of leveraging machine learning techniques to understand consumer behavior and enhance sales predictions in retail settings.

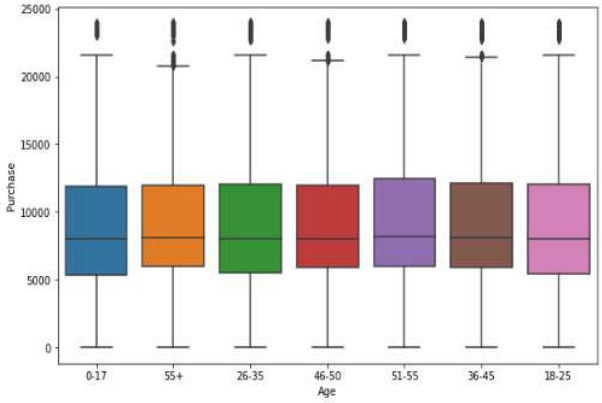


Figure 9. Outlier detection of purchase according to age

The Big Mart Sales Data comprises 12 attributes and encompasses 8523 products distributed across various cities. During exploratory data analysis (EDA), several nuances were identified within the dataset. To address these issues, feature engineering techniques were applied. Following these preparatory steps, we proceeded to build models using four machine learning algorithms: Linear Regression, Decision Trees, Random Forest, and XGBoost.

Upon evaluating the performance of these algorithms, Linear Regression emerged as the most suitable choice for the Big Mart dataset. Its Root Mean Square Error (RMSE) value was lower compared to the other algorithms, indicating superior predictive accuracy. Consequently, Linear Regression was applied to the test data to obtain prediction results.

The effectiveness of Linear Regression stemmed from a strong relationship observed between the attributes "item_mrp" and "item_outlet_sales." This relationship was further elucidated through a heatmap visualization, as depicted in Fig 10. In the heatmap, values closer to 1 indicate a strong correlation between two variables, while negative values signify weaker or negligible correlation. The heatmap analysis helped in understanding the interrelationship between attributes and informed the selection of Linear Regression as the preferred model for the Big Mart dataset.

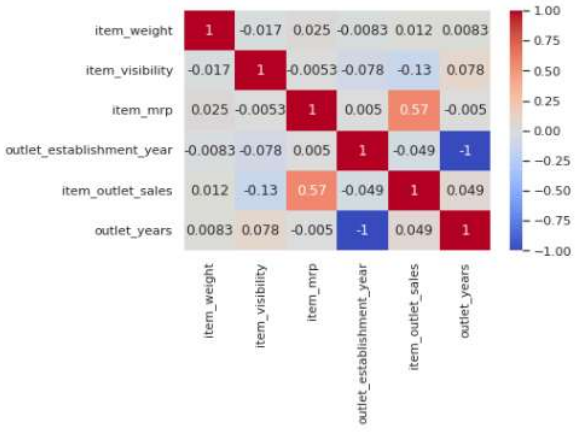


Figure 10. Heatmap

V. RESULTS AND ANALYSIS

Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are calculated using formulae

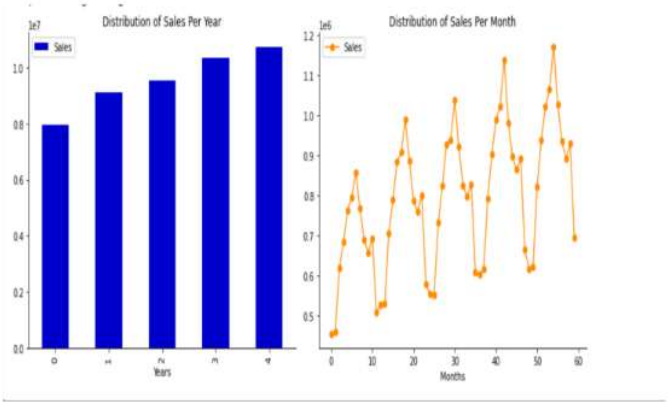


Figure. 11 Graph shows the distribution of sales per year and the distribution of sales per month respectively

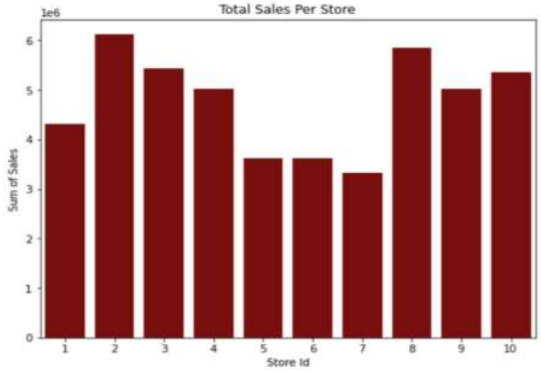


Figure 12 Shows total sales per store

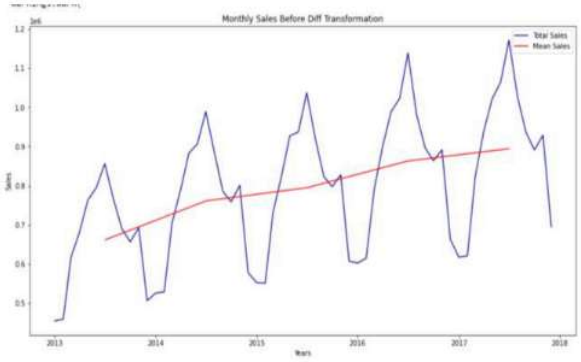


Figure 13 For seasonality forecasting we are doing this diff transformation; Graph shows monthly sales before diff transformation.

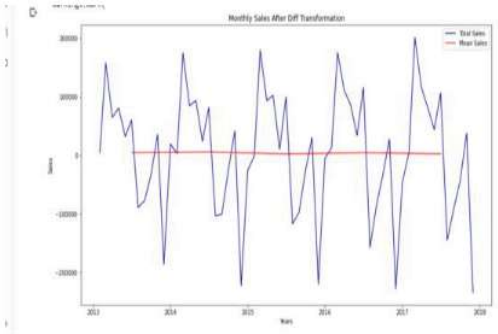


Figure 14 Result after diff transformation.

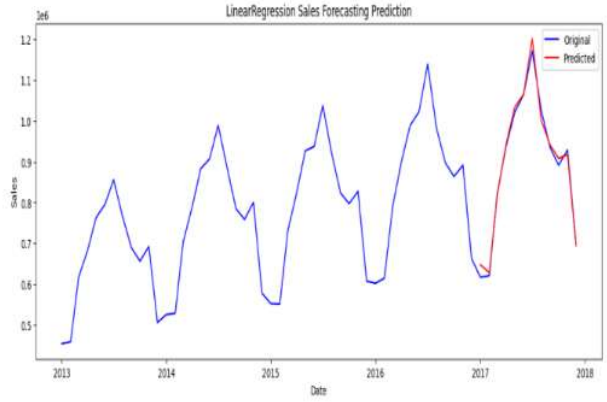


Figure 15 Linear regression sales forecasting prediction

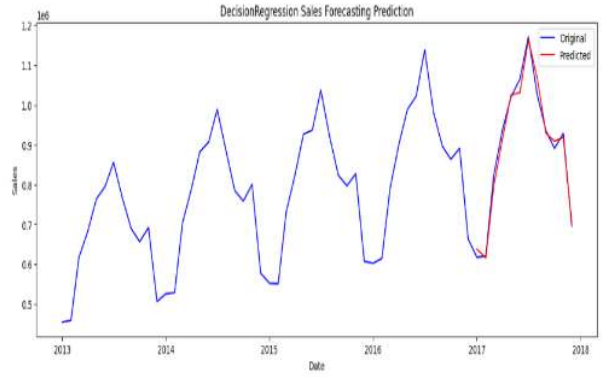


Figure 16 Decision Regression sales forecasting prediction

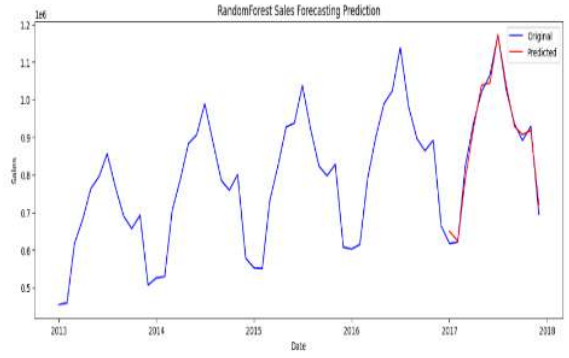


Figure 17 Random Forest sales forecasting prediction

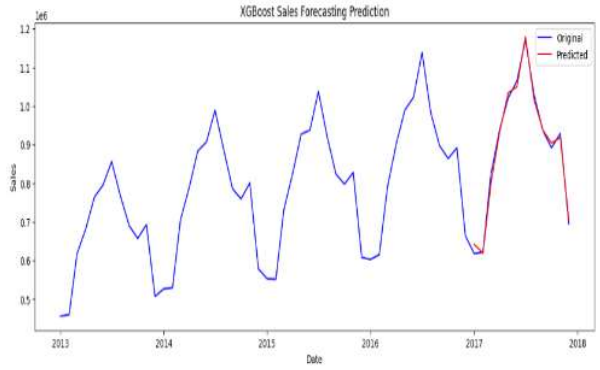


Figure 18 XGBoost sales forecasting prediction

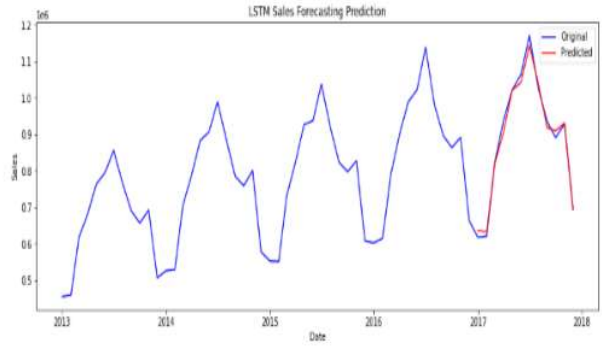


Figure.19 LSTM model

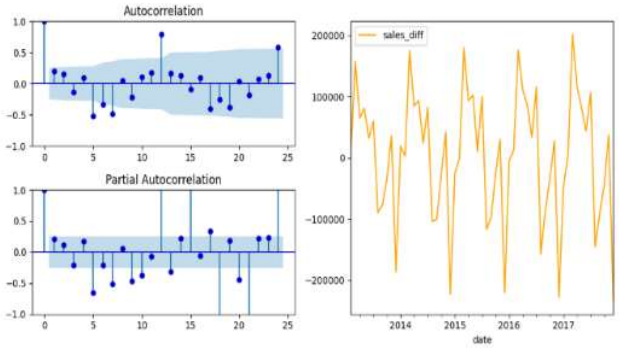


Figure.20.1 Arima model: Calculating Lags

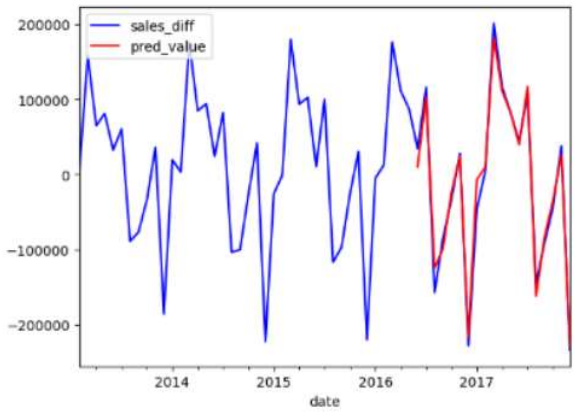
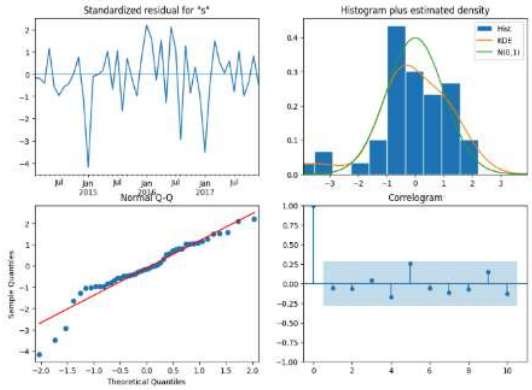


Figure.20.2 Arima model: Arima Sales forecasting prediction



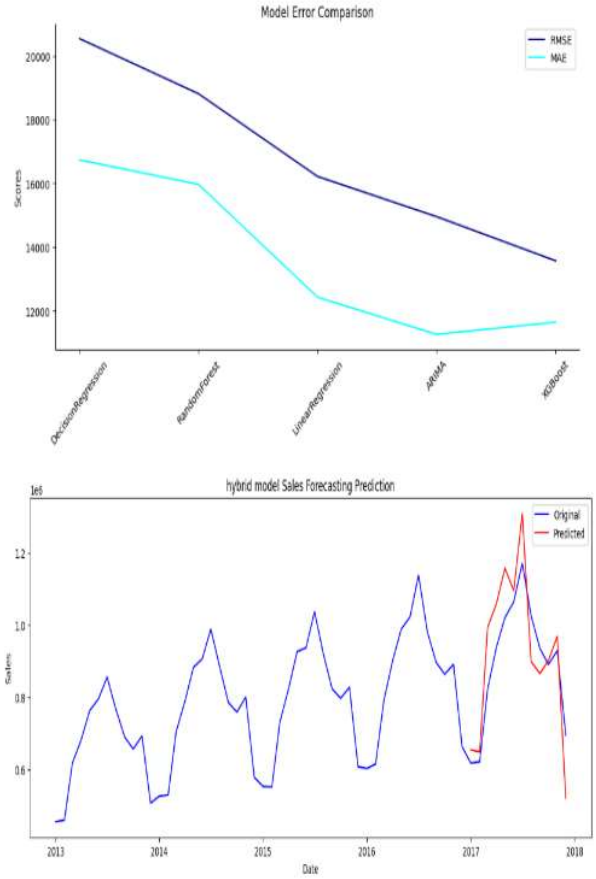


Figure.21 Hybrid Model - XGBoost & RandomForest

VI. CONCLUSION

In this paper, our primary objective was to predict sales for the year 2018 based on data from the years 2013 to 2017. We expanded our analysis by applying various machine learning algorithms to two additional datasets, alongside the store dataset used in this project. This allowed us to assess which algorithms perform well with different types of datasets.

Through our analysis, we found that XGBoost performs effectively with small to medium-sized datasets, especially those with structured data and few features. Given that our main dataset contains only three columns and 913,000 rows, indicating a small-sized dataset with limited features, XGBoost proved to be a suitable choice.

On the other hand, Random Forest excels with datasets containing numerous features and nonlinear relationships between features and the target variable. Since our Black Friday dataset contained many features with both categorical and numerical values, along with nonlinearity between these features and the target, Random Forest yielded superior prediction results compared to XGBoost.

Decision trees are useful when all features in the dataset are required for analysis and prediction. However, in cases where more accurate predictions are needed, Random Forest, which combines multiple decision trees, tends to be preferred over individual decision trees.

Linear Regression performs well when there is a strong relationship between the target variable and the feature used for prediction. Therefore, it is suitable for datasets where such strong relationships exist.

Ultimately, the choice of algorithm depends on the characteristics of the dataset. By employing appropriate algorithms, we can effectively forecast future sales. This prediction model enables businesses to accurately determine sales for specific items during particular periods, allowing for informed decision-making and action planning.

REFERE Figure NCES :

[1] Melvin Tom, Nayana Raju, Asha Issac, Jeswin James, Saritha R "Supermarket Sales Prediction Using Regression" International Journal of Advanced Trends in Computer Science and Engineering.

[2] Sunita Cheriyan, Saju Mohanan,Shaniba Ibrahim's , "Intelligent Sales Prediction Using Machine Learning Techniques ".

[3] H V Ramachandra, Balaraju G, Rajashekhar A,Harish Patil ,"Machine Learning application for Black Friday Sales Prediction Framework ."2021 International Conference on Emerging Smart Computing and Informatics (ESCI).

[4] Yiyang Niu "Walmart Sales Forecasting using XGBoost algorithm and Feature Engineering"2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)

[5] GopalKrishnan T,Ritesh Choudhary ,Sarada Prasad "Prediction of sales Value in online shopping using Linear Regression."2018 4th International Conference on Computing Communication and Automation (ICCCA).

[6] Xie Dairu, Zhang Shilong "Machine learning Model for Sales Forecasting by Using XGBoost" 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)

[7] Jingru Wang, "A hybrid machine learning model for sales prediction "2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI)

[8] Natalya V. Razmochaeva,Dmitry M. Klionsky, Nikita V. Popov "Comparative Analysis of Feature Extraction Algorithms in Investigation of Products Sales Data".

[9] Xiaojun Zhang, lisheng Pei, Xiaojun Ye, "Demographic Transformation and Clustering of Transactional Data for Sales Prediction of Convenience Stores ".

[10] Ren X, Guo H, Li S , et al. A Novel Image Classification Method withCNN-XGBoost Model[C] International Workshop on Digital Watermarking. 2017.

[11] Zhang D, Qian L, Mao B, et al. A data-driven design for fault detection of wind turbines using random forests and XGboos [J] IEEE Access 2018, 6: 21020-21031.

[12] Gumus M, Kiran M S. Crude oil price forecasting using XGBoost[C]//2017 International Conference on Computer Science and Engineering (UBMK). IEEE, 2017: 1100-1103.

[13] Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]//Proceedings of the 22nd international conference on knowledge discovery and data mining. 2016: 785-794.

[14] J. T. Hancock and T. M. Khoshgoftaar, "Survey on categorical data for neural networks," Journal of Big Data, vol. 7, no. 1, 2020.

[15] S. K. Sharma, S. Chakraborti and T. Jha, "Analysis of book sales prediction at Amazon marketplace in India: a machine learning approach," Information Systems and e-Business Management, pp. 261-284, 2019.

[16] Koutsoukas, K. J. Monaghan, X. Li and J. Huan, "Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data," Journal of Cheminformatics, 2017.

[17] Goodfellow, ı. Bengio and A. Courville, Deep Learning, MIT Press, 2016.

[18] Loureiro, ٧. Migueis and L. F. d. Silva, "Exploring the use of deep neural networks for sales forecasting in fashion retail," Decision Support Systems, vol. 114, pp. 81-93, 2018.

[19] Pavelkova, L. Homolka, J. Chytilova, ٧. M. Ngo, L. T. Bach and B.Denhing, "Passenger Car Sales Projections: Measuring the Accuracy of a Sales Forecasting Model," Ekonomicky casopis, vol. 66, pp. 227-249, 2018.

[20] Giri, S. Thomassey, J. Balkow and X. Zeng, "Forecasting New Apparel Sales Using Deep Learning and Nonlinear Neural Network Regression," 2019 International Conference on Engineering, Science, and Industrial Applications (ICESI), 2019.

[21] I.-F. Chen and C.-J. Lu, "Sales forecasting by combining clustering and machine-learning techni،ues for computer retailing," Neural Comput & Applic,2017.

[22] H. ı uan, C. ı ang and W. Xu, "A User Behavior-based Ticket Sales Prediction Using Data Mining Tools: An Empirical Stuy in an OTA Company," IEEE, 2014.

[23] ٧. Kotu and B. Deshpande, Predictive Analytics and Data Mining, Waltham: Elsevier Inc., 2015.

[24] Beheshti-Kashi, S., Karimi, H.R., Thoben, K.D., Lutjen, M. A survey on retail sales forecasting and prediction in fashion markets" Systems Science & Control Engineering 3(1), 154,161(2015)

[25] Smith, Oliver, and Thomas Raymen. "Shopping with Violence: Black Friday sales in the British context." Journal of Consumer Culture 17.3(2017): 677-694.

[26] Majumder, Goutam. "Analysis And Prediction Of Consumer Behaviour On Black Friday Sales." Journal of the Gujarat Research Society 21.10s (2019): 235-242.

[27] Challagulla, Venkata Udaya B., et al. "Empirical assessment of machine learning based software defect of prediction techniques." International Journal on Artificial Intelligence Tools 17.02 (2008): 389-400.

[28] Chu, C.W., Zhang, G.P.: "A comparative study of linear and Nonlinear models aggregate retail sales forecasting," International Journal of production economics 86(3), 217{231(2003)}