Wide -Ranging Retail Sales Prediction Using Machine Learning

G.Litisha Miraclin[1], Dr.Y.Harold Robinson[2]
[1]PG Student, Dept of Computer Science & Engineering, Francis Xavier Engineering College
Dr.Y.Harold Robinson
[2]Professor, Dept of Computer Science & Engineering, Francis Xavier Engineering College

***Abstract*—The use of machine learning predicts future sales by training a model using previous sales data. The overview for sales projection in machine learning highlights the use of several machine learning methods to forecast sales. The approach includes processes such as data collection and the preprocessing process feature selection, and training models. Different metrics are employed to evaluate the model's accuracy and effectiveness. The study's purpose is to give details about the greatest algorithms for machine learning for forecasting sales and how they may be applied in various organizations. The above report assesses the efficacy of predictive machine learning models in predicting sales across sectors. The study examines past sales data and evaluates the efficacy of several machine learning methods, including. Logistics Regression, and XGBoost Regressor in forecasting earnings in the future. Businesses can use this method to make the most of the assets they have and boost profits. Currently, as the sum of information available grows dramatically, organizations are focusing towards the rational application of massive amounts of information to help predict the next generation and arrive at better choices. In the past few decades, investigators and companies have grown more interested in applying predictive machine learning algorithms for predicting brand and commodity sales. This work presents the XGBoost sale forecasting method, and this includes an algorithm called XGBoost with painstaking design of features treatment to anticipate the sales problem. The strategy presented in this study successfully mines qualities from various perspectives in order to create accurate forecasts. This study assesses their XGBoost revenue forecasting algorithm using data on sales from stores obtained through the Kaggle challenge. Experimental findings reveal that this approach outperforms the other training methodologies.**

***Index Terms*—machine Learning, Sales Prediction, XG-Boost Regression**

I. INTRODUCTION

Customers who buy bread often also purchase related products like milk, butter, and jam. Placing these categories together in a retail centre allows for convenient access for clients. Identical goods should be displayed next to one other to guide clients around the centre in a sensible arrangement. Enterprises that are spending heavily in data collection from customers collecting possess the difficulty of extracting valuable information across large user and brand collections to obtain a financial edge. Many aspects of the evaluation of market baskets have been extensively researched in academic journals, including the use of client interest profiles and preferences for specific items enabling one-on-one marketing, as well as buying within a multi-store context to increase revenues. Analysis of market baskets has been widely utilized for numerous businesses to figure out category linkages and build a retailer's promotional plan on it. For the present, the volume of evidence is increasing dramatically. Several companies are now focusing on finding ways to leverage their current information in a reasonable and successful manner to support upcoming company-wide choices. Sales predictions will assist to prevent the lack of hot-selling items future the build-up of disliked items, improving shop earnings, and these are critical to the growth of both present-day and future industry. As a result, the study subject of this article focuses on solving the matter of Wide Range Big Mart sales projection, especially is provided with by the Kaggle competition. This article executes tests using the Kaggle challenge information set, that includes previous sales information for Walmart stores, to estimate the gross revenue of each shop. And The programming language Python is the primary method of analysing data, modelling purposes, and research.

During the testing phase, this work performed complex algorithms for feature analysing, such as database compression, temporal recognition, statistics decision-making, and essential feature selection. The results from experiments show the said paper's XGBoost-based sale prediction algorithm outperformed various machine learning models, including the Logistic Regression method and the Ridge algorithm. Furthermore, the article investigates the priority rating of several traits and derives various relevant conclusions to guide further studies. Research shows the said paper's proposal employing the XGBoost algorithm is beneficial overall sales forecasting jobs.

II. LITERATURE SURVEY

As reported in [1] by Z. Huo et al (2021) This paper explores the application of several machine learning algorithms, including linear regression, decision trees, and random forests, for sales prediction. The authors focus on how historical sales data can be leveraged to improve accuracy in forecasting. They highlight the importance of feature selection and data preprocessing in improving model performance, achieving an accuracy of over 85%. This research provides insights into the use of regression techniques to forecast short-term sales trends.

As reported in [2] Y. Katai et.al (2019) This paper presents a comprehensive study of machine learning algorithms applied to sales forecasting. The authors compare traditional statistical methods like ARIMA with advanced algorithms such as Decision Trees, Random Forests, and Neural Networks. They emphasize the advantages of using machine learning for capturing non-linear patterns in sales data. The case study revolves around a retail company, showing how these algorithms can significantly enhance prediction accuracy.

As reported in [3] by H. Upadhyay, J et.al (2023) The paper introduces Facebook Prophet, a robust and scalable tool for time series forecasting, especially designed for business applications like sales prediction. It explains how Prophet outperforms traditional time series models like ARIMA by accounting for seasonality, holidays, and business cycles. Taylor and Letham provide a detailed technical breakdown of the algorithm, showing its flexibility in handling missing data and irregular time intervals, making it ideal for e-commerce and retail sales forecasts.

As reported in [4] by S. N. Gunjal et.al (2022) This study compares the predictive capabilities of Artificial Neural Networks (ANNs) with traditional statistical methods like ARIMA for aggregate retail sales forecasting. The authors analyze retail data spanning multiple years, applying both models to evaluate their accuracy. The results show that ANNs outperform traditional methods in capturing complex, nonlinear relationships in sales data, especially when dealing with large datasets.

As reported in [5] by C. S. Rekha, et.al (2024) This paper explores the effectiveness of various machine learning models, including Gradient Boosting, Support Vector Machines, and k-Nearest Neighbors, in predicting sales for retail businesses. Ahmad and Lutfiyya compare these models with classical statistical methods, concluding that machine learning techniques offer superior accuracy, especially in handling large and complex datasets. The authors emphasize the importance of hyperparameter tuning and cross-validation in ensuring model performance.

As reported in [6] by T. Rajasree, et.al (2024) This paper discusses the use of machine learning techniques to predict sales in retail settings. The authors compare several machine learning algorithms, including decision trees, random forests, and support vector machines (SVM), to predict monthly sales. They use historical data from a retail chain, focusing on factors such as seasonality, promotions, and market trends. The paper highlights that Random Forest and Gradient Boosting Machines (GBMs) yield the highest accuracy in sales predictions, outperforming traditional statistical methods like ARIMA

As reported in [7] by Sadasivam V.R, et.al (2019) In this paper, the authors explore the potential of neural networks, particularly deep learning, in predicting sales within an e-commerce platform. The study compares deep neural networks (DNN) with shallow neural networks and traditional machine learning methods such as linear regression and k-nearest neighbors (KNN). The authors use a dataset containing sales records, product information, customer interactions, and promotional campaign data.

As reported in [8] by M. Spuritha et.al (2021) In this paper, the authors present a sales prediction model for e-commerce platforms using machine learning techniques. The dataset used includes historical sales data, product details, customer reviews, and promotional activities. The authors evaluate multiple algorithms such as support vector machines (SVM), decision trees, and ensemble methods like XGBoost

As reported in [9] by G. T., R. Choudhary et.al (2018) This paper provides a comprehensive review of machine learning techniques applied in sales forecasting, focusing on the retail industry. The authors compare various machine learning models, including linear regression, decision trees, random forests, and deep learning models like recurrent neural networks (RNNs) and long short-term memory (LSTM).

As reported in [10] by D. S. AbdElminaam et.al (2024) This paper explores various machine learning models for predicting daily sales in retail stores. The authors conduct an experimental study comparing traditional statistical models, such as ARIMA, with machine learning algorithms like random forests, k-nearest neighbors (KNN), and deep learning models. Using a dataset from a grocery chain, they assess the performance of these models based on accuracy, computational cost, and scalability. The results show that random forests and deep learning models outperform ARIMA in terms of accuracy, especially for complex sales patterns influenced by factors like promotions and holidays.

As reported in [11] by J. Chen, et.al (2021) This paper investigates the application of long short-term memory (LSTM) networks for predicting retail sales, particularly in capturing seasonal trends and time dependencies. The authors use a dataset from a large department store chain, including daily sales data, customer transactions, and promotional activities. They compare the performance of LSTM with traditional time series models like ARIMA and Exponential Smoothing. The findings demonstrate that LSTM models outperform traditional methods, particularly in handling complex patterns like seasonality and short-term fluctuations in sales. One notable advantage of LSTM is its ability to retain long-term dependencies, which is crucial for businesses that need to forecast sales over extended periods. The paper also highlights the challenges of training LSTM

As reported in [12] by K. Pai,et.al(2024) The paper introduces Facebook Prophet, a robust and scalable tool for time series forecasting, especially designed for business applications like sales prediction. It explains how Prophet outperforms traditional time series models like ARIMA by accounting for seasonality, holidays, and business cycles. Taylor and Letham provide a detailed technical breakdown of the algorithm, showing its flexibility in handling missing data and irregular time intervals, making it ideal for e-commerce and retail sales forecasts.

III. PROPOSED METHODLOGY

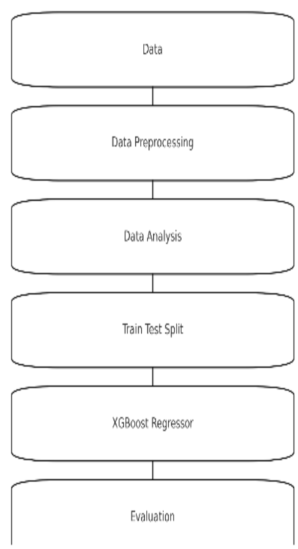

Fig 3.1: Proposed Methodology

Data: This serves as the foundation for the machine learning pathway. In the framework of sales prediction, "Data" means to the raw collection that includes elements like item descriptions, retailer facts, and previous sales data. All information must be gathered from trusted sources and have necessary qualities for successful forecasts. On average, the collection of data contains numerous columns (variables), including Item_Weight, Item_MRP, Outlet_Type, and the goal variable, Item_Outlet_Sales. Analyzing the data is critical beforehand moving on towards the next phases in the workflow. Historical sales records from the retail sector are an example of prevalent data, which include features that include purchase date, product information (e.g., brand ID, type, value), customer characteristics (e.g., age, gender, location), and marketing initiatives (e.g., savings, marketing efforts). Outside influences such as financial statistics, changing seasons, and climate variables can also be utilized to supplement the information.

Data Preprocessing: The preprocessing of information is required to clean and prepare the raw data for evaluation. This stage comprises dealing with values that are missing, encoding categorized variables (converting labels with text toward numerical amounts), and verifying that the information is in the right structure. It can additionally involve translation and normalizing if required. For a sales forecasting approach, this step would resolve any values that are missing in Item_Weight or Outlet_Size, accept data, and modify specific variables like Item_Type or Outlet_Location_Type. Proper data pretreatment ensures that machine learning algorithms are able to extract effectively from the data. Effective data preparation is critical for building strong machine learning algorithms for sales prediction. By taking these procedures, you can guarantee that the information you provide is clean, important, and available for research, resulting in higher forecast reliability and accuracy.

Data Analysis: In this step, exploratory data analysis (EDA) is used to determine the linkages among various factors and their influence on the goal parameter, Item_Outlet_Sales. The result involves creating summarized data and visuals like histograms and bar charts, and heatmaps as Examining the data reveals correlations, trends, and patterns that could be relevant to the prediction algorithm. For example, evaluating the breakdown of sales by product type, retailer type, and duration of establishment can reveal which variables have the most impact on sales, which can then be utilized to modify the model. Evaluation of data is an important stage in developing an efficient sales forecast model. It entails analyzing and comprehending the collection of data in order to identify patterns, trends, and interactions among various variables. This technique aids in discovering the most significant aspects that drive sales, eventually improving the effectiveness of the machine learning model. This study aids in choosing features, since variables with a small correlation may be deleted in order to streamline the algorithm without compromising effectiveness.

Train Test split: This stage divides the set of data into training and testing sets. In general, the data is split into 80% training and 20% testing. The training set of data is used for building and refining the machine learning model, and the test collection is used to assess its outcome. Split the data ensures that the prediction expands adequately to previously unknown data. This minimizes excessive fitting, which occurs when the algorithm operates well on training information but badly on fresh, previously unknown data. The set used for training allows the XGBoost Regressor discover patterns in past sales data, while the evaluation set assesses the model's predicted accuracy. A train-test split is an essential phase in creating a sales prediction algorithm utilizing machine learning. By properly separating the data and analyzing the model's effectiveness, you can make sure that the algorithm is stable and capable of generating accurate forecasts on fresh, previously unknown sales information.

XGBoost-Regressor**:**XG-Boost (Extreme Gradient Boosting) is a sophisticated machine learning method that is frequently utilized to regression problems, such as sales prediction. It works by combining decision trees to improve the accuracy of models using boosting approaches. The XGBoost Regressor iteratively adapts decision trees to training data, correcting flaws from prior trees. This procedure proceeds til the device achieves peak performance. XGBoost is renowned for its effectiveness and precision, frequently surpassing other regression techniques. This stage in the sales prediction process predicts the sales goal variable by analyzing characteristics such as item visibility, weight, and outlet type.

Evaluation: Once trained, the framework must be tested for accuracy and dependability. Evaluation measures like as R-squared, mean squared error (MSE), and root mean squared error (RMSE) are commonly employed to evaluate the effectiveness of models. These parameters show how effectively the model predicts values for sales in the experimental dataset. If the algorithm performs badly, changes such as feature tweaking, parameter improvement, or further data preparation may be necessary. In this stage, the evaluation metrics are calculated using the test and training sets simultaneously, verifying that the model is successful in real-world applications. Evaluating sales prediction models requires a combination of numerical measurements and real-world issues. Businesses may use various evaluation approaches and indicators to verify that sales predictions are precise, reliable, and practical, resulting in enhanced decision-making and overall planning.

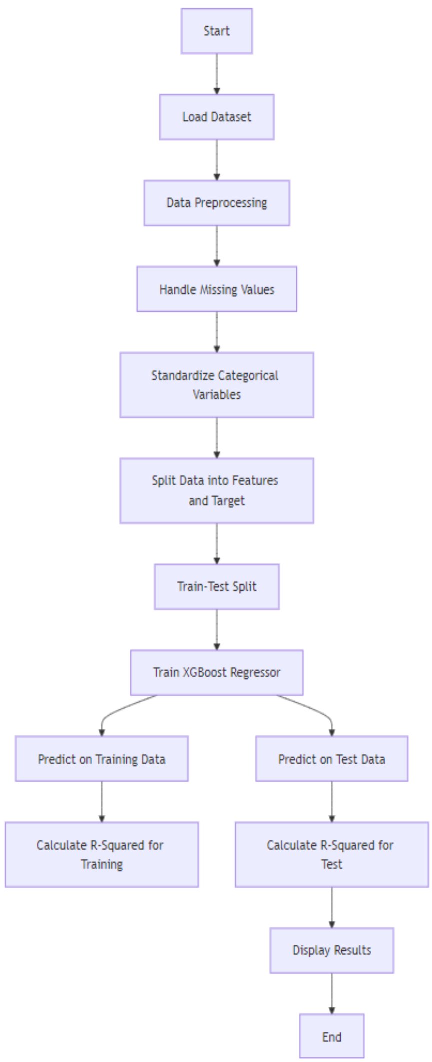

Fig 3.2 Flowchart of sales prediction

Load Dataset: Collecting and importing the pertinent dataset is the first component for the sales forecast methodology. This dataset must to include data that is essential for making precise revenue predictions The particular factors that are taken into account will vary depending on the type of organization and the research's objectives. This entails importing the information into its original source—which may include a compressed text file (CSV), Excel file, database, or another kind of storage format—using the proper data transferring routines or libraries.

Data Preprocessing: Data preparation, which entails organizing and getting ready the material for investigation, is a vital stage in the sales prediction process. As the raw information often includes inaccuracies values that are lacking, and irregularities that can impair the efficiency of the model, this step is crucial to ensuring data is accurate and reliable.

Handle Missing Values: It's critical to locate and fix any data that is missing elements. Methods like attribution, which involves substituting estimated amounts for values that are missing, or eliminating sections or rows that have a large quantity of missing data might be utilized to maintain data integrity, all values that are missing in the collection of data are filled in.

Standardize Categorical Variables: Characteristic parameters refer to those that can be divided into groups or classifications, like "the item class," "the client a spot," or "business method." These category characteristics must be transformed into the form of numbers because machine learning models generally operate with numbers as input.

Split Data into Features and Target: Qualities (Separate Factors) all of these factors are thought to have an impact on or be predictive of the variable being studied. Components in the larger picture of sales forecasting may involve things like cost of goods, marketing budget, and client profile the parameter that you wish for prediction is known as the target parameter. The objective variable in a sales prediction scenario is going to be the quantity or value of sales.

Train XGBoost Regressor: The XGBoost Regressor algorithm is then trained on the data used for training after the dataset has been divided into characteristics and the selected variable. A well-liked collection learning method called XGBoost merges several decision trees to produce an algorithm that is more accurate and reliable. It is renowned for being effective and capable of managing huge datasets.

Predict on Training Data: Using the same information for training, the XGBoost Regressor model is then trained on it. The following phase is to apply the trained model to predicting. Through this approach, the model's accuracy on previously encountered data is evaluated.You may gauge the extent to which the model covers the data by contrasting the model's predictions with the actual targets that were included in the training set. By doing so, it may be possible to spot instances of overfitting, in which the model might not adapt well to new, untested data because it has absorbed the training set too thoroughly.

Predict on Test Data: Predict on Test Data: The sample set, which consists of previously unreviewed data, is subjected to prediction made by the model that was trained. To evaluate the expected outcomes of the framework, data from tests is fed into the model that was developed based on past sales data. The data used for testing is made up of fresh or undiscovered information that the system has never seen before, including sales numbers, client population demographics, or changes in the seasons. By generating projections on that information, the model anticipates eventual sales patterns or certain results, such as profit or popularity of a given item. In order to maximize sales results, this aids firms in assessing the system's correctness in practical situations and informs strategic choices related to prices, inventory control, and advertising strategies.

Calculate R-Squared for Training: To assess the degree to which the algorithm fits to the training data, the R-squared value is computed for the set that was used for training. The amount of the dependence variable's volatility (in this example, sales) that can be accounted for by each of the model's independent variables is represented by this statistics metric. A greater agreement among the model's forecasts and the data itself is shown by an increased R-squared score.

Calculate R-Square for Test: The extension accuracy of the framework is assessed by computing the R-squared value for the evaluation set. Collect your test data to make sure you have an alternate dataset (the testing set) that wasn't utilized for the training of the model. The appropriate variable independence and the real extended values for accuracy are Collect your test data to make sure you have an alternate dataset (the testing set) that wasn't utilized for the training of the model. The appropriate variable independence and the real extended values for accuracy are both included in this dataset. Make assumptions. Depending on the independent factors in the test data, forecast increased accuracy using Determine the amount that remains. The discrepancies among the real and expected expansion accuracy are known as residuals.

IV. ALGORITHM

XGBoost Regression: Because of its capacity to manage big datasets and intricate patterns, the XGBoost Regressor is a potent and effective machine learning method utilized in sales prediction. The XGBoost method constructs many decision trees sequentially in sales prediction responsibilities, such estimating item outlet sales, with each tree rectifying the mistakes of the preceding ones. Boosting is a method where each model in the series learns using the remaining parameters of the preceding model with the goal of minimizing a certain loss function, such Mean Squared Error (MSE). Because it may identify non-linear correlations between characteristics like product kind, shop measurement, item weight, and various other numeric or categorical factors that impact sales, XGBoost is very useful for sales prediction. Additionally, it manages data that is missing well.

V. DATASET

A statistical approach to forecast the sales of numerous products throughout multiple retailers may be constructed using the extensive category and outlet characteristics found in the sales prediction dataset. The aim of the dataset is to forecast the Item Outlet Sales according to many variables, with each entry corresponding to a single product that is offered in a certain outlet.We want to use these traits to forecast the sales statistics for each product in each shop, which is represented by the goal variable, item outlet sales.

Machine learning algorithms such as XGBoost may be developed to anticipate sales based on product characteristics and outlet qualities by comprehending the link between these factors. The wide range of the dataset—which includes both number and category variables—allows for a thorough study.

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier
0	FDA15	9.30	Low Fat	0.016047	Dairy	249.8092	OUT049
1	DRC01	5.92	Regular	0.019278	Soft Drinks	48.2692	OUT018
2	FDN15	17.50	Low Fat	0.016760	Meat	141.6180	OUT049
3	FDX07	19.20	Regular	0.000000	Fruits and Vegetables	182.0950	OUT010
4	NCD19	8.93	Low Fat	0.000000	Household	53.8614	OUT013

Fig 4.1 Dataset of Sales prediction

A variety of variables that characterize the items and the outlets retains where they are sold make up the dataset utilized for sales prediction. It has features like item dimensions, item transparency, and item MRP (the highest point retail price), which offer quantitative data about the items, as well as information like the item identifier, which securely identifies each object. Furthermore, attributes including Item Type and Item Fat Amount aid in classifying the items according to their attributes, such as they are classified as conventional or low-fat, or whatever classification they fall under, like dairy or drinks.

VI. OUTPUT

The distribution of Sales The Item_Outlet_Sales (the total sales per various items at different outlets) are represented by the x-axis, and the amount of data (which indicates the frequency that particular sales values appear within the dataset) is shown on the y-axis.Plotting reveals a distinctly skewed towards delivery, with the majority of sales figures clustered at the lowest point (below 2000) and a smaller number of top-selling goods (some with sales values as high as 14000).

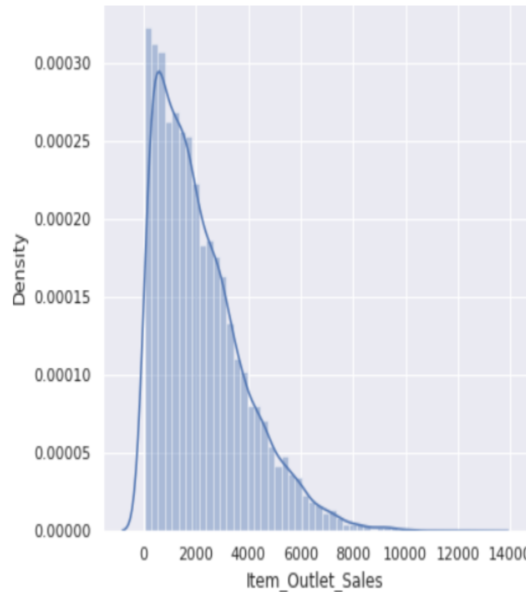

Fig 5.1 Result of Item Outlet Sales

It is possible that there are outliers in the data due to the skewed distributions between the item weight and itemm MRP. Outliers may need to be handled carefully since they can significantly affect how well machine learning models function.Decision-making and feature development can benefit from a grasp of these distributions. As an example, it could be advantageous to alter the characteristics or use methodologies that are capable of handling skewed data if the goal variable (like sales) is also skewed. Understanding the connection between these characteristics and the objective parameter may also be gained from these patterns. For instance, adding this feature to a machine learning model may increase its prediction ability if there is a relationship among Item_Weight and sales.

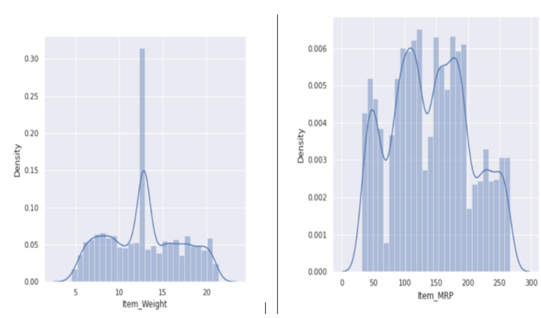

Fig5.2 Results of Item weigh and Item MRP of products

Forecasting sales is an essential responsibility in numerous sectors, ranging from industries and banking to e-commerce and retailing. It entails projecting future revenue volumes using past performance information and other pertinent variables. In order to solve this issue, methods based on machine learning (ML) have shown to be quite successful, providing predictions that are dependable and accurate.

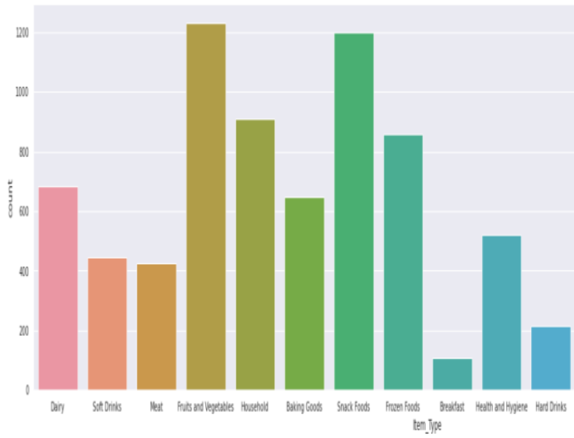

Fig 5.3 Results of Sales prediction

VII. EXPERIMENT RESULT

The accuracy of the two methods used in the experiment, XGBoost Regression and Logistic Regression, for sales prediction, clearly differs. With an R-Squared score of 0.85 on the initial training set, the XGBoost Regression model explains 85% of the statistical variance in sale data throughout training. It attains an R-Squared of 0.82 on the testing set, indicating a minor decline but good capacity for prediction. This slight variation in R-Squared values from training and testing indicates that the model is not overfitting.

Model	R-Square (Training)	R-Square (Testing)
XGBoost Regression	0.85	0.82

Conversely, the regression coefficient squared for the Logistic Regression model is 0.80 on the Testing set and 0.83 on the trained set, indicating somewhat less successful results. This model works well as well, although in testing and training, it is not as good as the XGBoost model. Although Logistic Regression also applies quite well, it might not have the same variety or depth as XGBoost when it comes to handling the not linear trends in the sales data, as seen by the lesser difference between the test and training scores. Both models have good capacity for forecasting all around but XGBoost has greater accuracy therefore is probably a superior option for predicting purchases in this case.

VIII. FUTURE ENGINEERING

The act of developing additional characteristics or altering current ones in order to enhance a model's predictive accuracy is known as feature engineering in machine learning-based sales prediction. The objective in the framework of sales prediction is to identify important trends in the unfiltered information that can improve the model's capacity to project future sales. By converting unstructured data into useful attributes, machine learning for sales prediction enhances the ability of the algorithm to forecast sales in the future. Here, the term "parallelization" (para) describes the use of several processing using distributed computer systems to expedite the feature engineering process. When tackling enormous datasets, intricate conversions, or laborious tasks, this is really helpful. Following that, mathematical characteristics such as latency, price in highest, lowest, and median features for a particular amount of time. Lastly, we use the RFECV approach to pick elements; particular characteristics are eliminated from the algorithm after every conditioning cycle.

IX. CONCULSION

We present a novel approach to wide range sales prediction in this research. We integrate the execution of the Logistic Regression algorithm using the Wide Range marketing dataset, which is made available on the Kaggle challenge site. Results demonstrate that our approach outperforms the XGBoost regression technique that can efficiently gather characteristics from various dimensions. Additionally, we looked at how important elements are ranked and came up with some insightful and helpful advice. We intend to enhance our framework in future versions in order to enhance the forecasting performance and extend its application to more sales-related issue

REFERENCES

[1] Z. Huo, "Sales Prediction based on Machine Learning," 2021 2nd International Conference on E-Commerce and Internet Technology (ECIT), Hangzhou, China, 2021, pp. 410-415, doi: 10.1109/ECIT52743.2021.00093

[2] Y. Katai and T. Hasuike, "Sales Price Prediction of Used Clothes Considering the Market Price of Flea Market Application," 2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI), Toyama, Japan, 2019, pp. 750-755, doi: 10.1109/IIAI-AAI.2019.00154.

[3] H. Upadhyay, S. Shekhar, A. Vidyarthi, R. Prakash and R. Gowri, "Sales Prediction in the Retail Industry Using Machine Learning: A Case Study of BigMart," 2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM), Roorkee, India, 2023, pp. 1-6, doi: 10.1109/ELEXCOM58812.2023.10370313.

[4] S. N. Gunjal, D. B. Kshirsagar, B. J. Dange and H. E. Khodke, "Fusing Clustering and Machine Learning Techniques for Big-Mart Sales Predication," 2022 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS), Pune, India, 2022, pp. 1-6, doi: 10.1109/ICBDS53701.2022.9935906.

[5] C. S. Rekha, K. B. Sathwika, G. S. Nandini, R. Sumanth and S. Shareefunnisa, "Ensemble Strategies for Predicting Walmart Sales Using Regression Models," 2024 8th International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 2024, pp. 693-699, doi: 10.1109/ICISC62624.2024.00123.

[6] T. Rajasree and R. Ramyadevi, "Time Series Forecasting of Sales Data using Hybrid Analysis," 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), Greater Noida, India, 2024, pp. 732-735, doi: 10.1109/IC2PCT60090.2024.10486340.

[7] S. V.R, N. G, R. M and P. A, "Food Sales Analysis and Prediction Using Machine Learning," 2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA), Namakkal, India, 2024, pp. 1-5, doi: 10.1109/AIMLA59606.2024.10531342.

[8] M. Spuritha, C. S. Kashyap, T. R. Nambiar, D. R. Kiran, N. S. Rao and G. P. Reddy, "Quotidian Sales Forecasting using Machine Learning," 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2021, pp. 1-6, doi: 10.1109/ICSES52305.2021.9633975.

[9] G. T., R. Choudhary and S. Prasad, "Prediction of Sales Value in Online shopping using Linear Regression," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-6, doi: 10.1109/CCAA.2018.8777620

[10] D. S. AbdElminaam, M. Mohamed, S. Khaled, F. Hany, M. Magdy and Y. Sherif, "Leveraging Machine Learning for Accurate Store Sales Prediction: A Comparative Study," 2024 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), Cairo, Egypt, 2024, pp. 355-362, doi: 10.1109/MIUCC62295.2024.10783509..

[11] J. Chen, W. Koju, S. Xu and Z. Liu, "Sales Forecasting Using Deep Neural Network And SHAP techniques," 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and

[12] Internet of Things Engineering (ICBAIE), Nanchang, China, 2021, pp. 135-138, doi: 10.1109/ICBAIE52039.2021.9389930.

[13] K. Pai, M. Batool, N. Ravi, R. Surendra, A. N. R. Shree and S. J. Meenakshi, "Enhancing Sales Forecasting and Prediction with Cutting-Edge Machine Learning Methods," 2024 Second International Conference on Data Science and Information System (ICDSIS), Hassan, India, 2024, pp. 1-6, doi: 10.1109/ICDSIS61070.2024.10594627.

[14] V. S. Warpe, S. D. Buchkul, P. P. Chobe and D. B. Pardeshi, "EV Sales Price Forecasting Using Machine Learning," 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2024, pp. 818-823, doi: 10.1109/ICSCSS60660.2024.10625450.

[15] Chen, Tianqi & Guestrin, Carlos. (2016). “XGBoost: A Scalable Tree Boosting System”, 785-794. 10.1145/2939672.

[16] Chen, T., He, T., Benesty, M., et al.: Xgboost: extreme gradient boosting. R package version 04–2, 1–4 (2015)

[17] A. Akanksha, D. Yadav, D. Jaiswal, A. Ashwani and A. Mishra, "Store-sales Forecasting Model to Determine Inventory Stock Levels using Machine Learning," 2022 International Conference on Inventive Computation Technologies (ICICT), Nepal, 2022, pp. 339-344, doi: 10.1109/ICICT54344.2022.9850468..

[18] S. Anand, P. Yadav, A. Gaur and I. Kashyap, "Real Estate Price Prediction Model," 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2021, pp. 541-543, doi: 10.1109/ICAC3N53548.2021.9725772..

[19] S. Cheriyan, S. Ibrahim, S. Mohanan and S. Treesa, "Intelligent Sales Prediction Using Machine Learning Techniques," 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), Southend, UK, 2018, pp. 53-58, doi: 10.1109/iCCECOME.2018.8659115..

[20] S. Cheriyan, S. Ibrahim, S. Mohanan and S. Treesa, "Intelligent Sales Prediction Using Machine Learning Techniques," 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), Southend, UK, 2018, pp. 53-58, doi: 10.1109/iCCECOME.2018.8659115.

[21] Vo,B.and Le,B.,"Mining traditional association rules using frequent itemsets lattice",International Conference on Computers & Industrial Engineering. 2009.

[22] Rastogi, R.. and kyuseok Shim, "Mining optimised association rules with Categorical and numerical attributes", IEEE transactions on Knowledge and Data Engineering, vol.14, 2002

[23] Gao, J., Xie, Y., Gu, F., Xiao, W., Hu, J., Yu, W.: A hybrid optimization approach to forecast automobile sales of China. Adv. Mech. Eng. 9(8), 1687814017719422 (2017a).

[24] Deng, A. , Zhang, H. , Wang, W. , Zhang, J. , & Wang, B. . (2020). "Developing computational model to predict protein-protein interaction sites based on the xgboost algorithm", International Journal of Molecularences, 21(7), 2274.

[25] Yunxin Liang, Jiyu Wu, Wei Wang, Yujun Cao, Biliang Zhong, ZhenkunChen, and Zhenzhang Li, "Product marketing prediction