International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-11 Issue-6, May 2022

Machine Learning Approach for Big-Mart Sales Prediction Framework

Sanjay. N Gunjal, D.B Kshirsagar, B.J Dange, H.E Khodke, C.S Kulkarni

Abstract*: The amounts of data predicted to increase at an exponential rate in the future. The modifications are essential to meet transaction speeds as well as the anticipated growth in data and customer behaviors. The information derived from prior data is extensively relied upon by the majority of companies. One of the primary goals of the suggested system is to identify a reliable sales trend prediction mechanism that is executed using machine learning techniques in order to maximize income. Sales forecasting advises managers about how to manage a company's employees, working capital and assets. It's a requirement for strategic planning and decision-making in the corporate world. Reasonable forecasts enable the company to increase market growth while increasing revenue generating. Operations, marketing, sales, production, and finance all use sales predictions as inputs in their decision-making processes. The concept of sales data and sales forecast has been examined in the suggested system. Machine learning algorithms such as GLL (Generalized Linear Model), GBT (Gradient Boosted Trees), and Decision Trees were used to develop the model, and the optimum model for prediction was established based on the results analysis. A best-fit prediction model for anticipating sales trends is offered based on a performance review. The effectiveness and accuracy of the prediction and forecasting approaches used are discussed in the findings. The Gradient Boost Algorithm has been demonstrated to be the best fit model for forecasting and predicting future sales. The sales projection is done using Gradient Boosted Trees, which predicts which product will be sold in what quantity in the future..

Keywords: Data Mining Techniques, Machine Learning Algorithms, Prediction, Reliability, Sales Forecasting.

I. INTRODUCTION

Forecasting sales is the process of predicting future sales based on past performance. Sales forecasting is critical for businesses that are entering new markets, providing new services or products, or expanding rapidly.

Manuscript received on 28 April 2022.
Revised Manuscript received on 21 May 2022.
Manuscript published on 30 May 2022.
* Correspondence Author
Dr. Gunjal Sanjay Nana*, Assistant Professor, Department of Computer Engineering, Sanjivani COE, Kopargaon Savitribai Phule Pune University, Pune (Maharashtra), India. Email: gunjalsanjay1982@gmail.com
Dr. D.B Kshirsagar, HOD and Professor, Department of Computer Engineering, Sanjivani COE, Kopargaon, Savitribai Phule Pune University, Pune (Maharashtra), India. Email: dbkshirsagar444@gmail.com
Dr. B.J Dange, Associate Professor, Department of Computer Engineering, Sanjivani COE, Kopargaon, Savitribai Phule Pune University, Pune (Maharashtra), India. E-mail: dangebapusaheb@sanjivani.org.in
Dr. H.E Khodke, Assistant Professor, Department of Computer Engineering Dept, Sanjivani COE, Kopargaon, Savitribai Phule Pune University, Pune (Maharashtra), India**.** E-mail: hekhodke@gmail.com
Dr. C.S Kulkarni, HOD and Associate Professor, Department of Computer Engineering, VPKBIET, Baramati, Savitribai Phule Pune University, Pune (Maharashtra), India. chaitanya.kulkarni@vpkbiet.org

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Forecasting is used by businesses to create a balance between marketing resources, sales planning, and supply capacity planning. Forecasting can be used to anticipate sales revenue at the product level, at the level of a single firm, or at the level of an entire corporation. The published paper focused on projecting sales at the product level[1]. Many sellers want the ability to adjust their promotion techniques in order to achieve the best balance of sales volume and profit, and sales volume prediction has become a popular request in recent years. When sales volume is overestimated, actual sales are lower than predicted, whilst underestimated sales volume might lead to more promotional costs and, as a result, poorer profit. As a result, accurate sales volume forecasting is critical for sellers when deciding on promotion techniques. The sales volume prediction challenge in this case can be defined as a technique for estimating future sales using the seller's historical data and future advertising efforts. Historical data includes sales volume, dynamic influencing factors such as historical promotion plans and weather information, as well as the seller's static properties[3]. Sales predictions are important inputs to many different organizational decision-making processes, including operations, marketing, sales, manufacturing, and finance. For organisations seeking investment funding to properly service their internal resources, predictive sales data is vital[2]. The research continues with a new viewpoint on how to select the best strategy for accurately forecasting sales. Despite the fact that the initial data set for this study included a large number of entries, the final data set utilised for analysis was significantly less than the original after non-usable data, redundant entries, and irrelevant sales data were deleted.To achieve optimum accuracy, this article used a range of forecasting approaches such as GLL (Generalized Linear Model), GBT (Gradient Boosted Trees), and Decision Trees. In Section I, we go over data mining techniques and prediction methodologies. Section II is a survey of diverse literature on sales projections. The topic of sales data and sales forecast was examined in Section III. The model was created using machine learning algorithms such as GLL (Generalized Linear Model), GBT (Gradient Boosted Trees), and Decision Trees, and the best model for prediction was determined based on the results analysis. Based on a performance evaluation, a best-fit predictive model for projecting sales trends is proposed.

The objectives of our work are as follows:

- To convert data into a format appropriate for Machine Learning algorithms, a number of preprocessing approaches areused.

© Copyright: All rights reserved.

- Determining which important features will have the greatest influence on product sales.
- Choose the best effective Machine Learning algorithm for sales forecasting.Choosing measures to assess the present Machine Learning algorithm's performance with others

II. RELATED WORK.

To date, a lot of work has been proposed in the domain of sale forecasting utilising machine learning. This section provides a brief overview of related sales forecasting research. The author built six clustering-based forecasting models for computer product sales forecasting using K-means, SOM, and GHSOM as three clustering methodologies, as well as an SVR and an ELM as two machine-learning techniques. Three computer merchants' real sales numbers for PC, NB, and LCD items were used as empirical data. When it came to predicting the sales of three different computer devices, the GHSOM-ELM model beat five other clustering-based forecasting models, one SVR, and one ELM.[4]. An approximate technique for exact greedy algorithms, data storage in in-memory units for parallel learning, cache-aware prefetching, and out-of-core processing are all included in XGBoost. Users can manage a larger dataset and execute it more faster with XGBoost[5]. The author utilised a two-layer model that included Linear Regression and Support Vector Machine to anticipate sales data from large shopping centres.[6]. The authors experimented with exploiting the point of sale (POS) as internal and even external data to improve the efficacy of demand forecasting by looking at a variety of scenarios. They looked into approaches including Bayesian Linear Regression, and Decision Forest Regression, Boosted Decision Tree Regression,[10]. The authors of the study used Random Forests, k-nearest neighbour, and XGBoost to investigate how clients arrive at eateries. They used two real-world data sets from separate booking platforms, as well as diverse restaurant features as input variables. According to the results, XGBoost is the best model for the dataset [9]. The weather has an impact on regular restaurant sales, according to the author. The accuracy of two Machine Learning techniques, XGBoost and neural network, was investigated. and found that the XGBoost technique outperformed the neural network technique. They also observed that factoring in meteorological variables improved the effectiveness of their model by 2-4 percentage points. They took into account a variety of criteria to increase accuracy, including date characteristics, sales history, and weather conditions[8]. The AdaBoost is a boosting method that is mostly used to improve model performance. It uses the output of the other weak learners to aggregate the results of those algorithms into a weighted sum before arriving at the final result. Whether using manual or real data, AdaBoost has the potential to significantly increase learning accuracy.[11]. Machine learning technique for predicting how much a buyer would spend on the next "Black Friday" deal. The decision has been made to use exploratory data analysis to find relevant patterns in the dataset. According to this study, when a user tries to forecast which product a client is more likely to buy based on their gender, age, and occupation, the user's gender, age, and occupation all play a role. Experiments demonstrate that our solution outperforms techniques like decision trees and ridge regression in terms of prediction accuracy[7]. Reliable prediction findings for just one commodity, according to the author, are irrelevant to vendors. For all commodities, it is vital to have a broad forecast. This research presents a unique trigger mechanism that might be used to improve prediction results for a wide range of commodities, associate selected commodities with a prediction model. Related categorization factors have been discovered. As basic categorization models, some classical prediction models are provided. The trigger model's findings are compared to single-model outcomes. In terms of accuracy, the trigger model appears to outperform a single model. Vendors can utilise the suggested method to properly estimate the sales of a range of commodities, which has commercial implications[2].

III. ARCHITECTURE AND MODELING

Using the retail store's sales data, the proposed study suggested the following several processes for projecting the sales of various categories. Figure 1 depicts the suggested system's architecture diagram. The many steps in the process are outlined below.

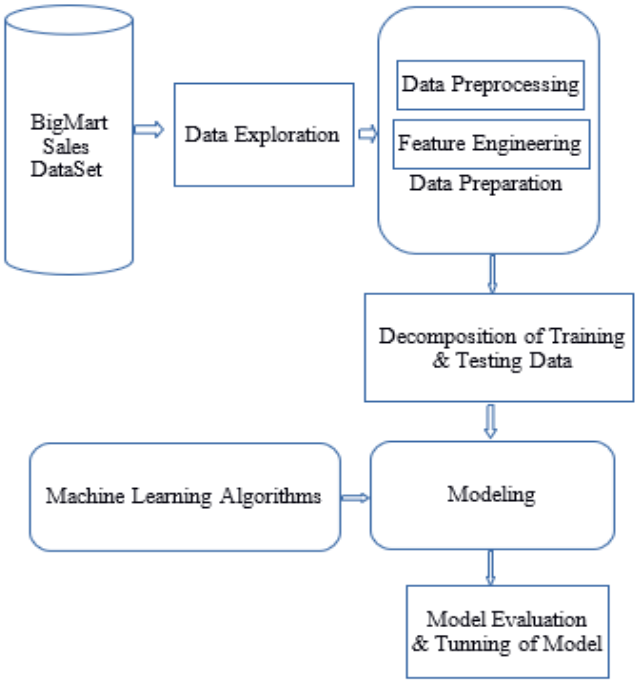


Figure 1 Architecture of sales predication framework

A. Hypothesis Generation

This is the most important step in the data analysis process. By examining the problem description, numerous hypotheses are created in this step. The assumptions are constructed in a way that favours the desired conclusion. Take a look at the following problem statement: "The dataset acquired includes sales data from 10 stores in various cities for 1559 goods in 2013."

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

The goal is to create a prediction model that will determine how much each product will sell in a specific store." As a result, the primary goal of this stage is to learn about a product's properties and the shops that might have an effect on sales. Forecasting is done on the basis of goods and retailers, therefore it's a bit more complicated [12-24]. The hypotheses can be divided into "Product Level Hypotheses" and "Store Level Hypotheses. "Brand, packaging, display area, usability, promotional offers, and store visibility are all important factors to consider, advertising, and other product-level assumptions all have the potential to affect sales, While city size, size of population, shop capacity, competitors, locality, consumer habits, advertisement, atmosphere, and other store-level assumptions can all influence sales. Branded products, for example, sell better than other products Customers have a higher level of trust in the brand as a result of this, which leads to increased sales. Similarly, we should expect larger sales from places that are well-maintained and managed by humble and nice employees because store walk-ins will be higher. As a result, we've come up with 15 hypotheses that help us comprehend the situation better.

B. Data Exploration

Our aim to increase accuracy by updating and implementing various models when we think about a business challenge. However, we will stress that there will come a time when we will be unable to improve the model's accuracy. Data exploration is used to solve these types of difficulties. The initial stage in data exploration is to investigate the data set and learn as much as possible about the accessible and speculated information. There are 6 hypothesised and present features in the dataset, 3 theorised but not present features in the dataset, and 9 postulated but not found features in the dataset, according to our analysis. The diagram in Figure 2 is the best representation of this.

Some values are missing from the columns "Outlet Size" and "Item Weight" in the dataset under consideration. In the data preprocessing stage, missing data values will be imputed. The variables in our dataset are divided into two categories: category variables and numerical variables. The numerical variables are described in Table 2 in general

Table 2.Numerical Variable Description

The count, Mean, Standard deviation, min and max of each of the column in the dataset is as below....

	Item_Weight	Item_Visibility	Item_MRP	Outlet_Establishment_Year
count	7060.000000	8523.000000	8523.000000	8523.000000
mean	12.857645	0.066132	140.992782	1997.831867
std	4.643456	0.051598	62.275067	8.371760
min	4.555000	0.000000	31.290000	1985.000000
25%	8.773750	0.026989	93.826500	1987.000000
50%	12.600000	0.053931	143.012800	1999.000000
75%	16.850000	0.094585	185.643700	2004.000000
max	21.350000	0.328391	266.888400	2009.000000

Two basic conclusions may be drawn from the table.

1. The minimum value for Item Visibility is 0. The value 0 creates difficulties because it shows that the product is not visible but is being sold. As a result, visibility should be larger than 0 in this case.
2. The options for setting up an outlet The years 1985 to 2009 are covered. These values are translated to reflect the age of a store during the pre-processing step. There are 1559 unique goods and 10 unique outlets based on the categorical characteristics in the dataset. We came to the following conclusions after analysing the dataset and looking at the frequency of various categories.
1. Certain "Low Fat" values are labelled "LF" and "low fat" in the Item Fat Content category, whereas a few "Regular" values are labelled "regular."
2. There are 16 subcategories in the Item Type category. Food to drinks, meat to canned food, home products to medications, and so on are examples of these commodities., implying that each store carries a diverse selection of goods.

********Exploatory Analysis******

First five records in your dataset are as follows:

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	\
0	NCP30	20.500	Low Fat	0.032835	
1	NCV30	20.200	Low Fat	0.110357	
2	FDT12	6.215	Regular	0.049722	
3	FDC08	19.000	Regular	0.103364	
4	FDN02	16.500	Low Fat	0.074245	

	Item_Type	Item_MRP	Outlet_Identifier	\
0	Household	40.2822	OUT045	
1	Household	62.0510	OUT010	
2	Baking Goods	224.5062	OUT045	
3	Fruits and Vegetables	226.1720	OUT013	
4	Canned	208.2638	OUT017	

	Outlet_Establishment_Year	Outlet_Size	Outlet_Location_Type	\
0	2002	NaN	Tier 2	
1	1998	NaN	Tier 3	
2	2002	NaN	Tier 2	
3	1987	High	Tier 3	
4	2007	NaN	Tier 2	

	Outlet_Type	Item_Outlet_Sales	Item_Date
0	Supermarket Type1	707.0796	04-01-2011
1	Grocery Store	126.5020	04-01-2011
2	Supermarket Type1	4514.1240	04-01-2011

Figure 2: Exploratory Analysis

C. Data Pre-processing

This phase, in most situations, imputes missing values and manages any dataset outliers. In the dataset, the Item Weight and Outlet Size columns are both blank. Item weight is a numeric variable, whereas outlet size is a categorial variable. The average weight of that particular item in the sample is used to impute Item Weight, which is subsequently substituted for the missing values. We can't calculate the average because Outlet Size is a categorical variable, therefore we utilise the mode technique to fill in the gaps. As a result, by identifying the size mode based on the outlet type, the missing values in the outlet size are determined.

D. Feature Engineering

We discovered a few quirks in the data during the data exploration stage. This stage takes care of those intricacies as well as establishing new variables from current variables so that our data is ready to be analyzed. We discovered that Item Visibility had a minimum value of 0 during the data exploration phase, which we were unaware of. We fixed the problem by considering it as a missing number and imputing the average visibility of the product.We looked at visibility as one of the store level hypotheses during hypothesis formation, with the idea that the higher the product's visibility in the store, the higher the sales. When a product's exposure at one store is compared to the product's average visibility across all stores, It's possible to figure out how essential the product was in that particular store in comparison to others. As a result, Item Visibility MeanRatio is generated as a new variable to record these values. We learned that Item Type has 16 separate categories during the data exploration phase; if these categories are grouped together by a similar attribute, they may be useful in the analysis. The Item Identifier variable, which contains the Unique ID of each object, exemplifies this feature.

© Copyright: All rights reserved.

The letters FD, DR, or NC in this variable stand for Food, Drinks, and Non-Consumables, respectively. As a result, we add a new column to the table named Item Type. The two columns Item Type and Item Identifier were used to categorise each product as Food, Drinks, or Non-Consumables. The year the business first opened is stored in the field Outlet Establishment Year. We make a new variable out of this that shows how long the store has been open. As a result, the ages ranged from 4 to 28. As we observed, there was some difference in the way categories were represented in the Item Fat Content variable. Replace LF with Low Fat and reg with Regular to fix this. Non-Consumables, on the other hand, have a fat level that is either Low Fat or Regular when the categories are switched. In Item Fat Content, Non-Consumables have been classed as Non-Edibles.

E. Segregation of Training and Testing Data

One of the most crucial processes in machine learning is feeding data into the algorithm and training it to recognize patterns. Once the algorithm has learned the pattern, it must be fed another dataset to determine the algorithm's level of knowledge. It is typical practise to divide available data into two subsets at a 4:1 ratio for training and testing reasons. However, we may need to re-adjust this ratio in order to achieve better results. Furthermore, the ratio may differ amongst regression algorithms.

F. Model Building and Prediction

Prediction is concerned with future events. Machine learning algorithms increase the system's intelligence without requiring operator interaction. "With sample data or previous experience, To improve the performance criterion, machine learning (ML) is used. All fields can benefit from machine learning techniques. To tackle many classification and clustering difficulties, machine learning employs statistics. Machine learning algorithms are divided into three categories. Oversight is divided into three categories: monitored, unsupervised, and semi-supervised. The three machine learning prediction algorithms studied in this research were the Generalized Linear Model (GLM), Decision Tree (DT), and Gradient Boost Tree (GBT). In this study, we used the training dataset to develop three machine learning methods, and the models were then tested for performance. The optimal algorithm for prediction is chosen based on performance accuracy.

1) Generalized Linear Model

The generalised linear model (GLM) is a technique for predicting data sets. It extends the concept of a general linear model (such as a linear regression equation). The GLM generalises linear regression by allowing a link function to link the linear model to the response variable and the magnitude of each measurement's variance to be a function of its predicted value.

I. mdl = stepwiseglm(tbl)

Intelligent stepwiseglm(tbl) starts with a constant model and Stepwise regression is used to add or delete predictors in a generalised linear model of a table or dataset array tbl. The last variable of tbl is utilised as the response variable in stepwiseglm. To arrive at a final model, Stepwiseglm employs forward and backward stepwise regression. The function searches for terms to add to or remove from the model at each stage based on the value of the 'Criterion' input.

II . mdl = stepwiseglm

stepwiseglm(X,Y) generates a generalised linear model of the y responses to a data matrix X.

III. mdl = stepwiseglm(,modelspec)

Using any of the input argument combinations in earlier syntaxes, stepwiseglm(,modelspec) defines the beginning model modelspec.

IV. mdl = stepwiseglm(,modelspec,Name,Value)

Using one or more name-value pair parameters, stepwiseglm(,modelspec,Name,Value) specifies additional choices. For example, you can give stepwiseglm categorical variables, the smallest or largest collection of terms to utilise in the model, the maximum number of steps to execute, and the criterion stepwiseglm uses to add or delete terms.

2) Decision Tree

Decision Tree Analysis is a predictive modelling approach that can be used in a wide range of industries. In general, decision trees are created using an algorithm that finds different ways to partition a data set based on specified criteria. It is one of the most widely used and useful supervised learning algorithms. For regression applications, Decision Trees is a non-parametric supervised learning approach. The goal is to construct a model that predicts the value of a target variable by learning fundamental decision rules from data features.

i. Read the sales data from the file system using the pandas module. Examine a couple of the dataset's records.
ii. Choose the variable that will be the goal. Future sales are the target variable here. The remaining factors are predictor variables.
iii. There are two parts to the dataset: training and testing. The test set is used to measure the model's performance, while the training data is used to train it.
iv. Apply the model to the supervised data you've selected.

3) Gradient Boosted Tree

Boosting is a method for transforming weak learner into strong ones. Each new tree in boosting is constructed from a modified version of the original data set. The objective is to improve upon the previous tree's forecasts. As a result, Tree 1 Plus Tree 2 is our new model. The prediction error of this updated 2-tree ensemble model is then calculated, and to forecast the updated residuals, a third tree is created. This technique is repeated a certain number of times. Following trees help us forecast observations that previous trees didn't predict correctly. The final ensemble model's predictions are the weighted sum of the predictions provided by the preceding tree models.

Loss Function:

loss function : Error between true and predicted sales values.
i. Calculate the average of the target table.
ii. Calculate the residuals.
residual = actual value - predicted value
iii. Construct the decision tree.
iv. Predict the target label using all of the trees within the ensemble.
Average price = Learning rate(0.1) * Residual predicted by decision tree
v. Compute the new residuals.

© Copyright: All rights reserved.

IV. DATASET DETAILS

It's a BigMart sales dataset with 1559 products spread over 10 locations in various cities. There are 8523 train and 5681 test records in all. Both input and output variables are present in the train dataset. There are 12 attributes in the dataset that are Identifier for this item: This is a one-of-a-kind product code. Item Weight: The item's weight. Item Fat Content: Whether or not the item has a low fat content. Item Visibility: In a store, the percentage of total display area allotted to a specific product. Item Type: This refers to the categorisation of the product. MRP (Maximum Retail Price): A product's maximum selling price (list price) The outlet is the product's unique shop ID. identifier. Establishment of the Outlet Year: The year the store first opened its doors. The outlet size refers to the size of the store in terms of square feet. Outlet city : The location of the shop in terms of city type. There's something for everyone, whether it's a small grocery store or a full-fledged supermarket. Item Outlet Sales are product sales at a certain retailer. This. This is the variable that has to be anticipated.

V. EXPERIMENTAL EVALUATION

Classification accuracy, accuracy within each class, and a confusion matrix, which provides the number of predictions for each class that may be compared to the occurrences of each class, are all used to assess the classification algorithms' accuracy. The Root Mean Square Error, Mean Square Error, and Absolute Error are used to generate the Error Rate, and Table III provides the average of the errors. This measure may be used to see if a forecast is on average correct or not. Table 3 shows the comparison analyses of the three algorithms based on prediction performance, and graph 3 shows the visualisation. According to the data, the Gradient Boost Algorithm had 95.84 percent overall accuracy, followed by Decision Tree Algorithms with about 58.46 percent overall accuracy and the Generalized Linear Model with 56.03 percent accuracy. Finally, the Gradient Boosted Tree technique is the best fit for the model based on the empirical evaluation of the three algorithms. Although the classification accuracy rate can approach 100%, the accuracy rate in the GBT model examined and reported in Table III was around 95.84 percent. The accuracy rate will improve if the GBT implementation is improved further with the use of a big data set.

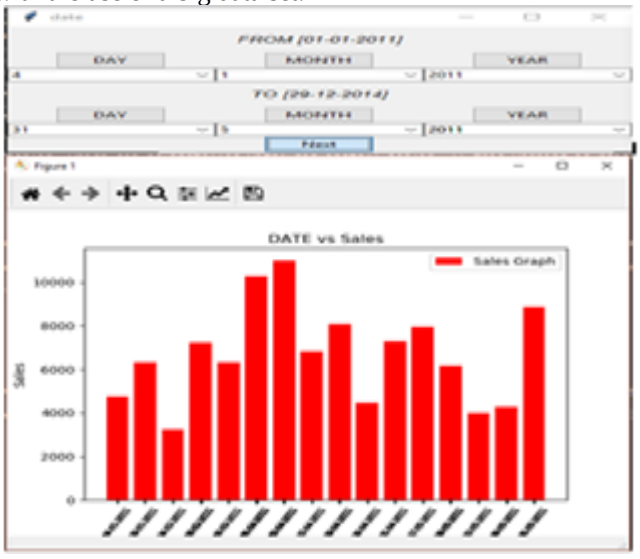
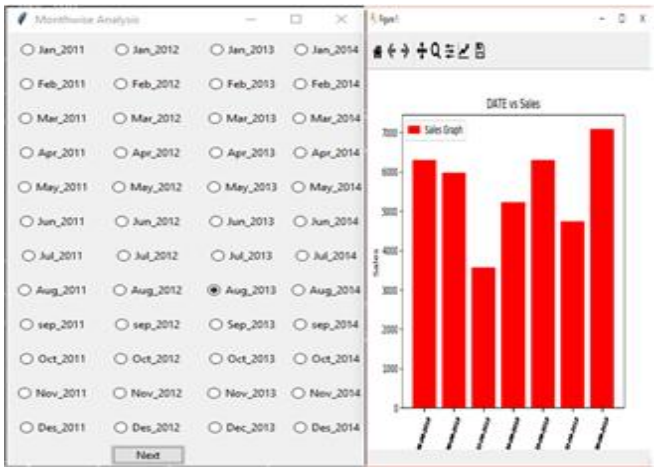


Figure 3: sales prediction of items at particular outlet..

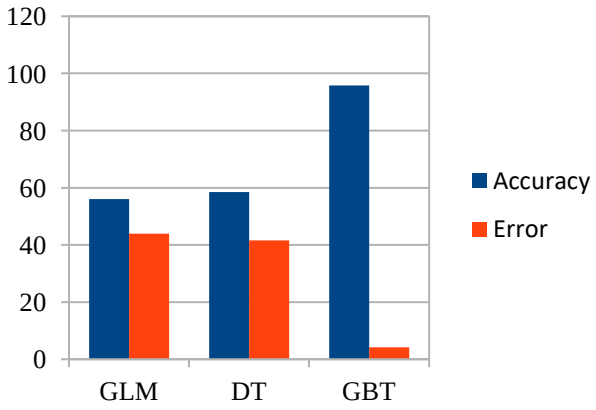
The following graph1 shows the predication of Sales from 01-01-2011 to 31-05-2011. Graph1 : Predication of Sales from 01-01-2011 to 31-05-2011. The following graph shows the monthly analysis of sales prediction.



Graph 1: Monthly analysis of sales prediction

Table 3: Performance Summary of ML Algorithms

Model Name	Performance Summary of ML Algorithms	
	Accuracy Rate (%)	Error Rate
GLM	56.03	43.97
DT	58.46	41.54
GBT	95.84	41.6



Graph2: Comparison of ML Algorithms

VI. CONCLUSION

In this research, we assess the performance of various algorithms on a data set of retail sales and analyse the algorithm that performs the best. The results of a comparison of various methodologies are presented. Furthermore, we discovered that our model with the lowest RMSE outperforms the competition. Cleaning and analysing data can be enhanced, and additional machine learning approaches can be employed to boost the model's accuracy. More accurate predictions may be made with a larger dataset. We must update the dataset with the types of attributes that exist in order to improve the data or obtain a better outcome.

Retrieval Number: 100.1/ijitee.F99160511622
DOI: 10.35940/ijitee.F9916.0511622
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

We need to utilise a properly balanced dataset with different values in each field to get a better outcome. We'll need a well balanced dataset with different values in each field to get a better outcome. The same procedures can be used, with the exception that the dataset must be updated. For the best results, large datasets are recommended. A best-fit prediction model for anticipating sales trends is offered based on a performance evaluation. The effectiveness and accuracy of the prediction and forecasting approaches used are discussed in the findings. The Gradient Boost Algorithm was the best-matched model in the study, with the highest accuracy in predicting and forecasting future sales.

ACKNOWLEDGMENT

The full session of this desk work consummation was an excellent incorporation giving me with unprecedented insight and progression into learning particular ideas of information science, huge information, noteworthy learning. As is authentically stated, for the gainful fruition of any work individuals ar the main fundamental resource. This paper would not be appeared without the enthusiasm of different of the individuals invloved. To begin with and in any case, I am grateful to my supervisor Dr. D..B Kshirsagar for his driving guidance and genuine undertakings in finishing the subject and work. He took a noteworthy interest in altering the minor bungles and guided me through my movement so distant.

REFERENCES

1. Akshay Krishna, Akhilesh V, Animikh Aich, Chetana Hegde (2018), " Sales-forecasting of Retail Stores using Machine Learning Techniques," IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions, IEEE Xplore,160-166. [CrossRef]
2. Sunitha Cheriyan, Shaniba Ibrahim, Saju Mohanan; Susan Treesa (2018), "Intelligent Sales Prediction Using Machine Learning Techniques," International Conference on Computing, Electronics & Communications Engineering (ICCECE),IEEE Xplore,53-58. [CrossRef]
3. Chenhui Lu, Shuo Feng, Jiahao Huang, Xiaojun Ye (2021), "A Multi-Task Prediction Framework for Sales Prediction," International Conference on Computing, Electronics & Communications Engineering (ICCECE),IEEE Xplore.194-19
4. I-Fei Chen, Chi-Jie Lu (2016), "Sales forecasting by combining clustering and machine-learning techniques for computer retailing," Neural Comput & Applications , Vol.18, 105–112
5. Xie dairu1,Zhang Shilong1 (2021), " Machine Learning Model for Sales Forecasting by Using XGBoost" IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE 2021),IEEE, 480-483 [CrossRef]
6. Punam, K., Pamula, R., & Jain, P. K. (2018,),"A two-level statistical model for big mart sales prediction", In 2018 InternationalConference on Computing, Power and CommunicationTechnologiesGUCON),IEEE, 617-620.machandra H V,Balaraju G,Rajashekar A, Harish Patil(2021),"Machine Learning Application for Black Friday Sales [CrossRef]
Prediction Framework", 2021 International Conference on Emerging Smart Computing and Informatics (ESCI),IEEE,57-61.
7. Takashi Tanizaki, Tomohiro Hoshino, Takeshi Shimmura, and Takeshi Takenakam(2019),"Demand forecasting in restaurants using machine learning and statistical analysis," *Procedia CIRP*, 79:679–683. [CrossRef]
8. Xu Ma, Yanshan Tian, Chu Luo, and Yuehui Zhang(2018)," Predicting future visitors of restaurants using big data", In 2018 International Conference on Machine Learning and Cybernetics (ICMLC), volume 1,IEEE, 269–274 [CrossRef]
9. Xinqing Shu, Pan Wang, "An Improved Adaboost Algorithm based on Uncertain Functions", Proc. of Int. Conf. on Industrial Informatics – Computing Technology, Intelligent Technology, Industrial Information Integration, Dec. 2015.
10. Wenjie Huang and Qing Zhang " A Novel Trigger Model for Sales Prediction with Data Mining Techniques"Data Science Journal,2015. [CrossRef]
11. S.N Gunjal, S K Yadav, D B Kshirsagar, "A Distributed Item Based Similarity Approach For Collaborative Filtering On Hadoop Framework", Advances in Parallel Computing (IOS Press),March 2020.pp.407-415. [CrossRef]
12. S. N. Gunjal, S. K. Yadav and D. B. Kshirsagar, "A hybrid scalable collaborative filteringbased recommendation system using ontology and incremental SVD algorithm"; 2020 International Conference on Smart Innovations in Design, Environment, Management,Planning and Computing (ICSIDEMPC), Aurangabad, India, 2020, pp. 39-45. [CrossRef]
13. A. S. Weigend and N. A. Gershenfeld, "Time series prediction: Forecasting the future and understanding thepast", Addison-Wesley, 1994
14. Zhang B, Tan R, Lin C J. Forecasting of e-commerce transaction volume using a hybrid of extreme learning machine and improved mothflame optimization algorithm[J]. Applied Intelligence, 2020: 1-14.
15. Dovžan, D., Logar, V., & Škrjanc, I. (2012, June). Solving the salesprediction problem with fuzzy evolving methods. In 2012 IEEEInternational Conference on Fuzzy Systems (pp. 1-8). IEEE. [CrossRef]
16. Rezazadeh, A. (2020). A Generalized Flow for B2B Sales Predictive Modeling: An Azure Machine-Learning Approach. Forecasting, 2(3), 267-283. [CrossRef]
17. Ching Wu Chu and Guoqiang Peter Zhang, "A comparative study of linear and nonlinear models for aggregate retails sales forecasting", Int. Journal Production Economics, vol. 86, pp. 217-231, 2003. [CrossRef]
18. O. Ajao Isaac, A. Abdullahi Adedeji, I. Raji Ismail, "Polynomial Regression Model of Making Cost Prediction In Mixed Cost Analysis", Int. Journal on Mathematical Theory and Modeling, vol. 2, no. 2, pp. 14 – 23, 2012.
19. Q.V. Le, "Building high-level features using large scale unsupervised learning," *2013 IEEE international conference on acoustics, speech and signal processing*, IEEE, 2013.
20. D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithm for Boltzmann machines," *Cognitive science*, vol. 9,issue 1, pp. 147-169, 1985. [CrossRef]
21. Jiang W, Zhang L. Geospatial data to images: A deep-learning framework for traffic forecasting[J]. Tsinghua Science and Technology, 2018, 24(1): 52-64. [CrossRef]
22. Zone-Ching Lin, Wen-Jang Wu, "Multiple LinearRegression Analysis of the Overlay Accuracy Model Zone", IEEE Trans. On Semiconductor Manufacturing, vol. 12, no. 2, pp. 229 – 237, May 1999. [CrossRef]
23. N. S. Arunraj, D. Ahrens, A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting, Int. J. Production Economics 170 (2015) 321-335P [CrossRef]
24. Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]//Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016: 785-794. [CrossRef]

AUTHORS PROFILE

Dr. S. N Gunjal, is working as an Assistant Professor in the Department of Computer Engineering, at Sanjivani College of Engineering, Kopargaon, Maharashtra, India. He has completed Ph.D from Shri JJT University Vidyanagari, Jhunjhunu Churu Road, Chudela, District - Jhunjhunu Rajasthan .His Area of interest is Machine Learning, Cloud computing , Data Mining.

© Copyright: All rights reserved.

Dr. D. B Kshirsagar, did his Ph.D in CSE from Shri Guru Gobind Singhji Institute of Engineering and Technology Nanded. Now he is Professor and HOD of Computer Engineering, at Sanjivani College of Engineering, Kopargaon, Maharashtra, India. He has more than 20 years of teaching experience in engineering college. His Area of interest is Machine Learning, Image Processing, Data Mining.

Dr. B.J Dange, is working as an Associate Professor in the Department of Computer Engineering, at Sanjivani College of Engineering, Kopargaon, Maharashtra, India. He has completed Ph.D from Shri JJT University Vidyanagari, Jhunjhunu Churu Road, Chudela, District -Jhunjhunu Rajasthan .His Area of interest is Machine Learning, Image Processing, Data Mining.

Dr. H.E Khodke, is working as an Assistant Professor in the Department of Computer Engineering, at Sanjivani College of Engineering, Kopargaon, Maharashtra, India. He has completed Ph.D from Shri JJT University Vidyanagari, Jhunjhunu Churu Road, Chudela, District - Jhunjhunu Rajasthan .His Area of interest is Machine Learning, Web Technology, Data Mining.

Dr. C.S Kulkarni, is working as an Associate Professor and Head of Dept. in the Department of Computer Engineering, at Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati, Maharashtra, India. He has completed Ph.D from Shri JJT University Vidyanagari, Jhunjhunu Churu Road, Chudela, District - Jhunjhunu Rajasthan .His Area of interest is Machine Learning, Cloud Computing, Data Mining.

Retrieval Number: 100.1/ijitee.F99160511622
DOI: 10.35940/ijitee.F9916.0511622
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)
© Copyright: All rights reserved.