Machine Learning Models for Analysing and Predicting Forecast Big Mart Sales.

Dr S.B.Thorat, Mr Amol Suryawanshi

Director[1], Assistant Professor[2]

Department of Computer Science
SSBES ITM College Nanded
suryakant_thorat@yahoo.com, suryawanshiamolv@gmail.com

Abstract:
The term "Machine Learning" refers to a group of techniques that allow software programmes to improve their predictive accuracy without being explicitly programmed to do so. Machine learning is based on the idea of creating models and deploying algorithms that can receive data, statistically evaluate it to predict a result, and then update their forecasts as more data becomes available. These models can be utilised in a variety of contexts and adapted to the leadership's objectives, enabling more targeted actions to be taken in the direction of the organization's aim. This article analyses the example of Big Mart, a one-stop shopping centre, to estimate the sales of various products and comprehend the effects of numerous factors on these sales. Big Mart has amassed a vast dataset, and by employing the appropriate methods for developing a predictive model, we can obtain highly accurate results that we can then use to inform our company's decisions and expand our operations.
Keywords: Machine Learning,Bigmart sales, XGBoost,Random Forest

1. Introduction

The everyday competition among retail businesses has intensified as a result of the rapid growth of international shopping malls and e-commerce. By offering limited-time discounts and promotions, markets fight for customers. This data is used to forecast future demand and improve stock management, shipping, and other back-end procedures. In order to outperform low-cost techniques of prediction, modern machine learning algorithms offer extremely complicated approaches for predicting or forecasting sales for any type of company. It is advantageous to have access to more accurate estimates for a number of reasons, including the improvement and development of more effective market marketing strategy. Big Mart is a major retail conglomerate with locations throughout the globe. Data scientists look at Big Mart's geography and product trends to find possible distribution hubs. Data scientists can improve their predictions of Big Mart sales by shop and product using computer-generated data. A lot of organisations are largely dependent on data, and precise market forecasting is essential. Forecasts must take into account a number of factors, such as consumer preferences and spending patterns. This study might be helpful for corporate financial management as well. This is a wonderful illustration of machine learning in action. Based on the results of data mining activities such data discovery, data transformation, feature development, model construction, and testing, we provide a sales estimate in this study. This procedure filters outliers, irregularities, and missing data from the raw data that a major market has received before examination. Following that, the data will be utilised to train an algorithm that will create a model.

2,Literature Survey

Described by Nikita Malik and Karan Singh The use of machine learning for sales forecasting has been discussed. She implemented a machine learning algorithm (linear regression, Random Forest, etc.). She has conducted research on a limited sample of products and discovered stores with ties to these items. There is a degree of precision between 70% and 80%.

The authors of A Prediction of Big-box Retailer Sales Based on Random Forests and Multiple Linear Regression, Kadam, H., Shevade, R., Ketkar, P., and Rajguru utilised Random Forest and Linear Regression for prediction analysis, which provides a lower level of precision. This issue is effectively resolved by the XG enhance Algorithm, which improves both accuracy and efficiency.

Wheelwright was written by S. Makridakis.

Hyndman, you get an A+. As R.J. explains Numerous forecasting methodologies and applications suffer from a lack of data and a brief shelf life. Consequently, it is possible to accurately predict outcomes using specific categories of data, such as historical data, from consumer-oriented markets with unpredictability in demand.

C. M. Wu, P. Patil, and S. Gunaseelan elaborate on the Black Friday Comparison of Machine Learning Algorithms for Multiple Regression. Using Neural Networks for Solr Algorithm Evaluation. To overcome this, it is inefficient to compare algorithms using complex models such as neural networks; instead, simplified algorithms can be used for prediction.

In a forthcoming article by Theresa Inedi, Dr. Venkata Reddy Medikonda, and K.V. Narasimha Reddy (March 2020), sales forecasting using Exploratory Machine Learning is described and discussed. They carried out the entire procedure by determining the correct processes, which included data collection, thesis generation to effectively comprehend flaws, data cleaning, and data processing. Various models, including Linear Regression, Decision Tree Regression, Ridge Regression, and a Random Forest model, were used to predict sales outcomes. They concluded that using multiple models produced more accurate predictions than using just one.

3. Methodology

In this study, we outline a model that follows the steps in Figure 1 to make reliable sales predictions from the Big Mart sales dataset. The ultimate form of the model depends on each method. Our model made use of data from the 2013 big-mart. We used ML classifiers such Decision trees, linear regression, Random forest, and Xgboost after standard preprocessing and missing value filling. The accuracy of Big Mart's sales projections is evaluated using both MAE and RSME. Root-mean-squared error (RSME) and lowest mean absolute error (MAE) were shown to be the best measures of accuracy for this model's predictions. The details of our suggested approach will be discussed below.

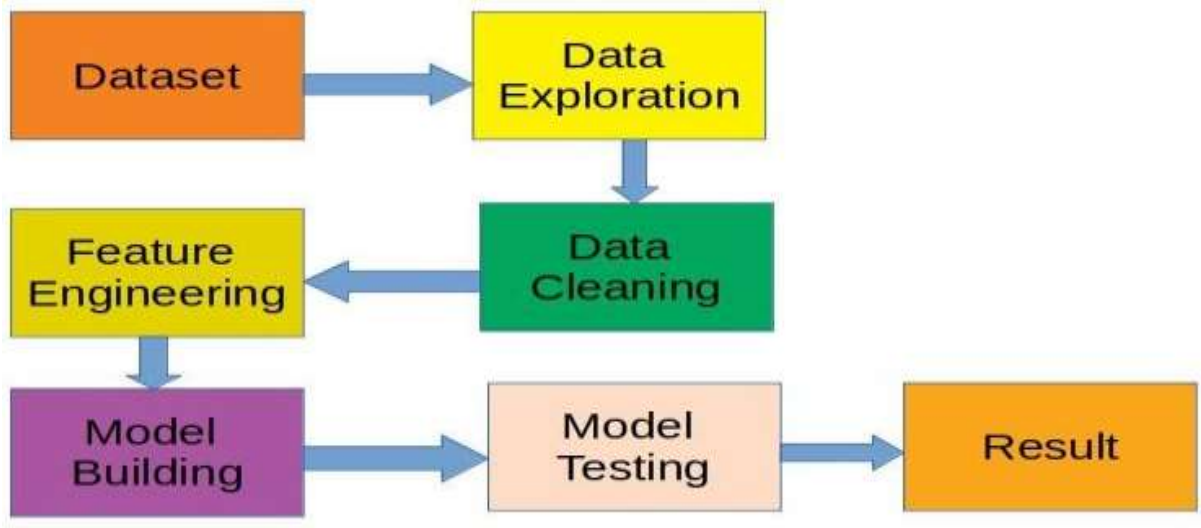


Fig. 1. System methodology

4.Algorithms employed

4.1 XGBoost Model Report

As an effective open-source implementation of the popular gradient boosted trees method, XGBoost has found widespread application. Gradient boosting is a supervised learning approach used to boost the prediction of a target variable by combining the outputs of numerous models with varying degrees of sophistication.

In gradient boosting regression, the weak learners take the shape of trees, with each input data point being assigned to a leaf on which a continuous score is kept. Minimising the regularised (L1 and L2) objective function of XGBoost (the regression tree functions) combines a convex loss function (based on the dissimilarity between the predicted and target outputs) with a penalty term for model complexity. To improve the accuracy of the final prediction, additional trees are added to the training process to predict the residuals or errors of previously trained trees. In order to minimise the loss while including more models, gradient boosting uses a gradient descent technique, hence the name.

4.2 Random forest Model Report

When it comes to achieving scalability, few algorithms are as effective as the Random Tree.

The random forest algorithm is quite effective in predicting future earnings. Predicting the results of machine learning initiatives is straightforward, both to implement and to understand. The hyper parameters of the random forest classifier are similar to those of the decision tree, making it an ideal tool for sales forecasting. The idea behind decision trees and tree models is the same. Figure 5 shows the connection between decision trees and random forest. For regression-based prediction problems, the random forest regressor class in the sklearn.ensemble package is a useful tool. The random forest regress or, or the n estimator’s parameter, is essential.

4.3 Decision tree Model Report

A decision tree's algorithm for choosing a dataset's category is executed from the tree's leaf node. By comparing the values of the root property with those of the record (real dataset) attribute, this algorithm begins at the root node and works its way up the branch.

After comparing the attribute value to the child nodes, the algorithm advances to the next node. Repeat this process until you reach a leaf node. To better grasp the big picture, consider the following algorithm:

First, as instructed by S, the entire dataset is located at the tree's root node.

In Step 2, we'll use Attribute Selection Measure (ASM) to zero in on the most helpful metric in the dataset.

Separate the S into subsets containing candidate values for the most desirable characteristics (Step 3).

Fourth, create the best attribute node in the decision tree.

The subsets of the dataset are used in Step 5 to generate new decision trees in a recursive fashion. Keep going until you get to a point where you can no longer divide the nodes into smaller groups; this is the leaf node.

4.4 Linear regression model Report

Using a set of independent variables to construct forecasts about a continuous dependent variable is one definition of regression. The technique is called parametric because it uses assumptions that can change depending on the circumstances.

$Y = \beta o + \beta 1X + \in (1)$

Equation shown in eq.1 is used for simple linear regression. These parameters can be said as:

Y - Variable to be predicted

X - Variable(s) used for making a prediction

βo - When X=0, it is termed as prediction value or can be referred to as intercept term

$\beta 1$ - when there is a change in X by 1 unit it denotes change in Y. It can also be said as slope term.

$\in$ -The difference between the predicted and actual values is represented by this parameter and

Equally stands for the worth of the remainder. There will always be a discrepancy between the real values and the anticipated values (irreducible error), therefore we can never fully rely on the results predicted by the learning process no matter how well the model is trained, tested, and validated. Dietterich also provides some alternate approaches that might be used to evaluate various learning algorithms.

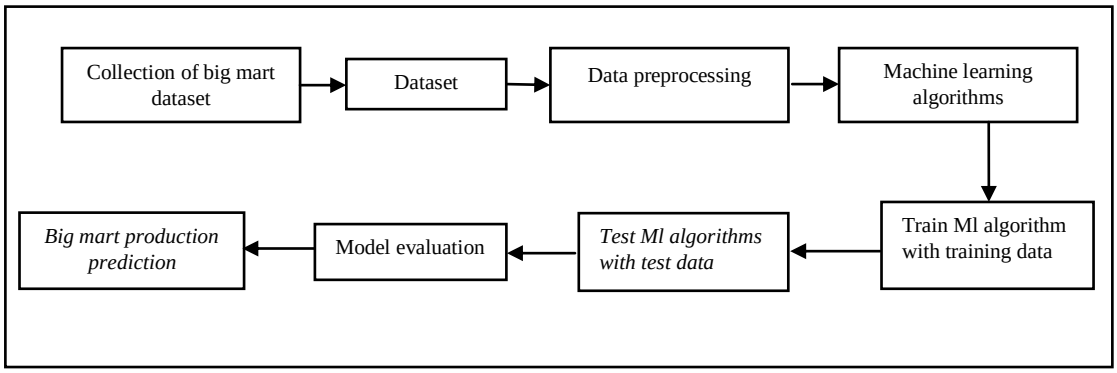
5. Metrics for modelling

One crucial metric during the estimating phase is the measuring of error. When evaluating the precision of a continuous variable, the RMSE and MAE are the most used measures of error. Using either MAE or RMSE, you may put a numerical value on how much off your model is on average when making predictions about that variable. The mean absolute error (MAE) is calculated by averaging the absolute differences between predictions and observations for each person in the test sample. To calculate RMSE, take the square root of the sum of all squared deviations between your prediction and the actual observation. Relative to R2, RMSE is an absolute measure of fit. The root-mean-squared error (RMSE) is a quadratic scoring rule that can be used to gauge the typical error associated with a variable. The accuracy of a regression model is improved when the root mean square error (RMSE) is small, whether the regression is linear or multivariate. As the RMSE ratio is calculated to be identical to the ratio between the train and test samples, it may be concluded that there is little to no difference between the two in this work. RMSE is a good indicator of the accuracy with which our model predicts responses, coupled with measuring precision and other necessary

characteristics. Extra data mining with outlier detection and strong leverage points could yield substantial gains. Another, conceptually simpler method, is to use ensemble learning to assemble multiple low-dimensional sub-models that can be verified by domain experts.The error measurement is an important metric in the estimation period. Root mean squared error (RMSE) and Mean Absolute Error (MAE) are generally used for continuous variable's
The error measurement is an important metric in the estimation period. Root mean squared error (RMSE) and Mean Absolute Error (MAE) are generally used for continuous variable's
The error measurement is an important metric in the estimation period. Root mean squared error (RMSE) and Mean Absolute Error (MAE) are generally used for continuous variable's
The error measurement is an important metric in the estimation period. Root mean squared error (RMSE) and Mean Absolute Error (MAE) are generally used for continuous variable's
The error measurement is an important metric in the estimation period. Root mean squared error (RMSE) and Mean Absolute Error (MAE) are generally used for continuous variable's
The error measurement is an important metric in the estimation period. Root mean squared error (RMSE) and Mean Absolute Error (MAE) are generally used for continuous variable's
The error measurement is an important metric in the estimation period. Root mean squared error (RMSE) and

5.1 Proposed architecture

Simply said, it's predicated on a Machine Learning model that could boost the efficiency of specialised computer vision programmes. The proposed method is grounded in particular configurations of the Machine Learning model, as shown in Figure



1.

Fig.2 proposed architecture

Proposed algorithm

Algorithm: Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms.(PABGS-MLA)

Inputs: big mart dataset details as P, machine learning models as M

Output: sales production as R

1. Start
2. Input sales dataset , P
3. Pre-processing
4. Splitting data
5. Extract features from training set()
6. For each model m in M
7. Train the model m
8. End For
9. For each model m in M
10. Use model for testing
11. Evaluate
12. Display results
13. End For
14. Save the model()
15. Predict the sales
16. Return R

Algorithm 3: Proposed algorithm

According to the first algorithm, a dataset is incorporated into the model. Several upgraded approaches can be set up in the model to boost performance. The algorithm's iterative process operates throughout multiple time periods, and the model is periodically updated.

6.Dataset description

Big Mart has 10 stores in various locations as of the 2013 data collection, each of which carried 1559 different products. It is possible to make judgements about how a product's characteristics affect sales if there is adequate data available. The data assumes the format seen in Fig.2 when the head () function is used on a dataset variable.

```
In [7]: df.head()
        #understanding rows and column
```

Out[7]:

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Size	Outlet_Locatic
0	FDA15	9.30	Low Fat	0.016047	Dairy	249.8092	OUT049	1999	Medium	
1	DRC01	5.92	Regular	0.019278	Soft Drinks	48.2692	OUT018	2009	Medium	
2	FDN15	17.50	Low Fat	0.016760	Meat	141.6180	OUT049	1999	Medium	
3	FDX07	19.20	Regular	0.000000	Fruits and Vegetables	182.0950	OUT010	1998	NaN	
4	NCD19	8.93	Low Fat	0.000000	Household	53.8614	OUT013	1987	High	

Figure 3:dataset

Each of the underlying patterns that the raw data may provide helps to understand the issue at hand and the topic of interest more thoroughly. Data must be handled carefully, though, as it could contain null values, redundant values, or different types of ambiguity, all of which call for pre-processing. Therefore, it is crucial to fully explore the dataset. This dataset has undergone pre-processing that entails analysis on the independent variables, such as checking for null values in each column and then replacing or filling them with supported acceptable data types. This ensures that analysis and model fitting go off without a hitch. The variable count for numerical columns and the modal values for categorical columns are two examples of the information supplied by the various representations made with the aid of the Pandas tools that you can observe. Prioritising which values to investigate in further detail can be done using the median value and the percentile rankings of the highest and least values in a set of numerical columns. Additionally, label processing and a one-hot encoding technique are used in model construction, both of which depend on the various data types contained in the different columns..

7. Results

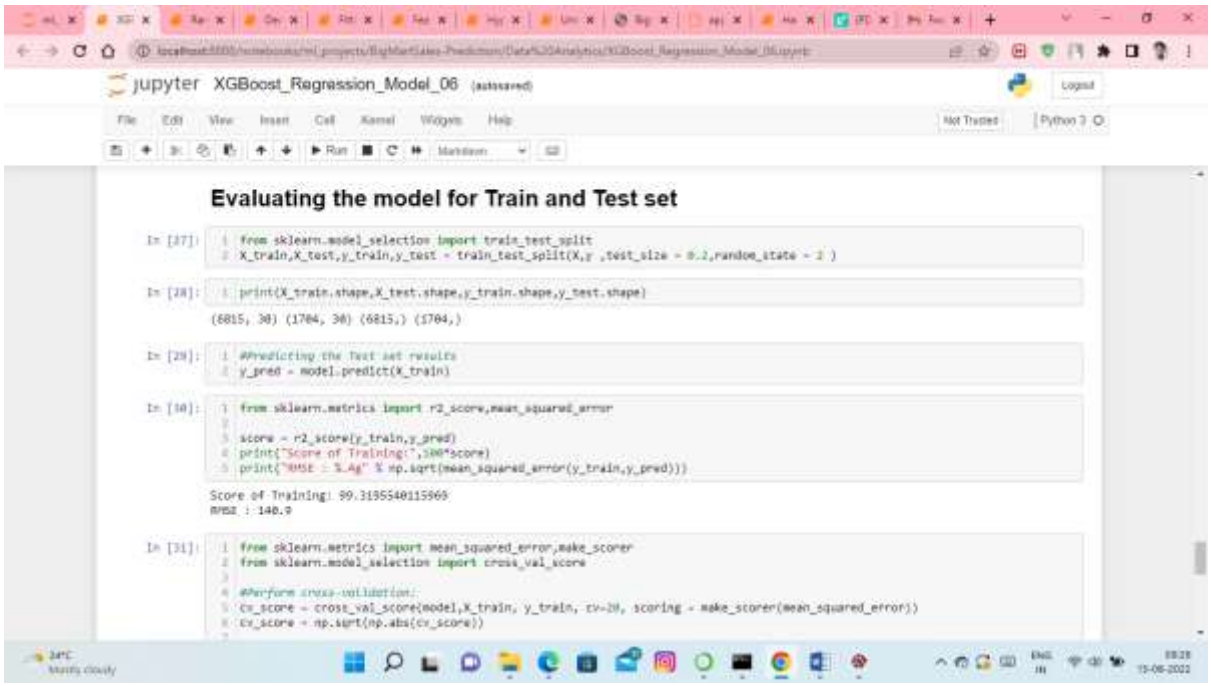


Figure 4:Model evalution

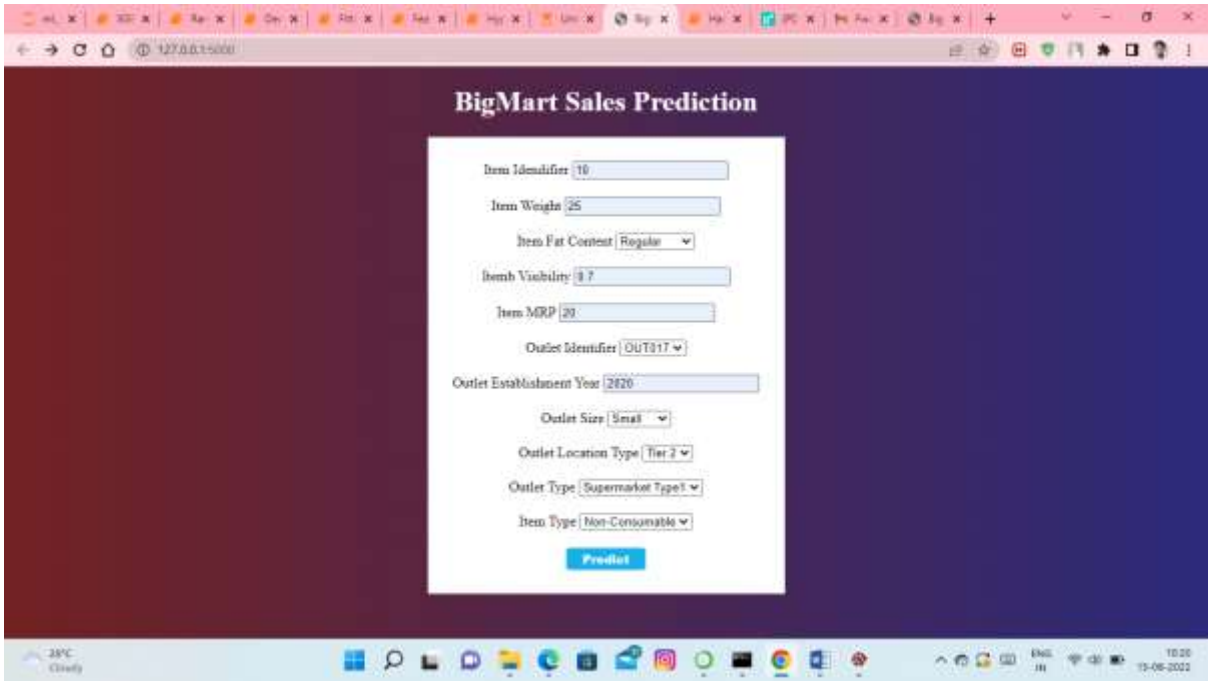


Figure 5: result page 1

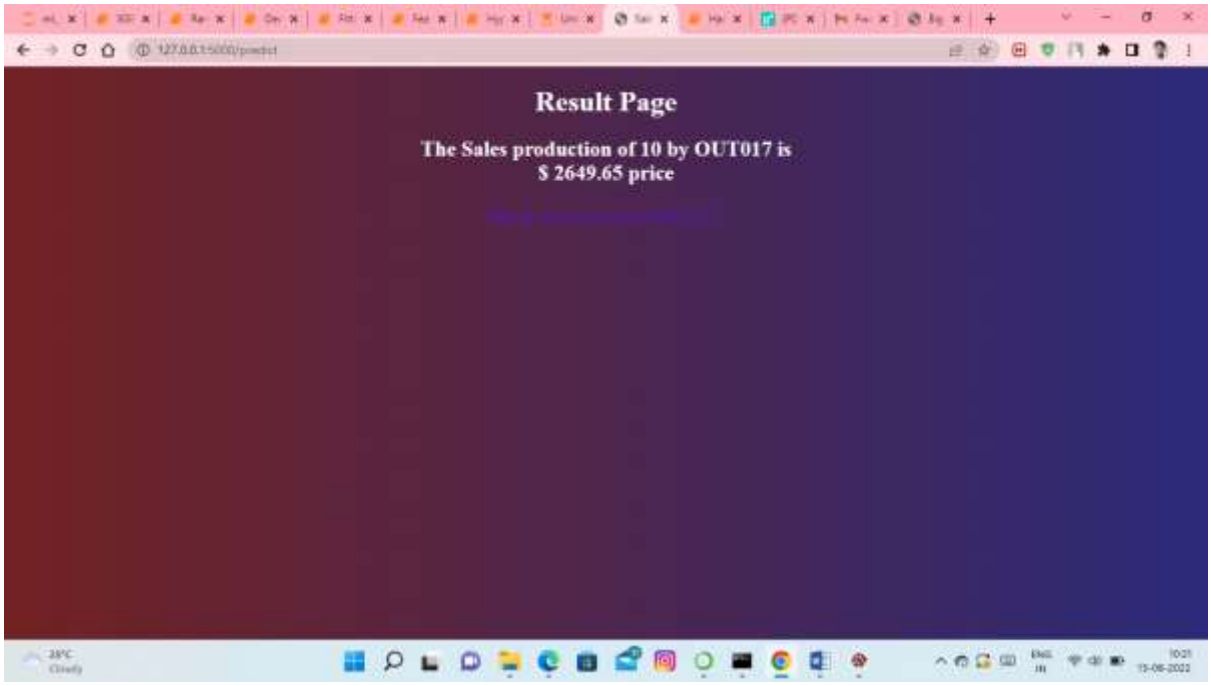


Figure 6:Pridictions page

7.Conclusion

Every shopping centre wants to know the client demands in advance in the modern, digitally connected world in order to avoid running out of sale items at all times. Companies or shopping centres are getting better at predicting daily changes in consumer demand or product sales. For precise sales forecasting, extensive research is being conducted at the

organisation level. The main markets want a more accurate prediction algorithm because a company's profit is closely correlated to how well it predicts sales, preventing losses for the business. In this study, we developed a predictive model for forecasting product sales from a specific outlet by modifying Gradient Boosting Machines as Xgboost approach and tested it on the 2013 Big Mart dataset. In comparison to other strategies like decision trees, linear regression, etc., experiments show that our technique produces more accurate predictions. A comparison of various models is then compiled.Additionally, it is found that our model, which has the lowest MAE and RMSE, outperforms other models.

8.References

[1] 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.

[2] Wang, Haoxiang. "Sustainable development and management in consumer electronics using soft computation." Journal of Soft Computing Paradigm (JSCP) 1, no. 01 (2019): 56.-2.

[3] Suma, V., and ShavigeMalleshwara Hills. "Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics." Journal of Soft Computing Paradigm (JSCP) 2, no. 02 (2020): 101- 110

[4] Giuseppe Nunnari, Valeria Nunnari, "Forecasting Monthly Sales Retail Time Series: A Case Study", Proc. of IEEE Conf. on Business Informatics (CBI), July 2017.

[5] 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.

[6] O. Ajao Isaac, A. AbdullahiAdedeji, 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.

[7] C. Saunders, A. Gammerman and V. Vovk, "Ridge Regression Learning Algorithm in Dual Variables", Proc. of Int. Conf. on Machine Learning, pp. 515 – 521, July 1998.IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 56, NO. 7, JULY 2010 3561.

[8] "Robust Regression and Lasso". HuanXu, Constantine Caramanis, Member, IEEE, and ShieMannor, Senior Member, IEEE. 2015 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration."An improved Adaboost algorithm based on uncertain functions".ShuXinqing School of Automation Wuhan University of Technology.Wuhan, China Wang Pan School of the Automation Wuhan University of Technology Wuhan, China.

[9] XinqingShu, 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] A. S. Weigend and N. A. Gershenfeld, "Time series prediction: Forecasting the future and understanding the past", Addison-Wesley, 1994.

[11] 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

[12] D. Fantazzini, Z. Toktamysova, Forecasting German car sales using Google data and multivariate models, Int. J. Production Economics 170 (2015) 97-135.

[13] X. Yua, Z. Qi, Y. Zhao, Support Vector Regression for Newspaper/Magazine Sales Forecasting, Procedia Computer Science 17 (2013) 1055–1062.

[14] E. Hadavandi, H. Shavandi, A. Ghanbari, An improved sales forecasting approach by the integration of genetic fuzzy systems and data clustering: a Case study of the printed circuit board, Expert Systems with Applications 38 (2011) 9392–9399.

[15] P. A. Castillo, A. Mora, H. Faris, J.J. Merelo, P. GarciaSanchez, A.J. Fernandez-Ares, P. De las Cuevas, M.I. Garcia-Arenas, Applying computational intelligence methods for predicting the sales of newly published books in a real editorial business management environment, Knowledge-Based Systems 115 (2017) 133-151.

Beheshti-Kashi, S., Karimi, H.R., Thoben, K.D., L¨utjen, M., Teucke, M.: A survey on retail sales forecasting and prediction in fashion markets. Systems Science & Control Engineering 3(1), 154–161 (2015)
2. Bose, I., Mahapatra, R.K.: Business data mininga machine learning perspective. Information & management 39(3), 211–225 (2001)
3. Chu, C.W., Zhang, G.P.: A comparative study of linear and nonlinear models for aggregate retail sales forecasting. International Journal of production economics 86(3), 217–231 (2003
Beheshti-Kashi, S., Karimi, H.R., Thoben, K.D., L¨utjen, M., Teucke, M.: A survey on retail sales forecasting and prediction in fashion markets. Systems Science & Control Engineering 3(1), 154–161 (2015)
2. Bose, I., Mahapatra, R.K.: Business data mininga machine learning perspective. Information & management 39(3), 211–225 (2001)
3. Chu, C.W., Zhang, G.P.: A comparative study of linear and nonlinear models for aggregate retail sales forecasting. International Journal of production economics 86(3), 217–231 (2003
[16] Pei Chann Chang and Yen-Wen Wang, “Fuzzy Delphi and back propagation model for sales forecasting in PCB industry”, Expert systems with applications, vol. 30,pp. 715-726, 2006.

[17] R. J. Kuo, Tung Lai HU and Zhen Yao Chen “application of radial basis function neural networks for sales forecasting”, Proc. of Int. Asian Conference on Informatics in control, automation, and robotics, pp. 325- 328, 2009.

[18] R. Majhi, G. Panda, G. Sahoo, and A. Panda, “On the development of Improved Adaptive Models for Efficient Prediction of Stock Indices using Clonal-PSO (CPSO) and PSO Techniques”, International Journal of Business Forecasting and Market Intelligence, vol. 1, no. 1, pp.50-67, 2008.

[19]Suresh K and Praveen O, "Extracting of Patterns Using Mining Methods Over Damped Window," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2020, pp. 235-241, DOI: 10.1109/ICIRCA48905.2020.9182893.

Beheshti-Kashi, S., Karimi, H.R., Thoben, K.D., L¨utjen, M., Teucke, M.: A survey on retail sales forecasting and prediction in fashion markets. Systems Science & Control Engineering 3(1), 154–161 (2015

[20] Shobha Rani, N., Kavyashree, S., &Harshitha, R. (2020). Object Detection in Natural Scene Images Using Thresholding Techniques. Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020, Iciccs, 509–515.