ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms

[1]B.N.S.GUPTA, [2]AKULA DEVI BHAVANI

[1](Assistant Professor), MCA, S.V.K.P & Dr K.S. Raju Arts & Science College

[2]MCA, scholar, S.V.K.P & Dr K.S. Raju Arts & Science College

ABSTRACT

Currently, supermarket run-centres, Big Marts keep track of each individual item's sales data in order to anticipate potential consumer demand and update inventory management. Anomalies and general trends are often discovered by mining the data warehouse's data store. For retailers like Big Mart, the resulting data can be used to forecast future sales volume using various machine learning techniques like big mart. A predictive model was developed using Xgboost, Linear regression, Polynomial regression, and Ridge regression techniques for forecasting the sales of a business such as Big -Mart, and it was discovered that the model outperforms existing models.

1.INTRODUCTION

Everyday competitiveness between various shopping centres as and as huge marts is becoming higherintense, violent just because of the quick development of global malls also online shopping. Each marketseeks to offer personalized and limited time deals toattract many clients relying on period of time, so that each item's volume of sales may be estimated for the organization's stock control, transportation andlogistical services. The current machine learning algorithm is very advanced and provides methods forpredicting or forecasting sales any kind of organization,extremely beneficial to overcome low – priced used for prediction. Always better prediction is helpful, both indeveloping and improving marketing strategies for themarketplace, which is also particularly helpful.

A great deal of work having been gotten really intended to date the territory of deals foreseeing. Aconcise audit of the important work in the field of big mart deals is depicted in this part. Numerous other Measurable methodologies, for example, withregression, (ARIMA) Auto-Regressive IntegratedMoving Average, (ARMA) Auto-Regressive Moving Average, have been utilized to develop a few dealsforecast standards. Be that as it may, deals anticipatingis a refined issue and is influenced by both outer and inside facto s, and there are two significant detriments to the measurable technique as set out in A. S. Weigendet A mixture occasional quantum relapse approach and(ARIMA) Auto-Regressive Integrated MovingAverage way to deal with every day food dealsanticipating were recommend by N. S. Arunraj andfurthermore found that the exhibition of the individual model was moderately lower than that of the crossovermodel.

E. Hadavandi utilized the incorporation of "Genetic Fuzzy Systems (FS)" and information gathering to conjecture the deals of the printed circuit board. In theirpaper, K-means bunching delivered K groups of allinformation records. At that point, all bunches were taken care of into autonomous with a data set tuningand rule-based extraction ability. Perceived work in the field of deals gauging was done by P.A. Castillo, Sales estimating of new distributed books was done in apublication market the executives setting utilizingcomputational techniques. "Artificial neuralorganizations" are additionally utilized nearby incomeestimating. Fluffy Neural Networks have been created with the objective of improving prescient effectiveness,and the Radial "Base Function Neural Network(RBFN)" is required to have an incredible potential foranticipating deals.

2.LITERATURE SURVEY

1) A comparative study of linear and nonlinear models for aggregate retails sales forecasting

AUTHORS: Ching Wu Chu and Guoqiang Peter Zhang The purpose of this paper is to compare the accuracy of various linear and nonlinear models for forecasting aggregate retail sales. Because of the strong seasonal fluctuations observed in the retail sales, several traditional seasonal forecasting methods such as the time series approach

and the regression approach with seasonal dummy variables and trigonometric functions are employed. The nonlinear versions of these methods are implemented via neural networks that are generalized nonlinear functional approximators. Issues of seasonal time series modeling such as deseasonalization are also investigated. Using multiple cross-validation samples, we find that the nonlinear models are able to outperform their linear counterparts in out-ofsample forecasting, and prior seasonal adjustment of the data can significantly improve forecasting performance of the neural network model. The overall best model is the neural network built on deseasonalized time series data. While seasonal dummy variables can be useful in developing effective regression models for predicting retail sales, the performance of dummy regression models may not be robust. Furthermore, trigonometric models are not useful in aggregate retail sales forecasting.

2) Sustainable development and management in consumer electronics using soft computation

AUTHORS: Wang, Haoxiang Combination of Green supply chain management, Green product deletion decision and green cradle-to-cradle performance evaluation with Adaptive-Neuro-Fuzzy Inference System (ANFIS) to create a green system. Several factors like design process, client 13 specification, computational intelligence and soft computing are analysed and emphasis is given on solving problems of real domain. In this paper, the consumer electronics and smart systems that produce nonlinear outputs are considered. ANFIS is used for handling these nonlinear outputs and offer sustainable development and management. This system offers decision making considering multiple objectives and optimizing multiple outputs. The system also provides efficient control performance and faster data transfer.

3) Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics

AUTHORS: Suma, V., and Shavige Malleshwara Hills There has been an increasing demand in the e-commerce market for refurbished products across India during the last decade. Despite these

demands, there has been very little research done in this domain. The real-world business environment, market factors and varying customer behavior of the online market are often ignored in the conventional statistical models evaluated by existing research work. In this paper, we do an extensive analysis of the Indian e-commerce market using data-mining approach for prediction of demand of refurbished electronics. The impact of the real-world factors on the demand and the variables are also analyzed. Real-world datasets from three random e-commerce websites are considered for analysis. Data accumulation, processing and validation is carried out by means of efficient algorithms. Based on the results of this analysis, it is evident that highly accurate prediction can be made with the proposed approach despite the impacts of varying customer behavior and market factors. The results of analysis are represented graphically and can be used for further analysis of the market and launch of new products.

4) Forecasting Monthly Sales Retail Time Series: A Case Study 14

AUTHORS: Giuseppe Nunnari, Valeria Nunnari This paper presents a case study concerning the forecasting of monthly retail time series recorded by the US Census Bureau from 1992 to 2016. The modeling problem is tackled in two steps. First, original time series are de-trended by using a moving windows averaging approach. Subsequently, the residual time series are modeled by Non-linear Auto-Regressive (NAR) models, by using both Neuro-Fuzzy and FeedForward Neural Networks approaches. The goodness of the forecasting models, is objectively assessed by calculating the bias, the mae and the rmse errors. Finally, the model skill index is calculated considering the traditional persistent model as reference. Results show that there is a convenience in using the proposed approaches, compared to the reference one. 5) Multiple Linear Regression Analysis of the Overlay Accuracy Model Zone AUTHORS: Zone-Ching Lin, Wen-Jang Wu The multiple linear regression method was used to analyze the overlay accuracy model and study the feasibility of using linear methods to solve parameters of nonlinear overlay equations. The methods of analysis include changing the number of sample points to derive the least sample number required for solving the accurate

estimated parameter values. Besides, different high-order lens distortion parameters were ignored, and only the various modes of low-order parameters were regressed to compare their effects on the overlay analysis results. The findings indicate that given a sufficient number of sample points, the usage of multiple linear regression analysis to solve the high-order nonlinear overlay accuracy model containing seventh-order lens distortion parameters is feasible. When the estimated values of low-order overlay distortion parameters are far greater than those of high-order lens distortion parameters, excellent overlay improvement can still be obtained even if the high-order lens distortion parameters are ignored. When the overlay at the four corners of image field obviously 15 exceeds that near the center of image field, it is found, through simulation, that the seventh-order parameters overlay model established in this paper has to be corrected by high-order lens distortion parameters to significantly improve the overlay accuracy.

3. EXISTING SYSTEM

Auto-Regressive Integrated Moving Average, (ARMA) Auto-Regressive Moving Average, have been utilized to develop a few deals forecast standards. Be that as it may, deals anticipating is a refined issue and is influenced by both outer and inside factors, and there are two significant detriments to the measurable technique as set out in A. S. Weigend et A mixture occasional quantum relapse approach and (ARIMA) Auto-Regressive Integrated Moving Average way to deal with every day food deals anticipating were recommend by N. S. Arunraj and furthermore found that the exhibition of the individual model was moderately lower than that of the crossover model. E. Hadavandi utilized the incorporation of “Genetic Fuzzy Systems (GFS)” andϖ information gathering to conjecture the deals of the printed circuit board. In their paper, K-means bunching delivered K groups of all information records. At that point, all bunches were taken care of into autonomous with a data set tuning and rule-based extraction ability. Perceived work in the field of deals gauging was done by P.A. Castillo, Salesϖ estimating of new distributed books was done in a publication market the executives setting utilizing computational techniques. “Artificial neural organizations” are

additionally utilized nearby income estimating. Fluffy Neural Networks have been created with the objective of improving prescient effectiveness, and the Radial “Base Function Neural Network (RBFN)” is required to have an incredible potential for anticipating deals.

DISADVANTAGES OF EXISTING SYSTEM:

Complex models like neural networks are overkill for simple problems likeϖ regression. Existing system models prediction analysis which gives less accuracy.ϖ Forecasting methods and applications contains Lack of Data and short lifeϖ cycles. So some of the data like historical data, consumer-oriented markets face uncertain demands, can be prediction for accurate result.

3.1 PROPOSED SYSTEM:

The objective of this proposed system is to predict the future sales from givenϖ data of the previous year's using Decision Tree Regression Another objective is to conclude the best model which is more efficient and givesϖ fast and accurate result by using Decision Tree Regression. To find out key factors that can increase their sales and what changes could beϖ made to the product or store's characteristics. Experts also shown that a smart sales forecasting program is required to manageϖ vast volumes of data for business organizations. We are predicting the accuracy for Decision Tree Regression. Our predictionsϖ help big marts to refine their methodologies and strategies which in turn helps them to increase their profit. The results predicted will be very useful for the executives of the company to know about their sales and profits. This will also give them the idea for their new locations or Centre‟s of Bigmart

ADVANTAGES OF PROPOSED SYSTEM:

Business assessments are based on the speed and precision of the methodsϖ used to analyze the results. The Machine Learning Methods presented in this research paper should provide an effective method for data shaping and decision-making. New approaches that can better identify consumer needs and formulateϖ marketing plans will be implemented. The outcome of machine learning algorithm will help to select the most suitableϖ demand prediction algorithm

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

and with the aid of which BigMart will prepare its marketing campaigns.

4. OUTPUT SCREENS

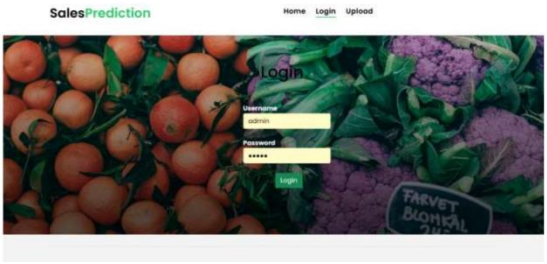

FIG 4.1 Log IN PAGE

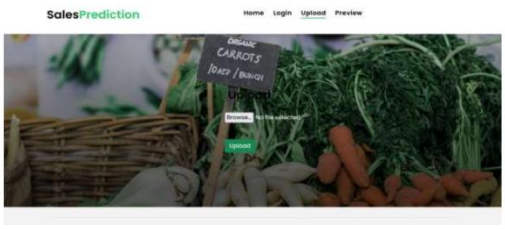

Fig 4.2 Data Collection From Kaggle

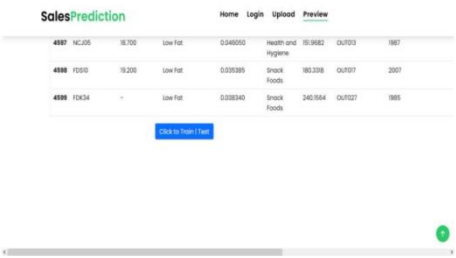

Fig 4.3 Prepocessed Dataset Of Bigmart Sales

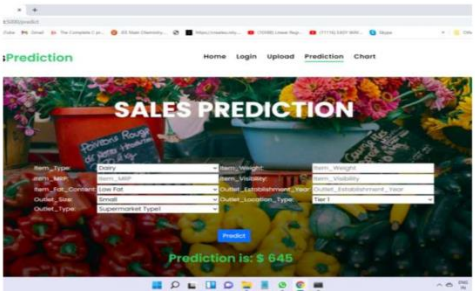

Fig 4.4 Prediction Of Sales Of A Product

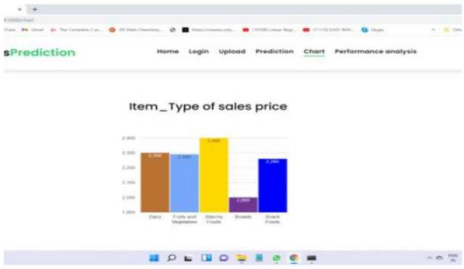

Fig 4.5 Comparision Graph On Sales Of Item Type

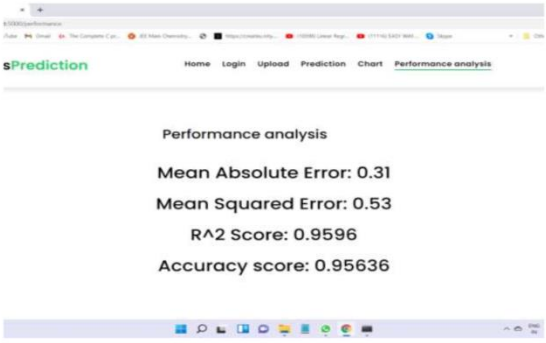

Fig 4.6 Performance Values Of Proposed Model

5. CONCLUSIONS

In this work, the effectiveness of various algorithms on the data on revenue and review of, bestperformance-algorithm, here propose a software to using regression approach for predicting the salescentered on sales data from the past the accuracy oflinear regression prediction can be enhanced with this method, polynomial regression, Ridge regression, andXgboost regression can be determined. So, we canconclude ridge and Xgboost regression gives the betterprediction with respect to Accuracy, MAE and RMSE than the Linear and polynomial regression approaches.In future, the forecasting sales and building a sales plancan help to avoid unforeseen cash flow and manage production, staff and financing needs moreeffectively.In future work we can also consider with theARIMA model which shows the time series graph

6. REFERENCE

[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 Shavige Malleshwara 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 41

[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]https://halobi.com/blog/sales-forecasting-five-uses/.

[6] Zone-Ching Lin, Wen-Jang Wu, “Multiple Linear Regression Analysis of the Overlay Accuracy Model Zone”, IEEE Trans. On Semiconductor Manufacturing, vol. 12, no. 2, pp. 229 – 237, May 1999.

[7] 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.

[8] 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.

[9] "Robust Regression and Lasso". Huan Xu, Constantine Caramanis, Member, IEEE, and Shie Mannor, Senior Member, IEEE. 2015 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration."An improved Adaboost algorithm based on uncertain functions".Shu Xinqing School of Automation Wuhan University of Technology.Wuhan, China Wang Pan School of the Automation Wuhan University of Technology Wuhan, China. 42

[10] 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.