

Sales Projection by using XGBoost, Ridge Regression, Polynomial Regression & Linear Regression Algorithms in Machine Learning

[1] Imran Bin Ibrahim, [2] Syed Adnan, [3] Shaik Sharf Uddin,

[4] Dr. Pathan Ahmed Khan

[1] BE Student, Dept. of Computer Science Engineering, ISL Engineering College

[2] BE Student, Dept. of Computer Science Engineering, ISL Engineering College

[3] BE Student, Dept. of Computer Science Engineering, ISL Engineering College

[4] Associate Professor, Dept. of Computer Science Engineering, ISL Engineering College

Article Info

Page Number: 1309-1315

Publication Issue:

Vol. 72 No. 1 (2023)

ABSTRACT

The BigMarts, the supermarket-run shopping malls, currently track sales data for each individual item to forecast possible consumer demand and revise inventory control. By mining the data store of the data warehouse, anomalies and broad trends are frequently found. The generated data can be utilised by retailers like BigMart to predict future sales volume using a variety of machine learning approaches. For projecting the sales of a company like BigMart, a predictive model was created utilising the XGBoost, Linear regression, Polynomial regression, and Ridge regression techniques. It is found that the model beats other models. To adapt the business model to anticipated outcomes, the sales estimate is based on BigMart sales for various stores. Through various machine learning techniques, the generated data may subsequently be utilised to forecast possible sales volumes for stores like BigMart. Price, outlet, and outlet location all Identifier's included in the estimated cost of the proposed system. Many networks make use of different machine learning techniques, including linear regression, using decision tree algorithms and an XGBoost regressor, BigMart sales may be accurately predicted. Finally, hyperparameter tweaking is employed to assist in selecting pertinent hyperparameters that enhance the algorithm and yield the maximum accuracy.

Keywords: Sales prediction, Continuous values prediction, Stacking, Machine Learning.

Article History

Article Received: 15 October 2022

Revised: 24 November 2022

Accepted: 18 December 2022

1.1 Introduction

Sales is a lifeblood of each and every company and sales forecasting plays a vital role in conducting any business. Good forecasting helps to develop and improve business strategies by increasing the knowledge about the marketplace [1]. Everyday competitiveness between various shopping centres as and as huge marts is becoming higher intense, violent just because of the quick development of global malls also online shopping. Each market seeks to offer personalized and limited-time deals to attract 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 and logistical services [2]. Manual infestation of this task could lead to drastic errors leading

to poor management of the organization, and most importantly would be time consuming, which is something not desirable in this expedited world[3].

The current machine learning algorithm is very advanced and provides methods for predicting or forecasting sales any kind of organization, extremely beneficial to overcome low – priced used for prediction. Every business needs to be good in profit, and profit does not mean that stock sales are at maximum but also avoid the extra stock. Every retailer must maintain the stock according to the requirements and check the flaws and drawbacks that lead the sales down. Therefore, the proposed study deals with the same problem by predicting the store's sales [2][4].

A major part of the global economy relies upon the business sectors, which are literally expected to produce appropriate quantities of products to meet the overall needs [3]. Today's business handles huge repository of data. The volume of data is expected to grow further in an exponential manner. Any forecast can be termed as an indicator of what is likely to happen in a specified future time frame in a field. [6]

To address the issue of deals expectation of things dependent on client's future requests in various Big Mart across different areas diverse Machine Learning algorithms like Linear Regression, Random Forest, Decision Tree, Ridge Regression, XGBoost are utilized for gauging of deals volume. [7][8] Machine learning is the domain where the machines gain the ability to outperform humans in specific tasks.[7] Therefore, the sales forecast indicates as to how much of a product is likely to be sold in a specified future period in a specified market at specified price.[6]

1.2 Scope of the Project

The objective here is to envisage the pattern of sales and the quantities of the products to be sold based on some key features gathered from the raw data we have. [3] A stronger prediction is always helpful in developing and enhancing corporate market strategies, which also help to increase awareness of the market. [7]. Here ridge and XGBoost regression gives the better prediction with respect to Accuracy, MAE and RMSE than the Linear and polynomial regression approaches.[10]

2. Literature Review

In a study of Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression (2018) Kadam, H., Shevade, R., Ketkar, P. and Rajguru

They specify that by using less accurate prediction analysis techniques, A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regressions utilised regression and Random Forest. We can work around this by using the XG boost technique, which is more precise and efficient.[5]

Sales Prediction System Using Machine Learning' In this paper, the objective is to get proper results for predicting the future sales or demands of a firm by applying techniques like Clustering Models and measures for sales predictions[3]. Intelligent Sales Prediction Using

Machine Learning Techniques' This research presents the exploration of the decisions to be made from the experimental data and from the insights obtained from the visualization of data. It has used data mining techniques[3][6]. 'Walmart's Sales Data Analysis - A Big Data Analytics Perspective' In this study, inspection of the data collected from a retail store and prediction of the future strategies related to the store management is executed. Effect of various sequence of events such as the climatic conditions, holidays etc. can actually modify the state of different departments so it also studies this effects and examines its influence on sales [3][5].

Forecasting methods and applications (2008) Makridakis, S., Wheelwrigth.S.C., Hyndman. R.J

In this study, Short life cycles and a lack of data are present in forecasting methodologies and applications. Because consumer-oriented markets encounter variable demand, some data, such as historical data, can be used to anticipate an outcome with accuracy.[5][6]

Comparison of Different Machine Learning Algorithms for Multiple Regression on Black Friday Sales Data (2018) C. M. Wu, P. Patil and S. Gunaseelan

In this paper, Multiple Regression Comparison Using Different Machine Learning Algorithms on Black Friday Sales Data Neural networks were used to compare various algorithms. To get around this Because it is inefficient to compare different algorithms using complex models like neural networks, we can make predictions using more straightforward algorithms.[5][3]

3. Proposed System

3.1 Proposed System Overview

In this project using Python, Flask and different ML algorithms & libraries we are predicting Future sales.

The proposed system & The suggested solution makes use of machine learning models to forecast sales. Following the creation of the prediction model, the most recent data is tested using the csv file. The model trains the previous data of a mart provided using the csv file and creates a prediction model using various different Machine learning algorithms like Linear Regression, Random Forest, and Decision Tree. A predicted value is given (RMSE, Mean, Standard, Minimum, Maximum). As soon as the data source is given, the data is pre-processed, which entails filling in all zeros and empty spaces with the column's mean value and clustering. Testing and training are then conducted using a variety of prediction models, including ML

algorithms and output is provided i.e the prediction

to the user. The further implementation mainly

involves Machine Learning algorithms.

LINEAR REGRESSION: The linear regression algorithm, often known as linear regressor, demonstrates a linear relationship between a dependent (y) and one or more independent (y) variables. Given that linear regression demonstrates a linear relationship, it may be used to determine how the dependent variable's value changes as a function of the independent

variable's value

$$y = a_0 + a_1x + \varepsilon$$

RIDGE REGRESSION: Ridge Regression is a technique used when multicollinearity (highly correlated independent variables) influences results. Although the least square estimates (OLS) in multicollinearity are objective, their variances are wide and differ from the actual value. Ridge regression eliminates standard errors by adding some bias to the regression calculations. In Ridge Regression, the Linear Regression Loss function is increased to penalise the parameter estimations in addition to minimising the number of square residuals.

XGBOOST: Extreme Gradient Boosting is referred to as XGBoost. The algorithm's implementation was designed to use memory and processing power efficiently. Boosting is a sequential procedure built on the ensemble principle. This incorporates a number of slow learners and raises prediction accuracy. At all moment t , model values are weighted based on how the previous instant t_1 affected that moment. The results that were incorrectly calculated are given a higher weight than the results that were calculated correctly. The XGBoost model uses this technique to internally implement stepwise ridge regression, which automatically selects the features and eliminates multi-collinearity.

4. Implementation

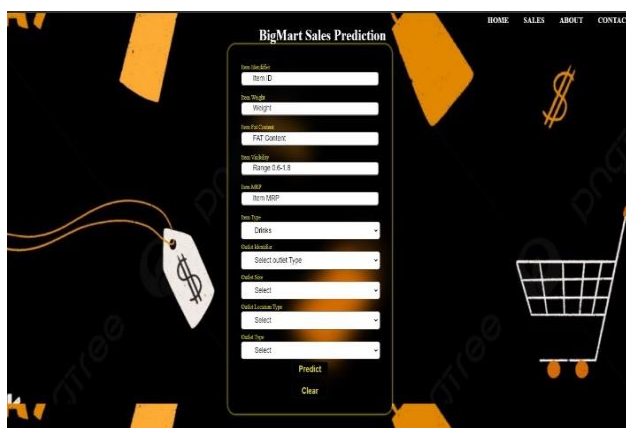


Fig 1: User Interface

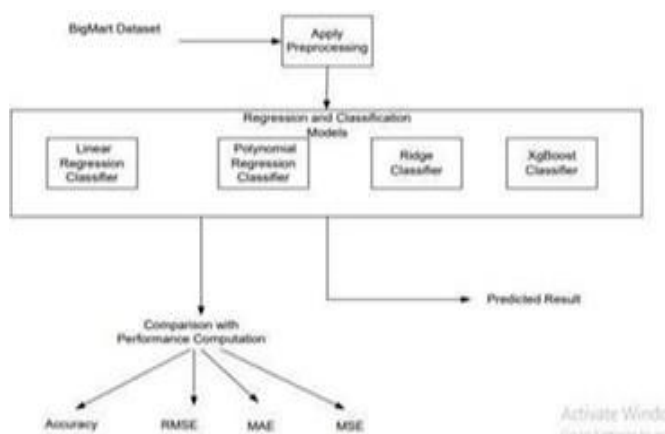


Fig 2: Working model



Attribute	Description
Item_Identifier	It is the unique product Id number.
Item_Weight	It will include the product's weight.
Item_Fat_Content	It will mean whether the item is low in fat or not.
Item_Visibility	The percentage of the overall viewing area assigned to the particular item from all items in the shop.
Item_Type	To which group does the commodity belong
Item_MRP	The product's price list

Outlet-Identifier	a distinct slot number
Outlet-Establishment Year	The year that the shop first opened its doors
Outlet-Size	The sum of total area occupied by the supermarket.
Outlet-Location	The kind of town where the store is situated
Outlet-Type	The shop is merely a supermarket or grocery store.
Item-Outlet-Sales	The item's sales in the original shop

[illegible]

Fig 4: Attributes & Sample Dataset

In conclusion, machine learning may be used to anticipate sales in a variety of industries and can give firms important information about customer behaviour, industry trends, and sales projections. Businesses may build precise models that accurately forecast future sales by using previous sales data and machine learning algorithms. This will allow them to streamline their processes, increase revenue, and gain a competitive edge. Machine learning has enormous potential for sales prediction, and as technology develops, firms may use this tool to make data-driven decisions and adjust to shifting market conditions. Overall, machine learning-based sales prediction is a potent tool that may assist companies in staying competitive and ahead of the curve in today's market.

The future work of this research could focus on expanding the current implementation to include more advanced Machine learning algorithms, Real-time analytics which further involves IoT-based devices. The sales prediction can be more personalized to predict according to their own preferences. Advanced Data Visualization could also be implemented to visualize different trends. Predictive Analytics could leverage the Machine learning algorithms to forecast future trends and patterns.

7. References

1. Punam, K., Pamula, R., & Jain, P. K. (2018, September). A two-level statistical model for big mart sales prediction. In *2018 International Conference on Computing, Power and Communication Technologies (GUCON)* (pp. 617-620). IEEE.
2. Ranjitha, P., & Spandana, M. (2021, May). Predictive analysis for big mart sales using machine learning algorithms. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1416-1421). IEEE.
3. Bajaj, P., Ray, R., Shedge, S., Vidhate, S., & Shardoor, N. (2020). Sales Prediction Using Machine Learning Algorithms. *International Research Journal of Engineering and Technology (IRJET)*, 7(06).
4. Ali, R. F., Muneer, A., Almaghthawi, A., Alghamdi, A., Fati, S. M., & Ghaleb, E. A. A. (2023). BMSP-ML: big mart sales prediction using different machine learning techniques. *Int J Artif Intell*, 12(2), 1-1x.
5. Kumar, N., & Choudhari, S. (2021). BigMart Sale Prediction using Machine Learning. *International Journal of Innovative Science and Research Technology*, 6(9).
6. Deepa, K., & Raghuram, G. (2021). Sales Forecasting Using Machine Learning Models. *Annals of the Romanian Society for Cell Biology*, 3928-3936.
7. Kumar, N., & Choudhari, S. (2021). BigMart Sale Prediction using Machine Learning. *International Journal of Innovative Science and Research Technology*, 6(9).
8. Lin, Z. C., Wu, W. J. (1999). Multiple linear regression analysis of the overlay accuracy model. *IEEE transactions on Semiconductor Manufacturing*, 12(2), 229-237.
9. <https://www.slideshare.net/LSURYAPRAKASHREDDY/big-mart-sales-pridiction-projectpptx>
10. Wang, H. (2019). Sustainable development and management in consumer electronics using soft computation. *Journal of Soft Computing Paradigm (JSCP)*, 1(01), 56.
11. Boyapati, S. N., & Mummidi, R. (2020). Predicting sales using Machine Learning Techniques.
12. Afreen Fathima, Samreen Jameel, Pathan Ahmed khan. (n.d.). *Accident-detection-and-alerting-system IEEE PROJECTS PAPERS*. Engpaper.com. Retrieved May 10, 2023, from *accident-detection-and-alerting-system IEEE PROJECTS PAPERS*. (n.d.). *Accident-detection-and-alerting-system IEEE PROJECTS PAPERS*.
13. Baig, M. S., Bari, D. R. M. A., & Khan, P. A. (n.d.). *Weapon detection using artificial intelligence and deep learning for security applications*. Ijarst.In. Retrieved May 10, 2023, from v(No date) *Weapon detection using artificial intelligence and deep learning for ...* Available at: <https://ijarst.in/public/uploads/paper/612441667180338.pdf> (Accessed: 12 May 2023).
14. Hafsa Fatima, Shayesta Nazneen, Maryam Banu, Dr. Mohammed Abdul Bar,” Tensorflow-Based Automatic Personality Recognition Used in Asynchronous Video Interviews”, *Journal of Engineering Science (JES)*, ISSN NO:0377-9254, Vol 13, Issue 05, MAY/2022
15. Fayaz, N., Fatima, H., Fatima, S., Ahmed Khan, P., & Student, B. E. (n.d.). *Diagnosis of diabetic patient and analyzing insulin dosage by using gradient boosting and logistic regression algorithm*. Jetir.org. Retrieved May 10, 2023, from (No date a) *Design and*

development of IOT based Virtual Doctor Robot - JETIR. Available at: <https://www.jetir.org/papers/JETIR2205241.pdf> (Accessed: 12 May 2023).

16. Mr. Pathan Ahmed Khan, Dr. M.A Bari,: Impact Of Emergence With Robotics At Educational Institution And Emerging Challenges”, International Journal of Multidisciplinary Engineering in Current Research(IJMEC), ISSN: 2456-4265, Volume 6, Issue 12, December 2021,Page 43-46
17. Hafsa Fatima, Shayesta Nazneen, Maryam Banu, Dr. Mohammed Abdul Bar,” Tensorflow-Based Automatic Personality Recognition Used in Asynchronous Video Interviews”, Journal of Engineering Science (JES), ISSN NO:0377-9254, Vol 13, Issue 05, MAY/2022
18. Mohammed Shoeb, Mohammed Akram Ali, Mohammed Shadeel, Dr. Mohammed Abdul Bari, “Self-Driving Car: Using Opencv2 and Machine Learning”, The International journal of analytical and experimental modal analysis (IJAEMA), ISSN NO: 0886-9367, Volume XIV, Issue V, May/2022
19. Mr. Pathan Ahmed Khan, Dr. M.A Bari,: Impact Of Emergence With Robotics At Educational Institution And Emerging Challenges”, International Journal of Multidisciplinary Engineering in Current Research(IJMEC), ISSN: 2456-4265, Volume 6, Issue 12, December 2021,Page 43-46
20. Kumar, A., Dhabliya, D., Agarwal, P., Aneja, N., Dadheech, P., Jamal, S. S., & Antwi, O. A. (2022). Cyber-internet security framework to conquer energy-related attacks on the internet of things with machine learning techniques. Computational Intelligence and Neuroscience, 2022 doi:10.1155/2022/8803586
21. Kumar, S. A. S., Naveen, R., Dhabliya, D., Shankar, B. M., & Rajesh, B. N. (2020). Electronic currency note sterilizer machine. Materials Today: Proceedings, 37(Part 2), 1442-1444. doi:10.1016/j.matpr.2020.07.064
22. Mandal, D., Shukla, A., Ghosh, A., Gupta, A., & Dhabliya, D. (2022). Molecular dynamics simulation for serial and parallel computation using leaf frog algorithm. Paper presented at the PDGC 2022 - 2022 7th International Conference on Parallel, Distributed and Grid Computing, 552-557. doi:10.1109/PDGC56933.2022.10053161 Retrieved from www.scopus.com
23. Mehraj, H., Jayadevappa, D., Haleem, S. L. A., Parveen, R., Madduri, A., Ayyagari, M. R., & Dhabliya, D. (2021). Protection motivation theory using multi-factor authentication for providing security over social networking sites. Pattern Recognition Letters, 152, 218-224. doi:10.1016/j.patrec.2021.10.002
24. Raghavendra, S., Dhabliya, D., Mondal, D., Omarov, B., Sankaran, K. S., Dhablia, A., . . . Shabaz, M. (2022). Development of intrusion detection system using machine learning for the analytics of internet of things enabled enterprises. IET Communications, doi:10.1049/cmu2.12530
25. Shahanawaj Ahamad, Mohammed Abdul Bari, Big Data Processing Model for Smart City Design: A Systematic Review “, VOL 2021: ISSUE 08 IS SN : 0011-9342 ;Design Engineering (Toronto) Elsevier SCI Oct