IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657
Volume 12, Issue 1, 2024

BIG MART SALES PREDICTION USING MACHINE LEARNING

Mr. P. Swetha, Professor, Department Of ECE SICET, Hyderabad
Jangili Pravalika, Kanduri Sreeja, Hanumanthu Uday Kiran, Kasula Sai Laxmi
UG Student, Department Of ECE, SICET, Hyderabad

Abstract:

Today, supermarkets and supermarkets track sales data for each product to predict future customer demand and impr ove inventory management. The data warehouse contains a lot of customer data and a product specification in the dat a warehouse. Also extract archived data from archives to check for anomalies and regular patterns. Data obtained thr ough machine learning from different stores like Big Mart can be used to predict future sales. In this paper, we prese nt a forecasting model that uses XG boost Regressor technology to predict the sales of a company like Big Mart and f ind that the model has better performance than the existing model.

I. Introduction

Now, executives at Big Mart, a major supermarket chain with stores across the country, have turned to all their data s cientists to help them create the model that can predict sales of any product for any store. . actually. Big Mart has coll ected sales data of 1,559 products from 10 stores in various cities since 2013.Based on this data, the company hopes t hat we can identify check products and stores that play an important role in sales. When used, this information will b e accurate to ensure the success of the business.

2. Research articles

1. Big box stores sell predictions based on random forests and multiple regression lines. Random forests and regressi on lines are used for predictions, but the accuracy is low. To overcome this problem, we can use the XG boost algorit hm, which will provide greater accuracy and efficiency.

2. Forecasting methods and applications suffer from lack of information and short life cycle. Therefore, some data, s uch as historical data, can be used to make predictions and obtain accurate results in situations where the customer's business focus is not expected.

3. Comparison of various machine learning algorithms for multiple regression on Black Friday sales data. Neural net works are used to compare different algorithms. To overcome this problem, it is not good to use complex models like neural networks to compare different algorithms, so we can use simple algorithms for prediction

4. Predicting Apparel Sales Using Feed Forward and Recurrent Neural Networks Use neural networks to predict sale s. Using neural network to predict weekly sales is not very efficient, so XG supports efficient work.

objective

– The purpose of this framework is to predict future sales based on data provided from previous years using machine learning techniques.

- Another aim is to conclude that the use of XG Boost Regressor is the best model to achieve better results and provi de fast and accurate results.âœ Learn key points that can increase sales and what changes can be made to the features of your product or store

ISSN 2347–3657
Volume 12, Issue 1, 2024

IV. SYSTEM ARCHITECTURE

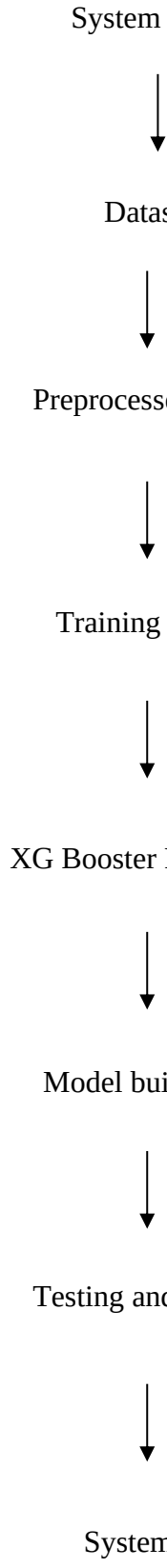

Fig:-Flow of XG Boost Regressor

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 12, Issue 1, 2024

V. PROPOSED MODEL

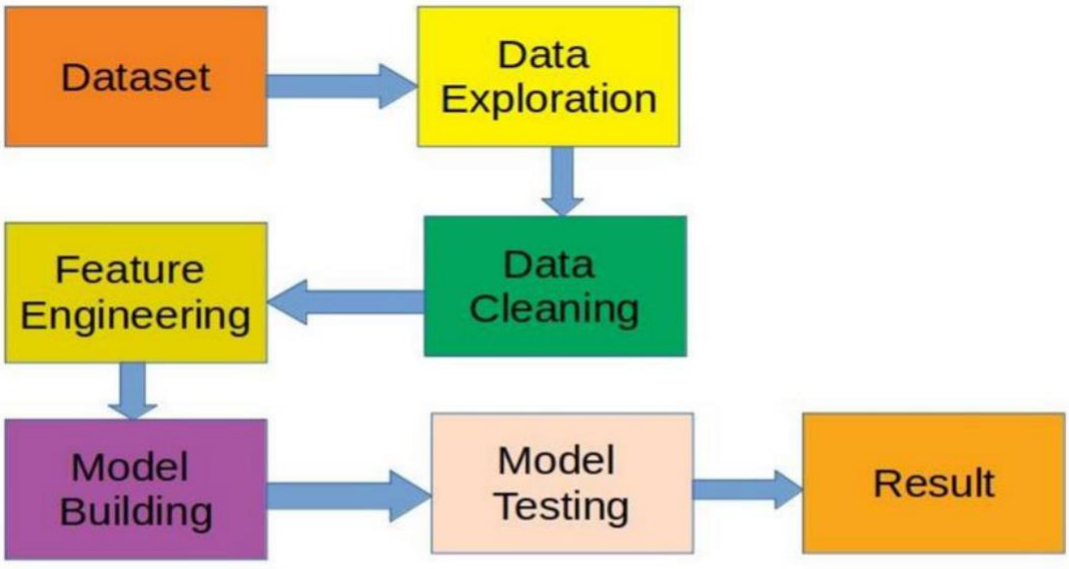

Fig:-Working procedure of proposed model

VI. PHASES IN MODEL

1. DATASET

Name	Type	Subtype	Description	Segment	Expectation
Item_Identifier	Numeric	Discrete	Unique Product ID	Product	Low Impact
Item_weight	Numeric	Continuous	Weight of product	Product	Medium Impact
Item_Fat_Content	Categorical	Ordinal	Wether the product is low fat or not	Product	Medium Impact
Item_Visibility	Numeric	Continuous	% of total display area in store allocated to this product	Product	High Impact
Item_Type	Categorical	Nominal	Category to which product belongs	Product	High Impact
Item_MRP	Numeric	Discrete	Maximum Retail Price (list price) of product	Product	Medium Impact
Outlet_Identifier	Numeric	Discrete	Unique Store ID	Store	Low Impact
Outlet_Establishment_Year	Numeric	Discrete	Year in which store was established	Store	Low Impact
Outlet_Size	Categorical	Ordinal	Size of the store	Store	High Impact
Outlet_Location_Type	Categorical	Ordinal	Type of city in which the store is located	Store	High Impact
Outlet_Type	Categorical	Ordinal	Grocery store or some sort of supermarket	Store	High Impact
Item_Outlet_Sales	Numeric	Discrete	Sales of product in particular store. This is the outcome variable to be predicted	Product	Target

ISSN 2347–3657
Volume 12, Issue 1, 2024

2. Data exploration

At this stage, the main information about the data was extracted from the data set. This is an a ttempt to analyze data from theory and existing data. This means that the voltage size and wei ght of the product meet the missing values
and the Recommended Product has a minimum value of zero, which is impossible. The sales center was established between 1985 and 2009. The results won't fit into this list. Therefore w e must change them according to the age of a particular socket. The database contains 1559 u nique items with 10 unique items for sale. The attribute object has 16 unique values. Althoug h there are two types of fat content, some are mistakenly labeled as regular instead of "regular " and lowfat instead of lowfat, LF. The response variable (e.g., inventory in the store) is positi vely skewed. Therefore, engine work was carried out on the Product Inventory in order to eli minate the ambiguity in the different answers.

3. Data cleaning

As can be seen in the previous section, the "Export Size" and "Product Weight" properties are not important. In our study, if the dimension does not have a value, we replace it with the mo del of that attribute, and if the value is missing, we replace it with the special attribute design ed for the weight object. The missing feature is the arithmetic mean, where changing the mea n and mode reduces the correlation between the predicted features. We assume that there is n o relationship between measured and predicted features for our model.

4. Feature Engineering

During data analysis some minor differences were found in the dataset. Therefore, this stage i s used to resolve all the nuances in the data and make it ready for the appropriate design. At t his stage we see that the visibility property has a value of zero, which does not make any sens e. Therefore, the average visibility of the object will be used for the zerovalued attribute. This ensures that all products sell well. Any category attribute differences are resolved by swappi ng all attribute categories for the appropriate attributes. Finally, we add a new "year" attribute to the database to determine the age of a particular store.

5. Model

After completing the previous stages, the information is now ready to create the design. Once building created, the model will be used as a predictive model to predict Big Mart's sales.
In our study, we propose a model that uses the XG boost regression algorithm

VI. Conclusion

We estimate the accuracy of the XG Boost Regressor. Our forecasts help large businesses adj ust their methods and strategies and thus increase their profits. Profit forecasts are useful for c ompany managers to understand their sales and profits. This will also give them ideas for new locations or Big-mart centres.

ISSN 2347–3657
Volume 12, Issue 1, 2024

VII. references

1. Makridakis, S., Wheelwright, S.C., Hyndman, R.J.: Forecasting methods and applications. John Wiley and Sons (2008).

2. Kadam, H., Shevade, R., Ketkar, P. and Rajguru. : “Sales forecasting based on random fore sts and multiple linear regression for large supermarkets.” (2018).

3. C. M. Wu, P. Patil and S. Gunaseelan: Comparison of different machine learning methods for multivariate Black Friday sales data (2018).

4. Das, P., Chaudhury: Predicting Footwear Retail Sales Using Feed Forward and Recurrent Neural Networks (2018)

5. Das, P., Chaudhury, S.: Comparison of various machine learning algorithms for multiple re gression on Black Friday sales data (2007)