SHOPPING MALLS SALES WITH DATA ANALYSIS USING ML

JAMES STEPHEN MEKA Principal, WISTM, Andhra University, (jamesstephenm@gmail.com)
K. HIMANTHA RAO Assistant Professor, WISTM (himantharaokolli@gmail.com)

Abstract-Shopping malls and Big Marts nowadays keep track of individual item sales data in order to estimate future customer demand and modify inventory management. In a data warehouse, the data stored fundamentally include a vast quantity of consumer data and individual item information. For businesses like Big Mart, the resulting data can be utilized to anticipate future sales volume using various machine learning approaches. In this paper, we use the XG boost technique to develop a predictive model for predicting the sales of a shopping malls, and discover that the model surpasses the competition.

Keywords— XG BOOST technique, machine learning, Sales Forecast, Linear regression, random forest.

1. INTRODUCTION

Only because of the rapid growth of the global economy is the competition between different shopping malls and huge stores becoming more serious and hostile day by day. On-line shopping and shopping malls are two of the most popular ways to shop. Each mall or supermarket tries to attract more customers by offering unique and limited-time discounts, so that the number of sales for each item may be forecasted for inventory management, logistics, and transportation, among other things. Present Machine learning algorithms are quite complex, and they provide techniques for foreseeing or forecast a company's future sales demand aids in the overcoming of the low cost of computing and storage systems. The topic of big-box store sales forecasting is addressed in this research.

A typical sales forecasting research can aid in the in-depth analysis of prior scenarios or conditions, and then the inferences regarding client acquisition, money inadequacy, and strengths may be implemented before making a budget and marketing plans for the future year. To put it another way, sales forecasting is based on historical resources. In-depth knowledge of the past is necessary for boosting and improving the likelihood of the marketplace regardless of any circumstances, particularly external events, which helps the firm to plan for future needs. In the retail industry, extensive research is being conducted in order to estimate future sales demand. The statistical approaches, also known as the classic

method, are the most basic and widely used methodology for predicting sales. However, these methods take substantially longer to predict sales. These approaches also couldn't handle non-linear data, thus machine learning techniques were used to overcome these limitations in traditional methods. Big Mart is a large grocery chain with locations around the country, and its current board has issued a challenge to all Data Scientists to assist them in developing a model that can accurately estimate sales per product for each store. Big Mart gathered sales data for a variety of products from ten sites across the country. The firm anticipates that by analyzing this data, we will be able to identify the products and stores that are most significant to their sales, and that we will be able to use that information to take the necessary actions to ensure their business future.

Many businesses rely substantially on their information and require market forecasting. Every shopping mall or the store strives to provide the individual with the present moment. Depending on the day, the proprietor can attract more customers. The goal of increasing business volume in all is as rated for stock management, logistics, and organization administration of transportation, and so on. To deal with the issue Deals issue Expectations of things based on client's upcoming requests from numerous Linear Regression is one of many Machine Learning methods. XG Boost, Random Forest, Decision Tree, Ridge Regression are used to determine the volume of sales.

The different applications of machine learning and the sorts of data they deal with are discussed. After that, the issue statement addressed in this study is defined. The approach used is then described, as well as the forecast outcomes obtained during implementation. Linear regression, K-Nearest Neighbors, Decision tree, and Random tree are examples of machine learning methods. Data Pre-processing, Dataset Training, and Machine Learning Model Implementation are the steps taken in this work, from dataset preparation to generating results.

2. LITERATURE REVIEW:

Big Mart data description:

Description: As the dataset, we have collected is from various sales. This dataset has been used to extract relevant information from the data. Item Fat, Item type, Item MRP, item weight, and other properties are included in these databases.

A Random Forest and multiple linear regression forecast for Big Mart sales

Description: A forest for Big Mart sales based on Random forests and multiple Linear

regression, with lower accuracy. To get around this, we can employ the XG boost algorithm, which is more accurate and efficient.

Existing system: With the rapid growth of worldwide malls and store chains, as well as an increase in the number of electronic payment clients, competitor organizations are becoming increasingly competitive. Every organization is attempting to attract more customers through personalized and short-term offers, making the prediction of future volume of sales of each item an important asset in the planning and inventory management of every organization, transportation service, and so on. Due to the low cost of computing and storage, sophisticated machine learning algorithms can now be used for this purpose.

Proposed system: We gathered sales data for different products from ten stores in various cities. The goal is to create a prediction model and determine the sales of each product in each store. Big Mart will use this model to try to identify the qualities of products and stores that are essential in growing sales.

3.METHODOLOGY

Artificial Intelligence includes **Machine Learning** as a subset. ML is the study of computer algorithms that improve themselves over time. ML is concerned with the research and development of algorithms that can learn from data and generate predictions based on that data. Machine learning can adjust actions and responses based on more data, making it more efficient, adaptive, and scalable. Machine learning
techniques are employed in the Big Mart sales prediction algorithms.

Machine learning algorithm:

Machine Learning aids task performance and productivity. When fresh data is introduced, it involves learning and self-correction.

Machine learning algorithms used include:

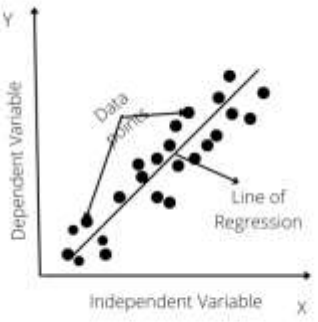
1. Regression linear.
2. Regression with random forests
3. Regression using Decision Trees

1. Linear regression:

Linear regression is one of the most important and often used regression techniques.It's one of the most basic regression techniques. Linear regression is used to predict sales, salary, age,

product price, and other continuous/real or numeric factors.

When applying the linear regression algorithm, a dependent (y) and one or more independent (y) variables have a linear relationship. In the linear regression model, the link between the variables is represented by a slanted straight line. Consider the following example:



Y stands for Dependent Variable in this case (Target Variable)

"Independent variable" is represented by X. (predictor Variable)

2.Random forest regression:

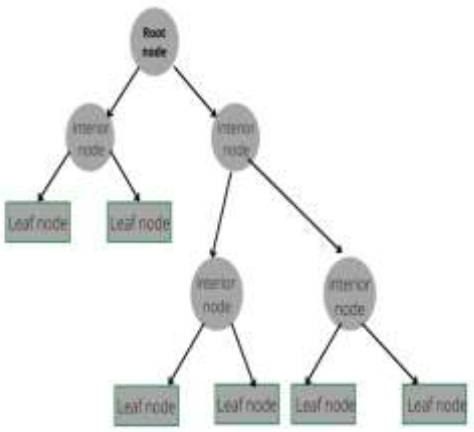
The random forest algorithm is a powerful sales forecasting tool. For anticipating the consequences of machine learning projects, it's straightforward to use and understand. Because they have decision tree-like hyperparameters, random forest classifiers are used in sales prediction. The tree model looks like a tool for making decisions. A random forest model is produced for each individual learner using a random collection of rows and a few randomly determined parameters. The final prognosis could be based on all of the predictions made by the pupils. In a regression problem, the final estimate might be the average of all previous projections.

Random forest selects data at random, forms a decision tree, and averages the results. It does not rely on any formulas.

3.Decision tree – regression:

A simple model with little bias can be used to construct a classifier model in a top-down fashion, with the root node being the first to be analyzed. It's a well-known model for machine learning. A tuple recursive classifier is another name for a decision tree. It's a powerful multivariable analysis method and a powerful data mining tool. In a variety of

fields, this approach displays the factors involved in achieving a specific objective, as well as the motives for achieving the goal and the means of execution. When a decision tree receives a data set with features as input, it will create a set of rules for prediction.There are three sorts of nodes in this tree-structured classifier. This tree-structured classifier has three types of nodes. a well-known model for machine learning. A tuple recursive classifier is another name for a decision tree. It's a powerful multivariable analysis method and a powerful data mining tool. for achieving the goal and the means of execution. When a decision tree receives a data set with features as input, it will create a set of rules for prediction. A tuple recursive classifier is another name for a decision tree. It's a powerful multivariable analysis method and a powerful data mining tool. In a variety of fields, this approach displays the factors involved in achieving a specific objective, as well as the motives for achieving the goal and the means of execution. When a decision tree receives a data set with features as input, it will create a set of rules for prediction. There are three sorts of nodes in this tree-structured classifier. This tree-structured classifier has three types of nodes. The Root Node is the graph's first node, and it represents the entire sample. It can be subdivided further into nodes. The internal nodes reflect the features of a data collection, while the branches represent the decision rules. Finally, the Leaf Nodes represent the outcome.



4.RESULTS AND DISCUSSION

All of these machine learning techniques will be used to create models and algorithms that can take in data and utilize statistical analysis to predict an output while also updating outputs as new data becomes available

5.REFERENCES:

[1].Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Kera's, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media

[2].Kadam, H., Shevade, R., Ketkar, P. and Raj guru.: "A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression." (2018).

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

[4].Das, P., Chaudhury: Prediction of retail sales of footwear using feed forward and recurrent neural networks (2018)

[5]. Cheriyan, Sunitha, Shaniba Ibrahim, Sanju Mohanan and Susan Treesa. " Intelligent Sales Prediction Using Machine Learning Techniques." In 2018 International Conference on Computing, Electronics and Communication Engineering (iCCECE), pp. 53 – 58, IEEE, 2018.

[6].Panjwani Mansi, Rahul Ramrakhiani, Hitesh Jumnani, Krishna Zanwar and Rupali Hande. "Sales Prediction System Using Machine Learning." No. 3243. Easy Chair,