Model for Predicting Prospective Big-Mart Sales Based on Grid Search Optimization (GSO)

Anantha Murthy, Puneeth B R, Harshitha G M, Neha Parveen, Balaji N and Keerthi Shetty
NMAM Institute of Technology, Nitte (Deemed to be University), Karnataka, India
Email: anantham2004@gmail.com, puneethbr9@gmail.com, harshithagm121@gmail.com, shkneha1999@gmail.com, balaji.hiriyur@gmail.com and keerthi.shetty1996@gmail.com

2022 International Conference on Artificial Intelligence and Data Engineering (AIDE) | 978-1-6654-9304-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/AIDE57180.2022.10059929

***Abstract*—Predicting sales before actual sales are critical for any retailing company, to sustain a thriving business, such as Big Mart or Mall Statistical models, for example, are examples of traditional forecasting models that are frequently employed as a strategy for forecasting future sales; however, these techniques take significantly longer to estimate sales and are incapable of dealing with non-linear data. As a result, both nonlinear and linear data are dealt with using Machine Learning (ML) methodologies. ML techniques can also be used to process vast amounts of data efficiently, such as the Big Mart dataset, which has a vast amount of client data and data item attributes. A store is looking for a model that can forecast precise sales so that may anticipate consumer demand and update sale stocks ahead of time. In this study, we offer a prediction model for estimating a company's sales, such as Big. We provide a Grid Search Optimization (GSO) approach in this research for optimizing parameters and selecting the optimal tuning hyperparameters for projecting future sales of a retailing firm like Big Mart, and we discovered that our model beats others.**

***Index Terms*—Xgboost, Grid Search, Sales Forecasting, Regression,Sales Prediction.**

I. Introduction

Forecasting sales has always been an important topic to concentrate on. All suppliers must use an effective and optimal forecasting strategy to keep marketing organizations effective. This activity's Manual material handling can lead to substantial errors resulting in bad organizational structure, and, more significantly, It would take too much time, which is unacceptable in today's fast-paced world. Business sectors' major purpose is to entice the intended audience. As a result, it is essential that the organization has previously proved its capacity to accomplish this goal using a prediction model. Big Mart is a worldwide retailer with stores all over the world. Big Mart trends are significant because data scientists exploit them and rely on them to discover potential hubs by product and area. Data scientists may use a computer to anticipate sales at Big Mart and look into different trends to achieve the greatest results, shop, and product. Many companies rely on data and require market estimates. Forecasting necessitates the analysis of data from several sources, such as consumer trends, purchasing behavior, and other characteristics. This research would also help businesses manage their finances more effectively, which is where machine learning can really light up. [1] Always more accurate forecasting is useful for creating and enhancing business plan for the industry, which is also very beneficial. [2] In this work, we anticipate sales utilizing several machine learning algorithms employing data mining methodologies such as discovery, data transformation, feature creation, model design, and testing. This approach comprises searching for missing data, abnormalities, and outliers in raw data received from a huge mart. Following that, an algorithm will be trained on the data to build a model. Data preparation is typically a required step. It changes old, irrelevant data into relevant content that works with a Data Mining method. [3] The data is then improved in order to obtain precise forecasts and acquire innovative and intriguing findings that add to our understanding of the task's data. Then, by using machine learning methods like the random forest and simple or multiple linear regression model, this may be utilised to predict potential purchases. [4] The following steps are involved in data pre-processing [5].

- Data Cleansing.
- Data Transformation.
- Data Reduction.

In this proposed system, we describe a model that employs the Grid Search Optimization (GSO) approach and an ensemble with the Extended Gradient Boosting algorithm. After improving the different parameters of the XGBoost Technique, GSO is used in this study to select the ideal parameters for the predictive model.

A. Problem Statement

"To understand Big Mart sales in order to determine what impact specific qualities of an item play and how they affect their sales". A predictive model might be built to assist Big-Mart to achieve this goal to evaluate the crucial components

that might boost sales in each shop and what changes to the product or location's features could be made.

II. Related Work

Many scholars have undertaken sales forecasting and analysis of sales forecasting, as detailed below:

- As demonstrated in this study, many vendors would benefit from forecasting a single transaction rate, indicating that the development of a setup that predicts a large number of events may benefit from the knowledge acquired. To make the prediction, the neural network technique is used. In this situation, they used Bayesian learning to gain insights. When making predictions, a Bayesian first constructs a model, chooses a prior, gathers data, evaluates the posterior, and finally chooses a model. [6], [7].
- This study explores the choices that should have been taken in light of research observations as well as information gleaned from visual analytics. It employed data mining techniques. In terms of predicting future transactions, The Gradient Boost approach was discovered to be the most precise [8].
- To anticipate sales, three modules were combined: tableau, hive, and R programming. You may have a better grasp of the income and make adjustments to the goalto increase the store's success by looking back at its historical data. By lowering the intermediate key feature, the diagram's key values are retrieved to bring all interim rates down. [9]
- The study's purpose is to provide relevant insights for anticipating a company's future sales or demands utilizing methods such as Clustering Models and sales forecasting metrics. As a result, the potential of the algorithmic methodologies is assessed and utilized in future studies. [10]
- Ait-alla et al., [11] developed a numerical strategy for reliable production scheduling for garment manufacturers. The authors concentrated on supporting decision-making on article distribution at various manufacturing companies, and they claim their model works substantially and can successfully deal with restrictions of unknown client requests.
- Kumari Punam, Rajendra Pamula and Praphula Kumar Jain in [12] A Two-Level Statistical Model for Big Mart Sales Prediction has developed a two-level technique to forecasting sales of goods that promise to increase effectiveness. It involves stacking algorithms, with one learning algorithm present in the top layer and one or more in the lower layer. This two-level modelling methodology surpasses the solitary model predictive technique, resulting in more accurate sales predictions.

III. Proposed System

The dataset of Big Mart sales goes through numerous Data Science processes in order to construct a model that predicts reliable results. In our project, we may train the model using any recent data set that is accessible, and in this case, we chose the Big Mart dataset from the year 2020. We present a GridSearch Optimization (GSO) approach for parameter optimization and selecting the best tuning hyperparameters, as well as an ensemble with Extended Gradient Boosting (XGBoost) techniques for estimating a retailer's future sales like Big Mart, and we discovered that our model outperforms the others. Figure 1 shows the structure diagram of the suggested model focusing on the various algorithms used in the dataset, where we calculate the R-square before deciding on the optimum yield algorithm. The following algorithms are utilized.

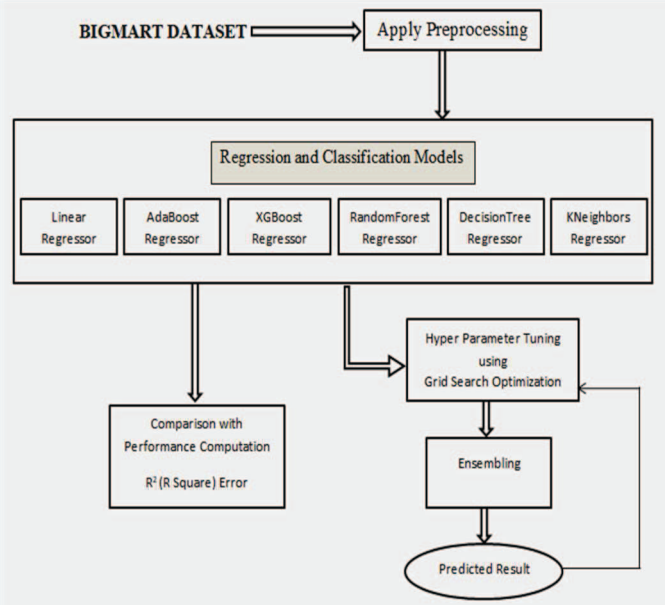

Figure 1. Shows the proposed Architecture Diagram

A. Implementation Platform and Language

Today, many programmers use Python, an interpreted, general-purpose language, to solve problems with domain-specific applications issue solving rather than dealing with system difficulties. It is sometimes referred to as the battery's internal programming language. It has several libraries for scientific purposes and queries, as well as a variety of third-party libraries to help with issue solving. The Python packages Numpy is used for scientific calculation, and Matplotlib is used for 2D charting were used in this work. In addition, the Python Pandas data analysis tool was employed, as was the Grid Search Optimization (GSO) approach for optimizing parameters and identifying the optimal tuning hyperparameters, and the ensemble employed Gradient Boosting techniques for sales forecasting. PyCharm IDE, which has tremendous community support, is adopted as a development platform, enabling for faster code generation.

IV. Exploratory Data Analysis

Data can be trained via exploratory data analysis so that it can be incorporated into testing and training data for feature engineering, data visualization, and other purposes. The exploratory analysis consists of two forms of analysis, univariate

(just one characteristic) and bivariate (two characteristics) analyses are performed on data to summarize and find trends. During the investigation, we discovered a few observations: the 'low fat' category is also referred to as 'LF' and 'Low Fat,' and 'reg' and 'Regular' correspond to the same group and maybe united as one. It is also found that the number of low-fat attributes is double that of other item kinds. It has also been discovered that the two most popular item types are fruits and snacks. We can see from the statistics that some of the goods, despite being labeled as low-fat and/or regular, are not edible. As a result of the investigation, it is possible to identify a link between product weight and sales, as well as item fat content and sales. A substantial proportion of sales are of goods with visibility less than 0.2. [13]

V. Methodology

A. Data Pre-Processing.

Data pre-processing is the preparation and adaption of raw data to a model for learning This is the first of many and most crucial stages in the creation of the output of a machine learning model. Real-world data contains noise and missing values in general, but it should not be used inefficiently, especially for machine learning models. Table 1 shows the Big Mart sales dataset that we got as part of our work. The dataset has 12 features with 8523 tuples of each feature, for a total of 102276 occurrences. Table 1 also includes the expectations or variables influencing a store's sales.

TABLE 1.
FEATURES OF DATASET AND EXPECTATIONS

Name	Type	Description	Segment	Expectation
Item_Identifier	Numeric	Unique product ID	Product	Low Impact
Item_Weight	Numeric	Weight of product	Product	Medium Impact
Item_Fat_Content	Categorical	Whether the product is low fat or not	Product	Medium Impact
Item_Visibility	Numeric	The % of total display area of all products in a store allocated to the particular product	Product	High Impact
Item_Type	Categorical	The category to which the product belongs	Product	High Impact
Item_MRP	Numeric	Maximum Retail Price (list price) of the product	Product	Medium Impact
Outlet_Identifier	Numeric	Unique store ID	Store	Low Impact
Outlet_Establishment_Year	Numeric	The year in which store was established	Store	Low Impact
Outlet_Size	Categorical	The size of the store in terms of ground area covered	Store	High Impact
Outlet_Location_Type	Categorical	The type of city in which the store is located	Store	High Impact
Outlet_Type	Categorical	Whether the outlet is just a grocery store or some sort of supermarket	Store	High Impact
Item_Outlet_Sales	Numeric	Sales of the product in the particular store. This is the outcome variable to be predicted.	Product	Target

B. Data Cleaning

In the prior section Element weight and outlet, size was discovered to be empty. In this instance, we use the mean value to fill in the blanks in the property for object weight and the mode value for the property's missing values for outlet size. Correlation between imputed attributes decreases when missing attributes are numerical, as does mean and mode replacement. In our model, we believe there is no link between the calculated and imputed properties.

C. Feature Engineering

Some anomalies in the dataset were detected during the data exploration phase. Before constructing the prediction model, this phase is utilized to correct any discrepancies found in the dataset. The Item visibility property was set to 0, which was nonsensical. As a consequence, the mean value item visibility of the product is utilized to replace zero value characteristics increasing the likelihood that all products will sell. All differences in categorical attributes are addressed by changing them to relevant ones. In some cases, non-consumables and fat content are not included. We provided a third fat content category: none to avoid such scenarios. The Item Identifier property revealed that the unique ID begins with either DR, FD, or NC. As a result, we create Item Type New, which is separated into foods, drinks, and non-consumables divided into three groups. Finally, we add an extra property to identify the age of a certain outlet and add the year to the dataset.

D. Feature Transformation

Our premise states that the more visible the product, the more likely it is to sell." Items that are less apparent are less likely to sell. We use feature transformation to replace a zero visibility item with Item Mean Visibility to get around this constraint. At this point, other theories can be used. The data is ready for model generation after all of the background work is performed.

E. Model Building

1) Linear Regression Algorithm: Regression is a parametric approach for forecasting a continuous or dependent variable. that is determined by a series of independent factors Due to the fact that various assumptions are made dependent on the data set, this strategy is considered to be parametric.

$$Y = o + 1X + \qquad (1)$$

For Simple Linear Regression, the equation indicated in Eq2 is employed. These parameters are known as
Y - Variables to be anticipated.
X - The variable(s) utilized to make a prediction.
0 -When X-0 is zero, it is also known as the forecast value or the intercept term.
1-Any change in X by I unit indicates a change in Y. It is also known as the slope word.
- This parameter indicates the difference between the expected and actual values, as well as the residual value. Regardless of how effectively the model has been trained, validated, and tested is always a difference between real and anticipated values, which is an irreducible error; hence, we cannot rely only on the learning algorithm's performance projected outcomes. Dietterich's alternative methods can be used to compare learning algorithms. [14]

2) Random Forest Algorithm: The random forest approach is a dependable machine learning strategy for forecasting sales. For anticipating the consequences of machine learning projects, it is straightforward to use and comprehend. Random forest classifier is utilized in sales prediction since it has

similar hyperparameters to the decision tree. The tree model and decision tool are the same things. Figure 2 depicts the relationship between the random forest and the decision tree. The sklearn. ensemble package's random forest regressor class is used to address the regression issues with random forest forecasts. The random forest regressor, also referred to as the parameter n estimators, is essential. Random forest is a meta-estimator that uses varied sub-samples of the dataset to fit on multiple decision trees.

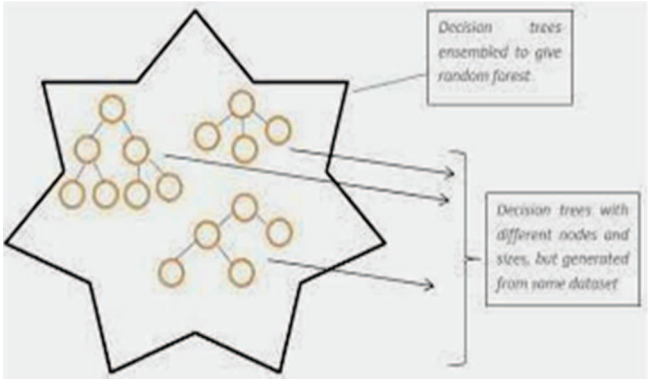

Figure 2. Relation between Decision Trees and Random Forest.

3) Adaptive Boosting Regressor: Adaptive Boosting is a collection of numerous decision trees, each of which is a poor learner and performs just marginally better than random guessing. However, the adaptive AdaBoost method conveys the gradient of previous pees to the next trees in order to improve the error of the prior tree. As a result, the future learning of trees at each phase develops a powerful learner. The final forecast is the weighted average of each tree's projections. Because of its high flexibility, AdaBoost is more resistant to outliers and noisy data, which is critical in our circumstances. Furthermore, the algorithm is designed so that prior tree information is transmitted to subsequent trees, allowing them to focus exclusively on difficult-to-predict training samples.

4) Decision Tree: A decision tree is a classifier using a tree structure the core nodes contain dataset attributes, each branch is a decision rule, and each leaf node is the result. It is advantageous for classification issues, however, it is also employed for regression difficulties.

5) K-Nearest Neighbours: KNN, which stands for K-Nearest Neighbors, is a fundamental machine learning technique for classification and regression issues. KNN also aims to acquire any equation from training data rather than depending on assumed conditions because it is non-parametric. Unlike most algorithms with complicated names, which might be perplexing as to what they actually represent, KNN is ratherobvious.

6) XG Booster: eXtreme Gradient Boosting is abbreviated as XG Boost. This is a powerful ensemble learning approach that makes use of a gradient boosting architecture. Decision trees are constructed in the form of a series in this concept. It combines weak learners in order to increase prediction accuracy. Assume that the outcomes in case X are weighted based on the output of the prior instance X-1. The correctly predicted results are given a lower weight, whereas the incorrectly predicted results are given a larger weight.

F. Hyperparameter Tuning

When a machine learning algorithm is applied to a given dataset, its performance is assessed depending on how well it generalises or how it responds to fresh, previously unexplored data. If the algorithm's performance is insufficient or could be improved, specific parameters must be adjusted. These variables are referred as "hyperparameters," and "hyperparameter tuning" is the act of modifying these hyperparameters to improve the effectiveness of the learning algorithm.' Algorithm training does not directly learn these hyperparameters. These variables are fixed before data training begins. They consider aspects such as learning rate, which indicates how rapidly the model is expected to learn and how complex the prototype looks to be.

Strategies of Hyperparameter

- GridSearchCV
- RandomizedSearchCV

In our proposed system we have used GridSearchCV Hyperparameter Tuning.

1) GridSearchCV: This is the traditional approach of doing hyperparameter optimization.Until a condition is discovered to be satisfied or the end is reached, it explores a certain subset of hyperparameters indefinitely.The GridSearchCV module in Scikit-learn can be used to accomplish this strategy.

2) RandomizedSearchCV: Instead of searching continually, RandomizedSearch searches for a specified sample of data at random (much like GridSearch). This decreases the hyperparameters' processing time. Random search may be implemented using the scikit-learn package's RandomizedSearchCV function. Before exploring for hyperparameters, they are specified.

G. Evaluation Metrics

Model evaluation is an important stage in creating an effective machine learning model As a result of this, creating a model and using it to provide metrics suggestions is crucial. It will take some time and effort to get excellent accuracy depending on the value acquired via metric improvements.

Evaluation metrics summarize the conclusions of one model. [15] The ability to differentiate between model outputs is a key element of the evaluation metrics. For the evaluation process, we employed the R (R-squared) statistic.

A metric that assesses how well a model correlates with the data is the coefficient of determination, or R2. It is a measurable statistic of how closely the regression line in the framework of regression matches the real data. It is crucial to employ statistical models whenever future forecasts or assumptions are being tested.

Although there are severalvarieties, this one is the most typical.

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (2)$$

SSREG-Sum Squared Regression
SSTOT Total Sum of Squares
The total sum of squares is the sum of all squared distances from the mean, and the sum squared regression is the sum of the residuals squared. It will only take numbers between 0 and 1 because it is a percentage.
According to the experimental results, our approach produces more accurate predictions with the highest R-Square for both the training and testing sets, as shown in Table 2.

TABLE 2.
CORRELATION OF R-SQUARE OF THE PROPOSED MODEL WITH OTHER MODELS.

MODEL R-SQUARE		
1	Linear Regression	0.510595
2	KNeighbors Regression	0.561299
3	Decision Tree	0.537335
4	Random Forest Regression	0.550172
5	AdaBoost Regression	0.529487
6	**XGBoost Regression**	**0.591471**

VI. Results And Discussion

The attribute having the lowest association with our goal variable is Item Visibility, as shown in the graph below. As a result, the lower the availability of the commodity, the higher the price in the shop. Item MRP yielded the most encouraging results.

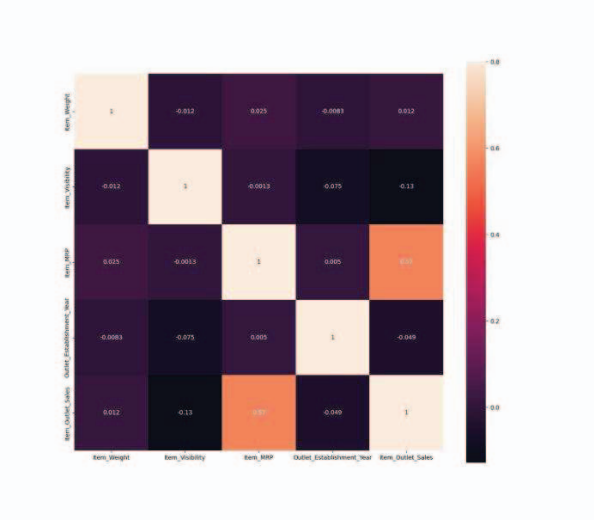

Figure 3. A graph to predict the correlation of variables with the target variable.

Data visualization showed that the tiniest sites produced the least sales. Although, in other circumstances, it was observed that a moderate site achieved the maximum sales despite being a type-1 location (There are three sorts of supermarkets: type-1, type- 2, and type-3) which is shown in Fig 4 and Fig 5.

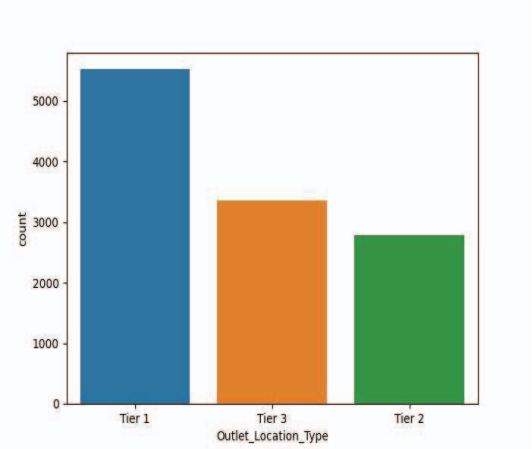

Figure 4. Impact of outlet location type on target variable item outlet sales.

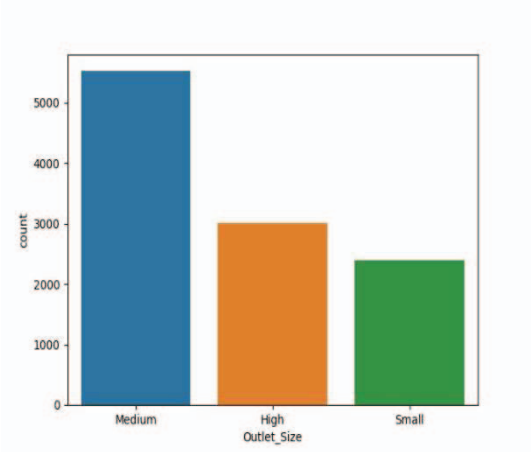

Figure 5. Impact of outlet size on target variable item outlet sale.

A. Prediction Analysis

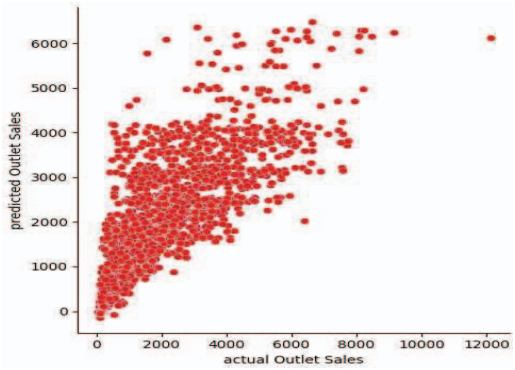

Figure 6: Scattered images of Actual outlet sales v/s Predicted outlet sales.

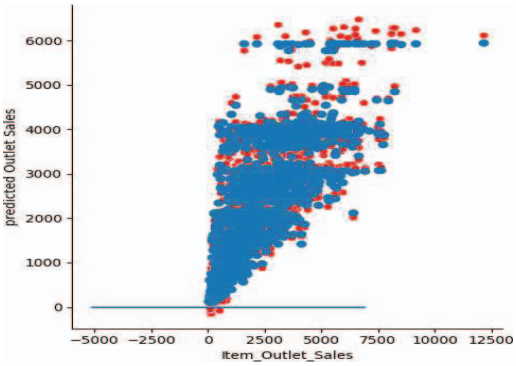

Figure 7: Test Prediction Random Forest Regressor.

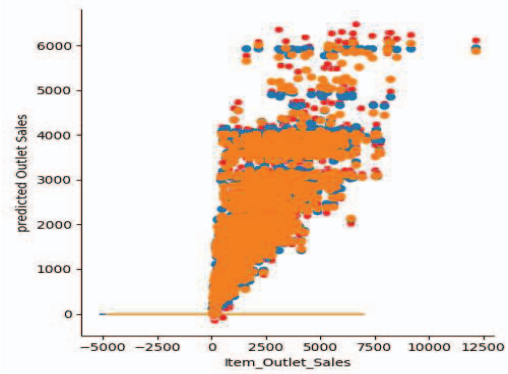

Figure 8: Test prediction with XGBoost Regressor.

VII. Conclusion

The goal of this framework is to estimate future sales based on the previous year's data using machine learning techniques. In this paper, we have discussed how various machine learning models are generated utilizing various techniques for example, linear regression and random forest, regressor, and XG Boost algorithms. These algorithms have been used to forecast the final result of sales. In this study, we created a predictive model in the Big Mart dataset utilizing ensemble techniques and the XGBoost algorithm to anticipate future sales of a certain Big Mart store or outlet.

VIII. Future Enhancement

To make the project more accessible, it could have collaborated into any device supported by an inbuilt intelligence through the Internet of Things (IoT). As more cases are investigated and additional inputs are provided to aid in the formulation of hypotheses, more accurate conclusions that are more representative of actual events can be drawn. Traditional techniques, when paired with effective data mining techniques and properties, can have a stronger and more favorable impact on the overall development of an organization's work. One of the project's key strengths is more expressive, accuracy-bound regression results. Furthermore, the adaptability of the proposed technique may be limited and strengthened by including variations at key phases of regression model building. Experiments are also necessary for reliable assessments of both precision and resource efficiency are required in order to analyze and optimize effectively.

References

[1] Naveenraj, R., Vinayaga Sundharam, R. (2013). Prediction Of BigMart Sales Using Machine Learning.
[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] Garc´ıa, S., Luengo, J., Herrera, F. (2015). Data preprocessing in data mining (Vol. 72, pp. 59-139). Cham, Switzerland: Springer International Publishing.
[4] Malik, N., Singh, K. (2020). Sales prediction model for Big Mart. Parichay: Maharaja Surajmal Institute Journal of Applied Research, 3(1), 22-32.
[5] Meghana, N., Chatradi, P., Chakravarthy, A., Kalavala, S. M., Ks, M. N. Improvizing Big Market Sales Prediction.
[6] Ragg, T., Menzel, W., Baum, W., Wigbers, M. (2002). Bayesian learning for sales rate prediction for thousands of retailers. Neurocomputing, 43(1-4), 127-144.
[7] Neal, R. M. (2012). Bayesian learning for neural networks (Vol. 118). Springer Science Business Media.
[8] Cheriyan, S., Ibrahim, S., Mohanan, S., Treesa, S. (2018, August). Intelligent sales prediction using machine learning techniques. In 2018 International Conference on Computing, Electronics Communications Engineering (iCCECE) (pp. 53-58). IEEE.
[9] Armstrong, J. S. (2008). Sales forecasting. Available at SSRN 1164602.
[10] Panjwani Mansi, Rahul Ramrakhiani, Hitesh Jumnani, Krishna Zanwar and Rupali Hande."Sales Prediction System Using Machine Learning." No. 3243. EasyChair, 2020.
[11] Ait-Alla, A., Teucke, M., Lütjen, M., Beheshti-Kashi, S., Karimi, H. R. (2014). Robust production planning in fashion apparel industry under demand uncertainty via conditional value at risk. Mathematical Problems in Engineering, 2014.
[12] 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.
[13] Dasari, A. K., Ramasubbaiah, B. Bigmart Sales Using Machine Learning With Data Analysis.
[14] Mitchell, T. M. (1999). Machine learning and data mining. Communications of the ACM, 42(11), 30-36.
[15] Krishna, A., Akhilesh, V., Aich, A., Hegde, C. (2018, December). Sales-forecasting of retail stores using machine learning techniques. In 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions(CSITSS) (pp. 160-166). IEEE.