Elevate Sales Profit Insight Analysis Using Machine Learning

P. Vanathi[1], V. Balaji[2], M. Bharath[3], Harsh Kumar Pandey[4], M. Jeeva[5]
[1]*Assistant Professor, Adhiyamaan College of Engineering, Hosur, Tamilnadu*
[2,3,4,5]*Student, Adhiyamaan College of Engineering, Hosur, Tamilnadu*

Abstract: The sales forecast is based on BigMart sales for various outlets to adjust the business model to expected outcomes. The resulting data can then be used to prediction potential sales volumes for retailers such as BigMart through various machine learning methods. The estimate of the system proposed should take account of price tag, outlet and outlet location. A number of networks use the various machine- learning algorithms, such as linear regression and decision tree algorithms, and an XGBoost regressor, which offers an efficient prevision of BigMart sales based on gradient. At last, hyperparameter tuning is used to help you to choose relevant hyperparameters that make the algorithm shine and produce the highest accuracy. Furthermore, leveraging advanced machine learning techniques enables the creation of robust models capable of capturing intricate patterns within BigMart's sales data. Through meticulous feature engineering, including variables like price, outlet size, location, and historical sales trends, these models can offer precise sales forecasts essential for informed decision-making. The incorporation of diverse algorithms, ranging from traditional linear regression to sophisticated ensemble methods like XGBoost, enhances the predictive prowess of the system. By harnessing the power of gradient boosting, the XGBoost regressor elevates forecasting accuracy by effectively capturing nonlinear relationships and interactions among variables.

Index Terms - Machine Learning Algorithms, Prediction, Reliability, Sales forecasting, Prediction model, Regression.

I. INTRODUCTION

Every item is tracked for its shopping centres and BigMarts in order to anticipate a future demand of the customer and also improve the management of its inventory. BigMart is an immense network of shops virtually all over the world. Trends in BigMart are very relevant and data scientists evaluate those trends per product and store in order to create potential centres. Using the machine to forecast the transactions of BigMart helps data scientists to test the various patterns by store and product to achieve the correct results. Many companies rely heavily on the knowledge base and need market patterns to be forecasted. Each shopping centre or store endeavours to give the individual and present moment proprietor to draw in more clients relying upon the day, with the goal that the business volume for everything can be evaluated for organization stock administration, logistics and transportation administration, and so forth. To address the issue of deals expectation of things dependent on client's future requests in various BigMarts across different areas diverse Machine Learning algorithms like Linear Regression, Random Forest, Decision Tree, Ridge Regression, XGBoost are utilized for gauging of deals volume. Deals foresee the outcome as deals rely upon the sort of store, populace around the store, a city wherein the store is located, i.e. it is possible that it is in an urban zone or country. Population statistics around the store also affect sales, and the capacity of the store and many more things should be considered. Because every business has strong demand, sales forecasts play a significant part in a retail centre. A stronger prediction is always helpful in developing and enhancing corporate market strategies, which also help to increase awareness of the market.

II. LITERATURE SURVEY

[1] '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.

[2] 'Applying machine learning algorithms in sales prediction'

This is a thesis in which several distinct procedures of machine learning algorithms are utilized to get better, optimal results, which are further examined for prediction task. It has made use of four algorithms, an ensemble technique etc. Feature selection has also been implemented using different tactics.

[3] '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. The potential of the algorithmic methods is estimated and accordingly used in further research.

[4] '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. Gradient Boost algorithm has been shown to exhibit maximum accuracy in picturizing the future transactions.

[5] 'Retail sales prediction and item recommendations using customer demographics at store level'

This paper outlines a sales prediction system along with the product recommendation system, which was used for the benefit of the group of retail stores. Consumer demographic details have been used for precisely designing the sales of each individual.

[6] 'Utilization of artificial neural networks and GAs for constructing an intelligent sales prediction system'

In the study, usage of deep neural network techniques is to know about their sales strategy regarding electronic components ahead in time. Some optimization algorithms are also used to maximize the efficiency of the system: like Genetic Algorithm.

[7]'Bayesian learning for sales rate prediction for thousands of retailers'

In this paper it is shown that from the prediction of the single one's rate of transactions, many vendors would benefit from it, that means the information obtained could be beneficial for the construction of a set-up that would estimate large number of outputs. The prediction uses neural network approach. Here they have practiced Bayesian learning to gain insights.

[8] 'Combining Data Mining and Machine Learning for Effective User Profiling'

This research describes the way of detecting suspicious behaviour by employing an automatic prototype. Several machine learning methodologies have been made in use for concluding this appropriate prototype. Here data mining and constructive induction techniques are merged to pull out the discrepancy found in the conducts of the owners of cell phones.

III. EXISTING SYSTEM

In the existing system, the approach to sales prediction at big malls and marts relies on traditional linear regression models. These models are widely used for their simplicity and ease of interpretation. The system incorporates fundamental statistical techniques to estimate the relationship between various features, such as product we and maximum retail price, and the corresponding sales Figures. The linear regression models offer a straightforward way to make predictions based on historical sales data, providing valuable insights into potential trends. However, the existing system may face challenges in capturing more complex relationships within the data, and its interpretability may be limited compared to more advanced machine learning models. The reliance on linear regression underscores a traditional but potentially less flexible methodology for sales prediction in a dynamic retail environment.

Disadvantages:

•Traditional Linear Regression

The existing system relies on traditional linear regression models, which may not capture the complex relationships present in the sales data. This can lead to limitations in predictive accuracy, especially when dealing with non-linear patterns.

•Limited Model Flexibility

Due to the reliance on linear regression, the existing system may lack the flexibility to adapt to diverse data patterns and may not perform optimally when faced with complex, non-linear relationships.

•Fixed Hyperparameter ConFigureuration

The hyperparameters in the existing system are often fixed and may not be optimized for the specific dataset. This can lead to suboptimal model performance, especially in scenarios where the relationship between features and sales is dynamic.

IV. PROPOSED SYSTEM

A) EXPLORATORY DATA ANALYSIS

It is beneficial to add test data to train data to explore data in every dataset and thus to merge train and test data with a view to data visualization, feature engineering. For the exploratory method, univariate analysis and bivariate analysis are to be conducted to obtain data information. Few observations have been made during the Univariate Analysis and are as follows: The categories 'LF', 'low fat', and 'Low Fat' are the same and 'reg' and 'Regular' are the same category. As a result, they can merge into one, and Low fats are almost twice that of regular items. The main sales in the Item Type column are Fruit and Snack. The variable goal is skewed to the right. These items are not consumable, but all items are labelled either as low fat or regular items. Through the study of Bivariate, a clear relationship between product weight and sales and between item fat content and sales has been found. A significant amount of sales is obtained from products with visibility below 0.2. Individuals have selected a low fat category over other groups. In the relationship between the item identifiers and the outlet size, the items are purchased more frequently as the outlet size increases. The exposure of the item means that more visible items have less sales.

B) DATA PREPROCESSING

The pre-processing of data is a method for preparing and adapting raw data to a model of learning. This is the first and significant step to construct a machine learning model. Real-world data generally contain noise, missing values and may not be used in an unusable format especially for machine learning models. Data pre-processing needs to be performed in order to purify data and adapt it to the machine learning model of a system which also makes a machine learning model more accurate and efficient. The first thing for data pre-processing is to collect the required dataset, and then check the missing values once the dataset is imported. Correcting missed values is necessary, or else the data would be difficult to access and maintain. Then calculate the mean of the column containing missing values to rectify the missed values, and substitute it with the measured mean. When the dataset is pre-processed, the dataset is separated into the dataset of train and test. Now, this dataset can be used to train a machine learning algorithm to predict Item Outlet Sales against a variety of items that will help retailers create personalized offers against specific products for customers.

C) FEATURE ENGINEERING

Feature Engineering is a method to exploit domain data understanding to construct functions that work with machine learning algorithms. When feature engineering is done correctly, the predictive capability of machine learning algorithms is enhanced by building raw data features that help facilitate the machine learning process. Feature engineering also includes the correction of inappropriate values. In the device dataset, the visibility of the item had a minimum value of 0 which is not acceptable, because the item should be accessible to all. And so it was replaced by the mean of the column. As Outlet Years, a new column is created so we must consider how long the store runs instead of the year it was formed. Item Type is another column in the dataset that has 16 categories and is combined under the Food, Drink and Non-Consumable category. Column Item fat content had various representations, which were divided into low fat and regular categories. Outliers present in Item Outlet Sales are often excluded for better performance.

D) EVALUATION METRICS

Evaluation of the model is the vital part of creating an efficient machine learning model. Therefore, it is important to create a model and get suggestions from it in terms of metrics. It will take and continue until we achieve good accuracy according to the value obtained from metric improvements. Evaluation metrics describe one model's results. The ability to distinguish between model outcomes is an important feature of the evaluation metrics. Here, we used Root Mean Squared Error(RMSE)metric for evaluation process. RMSE is given by following formula

RMSE=qPN i=1(PredictediActuali)2N.

Where, N is the Complete Number of Observations. RMSE is the most commonly used evaluation method for regression problems. The power of 'square root' causes this metric to display significant variation in percentages. The 'squared' aspect of this metric tends to deliver more stable outcomes that avoids the cancelation of positive or negative error values.

V. MODEL BUILDING

The dataset is now ready to fit a model after performing Data Pre-processing and Feature Transformation. The training set is fed into the algorithm in order to learn how to predict values. Testing data is given as input after Model Building a target variable to predict. The models are build using:

A. Linear Regression

B. Random Forest

C. XGBoost

For all models based on the above algorithms, 20 fold cross validation is used. Essentially cross validation provides an indication of how well a model is generalizing to the unseen results. Description of different algorithms used as follows:

A. Linear Regression

The most common and simplest statistical approach for predictive modelling is linear regression. Below is the linear regression equation: Y = _1X1 + _2X2 + ::::_nXn Where X1, X2,..., Xn are the independent variables, Y is the target variable and all the coefficients are the thetas. The magnitude of a coefficient as compared to the other variables determines the importance of the corresponding independent variable. This algorithm's basic principle is to match a straight line between the chosen training dataset features and a constant target variable, i.e. sales. The algorithm chooses a line which fits better with the data. Linear regression performs the task of predicting a dependent variable value (y) based on a given independent variable (x). This regression technique considers a linear relationship between x (input) and y(output). Some requirements for a successful linear regression model must be fulfilled by data. Some of those is the lack of multi-collinearity, i.e. the independent variables should correlate with each other.

B. Random Forest

Random Forest is a tree-based bootstrapping algorithm that combines a certain number of weak learners (decision trees) to construct a powerful model of prediction. For each person learner, a random set of rows and a few randomly selected variables are used to create a decision tree model. Final prediction may be a function of all the predictions made by the individual learners. In the event of a regression problem, the final prediction may be the mean for all predictions. With this algorithm RMSE:1069 is reached.

C. XGBoost

XGBoost stands for extreme Gradient Boosting. The implementation of the algorithm was engineered for the efficiency of computing time and memory resources. Boosting is a sequential process based on the principle of the ensemble. This incorporates a collection of low learners and improves the accuracy of predictions. Model values are weighted at any moment t, based on the effects of the preceding instant t1. The correctly calculated results are given a lower weight, and the wrong ones are weighted higher. With this algorithm, the XGBoost model implements the stepwise, ridge regression internally, which automatically chooses the features and removes the multi colinearity. RMSE:1052 is achieved with this algorithm.

VI. HYPERPARAMETER TUNING

Hyperparameter tuning selects an optimal range of hyperparameters for algorithm learning. A hyperparameter for this is a parameter the value of which is set before learning starts. Hyperparameters are not model parameters, and cannot be directly derived from results. By planning, System parameters shall be equipped when using gradient descent minimize the function to loss. Whilst the model parameters specify how input data can be translated to the desired output, the hyperparameters explain how the model is actually being structured. The best way to think of hyperparameters is like an algorithm 's settings which can be modified to maximize performance. Models can have multiple hyperparameters and can be treated as a test problem in order to find the right combination of parameters. While there are now many hyperparameter optimization / tuning algorithms, simple strategies: 1. Grid Search, and 2. Random search. However, computational methods for both grid search and random search tuning take a very long time, from an hour to a day. Because of its quickest calculation, thus, the Bayesian Optimization approach is used for hyperparameter tuning.

VII. BAYESIAN OPTIMIZATION

Bayesian methods, in contrast to random or matrix search, maintain track of previous test outcomes that they use to construct a probabilistic model mapping of hyperparameters to the likelihood of an objective function score:P(score jhyperparameters): The simple theory is to spend a little more time choosing the next hyperparameter and allow fewer calls to the objective function. The goal of Bayesian reasoning is to become "less accurate" by constantly updating the surrogate probability model after-objective function evaluation with more data than these methods do. Bayesian model-based approaches can find better hyperparameters in less time, since they purpose for determining the right range of hyperparameters based on previous experiment.

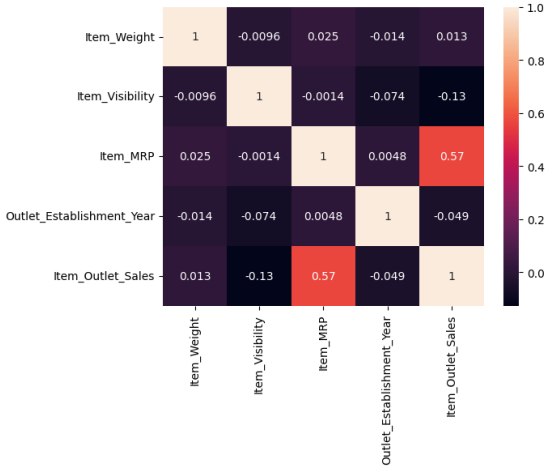

Figure 1: Heatmap of correlation coefficients.

VIII. RESULTS

i) Correlation

The heatmap that displayed the correlation between each variable was generated by the function nominal.association() from python library. It should be noted that the product ID and outlet ID, the two features used for identification, were not considered because they were not relevant to the objective of this study. The correlation was presented in Figure. 1. Except for MRP and outlet type, most of the features had a low correlation with the target variable, outlet sales. The relatively high correlation(0.21) between outlet size and outlet sales was, to some extent, due to their high correlation with outlet type. Based on the obtained correlation coefficients, the three least correlated features - weight(0.01), fat content (0.03), and year of establishment (-0.05) - were tested during model training to decide whether to be applied. Test results showed that both fat content and the year of establishment had a positive effect on the accuracy of the model. Thus, excluding weight, the remaining 8 features were used in model training.

ii) Model Performance

For each model, scatter plots with observed values on x-axis and predicted values on y-axis were used to illustrate the performance. Ideally, the model should have a high R2 score, which meant that the arrangement of the scatter points should be close to a straight line with a slope of 1.

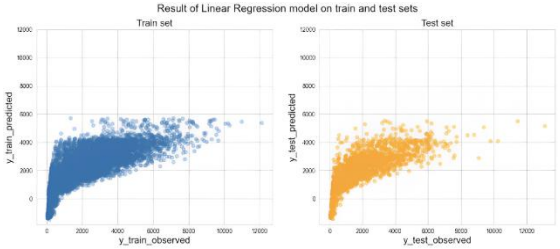

Figure 2: Scatter plots for results of linear regression.

The results of the linear regression model (seen in Figure. 2) did not appear to be linear. Instead, it was more like a curve with a positive but decreasing slope, which gave an upper limit that was around 6000 to predicted values. In addition, for part of the data with observed values less than 1000, the model's predictions were less than 0, which was inaccurate because the value of sales would never be negative. The reason for the unsatisfactory results might be related to the nature of the features used. Most of the features involved in training were categorical variables, which were not well suited for linear regression in the case of low correlation.

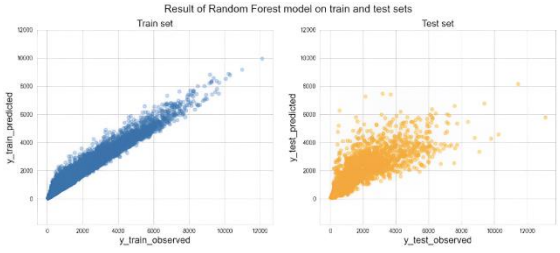

Figure 3: Scatter plots for results of random forest

As presented in Figure. 3, the result of Random Forest was more reasonable than that of the linear regression model. The model performed very well in the training set due to the characteristics of decision trees and was

therefore not informative, while the result for test set was not as perfect, but a linear trend still existed. From the density of data points, it could be seen that the random forest model performed better for data with observed sales between 0 and 6000, and it had a large error for data that had sale values greater than 6000. A possible reason for this phenomenon was that most of the samples used for training were concentrated between 0 and 6000, which made the model unable to effectively classify data beyond this interval.

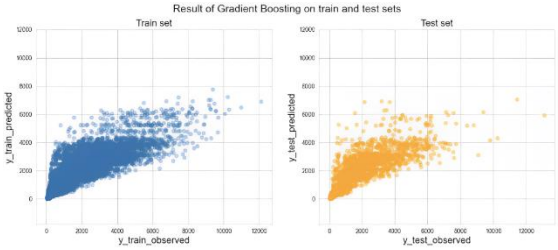

Figure 4: Scatter plots for results of gradient boosting

The results of Gradient Boosting were illustrated in Figure. 4. Also based on decision trees, Gradient Boosting performed worse than Random Forest in fitting the training set due to its additive nature. However, its performance in the test set was slightly better, i.e., the linear trend of data points was more obvious. The predicted values obtained from this model were lower than the true values. Besides, the same issue in the prediction for data with true values larger than 6000 existed as well. The errors and R2 scores of different models were generally consistent with the trends exhibited by the scatter plots, as presented in Table. 1.

	Linear Regression	Random Forest	Gradient Boosting
RMSE_train	1124.926	425.269	1029.046
MAE_train	834.714	293.638	727.225
R2_train	0.562	0.937	0.633
RMSE_test	1142.004	1133.550	1086.291
MAE_test	840.730	790.548	753.112
R2_test	0.567	0.574	0.574

Table 1: Errors and R2 scores of models

Overall, Gradient Boosting performed the best among the three models. It had the highest R2 score and lowest errors in testing. Random Forests performed slightly worse than Gradient Boosting, which Proceedings of the 2nd International Conference on Business and Policy Studies DOI: 10.54254/2754-1169/17/20231094 206 was consistent with the fact that gradient-boosted trees generally outperform random forests [9]. The linear regression model had relatively the worst accuracy, which could be determined by both the RMSE and MAE. Its RMSE was close to that of the other two models because there were more outliers in the results of them, which was because of the lack of accuracy in their prediction of data with large values.

IX. CONCLUSION

Experts also shown that a smart sales forecasting program is required to manage vast volumes of data for business organizations. Business assessments are based on the speed and precision of the methods 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 formulate marketing plans will be implemented. The outcome of machine learning algorithms will help to select the most suitable demand prediction algorithm and with the aid of which BigMart will prepare its marketing campaigns.

REFERENCE

[1] Singh Manpreet, Bhawick Ghutla, Reuben Lilo Jnr, Aesaan FS Mohammed, and Mahmood A. Rashid. "Walmart's Sales Data Analysis-A Big Data Analytics Perspective." In 2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), pp. 114-119. IEEE, 2017.

[2] Sekban, Judi. "Applying machine learning algorithms in sales prediction." (2019).

[3] Panjwani, Mansi, Rahul Ramrakhiani, Hitesh Jumnani, Krishna Zanwar, and Rupali Hande. Sales Prediction System Using Machine Learning. No. 3243. EasyChair, 2020.

[4] Cheriyan, Sunitha, Shaniba Ibrahim, Saju Mohanan, and Susan Treesa. "Intelligent Sales Prediction Using Machine Learning Techniques." In 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), pp. 53-58. IEEE, 2018.

[5] Giering, Michael. "Retail sales prediction and item recommendations using customer demographics at store level." ACM SIGKDD Explorations Newsletter 10, no. 2 (2008): 84-89.

[6] Baba, Norio, and Hidetsugu Suto. "Utilization of artificial neural networks and GAs for constructing an

intelligent sales prediction system." In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, vol. 6, pp. 565-570. IEEE, 2000.

[7] Ragg, Thomas, Wolfram Menzel, Walter Baum, and Michael Wigbers. "Bayesian learning for sales rate prediction for thousands of retailers." Neurocomputing 43, no. 1-4 (2002): 127-144.

[8] Fawcett, Tom, and Foster J. Provost. "Combining Data Mining and Machine Learning for Effective User Profiling."In KDD, pp. 8-13. 1996.

[9] S. Cheriyan, S. Ibrahim, S. Mohanan and S. Treesa, quot;Intelligent Sales Prediction Using Machine Learning Techniques,quot; 2018 International Conference on Computing, Electronics amp; Communications Engineering (iCCECE), Southend, United Kingdom, 2018, pp. 53-58.

[10] C. M. Wu, P. Patil and S. Gunaseelan, quot;Comparison of Different Machine Learning Algorithms for Multiple Regression on Black Friday Sales Data,quot; 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2018, pp. 16- 20.

[11] A. Krishna, A. V, A. Aich and C. Hegde, quot;Salesforecasting of Retail Stores using Machine Learning Techniques,quot; 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 2018, pp. 160-166.

[12] G. Nunnari and V. Nunnari, quot;Forecasting Monthly Sales Retail Time Series: A Case Study,quot; 2017 IEEE 19th Conference on Business Informatics (CBI), Thessaloniki, 2017, pp. 1-6.

[13] Kadam, H., Shevade, R., Ketkar, P. and Rajguru, S. (2018). A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression. International Journal of Engineering Development and Research, 6(4), pp. 1-2.

[14] T. Alexander and D. Christopher, quot;An Ensemble Based Predictive Modeling in Forecasting Sales of Big Martquot;, International Journal of Scientific Research, vol. 5, no. 5, pp. 1- 4, 2016. [Accessed 10 October 2019].

[15] G. Behera and N. Nain, quot;A Comparative Study of Big Mart Sales Predictionquot;, pp. 1-13, 2019. [Accessed 10 October 2019].

[16] S. Beheshti-Kashi, H. Karimi, K. Thoben and M. L¨utjen, quot;A survey on retail sales forecasting and prediction in fashion marketsquot;, Systems Science amp; Control Engineering, vol. 3, no. 1, pp. 154-161, 2014. Available: 10.1080/21642583.2014.999389 [Accessed 27 January 2020].

[17] A. Chandel, A. Dubey, S. Dhawale and M. Ghuge, quot;Sales Prediction System using Machine Learningquot;, International Journal of Scientific Research and Engineering Development, vol. 2, no. 2, pp. 1-4, 2019. [Accessed 27 January 2020].

[18] B. Pavlyshenko, quot;Machine-Learning Models for Sales Time Series Forecastingquot;, Data, vol. 4, no. 1, p. 15, 2019. Available: 10.3390/data4010015 [Accessed 27 January 2020].

[19] T. T. Joy, S. Rana, S. Gupta and S. Venkatesh, quot;Hyperparameter tuning for big data using Bayesian optimisation,quot; 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, 2016, pp. 2574- 2579.

[20] M. Wistuba, N. Schilling and L. Schmidt-Thieme, quot;Hyperparameter Optimization Machines,quot; 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Montreal, QC, 2016, pp. 41- 50.

[21] M. Wistuba, N. Schilling and L. Schmidt-Thieme, quot;Learning hyperparameter optimization initializations, quot; 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Paris, 2015, pp. 1-10.

[22] K. Punam, R. Pamula and P. K. Jain, quot;A Two-Level Statistical Model for Big Mart Sales Prediction,quot; 2018 International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, Uttar Pradesh, India, 2018, pp. 617-620.