Analytical Methods of Machine Learning Model for E-Commerce Sales Analysis and Prediction

2023 IEEE International Conference on Contemporary Computing and Communications (InC4) | 979-8-3503-3577-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/InC457730.2023.10263003

Anushka Xavier K
Department of CSE
CHRIST (Deemed to be University)
Bangalore, India
anushkaxavier@gmail.com

Chinthakunta Manjunath
Department of CSE
CHRIST (Deemed to be University)
Bangalore, India
chinthakunta.manjunath@christuniversity.in

Manohar M
Department of CSE
CHRIST(Deemed to be University)
Bangalore, India
manohar.m@christuniversity.in

Gurudas V R
Department of CSE
CHRIST(Deemed to be University)
Bangalore, India
gurudas.vr@christuniversity.in

N. Jayapandian
Department of CSE
CHRIST(Deemed to be University)
Bangalore, India
njayapandian@gmail.com

Balamurugan M
Department of CSE
CHRIST(Deemed to be University)
Bangalore, India
balamurugan.m@christuniversity.in

***Abstract*— In the commercial market, E-commerce sales show a significant trend and have attracted many consumers. E-commerce sales forecasting has a significant role in an organization's growth and aids in improved operation. Many studies have been conducted in the past using statistical, fundamental, and data mining techniques for better analysis and prediction of sales. However, the current scenario calls for a better study that combines the available information to propose different machine-learning techniques. The sole motive of the study is to analyze and determine different machine learning models to predict accurate results. The research observed that the Extreme Gradient Boosting model outperformed all other models and brought a good result. It produced an RMSE value of 0.0004 and Explained Variance score of 0.99. Decision Tree algorithm also shows an exemplary result.**

Keywords*— *Sales Prediction, Machine Learning, Boosting, Explained Variance, Extreme Gradient Boosting.

I. INTRODUCTION

E-commerce is expanding in today's financial world [1]. It made a paradigm shift that affected both marketers and customers. The COVID-19 crisis also hastened the growth of online shopping [2]. Customers can now purchase a variety of goods from their own residence, and businesses can continue to run despite the restrictions. Moreover, the world is evolving fast, and the business sectors are on edge with technology to meet market demands. The corporate industries are striving to satisfy customer requirements and simultaneously procure profit from their investment. Therefore, comprehending the future trend plays an important role. Sales prediction aids in estimating the upcoming sales of the firm in advance [3]. It helps in effective decision making and appropriate resource allocation. Thus, it is ideal for enhancing revenue and promoting the organization's growth.

Sales prediction anticipates future sales and provides a way to analyze the company's performance. Traditional forecasting methods relied heavily on expert employee suggestions or quantitative analysis of historical data. Machine learning approaches have been seen to outperform these strategies. Machine learning is the discipline where the system utilizes algorithms to excel humans [4]. It learns from the patterns hidden in the data and observes the result. These techniques find the optimum result with minimal human intervention. It has proven beneficial in sales forecasting by providing great insight into future sales. In our study, we envisaged supervised machine learning algorithms to yield optimum results in the prediction.

Our research proposes methodology to obtain predictive sales data with a minimal error rate, consequently escalating the accuracy. For this purpose, eight machine learning algorithms were built and contrasted. Advanced boosting models were utilized along with regression models to obtain impeccable results. RMSE value and variance score are used as the evaluation metrics to estimate the efficiency of the models.

The remaining paper is structured as follows: The following section looks at scientific research on sales forecasting. The third section defines the proposed methodology and the steps followed with it. Fourth section explains the experimental analysis and results. In addition, the fifth section concludes the study by speculating future possibilities. Finally, in the following section, references are included.

II. RELATED WORK

There have been several recent works in sales prediction that use machine learning and data analytics techniques. This section provides insight into the methods adopted by numerous researchers in sales prediction. The study suggests an unique deep learning strategy for to predict stock movement [5]. Two recurrent neural networks are combined in the model using the blending ensemble learning technique, which is followed by a fully connected neural network. The results demonstrate that blending ensemble deep learning model exceeds the leading prediction model currently in use. The study proposes a hybrid network combining Convolutional Neural Network and Bi-directional Long Short-Term Memory to predict ecommerce [6]. Various types of data are normalised via feature engineering. All the

comments are analysed using BiLSTM. With the information that feature engineering has provided, CNN is used to create predictions. The research was performed using clustering model techniques and estimation for sales [7]. Based on the review, the best-fit prediction model for the firm's marketing was suggested. Similarly, a study on Walmart sales using machine learning techniques was conducted [8]. The research focused on acquiring the best result using three main classification models. R^2 score and Mean Absolute Error (MAE) metrics with appropriate hyperparameters were used to evaluate the algorithms. The Random Forest algorithm was chosen as the best algorithm with a minimal error rate and high R^2 score.

Furthermore, a study on car sales forecasting using machine learning was conducted [9]. The article concentrates on the production of vehicle sales statistics from several sources. The researchers used a random forest algorithm for their survey, and according to its results, the price was the most important feature that significantly impacted the sale of a car. The study resulted in a reasonable accuracy rate that hovered above 85 percent. A survey on market data and sales prediction was illustrated [10]. The study investigates the conclusions from the experimental data and the insights gained via data visualization and mining methods. The article mainly deals with three machine learning algorithms (Generalized Linear Model, Gradient Boost and Decision Tree) of which Gradient Boost yields the best result. A notion to minimize the error rate of prediction using the Extra Gradient Boost algorithm along with the assistance of the SigOpt Bayesian Optimization technique [11].

This work suggests a revolutionary prediction model to yield an outstanding result by providing perfect accuracy and suppressing the error rate to a minuscule level. Advanced machine learning models were analyzed and deployed on a large dataset. Data transformation techniques such as label encoding and feature scaling were employed to transform the data. Also, hyperparameters were tuned to build an outstanding model.

III. Proposed Methodology

In our research, a five-tier approach is taken to address the issue of sales prediction. Firstly, the Global Superstore dataset was collected from an open database, Kaggle Repository [12]. It has 24 attributes showcase the customer, product, order, and sales details. The dataset consists of online retail store order information gathered from 147 countries. These data underwent a preliminary analysis, where the data was visualized and interpreted. Secondly, data pre-processing is carried out, where the data gets cleaned from all the noises and prepared for the forthcoming stages. Thirdly, feature transformation is performed to forge the data into a more understandable and reliable form. In the fourth stage, the data is branched into training and testing sets to decrease the complexity. Later in the fifth stage, the branched information is fed into different models that have been constructed and evaluated for the results. Figure 1 shows the architecture diagram.

The significant findings of our research are the following. Our proposed approach provides an ideal model for sales prediction in an e-commerce market. A variety of statistical methods were explored and compiled to create the best model for predicting sales. Data transformation techniques such as label encoding and feature scaling were employed on the dataset to provide an outstanding result. The XGBoost regression model surpassed all other models producing high accuracy on test data. Advanced regressive and tree models also performed well by producing a good accuracy for the test data. The RMSE value was nearer to zero showing a lower error rate, thus creating high accuracy.

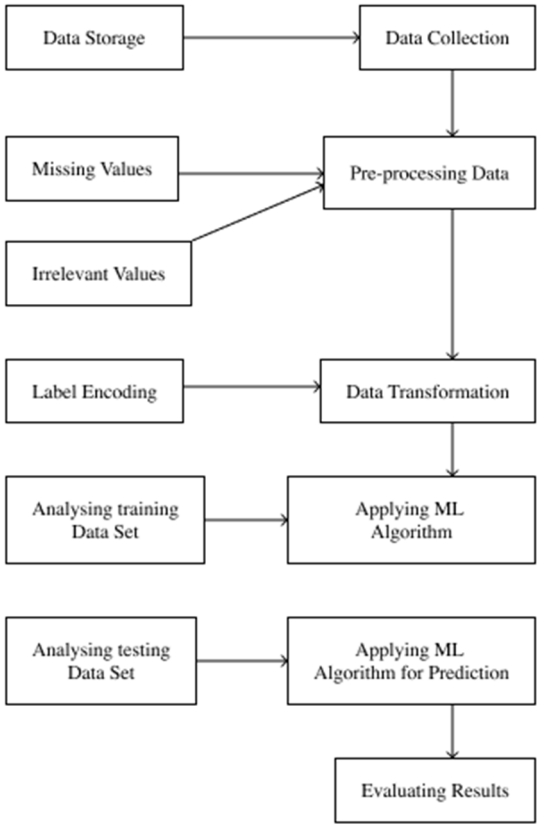

Fig. 1. System Architecture

A. Dataset Collection

The Global Superstore dataset was collected from the Kaggle repository [13]. Global Superstore is a collection of data from various New York-based online retailers collected and organized in the form of a dataset. The dataset constituted order data of Global Superstore gathered from 147 countries worldwide. The retail dataset consists of 51290 entries which is standard. The 24 attributes showcase the customer, product, order, and sales details. The customer information includes the customer name, id, segment, city, and state. The order information gives an idea about the order date and order id. Product id, category ,product name, and sub-category explain the product details. Quantity, sales, discount, and profit show an overview of the sales statistics. These particulars with suitable training are used for constructing the model. Table 1 shows the data description of the sales data set.

B. Data Pre-Processing

Data cannot be used in machine learning algorithms in its natural state because of the way it was gathered; hence the data must be synthesized before being employed; Pre-processing aids in this process. Data pre-processing ensures that the features are consistent, complete, and flawless. The steps that

make up data preparation are as follows. Firstly, the dataset and the necessary python libraries are imported. The Global Superstore CSV file is attached for the proceeding. Based on the data, a few categories (Order_Id, Customer_Id, Customer_Name, Row_Id, Ship_Date, Order_Date) were eliminated as these parameter does not contribute to final output.

Later missing values are handled to reduce the discrepancy of data. Postal_Code contains 41296 null values in our data, which are dropped to reduce the noise.

TABLE I. DATA DESCRIPTION

Attribute	Description	Data Type
Row_Id	Distinct Id	int64
Order_ID	Unique identifier for each order	Object
Order_Date	The ordered date	datetime64
Ship_Date	The shipped date	datetime64
Ship_Mode	The shipping mode	Object
Customer_ID	Unique identifier for each customer	Object
Customer_Name	The name of the customer	Object
Segment	The type of the consumer	Object
City	The city of the customer	Object
State	The state of the customer	Object
Country	The country of the customer	Object
Postal_Code	The postal code of the customer	float64
Market	The market of the sales	Object
Region	Region of the sales	Object
Product_ID	Unique identifier for each product	Object
Category	Category of the product	Object
Sub_Category	The sub-Category of the product	Object
Product_Name	The name of the product	Object
Sales	The sales made by the product	float64
Quantity	Quantity sold by each product	int64
Discount	Discount offered to each product	float64
Profit	The profit obtained from each sale	float64
Shipping Cost	The shipping cost of each product	float64
Order Priority	The order priority	Object

C. Data Transformation

Data transformation is a way of transforming the unstructured data into a suitable format for prediction. Converting the data assures the highest possible data quality, which is essential for proper interpretation. Additionally, it will improve the outcome. Label encoding is performed to convert the categorical features to numerical representations. This ensures that the dataset is enriched with adequate statistical data. Label encoding was performed on the categorical attributes such as Ship_Mode, City, Segment, State, Region, Country, Market, Product_ID, Category, Sub-Category, Product Name, and Order Priority. As a result, the model can learn a detailed representation of the data.

Furthermore, feature scaling is employed to convert data into a consistent and scalable size to improve precision and eliminate errors. It prohibits the algorithm from using a wide variety of data points, thus attaining better outcomes. Here, normalization using min-max scaling is utilized to bring down the features to a standard scale and more comparable form.

Lastly, the data is split into train and test sets. The training set was used to build a model that recognized the sales pattern, while the testing set was used to evaluate the model and assess its predictive skills. 70 percent of data was used in the test set, while 30 percent of samples were used to assess the model.

D. Algorithms

Algorithms are the procedure that is used to produce models for pattern recognition. A variety of algorithms are adopted to forecast accurate results. The following section describes the algorithms used in this research.

The K-Nearest Neighbors algorithm is a supervised learning approach by Fix [15,16]. It estimates the result based on the proximity of other available samples. The closeness of data is computed using the distance function, commonly the Euclidean distance.

$$\text{Euclidean Distance} = \sqrt{\sum_{i=0}^{n}(y_i - x_i)^2} \quad (1)$$

Equation 1 is the euclidean distance formula that estimates the distance between the points where x and y are points. Decision trees are Supervised machine learning algorithms proposed by Quinlan [17]. A decision tree regressor is a statistical method, for instance, classification based on attribute values. They usually begin from the cluster head and break into subgroups to build subtrees.

Random Forest is a kind of efficient ensemble machine learning approach for predictive modeling [18,19]. It is described as a combination of decision trees that uses bagging and boosting techniques to help deliver correct output. Jerome Friedman pioneered a new boosting algorithm named Gradient Boosting for statistical analysis [20]. A Gradient Boosting Algorithm is an ensemble learning technique that integrates predictions from several decision trees to provide the result. Extreme Gradient Boosting, commonly referred to as XGBoost, is another main boosting method. In XGBoost, trees are produced in sequential order, with each tree attempting to fix the mistakes of the preceding trees. It uses a gradient-based optimization algorithm to train the trees, which makes it more efficient than traditional gradient boosting methods. It uses regularization to prevent overfitting, which can lead to better generalization performance. It allows for custom loss functions and evaluation metrics and can handle missing values and categorical features. LightGBM, or Light Gradient Boosting Machine, is a Gradient boosting algorithm. When compared to other algorithms, the trees in Light gradient boost grow leaf-by-leaf. As the title implies, CatBoost is a boosting technique that can accommodate categorical features in data. They deal with the descriptive features in data on teir own. The adaptive Boosting Algorithm, commonly addressed as Adaboost, is an ensemble technique used in supervised learning. The Adaboost algorithm works in the same way as boosting does.

IV. EXPERIMENT SETUP AND PERFORMANCE MEASURE

The complete dataset was branched into train and test sets to minimize the complexity. The train set is used to fit the model, whereas the test set is used to assess the train set. The train and test split are 70% and 30%, respectively. Machine Learning algorithms such as Random Forest Regression, Decision Tree Regression, KNearest Neighbors, CatBoost, XGBoost algorithm, LightGBM, Gradient Boost, and AdaBoost algorithm have been used to anticipate the sales. To achieve an excellent result, the models were trained using appropriate hyperparameters. Table 2 showcases the model's parameters and its values.

Different metrics such as train and test sore accuracy, mean absolute error (MAE), variance score and root mean squared error (RMSE) and were adopted in our research. Explained Variance Regression Score is the disparity between a model and actual data measured using explained variance regression score. The best score is 1.0, while the lesser numbers are worse. It is equated using equation 2 where y is

each value in the dataset and $\hat{y}$ is the mean of all values in the dataset and var is the biased variance.

TABLE II. PARAMETER DESCRIPTION

Model	Hyperparameter
K Nearest Neighbors Regression	n_neighbors=9
Random Forest Regression	n_estimators = 100, n_jobs = -1
Gradient Boosting Regression	learning_rate=0.3, max_depth=9, verbose=False
Adaptive Boosting Regression	n_estimators = 100
Light Gradient Boosting Machine Regression	boosting_type='gbdt', max_depth=9, learning_rate = 0.5, feature_fraction = 0.8, min_data_in_leaf= 100, bagging_fraction= 0.3, metric= 'rmse', random_state=100, seed=4, objective='regression', num_leaves =60
Categorical Boosting Regression	learning_rate=0.3, max_depth=9, verbose=False
Extreme Gradient Boosting Regression	learning_rate=0.5, max_depth = 9, silent= 1, seed= 4, objective= 'reg:linear'

$$\text{Explained Variance} = 1 - \frac{\text{Var}\{y-\hat{y}\}}{\text{Var}\{y\}} \quad (2)$$

The standard deviation of the predicted errors is used to determine the Root Mean Square Error (RMSE). These errors define how the data points are grouped around the regression line. The RMSE value will differ based on the alignment of errors on this line. This factor determines how close the valid data is to the line. The RMSE value is evaluated using equation 3 where n is the total number of values and predicted and actual gives the predicted and actual values.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Predicted - Actual)^2} \quad (3)$$

Mean Absolute Error (MAE) is a widely used metric to evaluate the accuracy of regression models. It measures the average absolute difference between the predicted and actual values of the target variable. It is equated using equation 4.

$$MAE = (1/n) * \Sigma|yi - \hat{y}i| \quad (4)$$

where,n is the number of samples in the dataset, yi is the actual value of the target variable for the i-th sample, ŷi is the predicted value of the target variable for the i-th sample. Lower values of MAE indicates better model performance.

V. RESULT AND DISCUSSIONS

The train and test accuracy produced by Extreme Gradient Boosting regression were greater than all other models. This suggests that Extreme Gradient Boosting has a superior fit for the data. Extreme Gradient Boosting brought exemplary train and test accuracy. However, except KNN, other regression models churn out and have good accuracy in both the training and test scores of the data. This implies that the efficiency of Random Forest, Decision Tree, Gradient Boosting, CatBoost, LightGBM, and AdaBoost was equally good. KNN achieved a train score of 77% and a test score of 71%. The significant difference in these scores suggests that the model did not fully embrace the test and train data. Figure 2 depicts the test and train score.

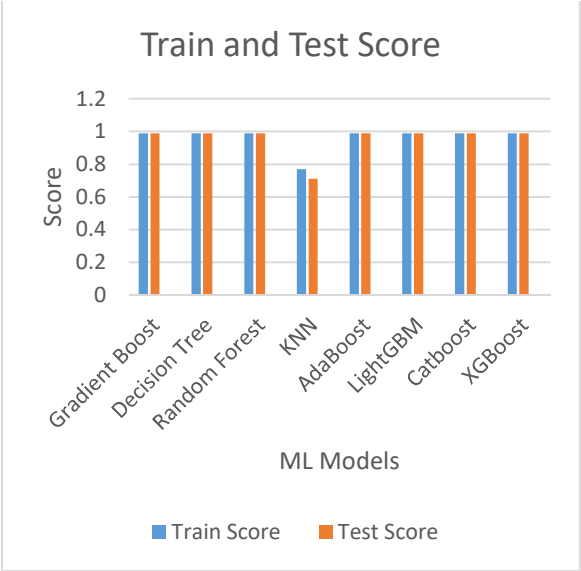

Fig. 2. Train and Test Scores

The XGBoost model's Root Mean Square Error (RMSE) is comparatively lower than other models. This implies that the model tried to achieve a reasonable prediction rate. The residual error of the Decision Tree and Random Forest ranges simultaneously. Also, CatBoost and LightGBM brings down the error to a comparable rate scaling 0.002. Additionally, it is evident that KNN has the highest error rate compared to other models indicating a frail prediction capability. Table 3 showcase the overall performance of each model using RMSE, Explained Variance Score and MAE.

TABLE III. PERFORMANCE OF EACH MACHINE LEARNING MODEL

S.No	ML Models	RMSE	Explained Variance	MAE
1	K Nearest Neighbors	0.078	0.49	0.062
2	Decision Tree	0.0004	0.99	3.082e-05
3	Random Forest	0.0006	0.99	0.0002
4	Gradient Boosting	0.006	0.99	5.557e-05
5	Adaptive Boosting	0.012	0.99	0.009
6	Light Gradient Boosting Machine	0.002	0.99	0.0009
7	Categorical Boosting	0.002	0.99	0.0009
8	Extreme Gradient Boosting	0.0004	0.99	2.492e-05

Figure 3 compares the RMSE Values of different models. The RMSE value of KNN is higher than all the models. The XGBoost and the Catboost is having the lowest RMSE when compared to other models. Explained Variance Score shows a value of 0.99 for the XGBoost Regression. However, all model except KNN depicts an unprecedented precision rate aligning to 0.99.

Overall, the XGBoost outperforms all other models in a good course of action. This regression model endeavor to

acquire high accuracy to the train and test score. Furthermore, it brought down the error to a negligible rate. Therefore, XGBoost Regression can be regarded as the pre-eminent model. It is also evident that all other models except KNN tried to yield the best result. When compared to other models, KNN succumbed to underperform. Figure 4 compares the explained variance regression score of different models.

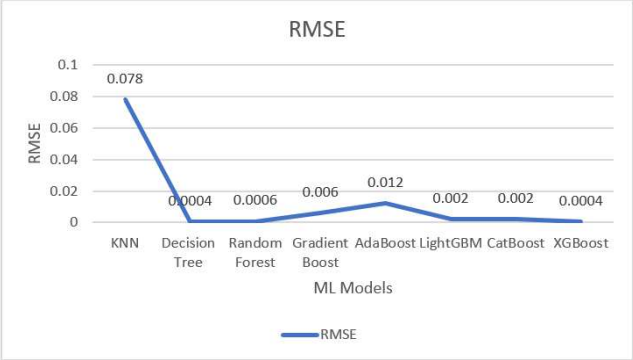

Fig. 3. Comparing RMSE Value of different Models

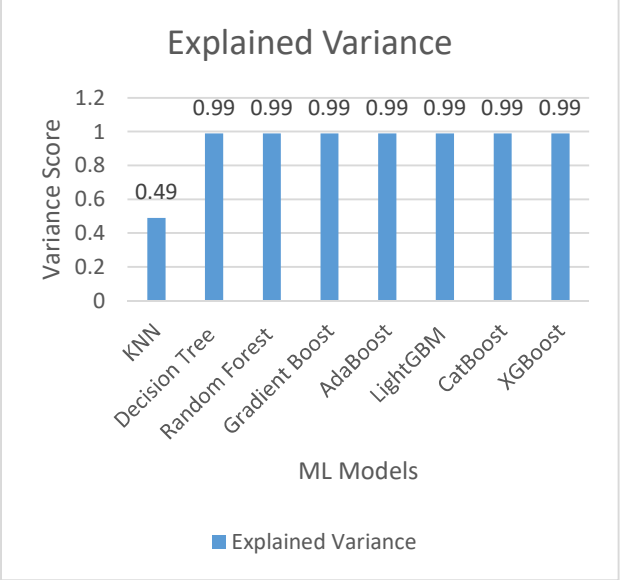

Fig. 4. Comparing RMSE Value of different Models

TABLE IV. COMPARISON OF EXISTING AND PROPOSED METHODOLOGY

Existing Study	Method	Performance Analysis
Panjwani et al. [7]	Random Forest Classifier	83.33% -Accuracy
Odegua [14]	Random Forest Algorithm	MAE-0.409178
Elias et al. [8]	Random Forest Algorithm	94%-Accuracy
Proposed Study	XGBoost Regression	RMSE-0.0004 MAE-2.492e-05

Table 4 shows the comparison between the existing and proposed studies. Panjwani et al., in their research, analysed the sales trend forecast of Bigmart using a machine learning algorithm. Random Forest Classifier, Linear Regression and Decision Tree Classifier were used for this purpose. The data was divided into 70 percent and 50 percent for training and testing. Overall the system produces an accuracy of 83 percent for Random Forest Classifier. Odegua employed a machine learning technique to forecast sales of a supermarket chain store. K-Nearest Neighbor, Random forest and Gradient Boosting were employed in his study. Random Forest Algorithm outperformed all other models producing an MAE of 0.4. Elias et al. forecasted Wallmart Sales using three classification models. Performance analysis was performed using the R^2 score and Mean Absolute Error (MAE) metrics. Random Forest Algorithm obtained 94 percent accuracy.

In our research, the XGBoost regression Regression model beat other models by delivering high train and test scores while lowering the RMSE value to 0.0004 and MAE of 2.492e-05

VI. CONCLUSION AND FUTURE SCOPE

Sales forecasting plays an important role in commercial world. It aids in anticipating the revenue and profit of an organization. It is an excellent approach to increase income and boost company growth. In this study, we developed a methodology to examine sales prediction. We proposed an innovative approach using a machine learning algorithm to prevent overstocking and lower product wastage. For this purpose, eight different algorithms were trained and employed for prediction. It includes KNN, Decision Tree Regression, Random Forest Regression, and ensemble boosting methods such as XGBoost, Light Gradient Boost, AdaBoost, Categorical Boost. The proposed method considers the performance of each regressive model over time and computes the model efficiency using different evaluation metrics. It was observed that XGBoost Regression excellently surpasses all other models. XGBoost Regression has been shown to procure perfect accuracy outrunning all other algorithms. It was evident that, except for KNN, all other algorithms attempted to achieve the best outcome.

For future enhancement, to mitigate the trade loss, inconsistency, and periodic tilt, more targeted and efficient tactics must be implemented to maximize profit and remain competitive. For the prediction, we intend to use deep learning techniques. Highly efficient and advanced neural network approaches could be implemented to examine the forecasting performance. Further evaluations can be performed in the future, and the bestselling products and price optimization strategy will be suggested.

REFERENCES

[1] Hendra, E. S. Rini, P. Ginting and B. K. F. Sembiring, "Impact of eCommerce service quality, recovery service quality, and satisfaction in Indonesia," 2017 International Conference on Sustainable Information Engineering and Technology (SIET), 2017, pp. 35-40, doi: 10.1109/SIET.2017.8304105.

[2] C. Zhan, C. K. Tse, Y. Gao and T. Hao, "Comparative Study of COVID-19 Pandemic Progressions in 175 Regions in Australia, Canada, Italy, Japan, Spain, U.K. and USA Using a Novel Model That Considers Testing Capacity and Deficiency in Confirming Infected Cases," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 8, pp. 2836-2847, Aug. 2021, doi: 10.1109/JBHI.2021.3089577.

[3] K. Rebane, M. Teichmann and K. Rannat, "Dynamics of the Public Satisfaction with Situation Management During COVID-19 Pandemic: Developments from March 2020 to January 2022," 2022 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA), 2022, pp. 112-114, doi: 10.1109/CogSIMA54611.2022.9830670.

[4] Q. Shen, "A machine learning approach to predict the result of League of Legends," 2022 International Conference on Machine Learning and Knowledge Engineering (MLKE), 2022, pp. 38-45, doi: 10.1109/MLKE55170.2022.00013.

[5] Li, Y., Pan, Y. A novel ensemble deep learning model for stock prediction based on stock prices and news. Int J Data Sci Anal 13, 139–149 (2022).

[6] H. Zhu, "A Deep Learning Based Hybrid Model for Sales Prediction of E-commerce with Sentiment Analysis," 2021 2nd International Conference on Computing and Data Science (CDS), Stanford, CA, USA, 2021, pp. 493-497, doi: 10.1109/CDS52072.2021.00091.

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

[8] N. Elias and S. Singh, "Forecasting of Walmart sales using machine learning algorithms," Research paper, Dept. of Electronics & Comm. Engineering, BMS Inst. of Technology & Management, Bangalore, India, 2018.

[9] Madhuvanthi, K. et al. (2019) 'Car sales prediction using machine learning algorithmns', International Journal of Innovative Technology and Exploring Engineering, 8(5), pp. 1039–1050.

[10] 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.

[11] M. Korolev and K. Ruegg, "Gradient boosted trees to predict store sales," Personal communication, 2015.

[12] Tableau.com.Online: https://www.tableau.com/sites/default/files/training/global_superstore.zip. (accessed Aug. 29, 2022).

[13] Galton, F. (1886). "Regression towards mediocrity in hereditary stature". The Journal of the Anthropological Institute of Great Britain and Ireland. 15: 246–263.

[14] R. Odegua, "Applied Machine Learning for Supermarket Sales Prediction," Project: Predictive Machine Learning in Industry, 2020.

[15] Fix, Evelyn; Hodges, Joseph L. (1951). Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties (PDF) (Report). USAF School of Aviation Medicine, Randolph Field, Texas.

[16] Y. Li, J. Shi, F. Cao and A. Cui, "Product Reviews Analysis of E-commerce Platform Based on Logistic-ARMA Model," 2021 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS), 2021, pp. 714-717, doi: 10.1109/ICPICS52425.2021.9524238.

[17] Quinlan, J. R. (1986). Induction of decision trees. Machine learning, 1(1):81–106

[18] Christoph Reinders, Bodo Rosenhahn, Learning convolutional neural networks for object detection with very little training data, (2019)

[19] Breiman, L. (2001). Random forests. Machine learning, 45(1):5–32.

[20] Friedman, J. H. (2001). Greedy function approximation: a Gradient boosting machine. Annals of statistics, pages 1189–1232.