Journal of Ecohumanism
2024
Volume: 3, No: 8, pp. 14457 – 14467
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i8.6762

A Comparative Review of Machine Learning Algorithms in Supermarket Sales Forecasting with Big Data

Sharmin Sultana[1], Farhana Karim[2], Habiba Rahman[3], Md Asikur Rahman Chy[4], Mukther Uddin[5], Md Nazibullah Khan[6], Md Ekrim Hossin[7], Evha Rozario[8]

Abstract

Consumers now have more options and comparison shopping when making purchases thanks to the new retail industry's quick expansion. This has increased rivalry among supermarket chains and continuously compressed profit margins. This research uses big data analytics to compare several ML approaches for forecasting grocery sales. The purpose of the research is to assess how well different ML algorithms predict sales. This paper uses supermarket sales data as the database, uses Python to conduct data preprocessing. By analyzing this data, including identifying anomalies and trends through data mining techniques, develop predictive models using advanced machine learning algorithms including DT, XGBoost, GB, and RF are employed to forecast sales volumes with greater accuracy than traditional methods across MAE, and R2-score. In evaluating model performance, the Extra Trees (ET) model exhibits superior accuracy with an R^2 of 0.94 and MAE of 1.96, compared to the existing models. Addressing challenges such as data sparsity, variability, and adapting to dynamic market conditions will be crucial. Additionally, future research could investigate novel methods for combining diverse forecasting approaches and refining models to better handle the complexities of real-world sales data.

Keywords: *Supermarket, sales prediction, business, data analysis, machine learning, extra tree, data mining.*

Introduction

Today, a lot of shopping malls, supermarkets, and retail stores, among others, maintain records of data on the sales of goods and commodities with different attributes, qualities, and dependent or independent features, as well as data about customers and assets [1]. Supermarkets, which provide convenience and a one-stop shop for a wide variety of goods, have become an essential element of contemporary living [2]. Supermarkets must now comprehend customer behavior and purchase habits in order to be competitive due to the growing level of competition in the business [3] [4]. Data analytics help supermarkets analyze consumer data and forecast sales trends, which in turn improves operations and allows them to make data-driven decisions [5]. Supermarkets may get valuable insights into customer behavior and preferences, inventory levels, and top-selling and underperforming items with an employ of data analytics. An analysis of consumer data allows supermarkets to develop segmented, personalized marketing strategies, which in turn boosts sales by building customer loyalty.

[1] School of Business, International American University, Los Angeles, CA 90010, USA. Email: sharminanis369@gmail.com. ORCID ID: https://orcid.org/0009-0005-7213-4504

[2] Harrison College of Business and Computing, Southeast Missouri State University, Cape Girardeau, Missouri 63701, USA. Email: fkarim2s@semo.edu. ORCID ID: https://orcid.org/0009-0001-4480-9907

[3] School of Business, International American University, Los Angeles, CA 90010, USA. Email: habiba.rahman1993@gmail.com. ORCID ID: https://orcid.org/0009-0009-8101-479X

[4] School of Business, International American University, Los Angeles, CA 90010, USA. Email: mdasikurrahmanchy21@gmail.com. ORCID ID: https://orcid.org/0009-0002-6550-7104 (Corresponding Author)

[5] School of Business, International American University, Los Angeles, CA 90010, USA. Email: muktheruddin1996@gmail.com. ORCID ID: https://orcid.org/0009-0009-0995-8761

[6] School of Business, International American University, Los Angeles, CA 90010, USA. Email: khanroyal2014@gmail.com. ORCID ID: https://orcid.org/0009-0008-4132-1413

[7] School of Business, International American University, Los Angeles, CA 90010, USA. Email: mdekrim@gmail.com. ORCID ID: https://orcid.org/0009-0005-1175-9076

[8] School of Business, International American University, Los Angeles, CA 90010, USA. Email: evhaaccabd@gmail.com. ORCID ID: https://orcid.org/0009-0006-3077-2011

Journal of Ecohumanism
2024
Volume: 3, No: 8, pp. 14457 – 14467
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i8.6762

The availability of massive volumes of data from a variety of sources, like as social media, consumer loyalty programs, and point-of-sale systems, has led to a rise in popularity of big data analytics in recent years. Supermarkets may use big data analytics to sift through this mountain of information and learn more about their customers' habits, tastes, and buying patterns. Supermarkets may get important insights into customer behavior, sales trends, and patterns by applying ML algorithms to the collected data and building predictive models.

Any company involved in retail, shipping, manufacturing, marketing, or wholesale must have a reliable sales forecast. It helps companies plan for the future of the organization, allocate resources better, and anticipate sales income [6][7]. The consumer data that Big Mart may have gathered from its dealings with other supermarkets is one of its most valuable assets. The data contains significant patterns and variables that may be analyzed using an ML algorithm. This allows for a very accurate forecast of sales [8][9]. ML models are taught using data to find repeating patterns, allowing them to make exact predictions about future occurrences. Machine learning models have emerged as indispensable tools in grocery sales analysis. An appealing option for sales trend modelling, it is both simple and easy to understand. Modern sales forecasting methods are required due to the complexities of customer behavior and the fast growth of supermarkets. This study aims to enhance decision-making in retail by comparing the effectiveness of various machine learning algorithms on big data. By identifying the most accurate models, the research seeks to optimize inventory management, reduce costs, and increase profitability in the competitive retail landscape.

Contribution and Aim of Paper

The primary contribution of this study is a comparative analysis of various ML models on sales dataset. Key contributions include:

- Conduct preprocessing for handling missing values, outlier detection, one-hot encoding, and feature scaling, to ensure data quality.
- Focused on selecting relevant features, improving model accuracy, and reducing the risk of overfitting.
- Evaluated the performance of RF, DT, GB, and XGBoost models in predicting supermarket sales, with MAE and R2 score.

Structure of Paper

This is the outline for the remaining sections of the paper. A review of the literature on grocery store sales forecasting is included in Section 2. The approach is described in depth in Section 3. Section 4 compares a findings, analysis, and discussion. The findings and future directions of the study are detailed in Section 5.

Literature Review

Some studies have attempted to model sales forecasting systems, but there is still a lot that hasn't been studied. Listed below are some works of literature that cover sales forecasting systems:

In [10], Three ML algorithms—XGBoost, lightGBM, and RF—are utilized to construct a forecasting model and extract data from 4 distinct feature sets. A three models' average score predictions for 'Sales' are distinct from one another. Out of the four feature set data types, Random Forest shows the best prediction effect with an Average Score81.25%, surpassing both XGBoost and lightGBM models. The "sub_cate_all" featureset data also yields an Average Score value higher than the other three, further supporting RF's superiority.

Volume: 3, No: 8, pp. 14457 – 14467
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i8.6762

In [11], examines the viability of several models for forecasting Walmart sales, including classic time series models, hybrid models combining time series and ML, and a model that relies only on ML. Walmart uses the Prophet model, which breaks out sales data by trend, season, and holiday, and the lightGBM model, a ML model, to train and evaluate its predictions. ML models seem to do a good job at predicting retail store sales, with RMSE0.694 for the Prophet model and 0.617 for a LightLGB model, according to a findings.

In [12], suggests a model for forecasting sales at Walmart using the XGBoost algorithm in conjunction with careful feature engineering processing. The outcomes of an experiment show that our method outperforms the other machine learning techniques. The RMSSE measure used in this study is 0.141 times lower than that of the LR method and 0.113 times lower than that of the Ridge algorithm.

Forecasting using machine learning techniques that are known to be effective are being used[13], specifically XGBoost, random forests, and linear regression. Crucially, the models outperform the existing Nordestão technique by a large margin and demonstrate effective performance across several testing weeks. In addition to the typical value of sales forecasting, Nordestão uses our work to assess how the COVID-19 pandemics have affected sales.

In[14], a Grid Search Optimization (GSO) strategy is proposed to choose the optimal tuning hyperparameters and optimise the parameters. They then combine this method with Xgboost techniques to anticipate the future sales of a retailer company, like Big Mart, and find that our model yields superior results.

For retailers like Big Mart[15], to anticipate future sales volume through the application of several machine learning techniques. In order to forecast the sales of a company like Big-Mart, a predictive model was constructed utilising Xgboost, Linear regression, Polynomial regression, and Ridge regression techniques. It was found that the model performs better than previous models.

Table 1 presents a comparative study of supermarket sales forecasting using various ML approaches. Where multiple study compares behalf dataset, approaches, performance, and limitation

Table 1: Comparative Study on supermarket sales prediction using machine learning approaches

Ref	Methodology	Dataset	Performance	Limitations & Future Work
[13]	Linear Regression, Random Forests, XGBoost	Nordestão sales data	Effective performance across multiple testing weeks, significant improvement over current approach	Evaluate impact of COVID-19 on sales
[14]	Grid Search Optimization (GSO) with XGBoost	Big Mart sales data	Better results compared to traditional models	Further optimization and tuning of parameters needed
16]	XGBoost, Linear Regression, Polynomial Regression, Ridge Regression	Big Mart sales data	Model outperforms existing models	Explore additional features and techniques for further improvement
(2021) [10]	Integrated Learning Algorithms (XGBoost, LightGBM, RF)	Global supermarket marketing data (4 years)	Highest score from Random Forest: 81.25%	Further evaluation needed across various types of feature sets and models.
[11]	Hybrid Models (Prophet for time series + LightGBM)	Walmart sales data (2011-2016)	RMSE: Prophet: 0.694; LightGBM: 0.617	Exploring additional hybrid models and features could be beneficial.

Journal of Ecohumanism
2024
Volume: 3, No: 8, pp. 14457 – 14467
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i8.6762

[12]	XGBoost Model with Meticulous Feature Engineering	Walmart sales data (Kaggle competition)	RMSSE: 0.141 (vs. Logistic Regression); 0.113 (vs. Ridge)	Potential for further feature engineering and analysis of feature importance.

Research Gaps

There has been progress in sales forecasting models in individual studies, but there is still a lack of study on how to combine different information and approaches to make stronger predictions. The generalizability of standard statistical models and machine learning algorithms is questionable because they are frequently dependent on specific datasets or feature sets, even though they have demonstrated encouraging outcomes. Possible directions for further study include integrating time series analysis with advance ML methods to create hybrid models that make the most of different perspectives. Sales forecasting models may be even more accurate and useful if they used real-time data and investigated domain-specific characteristics. Dealing with data sparsity, variability, and the ever-changing market conditions would be part of this process.

methodology

This study's overarching goal is to compare and contrast the ML algorithms used for big data sales forecasting in supermarkets. This study's methodology included several organized procedures for predicting sales at supermarkets using machine learning models. To begin, the supermarket sales datasets. The dataset had 8,523 observations. The data flows in various steps and phases that are displayed in the figure 1 data flow diagram. As part of the data preparation process, dealt with missing values, identified outliers, applied one-hot encoding to categorical data, and normalized the data using the min-max scaler before scaling features. The purpose of feature selection is to remove extraneous elements from the model in order to boost its performance. What followed was a division of the dataset into training and testing sets using an 80/20 rule. To make predictions, several machine learning models were utilized, such as Extra tree, DT, and RF. R^2 and MSE were among the measures utilized to evaluate a models' accuracy and predictive potential.

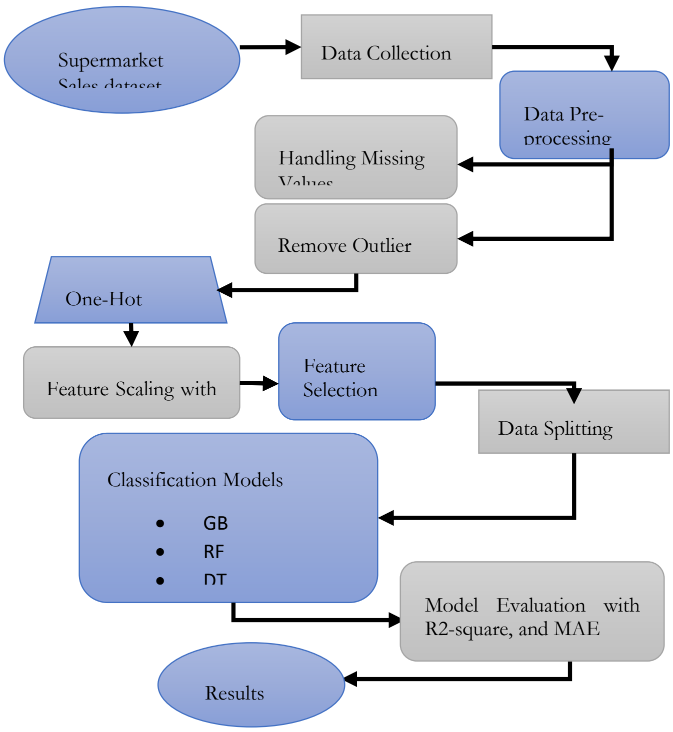

Figure 1: Data flow diagram for sales prediction

Journal of Ecohumanism
2024
Volume: 3, No: 8, pp. 14457 – 14467
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i8.6762

The whole steps of the data flow diagram in detailed explanation are provided below:

Data Collection

A sales data utilized in this research's experiments comes from a worldwide brand BigMart supermarkets. There are 12 characteristics (columns) and 8523 observations (rows) in the dataset. Every one of these observations pertains to a distinct product or object with its own unique identifier. A value is assigned to each item for each of the twelve variables.

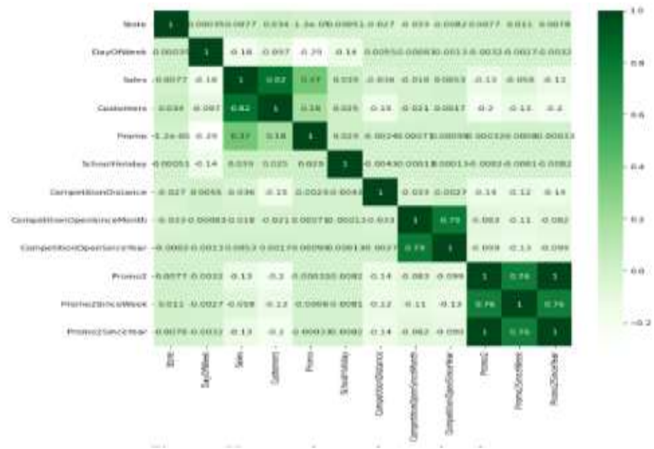

Figure 2: Heat map for correlation of attribute

The following figure 2 shows the correlation of attributes. The heat map makes use of varying shades of green to depict the relationships between attributes; a darker shade of green signifies a higher positive link. Perfect correlations are shown by the diagonal, and the map is useful for rapidly determining the strength of associations in a dataset.

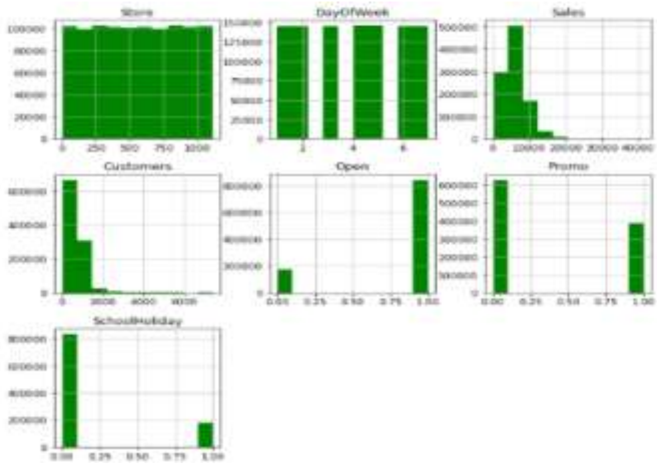

Figure 3: Histogram of attribute

Figure 3 histogram illustrate various dataset attributes: "Store" and "Sales" show diverse values, "Day of Week" has seven bars, "Customers" follows the "Sales" pattern, and "Open" and "School Holiday" depict binary data, highlighting store status and holiday frequency.

Journal of Ecohumanism
2024
Volume: 3, No: 8, pp. 14457 – 14467
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i8.6762

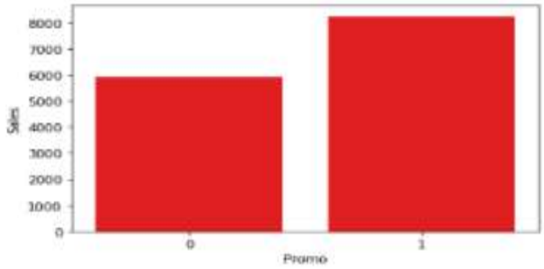

Figure 4: Analyzing the Sales variable

The above figure 4 shows that sales are higher with promotions ('Promo' value 1) compared to without them ('Promo' value 0), highlighting the effectiveness of promotions in boosting sales.

Data Preprocessing

A clean dataset may be created by transforming raw data via data pre-processing. The data is not obtained in a processed form, which makes it unreasonable for examination, when it is collected from numerous sources. The steps used to prepare the dataset for input into the algorithm are known as preprocessing. The following preprocessing methods are described in below:

- **handling missing values:** There are several ways to address missing values whether to use sophisticated imputation techniques, exclude cases with incomplete data, or replace missing values with zeros or the mean [16].

- **Outlier detection:** The occurrence of outliers, or data points that differ greatly from the rest of the dataset, is usually the result of inaccurate data input, incorrect observations, or very extreme data points.

One-hot Encoding

When working with categorical data, one-hot encoding is a typical approach in ML. Unfortunately, machine learning models only work with numerical input variables, so you'll have to modify any categorical variables during pre-processing. In categorical data, you may find nominal or ordinal values [17]. This method takes the original category column and inserts a new column for every single value.

Features Scaling with minmax

Data features or independent variables may have their range normalized via feature scaling. According to Equation 1, the Min-max Scaler transforms an attribute's scale by shifting its values along the X-axis, bringing the transformed attribute within the range of [0, 1].

$$x_i' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \dots \dots (1)$$

The range of the attribute is used as a scaling factor in this approach, while a lowest value of a property is used as the translational term.

Feature Selection

When building a machine learning model, feature selection is practically the first stage. The need of this procedure arises from the fact that data sets might be somewhat vast and likely include elements that are not essential to the target variable's prediction process; including these features into the set of predictors could adversely impact the model's performance [18].

Journal of Ecohumanism
2024
Volume: 3, No: 8, pp. 14457 – 14467
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i8.6762

Data Splitting

Dividing the original data set into three smaller sets—trained, validation, and testing datasets—before beginning to create any models is a standard procedure in ML. The data was broken up as follows: 20% went towards testing sets and 80% went towards training sets.

ML Regressor Models

For the supermarket sales prediction, select various machine learning models including RF, DT, GB, and xgboost that are explained in below:

Random forest (RF) Regressor

Classification and regression problems are well-suited to the ensemble learning approach known as Random Forest. It takes a portion of the input data and features at random, uses that data to train several decision trees, and then averages or votes on the best predictions. Its high dimensionality tolerance, resistance to overfitting, and prevalence in image classification, anomaly detection, and consumer segmentation make it a popular choice. All things considered, it is an effective and flexible method that yields reliable findings and comprehensible outputs [19].

Extra Tree (ET) Regressor

An improved method that originated with the RF model is the ETR technique. With the traditional top-down approach in mind, the ETC algorithm builds a set of unpruned choices or regression trees. Bootstrapping and bagging are the two procedures used by the RF model to carry out the regression, respectively. With the use of a randomly selected subset of the training dataset, a collection of DT is generated throughout the bootstrapping process. Choosing the optimal subset and assigning a value to it completes the decision-making process.

A uniform independent distribution vector θ is allocated before the formation of the Gth forecasting tree in the RF model, which is a series of decision trees, and the Gth tree is represented as G(x, θr). The trees of G(x) that make up an ensemble, or forest, are averaged and combined to get Equation (2):

$$G(x, \theta_1, \dots . \theta_r) = \frac{1}{R}\sum_{r=1}^{R} G(x, \theta_r) \dots \dots (2)$$

The ETR and RF systems differ from one another in two important ways. The ETR first separates nodes by randomly selecting a subset of all the cutting points. Secondly, to reduce bias, it cultivates the trees using all of the learning samples. The ETR method's splitting procedure is controlled by two parameters, k and nmin. k is the number of characteristics randomly chosen within a node, and nmin is the minimum sample size needed to identify and separate nodes. Furthermore, the attribute selection intensity and the average output noise strength are determined by k and nmin, respectively. The ETR model's accuracy is increased and overfitting is decreased by these two parameters [20][21].

Decision tree (DT) Regressor

The term "decision tree" describes a ML technique for classification and regression. By repeatedly splitting the data into subsets based on chosen characteristics, it constructs a decision model resembling a tree. This process continues until a stopping condition is met. A possible outcome is represented by each branch, and a feature-based decision is shown by each node. Decision trees are a simple tool for handling and interpreting numerical and category data.

Journal of Ecohumanism
2024
Volume: 3, No: 8, pp. 14457 – 14467
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i8.6762

Evaluation metrics

To evaluate a quality of ML models, use an evaluation error matrix. Metrics like as R2- square, and need to be measured in order for us to be able to compare different algorithms.

R2-score: An improved fit between the model and the data is shown by higher R2 values, which range from 0 to 1. A model is said to be completely predictive when its R2 value is 1; when it is 0, no variability in the response data around its mean is explained by the model. R2 is computed using the following formula: (3)

$$R^2 = 1 - \frac{SSR}{SST} \dots \dots . (3)$$

Where,

SSR (Sum of Squared Residuals) is the total squared discrepancies between the actual and projected values.

SST (Total Sum of Squares) is a total of a squared deviations among a target variable's mean and an actual value.

Mean square error (MSE): To calculate the difference among a regression problem's actual and projected values, one often used metric is the MSE [24]. The mean squared deviation between the actual and expected values is used to calculate this statistic. Having a lower MSE value indicates better performance. Applying the above formula yields the MSE (4):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \dots \dots (4)$$

Where,

Yi is the actual value,

Yi is the predicted value, ˆ

n is the number of observations

Results and Discussion

This section serves as a summary of the research's results and a graphic representation of its progress. Different ML algorithms have been utilized to forecast a sales of various retail business datasets, and several evaluation techniques including MAE, and R2 are computed and assessed.

Journal of Ecohumanism
2024
Volume: 3, No: 8, pp. 14457 – 14467
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i8.6762

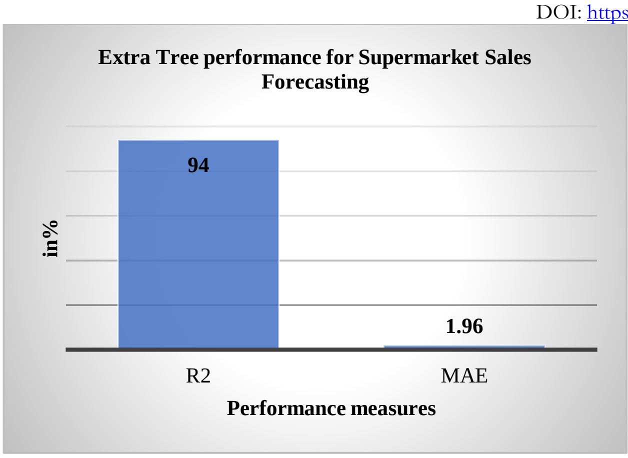

Figure 5: Bar graph of Extra Tree Regressor performance for Supermarket Sales Forecasting

The performance of the Extra Tree (ET) classifier in Supermarket Sales Forecasting is depicted in Figure 5 as a bar graph. The model achieved an R^2 score0.94, indicating a strong correlation among a forecasted and actual sales values, showcasing a model's effectiveness in capturing the data's variance. Additionally, the model's Mean Absolute Error (MAE) is reported as 1.96, suggesting a low average error in the predictions. This combination of high R^2 and low MAE demonstrates the Extra Tree classifier's robustness and accuracy in forecasting sales, making it a reliable choice for such predictive tasks.

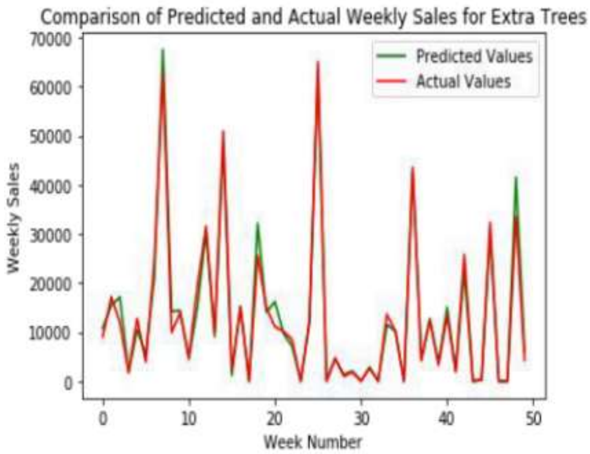

Figure 6: Actual and predicted performance of ET

The graph in Figure 6 shows a comparison between the actual and predicted weekly sales for the ET model over a period of 50 weeks. Actual sales figures are shown by the red line, while forecasted values are shown by the blue line. Both lines show fluctuations, with some weeks having significant peaks and drops in Supermarket Sales Forecasting.

Table 2: Comparative analysis for model performance with performance measures

Models	**R2**	**MAE**
ET	**0.94**	**1.96**
RF[22]	0.71	13.5
DT[23]	0.31	23.3

Journal of Ecohumanism
2024
Volume: 3, No: 8, pp. 14457 – 14467
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i8.6762

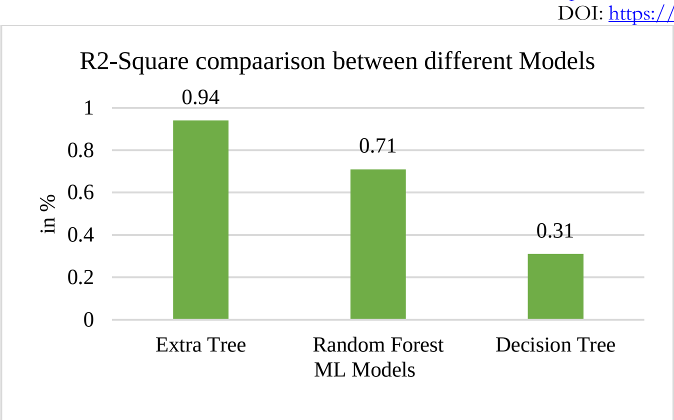

Figure 5: Bar Graph for comparison of R2-score for Supermarket Sales Forecasting

The following figure 5 shows the R2 comparison of different models for supermarket sales forecasting. In comparing the R^2 scores, XGBoost achieved the highest with 0.94, indicating superior predictive accuracy. RF had a moderate R^2 of 0.71, while the DT t R^2 of 0.31, reflecting the least accuracy among the models for supermarket sales forecasting.

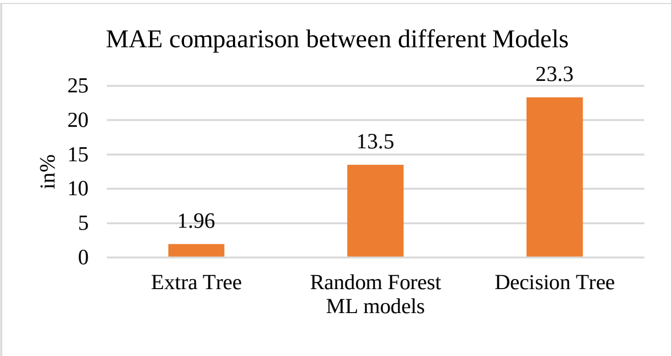

Figure 6: Bar Graph for comparison of MAE for supermarket sales forecasting

The following figure 6 shows the MAE comparison of different models. In comparison, the ET model is the most accurate with an MAE of 1.96, followed by RF at 13.5. The DT has a higher MAE of 23.3 for supermarket sales forecasting. The Extra tree model is perform well in compare to other ML models for forcasting.

Conclusion and Future Study

Supermarkets are a particularly saturated industry now due to their fast expansion. The supermarket is experiencing severe sales issues as a result of its fast expansion. Consequently, grocery shops and other retail establishments need a reliable sales forecast system in order to circumvent this issue. Applying EDA and feature engineering techniques, this article examined the sales data set from supermarkets. Three distinct methods have been utilised: RF, DT, and Extra Tree. The regressor approaches' efficacy has been assessed using mean square error and R2-score. According to the results of this research, the Extra tree may assist management in retaining sales by predicting them with an R2-score of 0.94%. The outcome proves that, according to all performance assessment parameters, the Extra tree is the best classifier for this dataset. Future research in this area should primarily focus on the ethical ramifications of gathering and analysing consumer data, the effectiveness of different ML-DL, hybrid algorithms and data sources for sales forecasting, and the impact of different marketing methods on sales.

Journal of Ecohumanism
2024
Volume: 3, No: 8, pp. 14457 – 14467
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i8.6762

References

[1] B. Kumar Jha and S. Pande, "Time Series Forecasting Model for Supermarket Sales using FB-Prophet," in Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021, 2021. doi: 10.1109/ICCMC51019.2021.9418033.

[2] I. Theresa, V. R. Medikonda, and K. V. N. Reddy, "Prediction of big mart sales using exploratory machine learning techniques," Int. J. Adv. Sci. Technol., 2020.

[3] J. Tian, Y. Zhang, and C. Zhang, "Predicting consumer variety-seeking through weather data analytics," Electron. Commer. Res. Appl., 2018, doi: 10.1016/j.elerap.2018.02.001.

[4] L. R. Berry, P. Helman, and M. West, "Probabilistic forecasting of heterogeneous consumer transaction–sales time series," Int. J. Forecast., 2020, doi: 10.1016/j.ijforecast.2019.07.007.

[5] J. Thomas, "Enhancing Supply Chain Resilience Through Cloud-Based SCM and Advanced Machine Learning: A Case Study of Logistics," J. Emerg. Technol. Innov. Res., vol. 8, no. 9, 2021.

[6] K. K. San, "Performance Analysis of Regression Models Using Myanmar Sales Data," Int. Res. J. Mod. Eng. Technol. Sci., 2020.

[7] R. K. Vinita Rohilla, Sudeshna Chakraborty, "Car Auomation Simulator Using Machine Learning," Proc. Int. Conf. Innov. Comput. Commun., 2020, [Online]. Available: https://scholar.google.com/citations?view_op=view_citation&hl=en&user=zlcFgwEAAAAJ&citation_for_view=zlcFgwEAAAAJ:2osOgNQ5qMEC

[8] J. W. Taylor, "Forecasting daily supermarket sales using exponentially weighted quantile regression," Eur. J. Oper. Res., 2007, doi: 10.1016/j.ejor.2006.02.006.

[9] S. Raizada and J. R. Saini, "Comparative Analysis of Supervised Machine Learning Techniques for Sales Forecasting," Int. J. Adv. Comput. Sci. Appl., 2021, doi: 10.14569/IJACSA.2021.0121112.

[10] Z. Mei and L. Mingjie, "Research on Supermarket Marketing Data Analysis Based on Business Intelligence," in Proceedings - 2023 6th International Conference on Information and Computer Technologies, ICICT 2023, 2023. doi: 10.1109/ICICT58900.2023.00011.

[11] H. Jiang, J. Ruan, and J. Sun, "Application of Machine Learning Model and Hybrid Model in Retail Sales Forecast," in 2021 IEEE 6th International Conference on Big Data Analytics, ICBDA 2021, 2021. doi: 10.1109/ICBDA51983.2021.9403224.

[12] Y. Niu, "Walmart Sales Forecasting using XGBoost algorithm and Feature engineering," in Proceedings - 2020 International Conference on Big Data and Artificial Intelligence and Software Engineering, ICBASE 2020, 2020. doi: 10.1109/ICBASE51474.2020.00103.

[13] F. M. De Almeida, A. M. Martins, M. A. Nunes, and L. C. T. Bezerra, "Retail sales forecasting for a Brazilian supermarket chain: An empirical assessment," in Proceedings - 2022 IEEE 24th Conference on Business Informatics, CBI 2022, 2022. doi: 10.1109/CBI54897.2022.00014.

[14] G. Behera and N. Nain, "Grid search optimization (GSO) based future sales prediction for big mart," in Proceedings - 15th International Conference on Signal Image Technology and Internet Based Systems, SISITS 2019, 2019. doi: 10.1109/SITIS.2019.00038.

[15] R. P and S. M, "Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms," in 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1416–1421. doi: 10.1109/ICICCS51141.2021.9432109.

[16] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, and O. Tabona, "A survey on missing data in machine learning," J. Big Data, 2021, doi: 10.1186/s40537-021-00516-9.

[17] S. Okada, M. Ohzeki, and S. Taguchi, "Efficient partition of integer optimization problems with one-hot encoding," Sci. Rep., 2019, doi: 10.1038/s41598-019-49539-6.

[18] V. Rohilla, S. Chakraborty, and R. Kumar, "Deep learning based feature extraction and a bidirectional hybrid optimized model for location based advertising," Multimed. Tools Appl., 2022, doi: 10.1007/s11042-022-12457-3.

[19] V. Rohilla, D. S. Chakraborty, and D. R. kumar, "Random Forest with Harmony Search Optimization for Location Based Advertising," Int. J. Innov. Technol. Explor. Eng., 2019, doi: 10.35940/ijitee.i7761.078919.

[20] V. John, Z. Liu, C. Guo, S. Mita, and K. Kidono, "Real-time lane estimation Using Deep features and extra trees regression," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2016. doi: 10.1007/978-3-319-29451-3_57.

[21] G. Mishra, D. Sehgal, and J. K. Valadi, "Quantitative Structure Activity Relationship study of the Anti-Hepatitis Peptides employing Random Forest and Extra Tree regressors," Bioinformation, 2017, doi: 10.6026/97320630013060.

[22] M. . Abdullahi, G. I. . Aimufua, and U. A. Muhammad, "Application of Sales Forecasting Model Based on Machine Learning Algorithms.," 2021. doi: 10.22624/aims/isteams-2021/v28p17.

[23] J. Sekban, "Applying Machine Learning Algorithms in SALES PREDICTION," 2019.