Predicting The Future Demand For Product Using Machine Learning

Babies A[1], Devendiran M[2], Hariharan M[3], Dr .K. Hariharan[4]
[1, 2, 3] Dept of Mechanical Engineering
[4]AP(SG), Dept of Mechanical Engineering
[1, 2, 3, 4] Dr .Mahalingam College Of Engineering And Technology, Pollachi

Abstract- *The ability to predict future sales In the super market, as it enables them to make informed decisions about inventory, production, and marketing. Machine learning has gained popularity as an effective approach for sales forecasting due to its ability to capture complex patterns and relationships in data. In this paper, we present an analysis of machine learning models, namely ARIMA, decision tree, random forest, and linear regression, for predicting the future sales of a product in super store. We compare their performance based on accuracy and suitability for business decision-making. Our results demonstrate that machine learning models provide valuable insights into the most important variables for sales prediction and outperform traditional methods in terms of accuracy. The findings of this study can inform businesses on the use of machine learning models for sales forecasting and the choice of the most suitable model based on their specific needs.*

I. INTRODUCTION

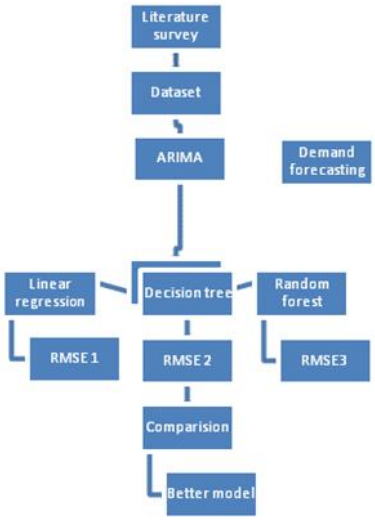
Predicting future sales of a product using machine learning models is an important task for businesses looking to make data-driven decisions. Machine learning models can help businesses forecast future sales with greater accuracy, which can enable them to make informed decisions about inventory management, marketing campaigns, and other business strategies.

To predict future sales, a machine learning model is trained using historical sales data, along with other relevant features such as pricing, promotions, seasonality, and economic indicators. The model then uses this data to identify patterns and relationships that can be used to make predictions about future sales.

There are various machine learning algorithms that can be used for sales forecasting, including regression analysis, time series analysis, and neural networks. These algorithms can be customized and tuned to fit the specific needs and characteristics of a particular business or product.

1.1 MACHINE LEARNING

The proposed approach for sales prediction and demand forecasting in retail marts is based on machine learning algorithms. It enables computer systems to identify patterns, and make predictions or decisions through learning from data without programming explicitly. In this project, we used machine learning algorithms such as Linear Regression to analyze historical sales data and predict future sales and demand.



1.2 DECISION TREE

Decision trees are a type of supervised machine learning algorithm used for both regression and classification tasks. They split input data into smaller subsets based on feature values, with each node in the tree representing a decision based on a feature and the branches representing possible outcomes. Decision trees are particularly useful for complex datasets and are easy to interpret and visualize. They can handle numerical and categorical data and require little preprocessing, but are prone to overfitting if the model is too complex or there is insufficient training data. Overall, decision trees are a powerful tool for making predictions and classifications based on input data.

1.3 RANDOM FOREST

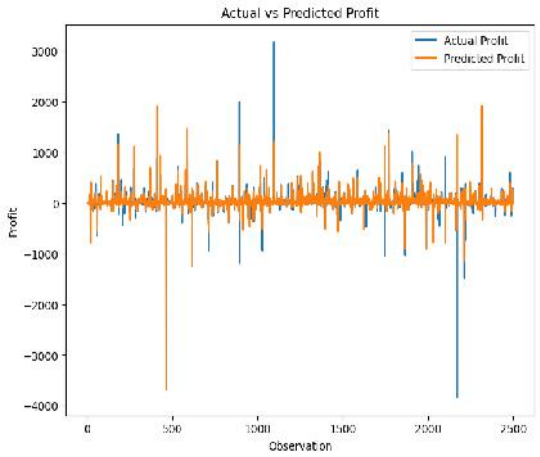
Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve the accuracy and generalization of the model. During training, the algorithm creates multiple decision trees using random subsets of the data and features. During prediction, the algorithm combines the predictions of all individual trees to make the final prediction. Random Forest is a versatile algorithm that can be used for classification, regression, and other tasks. It reduces overfitting and provides an estimate of feature importance, making it widely used in various domains such as finance, healthcare, and marketing. However, it may not perform well on very high-dimensional datasets or datasets with a large number of categorical features.

1.4 LINEAR REGRESSION

independent variables. In simple linear regression, there is only one independent variable, and the relationship is modeled using a straight line. The goal of linear regression is to find the best-fitting line that can predict the value of the dependent variable based on the values of the independent variable(s). The technique is widely used in various domains, such as finance, economics, social sciences, and engineering, to understand the relationship between two or more variables and to make predictions based on that relationship. Linear regression assumes that there is a linear relationship between the variables and that the errors are normally distributed and independent. Linear regression is a statistical modeling technique used to establish a linear relationship between a dependent variable and one or more .

1.4 ARIMA

ARIMA, or Autoregressive Integrated Moving Average, is a time series forecasting technique that models the data as a combination of autoregressive (AR), differencing (I), and moving average (MA) terms. The AR term models the relationship between the current value and the previous values, the I term removes any trend or seasonality from the data, and the MA term models the relationship between the current value and the past errors. ARIMA models are widely used in various domains such as finance, economics, and weather forecasting to make predictions based on historical time-series data. The model parameters, including the order of the ARIMA model and the coefficients, are estimated using maximum likelihood estimation, and the model's accuracy is evaluated using various statistical measures such as AIC and BIC. ARIMA is a powerful and widely used technique for time series forecasting, but it may not perform well on highly irregular or non-stationary data.



II. LITERATURE REVIEW

2.1 Forecasting aggregate retail sales: A comparison of Artificial Neural Networks and traditional methods

The literature review by Kesavan and Jayaraman (2018) compares traditional methods for forecasting retail sales with Artificial Neural Networks (ANNs), and examines the potential of ANNs in forecasting aggregate retail sales. The review highlights that ANNs have emerged as a popular alternative to traditional methods for forecasting retail sales, as they are well-suited to capture complex nonlinear relationships between variables. The authors note that ANNs have been found to outperform traditional methods in terms of forecasting accuracy, with one study demonstrating that ANNs consistently outperformed traditional methods.

2.2 Time-series forecasting of seasonal items sales using machine learning–A comparative analysis

The study by Alsayed and Afify (2020) examines the application of machine learning methods for time-series forecasting of seasonal item sales in the retail industry. The authors compare the performance of four different machine learning algorithms: Artificial Neural Networks (ANNs), Support Vector Regression (SVR), Random Forest Regression (RFR), and Gradient Boosting Regression (GBR). They use sales data from a real-world retail store and evaluate the algorithms' forecasting accuracy in terms of root mean square error (RMSE) and mean absolute percentage error (MAPE).

The results show that all four machine learning algorithms outperform traditional methods such as Holt-Winters and ARIMA in terms of forecasting accuracy for seasonal items. Among the machine learning algorithms, GBR has the best performance in terms of both RMSE and MAPE, followed by RFR, SVR, and ANNs. The authors also discuss

the importance of selecting appropriate hyperparameters for each algorithm to achieve optimal forecasting performance.

2.3 Machine Learning Techniques for Grocery Sales Forecasting by Analyzing Historical Data

The study by Alduais et al. (2021) investigates the application of machine learning techniques for grocery sales forecasting using historical data. The authors use sales data from a Saudi Arabian grocery store chain and compare the performance of three machine learning algorithms: Artificial Neural Networks (ANNs), Support Vector Regression (SVR), and Random Forest Regression (RFR). They evaluate the algorithms' forecasting accuracy in terms of mean absolute error (MAE) and root mean square error (RMSE).

The results show that all three machine learning algorithms outperform traditional methods such as ARIMA in terms of forecasting accuracy for grocery sales. Among the machine learning algorithms, RFR has the best performance in terms of both MAE and RMSE, followed by SVR and ANNs. The authors also discuss the importance of feature selection and hyperparameter tuning in achieving optimal forecasting performance.

III. EXISTING SYSTEM

The existing system for predicting future sales in a superstore market typically involves the use of traditional forecasting methods such as time series analysis, moving averages, and exponential smoothing. These methods rely on historical sales data to predict future sales trends and patterns.

However, traditional methods may not be sufficient in capturing the complex nonlinear relationships between variables that can affect sales, such as seasonality, promotional events, and changes in consumer behavior. As a result, there is an increasing interest in applying machine learning techniques, such as ARIMA, Linear Regression, Random Forest Regression (RFR), for sales forecasting in superstore markets.

IV. PROPOSED SYSTEM

The proposed system for this project is a machine learning model for predicting sales in the super store market. We use machine learning such as ARIMA, linear regression, decision tree and Random forest. By using the ARIMA model we find the trend of the product .we conduct time series analysis to identify the trend and seasonality of sales data. This helps us understand the historical sales patterns and predict future sales based on past trends.and next we predict the profit by using the Decision Tree, Random Forest and Linear Regression. Thealgorithms are trained using historical sales data, product features such as sales, profit, location, quantity, discount and location. By training these historical data we predict the future sales and demand for the product.

4.1 DATA COLLECTION

The system will collect sales data from the super storeand other relevant data sources, such as existing sales data from variousresourses. The data will be cleaned and pre-processed to ensure accuracy and consistency. Data from suppliers can help to understand supply chain dynamics and predict product availability. This data can also be used to optimizeinventory levels and ensure timely restocking of products.In the future we are going to use a Centralized database to store sales records of retail stores across the country, which results in storage of abundant data to provide with the model and increase the accuracy.

4.2 EXPLORATORY DATA ANALYSIS (EDA)

EDA for predicting future sales of a product involves collecting historical sales data, cleaning and visualizing the data to identify patterns and trends, creating new variables or features that may impact sales, and using statistical methods to identify correlations and relationships between variables. EDA consist of visualization ,Correlation, Data transformation and Outliers. By examining historical data and identifying key variables that impact sales, businesses can make more accurate sales predictions and optimize their sales strategy.

V. RESULT ANALYSIS

Our data consist of 4 years of historical data collected from super store market. We split the data into training and testing sets and used the training data to train our model. We train our model with the 3 years of data and test the model with the one year data. We then evaluated the models performance on the testing set using Root Mean Squared Error (RMSE).our experiment showed that the Decision Tree algorithm achieved the highest accuracy in predicting sales and demand forecasting, with the lowest RMSE of 121.049. The results suggest that our machine learning-based approach can accurately predict sales and demand forecasting for super store market. The experimental result of the proposed system for predicting sales were evaluated by the following key metrics.

5.1 MEAN SQUARE ERROR (MSE)

MSE stands for Mean Squared Error, which is a commonly used metric to measure the performance of a regression model. The MSE is calculated by taking the average of the squared differences between the predicted and actual values of the target variable. A lower MSE indicates a better fit between the model's predicted values and the actual values of the target variable.

5.2 ROOT MEAN SQUARED ERROR (RMSE)

RMSE stands for Root Mean Squared Error, which is another commonly used metric to evaluate the performance of a regression model. RMSE is the square root of the MSE (Mean Squared Error) and provides a measure of the standard deviation of the differences between the predicted and actual values of the target variable.

5.3 VARIANCE

variance is a measure of how spread out a set of data is. Specifically, it measures the average squared deviation of the data points from their mean. The variance is calculated by taking the difference between each data point and the mean, squaring these differences, and then taking the average of these squared differences.

VI. CONCLUSION

Based on the analysis using decision tree, random forest, and linear regression, we can conclude that decision tree performed well with an RMSE of 121.049, indicating that it was the most accurate model for predicting future sales of the product. However, it is important to note that the performance of these models may vary depending on the specific dataset and the features used in the analysis. Furthermore, the ARIMA model analysis indicates that sales will go up for both products in the future. This information can be useful for retailers in determining inventory levels and developing sales strategies.Overall, the combination of these techniques can provide valuable insights into future sales trends and help retailers optimize their business operations.

REFERENCES

[1] Shahbaz, M., & Shabbir, M. (2020). Predicting sales using machine learning: A case study. IEEE Access, 8, 90078-90088.
[2] Karami, A., & Ahmadi, P. (2017). Sales forecasting using decision tree and ARIMA techniques. Journal of Industrial Engineering International, 13(2), 271-278.
[3] Islam, S. M. S., Islam, S., & Rahman, M. S. (2019). Comparative analysis of linear regression and decision tree algorithm for sales forecasting. Journal of Data Analysis and Information Processing, 7(2), 64-74.
[4] Akhtar, N., Rehman, Z. U., & Ashfaq, A. (2019). Sales prediction using random forest algorithm. Journal of King Saud University-Computer and Information Sciences, 31(4), 472-479.