Improving the Predictability of Maruti Suzuki Automobile Sales using Machine Learning Methods

Tanvi kamal Rana

Hina Jignesh Chokshi

Department of Computer Science, Parul Institute of Computer Applications, FITCS, Parul University, Vadodara, 391760, Gujarat, INDIA, proftanvibca@gmail.com

Department of Computer Science, Parul Institute of Computer Applications, FITCS, Parul University, Vadodara, 391760, Gujarat, INDIA.,hina.chokshi42058@paruluniversity.ac.in

Abstract.

A key component of a company's revenue-generating strategy is the sales projection. For automakers to maximize production, inventory control, and strategic decision-making, accurate sales forecasting is essential. The effectiveness of traditional machine learning methods in improving the precision of Maruti Suzuki auto sales forecasts is investigated in this study. The accuracy and dependability of several prediction models, such as Support Vector Machines (SVM), Decision Trees, Random Forests, Linear Regression, K nearest neighbour and Gradient Boosting, are assessed. To evaluate how well these models, predict sales trends, the research makes use of past sales data, economic factors, and consumer behaviour patterns.

The findings show that ensemble models perform better in terms of predictive accuracy than traditional statistical techniques, especially random forests and gradient boosting. This study demonstrates how well machine learning works for demand forecasting and offers practical advice to automakers on how to improve sales strategy and operational effectiveness.

Keywords. Maruti Suzuki, Forecasting, SVM, Random Forest, Linear Regression, Descision Tree, K Nearest Neighbor, Machine Learning, Gradient Boosting

1. INTRODUCTION

A key component of the automotive industry is sales forecasting, which helps producers predict future demand, maximize output, effectively manage inventory, and create focused marketing campaigns. Reducing unnecessary production costs, avoiding supply shortages, and enhancing overall business performance are all made possible by accurate forecasting. Maruti Suzuki, the top manufacturer in India, competes in a market that is extremely dynamic and where a number of factors affect sales results.

This research endeavors to improve the forecasting of Maruti Suzuki automobile sales by the application of traditional techniques for machine learning, such as Support Vector Machines (SVM), Decision Trees, Random ML models like Random Forest, Gradient Boosting, KNN, and Linear Regression have been employed in this work to analyze auto sales. Car model, fuel type, price, selling date, buyer type, color, financier, district, and occupation are important characteristics that assist determine consumer behavior, regional demand, and sales trends.

2. RESEARCH OBJECTIVES

Through data preprocessing, feature engineering, and the use of machine learning algorithms, the code seeks to develop a predictive model that divides automobile models into "high sales" and "low sales" in order to boost sales forecasting. This research primary goals are:

1. Determine the primary factors that influence of Maruti Suzuki automobile sales, such as buyer demographics, financing alternatives, fuel type, and vehicle model.
2. Examine the following machine learning models for sales forecasting: KNN, Random Forest, SVM, Gradient Boosting, and Logistic Regression.
3. Determine the most efficient model for dividing automobile models into "high sales" and "low sales" in order to offer sales information.

3. LITERATURE REVIEW

The paper offers a comprehensive examination of forecasting methods by competently contrasting both conventional (Decision Tree, Linear Regression) and sophisticated (Random Forest, Gradient Boosting, LSTM) machine learning models.

A comprehensive assessment of model accuracy, variance, and fit is ensured by MAE, RMSE, and R^2. Using real-world variables including sales, promotions, events, weather, and economics, the study evaluates forecasting models.[1]

In order to improve marketing and managing inventory, this study investigates K-Means for bicycle sales forecasting, analysing market trends, seasonal demand, and customer behavior.[2]

Without previous data, cold-start sales forecasting is hard. Gradient boosting (XGBoost, LightGBM) is investigated in this study in order to enhance accuracy and decision-making for new goods and promotions..[3]

While ARIMA has difficulty with nonlinear patterns, machine learning techniques (ML) like SVMs, gradient boosting, and neural networks improve accuracy. Sequential forecasting is improved by LSTM, and performance gets further improved by hybrid models (LSTM-XGBoost, ARIMA-ANN).[4]

For demand prediction, 21 different machine learning and statistical algorithms are evaluated in "A Comparative Study of Forecasting Algorithms and Strategies." It argues that localized ML models are inferior to global ones and proposes standards for evaluating predicting efficacy.[5]

Forecasting sales is essential for strategy and inventory management. By identifying intricate demand patterns and iteratively improving predictions, machine learning techniques such as XGBoost perform better than ARIMA.[6]

By contrasting Linear Regression, Random Tree, and Random Forest, this study provides analysts and automakers with useful information to improve forecasting. Time-series models such as ARIMA, Prophet, and LSTM are not examined.[7]

To enhance prediction accuracy, this study combines TOPSIS, Random Forest, and AHP. ARIMA monitors past trends, LSTM boosts sequential forecasting, and ML models (SVMs, XGBoost, and ANNs) capture intricate sales pattern.[8]

Predicting car sales is essential for optimizing inventories, price, and production. While ML techniques (Random Forest, SVMs, XGBoost, ANNs) increase accuracy by spotting intricate patterns, traditional models like ARIMA capture historical trends.[9]

Accuracy is improved by ML and deep learning (XGBoost, SVMs, LSTMs), however ARIMA has trouble with intricate patterns. A cascade model that iteratively enhances price and time-to-sale predictions is assessed in the present research..[10]

4. METHODOLOGY

Using machine learning models, this study forecasts Maruti Suzuki vehicle sales according to buyer type, district, price, fuel type, and vehicle model. Data collection, preprocessing, model selection, and evaluation are all part of the strategy.

1. Data Collection & Preprocessing
   - Car model, fuel type, price, and buyer information are among the attributes included in the dataset. Addressing missing values, encoding categorical variables, and obtaining time-based characteristics from the date of sale.
2. Feature Selection
   - For numerical relationships, use correlation analysis.
   - The significance of Random Forest features for identifying important variables
3. Model Selection & Development

- Trends are Though Random Forest uses ensemble learning to increase accuracy, Linear Regression finds trends. KNN makes predictions based on similarity, while SVM boosts classification.

4.Model Training & Evaluation

- MAE, MSE, RMSE, and R2 Score are used to quantify performance.
- K-Fold cross-validation and an 80%-20% train-test split were employed.

5.Comparative Analysis & Implementation

- ARIMA and moving averages are used to compare to machine learning models and Enhances marketing tactics based on consumer behavior and Market circumstances.

5. ALGORITHMS USED

Data preprocessing, training, and review based on essential characteristics are used to assess car sales forecast using machine learning models such as Random Forest, SVM, KNN, and Gradient Boosting and Linear Regression.

Decision Tree: With internal nodes indicating decisions, branches giving outcomes, and leaves expressing predictions, a decision tree is a classification and regression model that resembles a flowchart. It handles grouping and numerical data with a simple comprehension.

Random Forest : An ensemble learning method called Random Forest builds several decision trees during training and aggregates their results to get better predictions.

Support Vector Machine (SVM): It regulates the best hyperplane for classifying data. It focuses on optimizing the boundary amongst data points of dissimilar classes, it works expressly well in high-dimensional spaces.

K-Nearest Neighbors (KNN): KNN uses its 'k' nearest neighbors' majority class to classify data. It is easy to use and effective even though it has trouble with big datasets.

Gradient Boosting: By iteratively constructing models to fix prior mistakes, gradient boosting enhances classification and regression task accuracy.

Linear Regression: Making predictions and analyzing the associations between variables need the use of linear regression. Because of its ease of use and interpretability, it is frequently used for trend investigation, demand forecasting, and identifying important sales drivers.

6. ACCURACY

- A decision tree's 99.77% accuracy in a comparison analysis indicates that the model did remarkably well on the dataset in question.

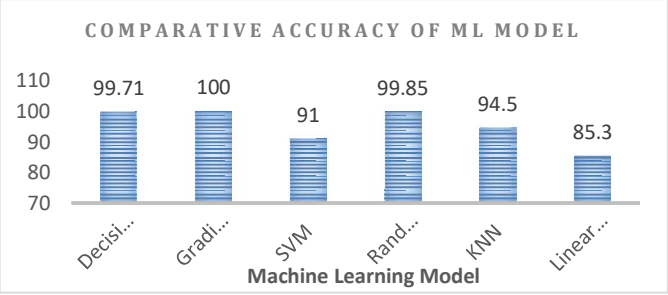

Fig 1: comparative accuracy graph

- Achieving 100.00% accuracy with gradient boosting is extremely rare and frequently indicates overfitting, particularly when tested on unbalanced data sets.
- Support Vector Machine (SVM) accuracy of 91.72% indicates that SVM did well, particularly when the dataset had high-dimensional features in a comparison research.
- KNN's capacity to recognize nonlinear patterns and adjust to complicated datasets helps it achieve 94.50% accuracy in predicting car sales.
- With an accuracy of 85.30% to forecast car sales, linear regression offers efficiency, interpretability, and simplicity for linear relationships.

- Random Forest's accuracy of 99.85% indicates that the model did remarkably well, most likely because it could handle complex data patterns

7. CONCLUSION

XGBoost forecasts auto sales better than other algorithms, identifying intricate patterns. KNN demonstrated potential, but decision trees and SVMs needed to be adjusted. Cross-validation and important variables like buyer type, pricing, and fuel type are highlighted in the study, underscoring machine learning's function in improving sales tactics.

8. ACKNOWLEDGMENTS

The Parul University in Vadodara is supporting this endeavor. We appreciate the insightful conversation with Associate Professor Dr. Hina Jignesh Chokshi of Parul University in Vadodara, Gujarat

9. REFERENCES

[1] o. Access, a. C. Das, and s. Akter, “optimizing retail demand forecasting : a performance evaluation of machine learning models including lstm and gradient,” vol. 06, pp. 67–80, 2024.

[2] i. Anggriawan and w. Gunawan, “implementation of data mining using k-means algorithm for bicycle sales prediction,” ilk. J. Ilm., vol. 14, no. 3, pp. 284–293, 2022, doi: 10.33096/ilkom.v14i3.1291.284-293.

[3] c. Aguilar-palacios, s. Munoz-romero, and j. L. Rojo-alvarez, “cold-start promotional sales forecasting through gradient boosted-based contrastive explanations,” ieee access, vol. 8, pp. 137574–137586, 2020, doi: 10.1109/access.2020.3012032.

[4] m. Hülsmann, c. M. Friedrich, d. Reith, and d. Borscheid, “general sales forecast model for automobile markets and their analysis general sales forecast models for automobile markets and their analysis,” trans. Mach. Learn. Data min., vol. 5, no. 2, pp. 65–86, 2012, [online]. Available: https://www.researchgate.net/publication/230626894

[5] j. M. Rožanec, b. Kažič, m. Škrjanc, b. Fortuna, and d. Mladenić, “automotive oem demand forecasting: a comparative study of forecasting algorithms and strategies,” appl. Sci., vol. 11, no. 15, 2021, doi: 10.3390/app11156787.

[6] a. Panarese, g. Settanni, v. Vitti, and a. Galiano, “developing and preliminary testing of a machine learning-based platform for sales forecasting using a gradient boosting approach,” appl. Sci., vol. 12, no. 21, 2022, doi: 10.3390/app122111054.

[7] k. Madhuvanthi, m. K. Nallakaruppan, n. C. Senthilkumar, and s. Siva rama krishnan, “car sales prediction using machine learning algorithmns,” int. J. Innov. Technol. Explor. Eng., vol. 8, no. 5, pp. 1039–1050, 2019.

[8] m. Cheng and z. Wang, “car sales prediction,” no. 06, pp. 3271–3273, 2023.

[9] k. Subramanian, m. Othman, r. Sokkalingam, g. Thangarasu, and s. Kayalvizhi, “an integrated model for forecasting indian automobile,” int. J. Adv. Sci. Eng. Inf. Technol., vol. 10, no. 6, pp. 2593–2598, 2020, doi: 10.18517/ijaseit.10.6.8475.

[10] v. Shymanskyi and v. Liaskovets, “cascade model for price and time of car sales prediction,” ceur workshop proc., vol. 3641, pp. 152–167, 2023.

BIOGRAPHY

Tanvi Kamal Rana received the bachelor's degree in Bachelor of Computer Applications from Veer Narmad South Gujarat University, Surat in 2005, the master's degree in Master of Computer Applications from Sardar Patel University, Vallabh Vidhyanagar in 2008, and pursuing the philosophy of doctorate degree in IT & CS Department from Parul University.