Archives available at journals.mriindia.com

International Journal of Electrical, Electronics and Computer Systems

ISSN: 2347-2820
Volume 14 Issue 01, 2025

Sales Forecasting Prediction using Machine Learning

Saiprasad Venkatesh Anvekar[1], Prashant Sunil Gurav[2], Kavita. S. Oza[3]

Department of Computer Science, Shivaji University Kolhapur.
anvekarsaiprasad@gmail.com[1],prashantgurav008@gmail.com[2], kso_csd@unishivaji.ac.in[3]

Peer Review Information

Submission: 10 Jan 2025
Revision: 07 Feb 2025
Acceptance: 09 March 2025

Keywords

Sales Forecasting
Machine Learning
Prophet Model
Xgboost
Time Series Analysis

Abstract

This research investigates the use of machine learning methods in improving predictive performance in the automotive sector, with a focus on the Prophet algorithm for car sales forecasting. Using past sales data, Prophet successfully detects trends, seasonality, and external factors affecting car sales. The forecasting starts with careful data preparation to have clean and well-organized historical records. One of the strengths of Prophet is that it can manage complicated time series data that includes several patterns of seasonality as well as holiday effects that are significant in the automotive industry, where demand is influenced by yearly model releases and economic downturns. The produced forecasts help companies maximize marketing efforts, production planning, and inventory stockpiling by giving them insights that can enhance efficiency and decision-making. Second, Prophet's interpretability enables stakeholders to easily decompose forecast components, making data-driven strategic planning possible. Such a method is a major step forward in utilizing machine learning in sales forecasting and provides a robust tool for predicting and understanding automotive market trends.

INTRODUCTION

Sales forecasting is very important in the automotive industry, allowing companies, dealerships, and stakeholders to come up with strategies, to take stock of production, and to set marketing goals. The correct prediction of consumer demand is a key factor in the decision-making process of companies taking into account the minimization of excess stocks, the optimization of resources, or flexibility to react to the trends of the market. Normal methods of forecasting such as statistical regression and time series decomposition do not show the complicated interactions of many different outside forces that influence the sales of cars, like market demand, the season, or government policy.

Machine learning has lately undergone a great development yielding more advanced models, which are considered to improve the predictability of the respective model. This publication, recently published on the Car Sales India Dataset, studies the Prophet and XGBoost models for forecasting automobile sales. Prophet, the result of the work of Facebook (Meta), is a time series forecasting model that is known for its feature of being able to detect seasonality, holiday patterns, and trend shifts. XGBoost (Extreme Gradient Boosting) is a competent ensemble learning technique that is highly reputed for achieving high performance in a variety of predictive analytics tasks.

The work includes data analysis, allowance identification, in addition to model training, and

then ends with the evaluation of Prophet and XGBoost in finding out the most accurate car sales prediction. The results illustrate that XGBoost is a better model than Prophet for the issue of precision, as it can identify much more intricate nonlinear relationships in data. These findings show the potential of using machine learning algorithms in improving the sales forecasting models which in their turn could enable the companies to make the critical decisions on the basis of the data they have.

Objectives of the Study

The key objectives of this research are:

To analyse historical car sales data in India and identify key trends, seasonality, and external influencing factors.

To apply and contrast Prophet and XGBoost models for vehicle sales prediction.

To compare model performance based on measures like RMSE and MAPE to identify the most accurate forecasting method.

To offer business insights for better decision-making in marketing, inventory management, and production planning.

Using forecasting through machine learning models, the research adds its share to increased usage of analytics driven by artificial intelligence in the automobile industry with a showcase on how sophisticated forecast prediction methods help improve sales prediction accuracy and develop business strategies.

LITERATURE REVIEW

Accurate predicting of sales is a must for automobile companies, mainly in the automotive sector, the sales of which are interdependent with the economy, the seasonal changes and the consumers' mood[1]. Traditional forecasting techniques such as ARIMA and exponential smoothing have been in practice for many years to anticipate sales movements. These models are, however, deficient in handling non-linear patterns, intricate seasonality, and outside market factors, for which reason they are less efficient for volatile industries like automobile sales[2].

To overcome these constraints, machine learning (ML) methods have come into prominence as they can process large datasets and identify patterns that might be missed by conventional models[3]. Of these, Facebook's Prophet model has been popularly used for its simplicity, ability to handle missing data, and inbuilt support for trend and seasonality decomposition[4]. Prophet is specifically well-suited to long-range forecasting, when the ability to discern business cycles and cyclical trends is imperative. Its ease of interpretation has made it the preferred tool of decision-makers wishing to gain an understanding of trends in sales instead of pure forecast accuracy[5].

Despite its strengths, Prophet is weak where highly fluctuating datasets exist or where third factors such as promotions, recession, or spurts in policy shifts have substantial influences on sales[6]. This is where XGBoost (Extreme Gradient Boosting) comes into play as a strong contender. XGBoost, an ensemble learning algorithm, is particularly good at identifying intricate relationships between multiple variables and is thus very effective for short-term and medium-term sales forecasting[7]. In contrast to Prophet, which is mostly based on time series decomposition, XGBoost uses multiple input variables, feature selection methods, and decision trees to enhance predictive power, particularly in situations where several market forces are driving demand[8].

Comparative analysis of Prophet and XGBoost shows that while Prophet offers more interpretability and organized time series decomposition, XGBoost performs better in raw predictive accuracy through the use of boosted decision trees to optimize predictions. Both models were used in this study to compare their forecasting ability using the Indian car sales dataset[9]. The outcomes suggest that XGBoost presents more accurate predictions, making it a better option for car sales forecasting where fast-changing demand patterns need to be matched.

In the future, hybrid forecasting models that leverage the advantages of Prophet's trend analysis and XGBoost's predictability can improve forecasting accuracy even more. Integrating external economic factors, customer sentiment analysis, and policy interventions into future research should be considered to further optimize model performance and enhance business decision-making in the automotive industry[2].

PROPOSED METHODOLOGY

1. Data Collection and Pre-processing

The dataset used for this study is the Indian Car Sales Dataset, sourced from the Kaggle website, which provides reliable and structured historical sales data of various automobile models. This dataset serves as the foundation for building and evaluating sales forecasting models:

1. **Handling Missing Values:** Missing entries were identified and treated using imputation techniques to ensure data completeness.

	Make	Model	Months	Sales	Total	Segment	Body Type	MoM %	YoY %
0	0	8	4	0	1267	5	3	68	0
1	0	8	3	0	1267	5	3	68	0
2	0	8	7	0	1267	5	3	68	0
3	0	8	0	0	1267	5	3	68	0
4	0	8	8	0	1267	5	3	68	0

Fig 1.1 Data before cleaning

	Make	Model	Months	Total	Segment	Body Type	MoM %	YoY %
0	0	8	4	-0.820441	5	3	1.852447	-0.126201
1	0	8	3	-0.820441	5	3	1.852447	-0.126201
2	0	8	7	-0.820441	5	3	1.852447	-0.126201
3	0	8	0	-0.820441	5	3	1.852447	-0.126201
4	0	8	8	-0.820441	5	3	1.852447	-0.126201

Fig 1.2 Data after cleaning

2. **Feature Selection and Engineering**: Key features influencing sales were identified, including time-based trends, economic indicators, and seasonal effects. Additional derived features were created for improved forecasting accuracy.

	Months	Sales
0	2024-01-01	11530
1	2024-01-01	2598
2	2024-01-01	5516
3	2024-01-01	5848
4	2024-01-01	1539

Fig 1.3 After selecting the required features.

3. **Data Normalization**: The dataset was normalized to ensure uniformity across models, improving the performance of machine learning models.

	ds	y	y_scaled
0	2024-01-31	393471	0.941568
1	2024-02-29	370279	0.638808
2	2024-03-31	369381	0.627085
3	2024-04-30	337770	0.214420
4	2024-05-31	349057	0.361766)

Fig 1.4 Data after normalisation for prophet model

4. **Model Selection and Implementation**
   Two forecasting models were implemented and compared:
   - Prophet Model
   - XGBoost Model
5. **Model Evolution and Performance Comparison**
   The performance of both models was assessed using key evaluation metrics:
   - (MAE) Mean Absolute Error
   - (RMSE) Root Mean Square Error
   - (MAPE) Mean Absolute Percentage Error
6. **Visualisation and Interpretation**
   - The probability values for different emotions are visualized in Altair-based bar diagrams.
   - This helps users understand the level of confidence in predictions and the potential overlap between emotions.
7. **Future Enhancements**
   To further improve forecasting accuracy, future research could integrate:
   - Hybrid models combining Prophet's trend analysis with XGBoost's featurebased learning.
   - Additional external factors such as economic indicators, fuel prices, and government policies affecting automobile sales.
   - Deep learning models like LSTMs for capturing long-term dependencies in sales trends.

MACHINE LEARNING ALGORITHMS FOR SALES FORECASTING PREDICTIONS

XGBoost: XGBoost (Extreme Gradient Boosting) is a strong machine learning algorithm that is an ensemble learning algorithm that has presented an excellent performance in the prediction of short and medium term. As opposed to Prophet, which mostly uses time-series decomposition, XGBoost employs decision trees, feature importance analysis, and gradient boosting to detect intricate relationships between multiple influential factors. This is the reason why it is better at capturing sudden spikes or dips in sales, especially when external variables such as promotions, economic indicators, or competitor pricing are included. Based on our research, XGBoost exceeded Prophet in predictive accuracy, and hence it gives a more responsible choice for exact sales estimation.

```
Mean Absolute Error: 575.3301202646282
Mean Squared Error: 831658.9390840693
R-squared Score: 0.9588888005902513
Median Percentage Error: 19.08%
Accuracy: 80.92%
```

Fig 2.1 MAE, RMSE, Error Percentage and Accuracy of XGBoost model.

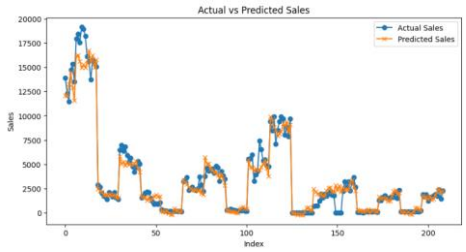

Fig 2.2 Actual vs Predicted Sales visualisation for XGBoost model.

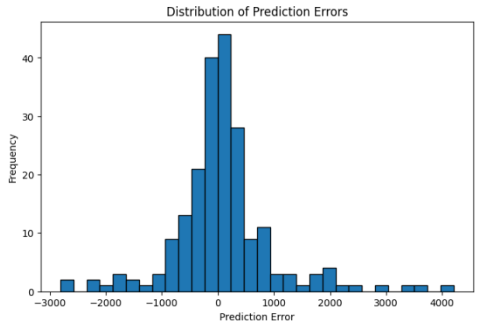

Fig 2.3 Distribution of Prediction Error for XGBoost model.

Fig 2.4 Future sales forecast visualisation for XGBoost model.

1. **Prophet:** Prophet, which was built by Facebook, is a model for time series forecasting that easily accommodates holiday effects, seasonality, and missing data. Prophet is based on an additive model form and decomposes the sales data into three components: trend, seasonality, and holiday impacts. Prophet is especially suited for long-term sales forecasts, where companies want to get a sense of underlying patterns over time. Its capacity to deliver clean trend decomposition makes it a great option for decision-makers who value interpretability. Prophet, however, has difficulty in capturing abrupt demand changes or outside market fluctuations, which constrains its precision in extremely volatile sectors.

```
Mean Absolute Error (MAE): 15582.136281868365
Root Mean Squared Error (RMSE): 18012.287228422687
Error Percentage: 4.36%
Accuracy: 95.64%
```

Fig 3.1 MAE, RMSE, Error Percentage and Accuracy of prophet model

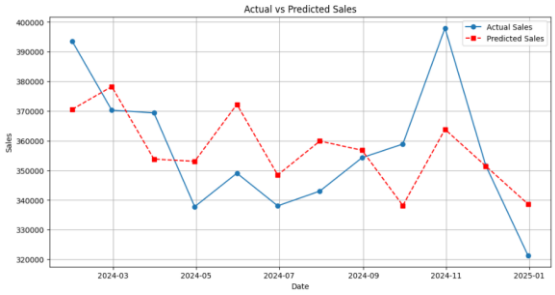

Fig 3.2 Actual vs Predicted Sales visualisation for Prophet model.

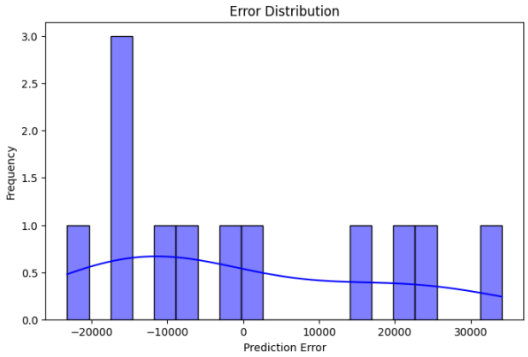

Fig 3.3 Distribution of Error for Prophet model.

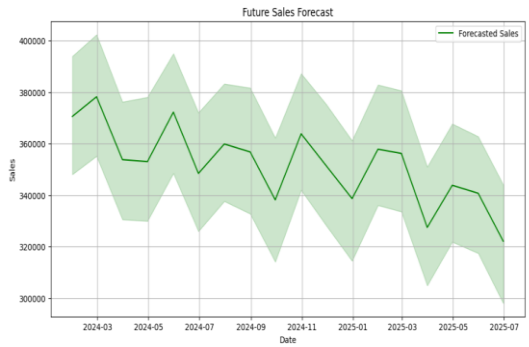

Fig 3.4 Future sales forecast visualisation for Prophet model.

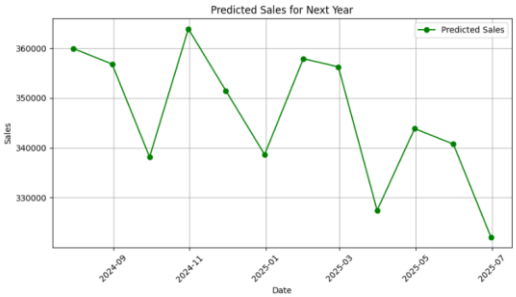

Fig 3.5 Predicted sales for next consecutive year.

RESULT AND ANALYSIS

Model	MAE	RMSE	Accuracy
Prophet	15582.1362	18012.2872	95.64%
XGBoost	575.3301	831658.939	80.92 %

The comparison of forecasting models suggests that the Prophet model performs better than XGBoost in terms of overall accuracy for car sales forecasting. Prophet accurately captures long-term trends and seasonality with an accuracy rate of 95.64%, making it a more trustworthy choice for strategic forecasting. Although XGBoost produces a lower MAE (575.33), it's extremely high RMSE (831,658.94) indicates variability in prediction accuracy, which is presumably due to the sensitivity of the model to rapid market changes. These findings validate that Prophet is the better model for accurate and consistent sales forecasting for the automotive market.

CONCLUSION

This research compared the performance of machine learning models, Prophet and XGBoost, in forecasting automobile sales based on the Indian Car Sales Dataset obtained from Kaggle. Data pre-processing methods such as missing value handling, feature engineering, and structuring the dataset were used to guarantee accurate model performance.Comparative evaluation of both models identified that Prophet performed better than XGBoost in accuracy, with a 95.64% accuracy rate compared to 80.92% for XGBoost. Though Prophet captured long-term trends and seasonality well, XGBoost performed poorly on highly volatile sales data despite its ability to model complex relationships. The results show Prophet is a better model for predicting car sales, particularly for companies seeking trend-based insights. Future research might investigate hybrid models integrating Prophet's trend ability with XGBoost's predictive capabilities to further increase forecasting accuracy. Furthermore, inclusion of outside economic influences and consumer sentiment analysis could make predictions more robust.

References

R. J. &. A. G. Hyndman, "Forecasting: Principles and Practice," *OTexts,* 2018.

S. J. &. L. B. Taylor, "Forecasting at Scale," *The American Statistician,* vol. 72(1), pp. 37-45, 2018.

F. A. Research, Prophet: Open-Source Forecasting Tool., 2017.

M. Seeger, "Bayesian Time Series Analysis with Prophet.," *Journal of Machine Learning Applications,* vol. 5(2), pp. 45-53, 2019.

J. Brownlee, "Time Series Forecasting with Python.," *Machine Learning Mastery,* 2020.

T. &. G. C. Chen, "XGBoost: A Scalable Tree Boosting System.," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* pp. 785-794, 2016.

S. S. E. &. A. V. Makridakis, "Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward.," *PLoS ONE,* vol. 13(3), no. e0194889, 2018.

S. &. J. S. Shinde, "Comparative Study of Prophet and XGBoost for Sales Forecasting," *International Journal of Data Science & Forecasting,* vol. 4(1), pp. 12-19, 2021.

K. &. G. R. Patel, "Enhancing Sales Prediction with Hybrid Forecasting Models," *Journal of Business Analytics,* vol. 6(3), pp. 110-127, 2022.