International Journal of Engineering Trends and Technology Volume 73 Issue 8, 101-111, August 2025
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I8P108

Review Article

Advancements in Sales Forecasting: A Critical Evaluation of Machine Learning Techniques and Approaches

Tanvi kamal Rana[1], Hina Jignesh Chokshi[2]

[1,2]*Department of Computer Science, Parul Institute of Computer Applications, FITCS, Parul University, Vadodara, Gujarat, India.*

[2]*Corresponding Author : hina.chokshi42058@paruluniversity.ac.in*

Received: 14 March 2025 Revised: 05 June 2025 Accepted: 22 July 2025 Published: 30 August 2025

Abstract - *Sales forecasting is crucial in the automotive sector, enabling organizations to optimize inventory, production, and financial planning. Traditional forecasting methods often fail to capture complex sales patterns, leading to inaccurate predictions. Machine learning techniques have suggestively enhanced forecasting accuracy using algorithms that discover hidden patterns. This work inspects key machine learning methods for sales forecasting, especially on models like XGBoost, Random Forest, Support Vector Machines, and deep learning. It discovers the role of feature engineering, data preprocessing, and model selection in enhancing predictive performance. Additionally, the research evaluates these models' flexibility and mathematical competence, especially concerning big automotive datasets. A relative study highlights their strengths and limits in terms of accuracy, interpretability, and real-world application. Regardless of their advantages, machine learning models face challenges such as overfitting, data scarcity, and dependence on external market issues like economic patterns and consumer behavior. This work confers strategies to progress prediction reliability, including hybrid modeling, ensemble learning, and macroeconomic factor integration. The growth of adaptive models that can expect in real time and make choices automatically should be the main objective of future research. This work thoroughly examines how refined machine learning techniques affect sales forecasting accuracy in the automotive sector.*

Keywords - *Automotive sector, Machine Learning, Predictive analytics, Sales forecasting, XGBoost.*

1. Introduction

Sales forecasting is an unstable basis of the automobile industry, allowing companies to enhance production planning, inventory management, and financial strategy. Precise forecasting allows firms to make more up-to-date decisions, cut operational costs, and react proactively to market shifts. For occurrence time series analysis and statistical models, traditional forecasting methods often fail to detect involved patterns in previous sales data, resulting in specious projections. The amplified accessibility of enormous datasets, coupled with developments in artificial intelligence, has paved the way for machine learning methods to increase forecasting accuracy. Machine learning methods, such as aspects that traditional models miss.

A key constituent of business decision-making, sales estimating has an impact on marketing policies, financial planning, manufacturing scheduling, inventory management, and more. Tumbling expenses, increasing success, and enhancing possessions are all made possible by accurate sales forecasts. Despite their extensive usage, traditional predicting measures-such as statistical models like moving averages, exponential smoothing, and ARIMA-regularly fail to identify the involved, nonlinear patterns that originate in actual sales data.

In view of their assistance, ML-based sales forecasting models challenge numerous complications, with overfitting, data sparsity, computational complexity, and reliance on external market issues. A well-organized forecasting model must carefully choose the proper features, deal with missing or noisy data, and select the best algorithm for the work. Furthermore, combining domain knowledge, financial metrics, and hybrid modelling practices can increase prediction consistency and interpretability.

Fresh inquiries have emphasized the efficiency of ML in sales forecasting, particularly in the marketing and automotive trades, where demand swings owing to a variety of external sources. Research demonstrates that machine learning models beat traditional forecasting practices in accuracy and efficiency, particularly when working with enormous datasets.

However, ML-based forecasting still faces difficulties such as overfitting, data quality concerns, and computational difficulty. A mixture of hybrid models, real-time data processing, and economic constituent analysis can all help increase prediction reliability [1].

Sales estimating has advanced meaningfully since the introduction of Machine Learning (ML), which uses data-driven methods to increase forecast correctness. ML models may take into reflexion seasonality, uncover unseen patterns, and assimilate a variety of manipulating elements, including market trends, patron behavior, and saleable data. Administered learning methods such as decision trees, regression algorithms, and deep learning constructions similar to Long Short-Term Memory (LSTM) and persistent neural networks (RNNs) have been thoroughly studied for sales forecasting. Furthermore, hybrid policies that combine ML and statistical methods have shown heartening outcomes in forecasting stability [2, 3].

This paper critically evaluates the advancements in sales forecasting by reviewing various ML techniques and their applications. It provides a proportional examination of dissimilar models, emphasizing their strengths, limitations, and practical suggestions. The study also discusses key challenges in implementing ML-based forecasting, including data quality, model interpretability, and computational complexity. Furthermore, emerging trends such as explainable AI (XAI), reinforcement learning, and big data integration are explored to provide insights into the future direction of sales forecasting [2, 4].

The accuracy and dependability of machine learning-based sales forecasting are impacted by a number of significant issues. Since missing, inconsistent, or incomplete sales data can lower model efficacy, data availability and quality are among the main problems. Furthermore, past sales data frequently overlooks outside factors like changes in the economy, rivalry, and new market trends. Feature selection and engineering can provide challenges because it takes domain expertise to identify pertinent variables like seasonality, marketing, and customer behavior. While adding external factors like macroeconomic indicators is still challenging, high-dimensional datasets might result in overfitting and higher processing expenses [5].

Since no single machine learning model is consistently effective across all datasets, model selection and performance variability present another major difficulty. While sophisticated models like LSTM and Boost necessitate considerable hyperparameter tuning to attain optimal results, traditional models like ARIMA perform well for short-term trends but struggle with nonlinear relationships. The challenges are exacerbated by computational complexity and scalability; deep learning models require significant processing power, which drives up the cost of real-time forecasting. Large datasets, therefore, require more storage and training time, necessitating cloud-based solutions and posing privacy and data security issues [6, 7].

Because cultured models function as "black boxes," confining decision-making transparency, ML-based sales forecasting presents several difficulties, including an absence of interpretability and explainability. Outside conflicts make forecasts more difficult, such as changing consumer behaviour and financial crises. Prejudices in training data can generate unfair results, while ethical and data confidentiality issues present regulatory obstacles. Furthermore, because most machine learning models rely on static data and need to be retrained frequently, actual forecasting is still challenging. Though their acceptance is still in its beginning, emerging methods like online learning and strengthening learning provide promising replacements. Deciding these issues is indispensable to improving the accuracy and usefulness of ML-based sales forecasting [7].

The enhancements in sales forecasting using Machine Learning (ML) styles are judgmentally inspected in this review paper, with an emphasis on how they affect decision-making and prediction accuracy. It starts by emphasizing the importance of sales forecasting and conflicting ML-based techniques with conventional statistical practices. Along with the implication of feature engineering and outside variables like seasonality and economic trends, a variation of machine learning methods-including supervised, unsupervised, and hybrid models-is observed. The advantages and disadvantages of ML models are strongminded by comparing them using performance measures. It includes a detailed discussion of the main obstacles, including problems with data quality, model interpretability, computing complexity, ethical considerations, and real-time forecasting. In order to improve forecasting accuracy, the study also looks at potential future possibilities, such as big data integration, reinforcement learning, and Explainable AI (XAI). This paper thoroughly explains ML-driven sales forecasting, its difficulties, and new potential before concluding with observations and suggestions [8].

2. Theoretical Background

In business strategy, sales forecasting is essential because it helps firms make well-informed choices about financial planning, resource allocation, and inventory control. In contrast to contemporary Machine Learning (ML) techniques, which capture intricate, nonlinear interactions, traditional forecasting methods rely on statistical models that presume historical trends will persist in the future.

2.1. Traditional Sales Forecasting Methods

Because of their ease of use and interpretability, traditional statistical approaches have been widely employed for sales forecasting. Finding patterns in past sales data and extrapolating trends are the main goals of these models.

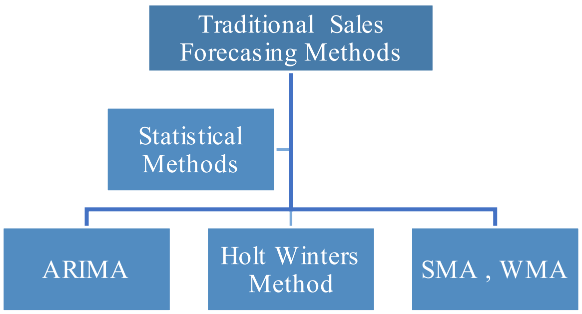

Fig. 1 Statistical methods for sales forecasting

Statistical techniques are frequently employed for time-series forecasting, especially in sales prediction. These techniques forecast future trends by using patterns in historical data. Above Figure 1 shows three primary statistical methods: Moving Averages (SMA, WMA), Exponential Smoothing (Holt-Winters Method), and Autoregressive Integrated Moving Average (ARIMA).

2.1.1. Autoregressive Integrated Moving Average (ARIMA)

ARIMA is a time-series forecasting model that captures linear trends and seasonality. It consists of Autoregression (AR) for predicting future values, Differencing (I) to remove trends, and Moving Average (MA) to correct forecast errors. While effective for short- to medium-term forecasting, ARIMA struggles with nonlinear patterns and volatile data.

2.1.2. Exponential Smoothing (Holt-Winters Method)

The Holt-Winters Method, habitually recognized as exponential smoothing, gives more weight to fresh information while lowering the weights of earlier observations. It is made up of three parts: Seasonality (recurring patterns), Trend (direction over time), and Level (baseline value). This approach is frequently used to estimate retail sales since it is good at predicting data with obvious seasonal fluctuations.

2.1.3. Moving Averages (SMA, WMA)

Short-term forecasting fluctuations are smoothed using moving averages. The Weighted Moving Average (WMA) gives more weight to recent data, whereas the Simple Moving Average (SMA) computes the average over a predetermined time period. Despite being extensively utilized in commercial and financial forecasting, they have trouble grasping intricate trend patterns and seasonality.

2.2. Machine Learning Approaches

Methods of Machine Learning for Sales Forecasting Due to its ability to automatically identify data patterns without explicit programming, machine learning techniques have become strong substitutes for conventional approaches. These strategies include hybrid models that integrate the advantages of both statistical and machine learning techniques and supervised and unsupervised learning approaches.

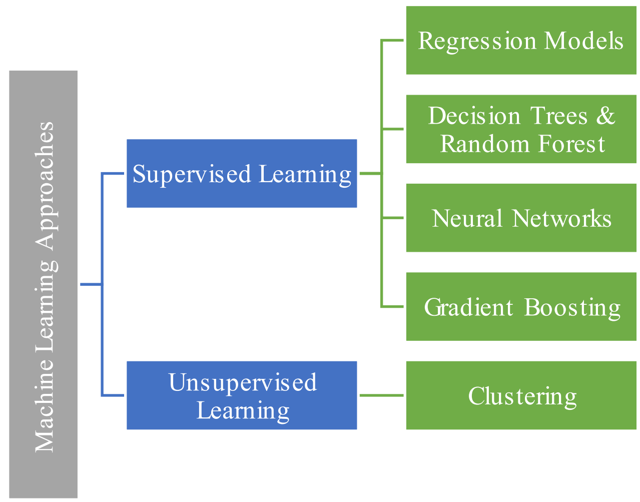

Fig. 2 Machine learning approach

2.2.1. Supervised Learning

Sales forecasting makes extensive use of supervised learning, in which models learn from past data to forecast future patterns.

- Regression models, such as Ridge, Lasso, and Linear Regression, examine the connections between sales and influencing variables in order to generate numerical forecasts.
- Decision trees and random forests are tree-based models that reduce overfitting and increase accuracy by dividing data into meaningful subsets.
- Gradient boosting (XGBoost, LightGBM) gradually improves forecasts for increased accuracy.
- Neural networks (MLP, RNN, LSTM) are deep learning models that can recognize complicated patterns and progressive relationships in sales data.

2.2.2. Unsupervised Learning

Without predetermined labels, unsupervised learning aids in revealing hidden patterns in sales data.

Clustering (K-Means, DBSCAN)

Similar sales trends are grouped using clustering algorithms like K-Means and DBSCAN, which help with demand forecasting and market segmentation. DBSCAN finds anomalies and dense areas, whereas K-Means creates clusters based on similarity. These techniques assist companies in improving sales forecasts and marketing strategy optimization.

3. Machine Learning Based Workflow for Sales Forecasting

Data collection, which includes collecting ancient sales data such as invoice date, buyer type, financier, automobile model, and booking amount, is the first step in the sales forecasting workflow. Brainwashing categorical variables, resolving missing values, and standardizing numerical appearances are the next steps in data preprocessing, which assurances data quality. After that, trends, seasonality, and outliers are originate using statistical methods and imagination in investigative data analysis, or EDA [7].

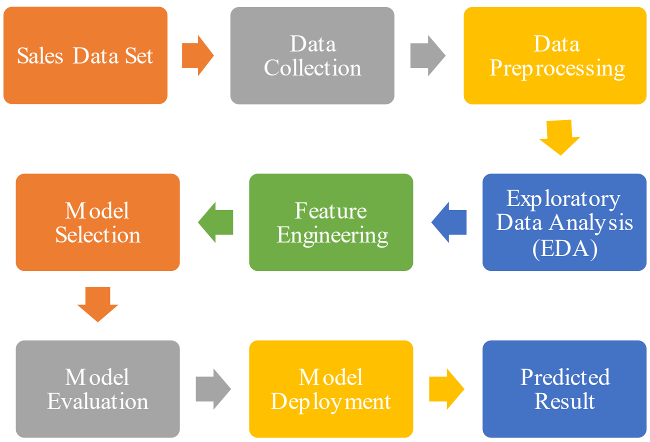

Fig. 3 System workflow [9]

A machine learning-based sales forecasting system's workflow is shown in Figure 3 in an organized, step-by-step manner [6].

- Sales Data Set: The first step in the procedure is collecting past sales information, which includes facts on the goods, buyer type, funder, and sales date.
- Data Collection: Databases, CRM systems, and external APIs are some of the sources from which the gathered raw data is assembled.
- Data Preprocessing: Cleaning the data by handling missing values, training unqualified variables, normalizing numerical topographies, and eliminating outliers.
- Exploratory Data Analysis (EDA): Analyzing patterns, seasonality, trends, and relationships in the data using visualizations and statistical techniques
- Feature Engineering: constructing significant elements (such as lag-based variables and moving averages) to improve the predictive accuracy of the model.
- Model Selection: Selecting suitable machine learning models for forecasting, such as LSTM, Random Forest, XGBoost, or ARIMA.
- Model Evaluation: Evaluating the model's performance by calculating prediction accuracy using R2, MAE, and RMSE measures.
- Model Deployment: Putting the learned model to use in a real-world setting to produce predictions.
- Predicted Result: Demand planning, Stock supervision, and corporate decision-making can all be assisted by the sales forecasts provided in the last stage.

4. Performance Metrics

To ensure accuracy, dependability, and commercial effect, measuring how glowing machine learning models perform in sales forecasting is essential. Different models call for different evaluation metrics depending on the type of predictions (numerical or time-series). A thorough explanation of the evaluation criteria applied to several ML models in sales forecasting can be found below.

Table 1 summarizes the evaluation metrics used for dissimilar machine learning models in sales prediction. The main metrics regression-based models (Linear Regression, Ridge, Lasso) use to gauge prediction accuracy are RMSE, MAE, R2, and MAPE. These procedures are also used by tree-based models (Decision Trees, Random Forest, XGBoost), but they additionally use Feature Importance to examine important sales factors.

Complex temporal patterns are handled by sMAPE, MSLE, and DTW, which are essential to deep learning models (MLP, RNN, LSTM). Time-series models (ARIMA, SARIMA) employ additional statistical metrics such as AIC, BIC, and Durbin-Watson Statistic for model selection and trend detection. Hybrid models combine different metrics (ARIMA-LSTM, RF-ARIMA) to balance interpretability and accuracy.

Businesses can increase the accuracy of their sales forecasting models and make data-driven choices by choosing the right assessment metric.

Table 1. Comparison of performance metrics for sales forecasting models

ML Model Type	Evaluation Metrics	Purpose
Regression-Based Models (Linear Regression, Ridge, Lasso)	RMSE (Root Mean Squared Error)	Measures the standard nonconformity of prediction errors. Lower RMSE indicates better accuracy.
	MAE (Mean Absolute Error)	Measures the typical absolute error amongst predicted and actual values.
	R^2 Score (Coefficient of Determination)	Evaluates how well the model explains variance in sales. A higher R^2 is better.

	MAPE (Mean Absolute Percentage Error)	It expresses error as a percentage and is useful for business forecasting.
Tree-Based Models (Decision Trees, Random Forest, XGBoost, LightGBM)	RMSE & MAE	Standard metrics for measuring error in tree-based models.
	MSE (Mean Squared Error)	Penalizes large errors more heavily, making it sensitive to outliers.
	R² Score	Determines how well the model fits the data.
	Feature Importance	Identifies key factors influencing sales predictions.
Deep Learning Models (MLP, RNN, LSTM)	RMSE, MAE, MSE	Standard loss functions for time-series forecasting.
	sMAPE (Symmetric MAPE)	Adjusted version of MAPE, reducing the impact of extreme values.
	MSLE (Mean Squared Logarithmic Error)	Suitable for exponential growth sales data.
	DTW (Dynamic Time Warping)	Measures similarity between actual and forecasted sales trends.
Time-Series Models (ARIMA, SARIMA, Holt-Winters)	RMSE, MAE, MAPE	Standard accuracy measures for time-series forecasting.
	AIC (Akaike Information Criterion)	Helps select the best model while avoiding overfitting.
	BIC (Bayesian Information Criterion)	Similar to AIC, but applies stronger penalties for complexity.
	Durbin-Watson Statistic	Checks for autocorrelation in residuals.
	MFE (Mean Forecast Error)	Determines whether the model consistently under-predicts or over-predicts.
Hybrid Models (ARIMA-LSTM, RF-ARIMA)	RMSE & MAE	Standard error metrics for hybrid models.
	R² Score	Measures how well both statistical and ML components explain variance.
	Weighted Error Metrics	Balances multiple evaluation criteria for improved accuracy.

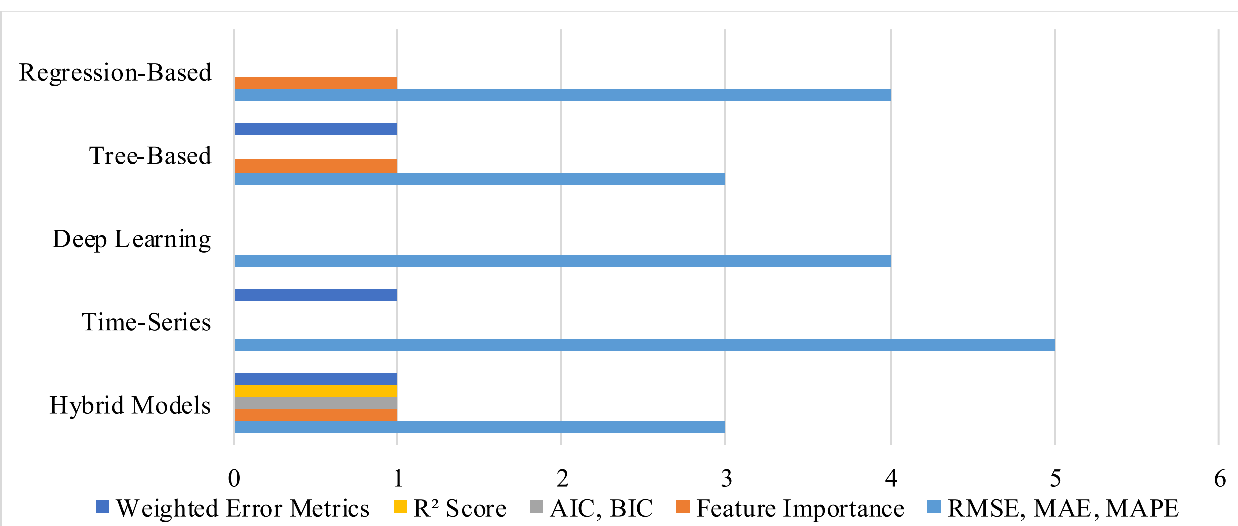

Fig. 4 Comparison of performance metrics

5. Literature Survey

Optimizing corporate operations requires accurate sales forecasting, yet complicated, nonlinear sales trends are typically difficult for standard methods to handle. Through sophisticated models like decision trees and LSTMs, Machine Learning (ML) improves accuracy; yet, issues with feature selection, data quality, and real-time adaptation still exist. The main challenge is determining the best machine learning strategies while considering interpretability and outside disturbances. In order to improve forecasting accuracy and decision-making, this paper assesses current machine learning techniques, their drawbacks, and potential avenues for future research.

The study evaluates the efficiency of Machine Learning (ML) in sales forecasting, mainly in the retail sector. The research objective is to recognize the most careful forecasting method by comparing time series models, artificial neural networks, and ML algorithms. The study examines numerous forecasting techniques using An Integrative Literature Review (ILR) methodology. The conclusions highlight that ML models outperform traditional methods, especially when including exogenous and endogenous variables. Evaluation limitations include accuracy metrics such as RMSE. The study's strength lies in classifying hidden demand designs, and the upcoming examination should focus on enhancing ML implementation for energetic market conditions [1].

In [2], the appraisal of statistical and Machine Learning (ML) techniques aims to increase forecasting accuracy. The study uses benchmarking. Practise and examine 100,000 time series from the M4 opposition dataset. ML-based methods, statistical models, and hybrid policies that incorporate both are used in implementation. Assessment actions like sMAPE and OWA are used to evaluate methods like collective learning, Recurrent Neural Networks (RNN), and exponential smoothing. The study's strength is its finding that hybrid models perform better than standalone machine learning techniques. Future studies should examine attractive prediction intervals and hybrid techniques for enhanced indecision approximation in sales forecasting.

In [3], the main goal is to employ Machine Learning (ML) events to raise sales forecasting accuracy. The objective of the study is to use relative analysis to determine which machine learning algorithm performs the best using a variety of techniques, including LSTM, XGBoost, ARIMA, Random Forest, Decision Trees, and Linear Regression. RMSE, R-squared, and MAE are assessment parameters. Showing Boost's advantage in structured datasets is the paper's strongest point. In order to improve analytical accuracy in dynamic sales conditions, future research should investigate hybrid models and real-time forecasting improvements.

In [4], the study evaluates the extrapolative influence of Machine Learning (ML) and conservative time series models for sales. Comparing ARIMA and Exponential Smoothing with cutting-edge machine learning models like XGBoost, LightGBM, and DeepAR is the aim of the study. Past sales data from the M5 competition and Kaggle datasets are analyzed as part of the process. Data preprocessing, model training, and validation with RMSE, MAE, and MAPE as evaluation parameters are all part of the operation. The study highlights the higher accuracy of ML models while pointing out their figuring requirements. Future research should quintessence on hybrid methods that strike a co-operation between efficiency, interpretability, and accuracy for changing corporate situations.

The work aims to insult Machine Learning (ML) to advance customer separation and demand forecasting in the vehicle rental sector. In order to recover the convoy operation and target marketing, the research aims to combine demand forecasts and consumer subdivision. Analysis of past rental data, the use of decision trees, random forests, clustering (DBSCAN, Agglomerative), and forecasting models such as ARIMA and Holt-Winters are all part of the process. Accuracy, silhouette coefficient, and MAE are evaluation parameters. Future research should examine real-time active pricing and hybrid forecasting models for enhanced business strategies, as the study's interpretable framework is one of its toughest points [10].

In [11], the objective of the work is to apply Machine Learning (ML) methods to forecast the price of used cars in Bangladesh. The study aims to create a precise forecasting model that will help consumers make wise choices. The methodology includes web scraping, preprocessing, and exploratory data analysis. Regression models are implemented, including Linear Regression, LASSO, Decision Tree, Random Forest, and XGBoost. RMSE, MAE, and R2 scores are examples of evaluation parameters. The study's strength is that it used XGBoost to achieve an accuracy of above 91%. For wider applicability, future studies should concentrate on growing datasets and deploying models in present.

In order to maximize energy management and infrastructure planning, the study focuses on predicting the demand for Electric Vehicle (EV) accusing. The study aims to assess Machine Learning (ML) and probabilistic approaches to demand forecasting. The methodology employs a systematic review approach, applying PRISMA criteria to analyze recent studies. Comparing ML-based techniques (LSTM, XGBoost, GNNs) and probabilistic models (Markov Chains, Monte Carlo) is part of the implementation process. The evaluation metrics are MAE, MAPE, and RMSE. The study excels at handling stochastic demand unpredictability and data scarcity. Future studies should investigate hybrid models and real-time demand forecasts for grid stability and energy efficiency [12].

In [13], the chief goal of the research is to apply the XGBoost algorithm to forecast passenger automobile sales. The objective of the work is to practise machine learning procedures to increase the accuracy of sales forecasting. Data preparation, feature selection utilizing correlation analysis and information gain, and model training with XGBoost are all part of the methodology. A sliding window method is used in employment to record past trends. The evaluation metrics are

LDSR, MAPE, and RMSE. The study's strength is its actual computation and great forecast accuracy. Future studies should investigate including economic indices such as GDP and CPI to further improve forecasting models.

The objective of the research effort is to use machine learning, namely XGBoost, to improve sales forecasting accuracy. The study aims to create an effective forecasting model for large-scale retail sales. Walmart's 1913-day sales dataset is used for feature engineering, outlier filtering, and model training. XGBoost regression, price and lag features, and time-based aggregation are all implemented. RMSSE, one of the evaluation parameters, indicates a 16.3% improvement above linear regression. The study's strength is its minimal computing cost and great accuracy. Real-time forecasting and hybrid models for dynamic retail environments should be investigated in future studies [14].

In [15], the work aims to apply machine learning (ML) to improve firm-level automotive demand forecasts. The study aims to generate a hybrid machine learning model that incorporates both exogenous (market-level) and endogenous (firm-level) elements. The process includes gathering data from South Korean companies between 2011 and 2020 and using machine learning algorithms such as Stochastic Gradient Descent, Random Forest, Neural Networks, and Linear Regression. Feature selection using the RReliefF method is part of the implementation. R2, MAE, and RMSE are evaluation parameters. The study's hybrid input technique is its strongest point. Future studies should examine multi-market applicability and real-time forecasting.

The study employs a variety of machine learning approaches to forecast sales of Electric Vehicles (EVs). By joining Gradient Boosted Regression Trees (GBRT) with Long Short-Term Memory (LSTM), the research aims to improve sales forecasting accuracy. Preprocessing, model training, and data collection from international EV markets are all part of the methodology. The operation uses GBRT to improve predictions and LSTM to capture the greatest temporal trends. RMSE and MAE are assessment parameters. The study's hybrid technique, which recovers forecast toughness, is its strongest point. Future studies should examine real-time prediction and include economic considerations for improved market insights [16].

In [17], the primary goal of the work is to apply Machine Learning (ML) methods to approximate used car prices. The study aims to compare the efficiency of the Random Forest and Decision Tree algorithms for exact price prediction. The methodology includes preprocessing, model training, and data assembly from Kaggle. Rapid Miner is used for model assessment and feature selection during application. Accuracy is one of the evaluation metrics; Random Forest and Decision Tree both realized 72.13% and 67.21%, respectively. The improved accuracy of Random Forest is the study's toughest point. For better forecasts, future studies should inspect outside variables, including the state of the economy and changing prices.

In [18], the examination uses Artificial Neural Networks (ANNs) to approximate consumer acquiring power for automobile sales. The research aims to create an ANN-based model for forecasting learning in auto sales. The approach entails assembling financial information from customers, preprocessing it, and then training ANNs. Feature collection and model optimization, applying past sales data and economic indicators, are part of employment. Accuracy scores and RMSE are evaluation criteria. The strength of the study is ANN's capacity to manage intricate patterns. Future studies should examine hybrid deep learning models and outside economic factors for enhanced prediction accuracy.

By adding statistical and Machine Learning (ML) models, the research aims to improve the forecasting of motorized sales in India. The aim of the study is to recover forecast accuracy by combining Backpropagation Neural Networks (BPNNs) with the Adaptive Multiplicative Triple Exponential Smoothing Holt-Winters (AHW) approach. Preprocessing, model training, and gathering previous sales data are all part of the methodology. For increased accuracy, AHW and BPNNs are hybridized at the time of employment. Prediction accuracy (0.9746) and RMSE are evaluation parameters. The study's hybrid method is its strongest point, but for improved forecasts, future research should look at real-time forecasting and the incorporation of economic factors [19].

In [20], Fuzzy neural networks are used in the study to forecast sales for the Indian auto sector. The aim of this research is to enhance forecasting accuracy through the combination of Fuzzy logic and Backpropagation Neural Networks (FBNN). Maruti Suzuki sales data is gathered, preprocessed, and then used to train the FBNN model. The implementation includes examining historical sales data, inflation rates, and gas costs. The suggested model performs better than multiple regression procedures, according to evaluation metrics such as Mean Squared Error (MSE). Although future research could include more economic elements and real-time predictive analytics, the study's strength is its higher accuracy.

In [21], using Machine Learning (ML) with contrastive explanations, the study aims to improve cold-start promotional sales forecasts. The study aims to improve prediction accuracy while maintaining the model's interpretability. Gradient Boosted Decision Trees (GBDT) with contrastive explanations are used in the methodology to enhance decision-making. The process of implementation entails selecting features, training models using past promotional sales data, and producing outputs that can be understood. These evaluation parameters are forecast bias, RMSE, and MAE. The study's strength is its explanatory ability, which helps to close the gap between business insights and machine learning projections. Deep learning integration should be investigated in future studies for improved dynamic promotional predictions.

The research focuses on Machine Learning (ML) methods for predicting auto sales. The aim of the project is to combine several machine learning approaches to create an optimal forecasting model. The Analytic Hierarchy Process (AHP) is used in the methodology to choose features and train models such as Random Forest, Random Tree, and Linear Regression. Clustering, model evaluation, and dataset preprocessing are all part of the implementation. RMSE, accuracy, and feature importance ranking are evaluation parameters. Combining statistical and machine learning techniques is the study's strongest point. Future studies should investigate hybrid deep learning models and real-time sales forecasting for improved predicted accuracy [22].

In [23], ARIMA, SARIMA, and LSTM forecasting techniques for auto sales are compared in this research. The aim of the work is to identify the best model for forecasting automobile sales. The procedure includes gathering historical sales data, preparing it, and using deep learning and statistical models. Adjusting hyperparameters such as learning rate and neuron count is part of the implementation. RMSE and MAE are among the evaluation measures; LSTM outperformed ARIMA in accuracy by 92%. The study's strength is deep learning's capacity to manage nonlinearity; nonetheless, future research should investigate hybrid models for increased accuracy. In [24], the primary aim of this research is to apply Machine Learning (ML) to decipher used automobile market consumer behavior. The aim of the study is to find the main determinants of buyer preferences. The process entails gathering data from both offline and online sources, preparing it, and using machine learning models such as Random Forest, Decision Trees, and Linear Regression. Model evaluation and feature selection are part of the implementation. Price sensitivity, mileage thresholds, and brand influence are some of the evaluation characteristics. The study's ability to identify trust determinants in purchasing decisions is one of its strongest points. Future studies should examine psychological aspects and digital trust in relation to online auto purchasing.

In [25], K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) regression are associated in the study for forecasting used automobile prices. The aim of the study is to evaluate and identify the most accurate price forecasting model. Data gathering via Kaggle, preprocessing, and exploratory data analysis are all part of the methodology. KNN and SVM models are trained utilizing important vehicle features as part of the implementation. KNN achieved 83% accuracy in terms of accuracy, precision, and F1-score, which are evaluation metrics. Future research should examine hybrid models and deep learning techniques for increased prediction accuracy, although the study's comparative approach is its strongest point.

5.1. Key Findings

Key findings from numerous studies examining the use of Machine Learning (ML) models in sales forecasting across various industries, including retail, automotive, electric cars, and rental services, are compiled in the table. Table 2 summarizes key findings from the literature review on machine learning-based sales forecasting, highlighting the superiority of ML models over traditional forecasting methods. Studies show that XGBoost, LSTM, and Random Forest consistently outperform techniques like ARIMA and Exponential Smoothing, especially when handling complex sales data. Hybrid models integrating ML with statistical methods further enhance prediction accuracy and robustness. Explainable AI (XAI) techniques, such as Gradient Boosted Decision Trees (GBDT) with contrastive explanations, improve model interpretability for business decision-making. The inclusion of economic factors like fuel prices and inflation rates strengthens sales predictions, particularly in the automobile and retail sectors. Real-time forecasting using IoT data and advanced deep learning models is a key area for future research, ensuring dynamic adaptability in market-driven industries.

Table 2. Summary of key findings from literature review

Study	Findings
[1] ML in Sales Forecasting	ML models outperform traditional forecasting methods, especially when incorporating exogenous and endogenous variables.
[2] Benchmarking ML for Forecasting	Hybrid models perform better than standalone ML techniques in large-scale forecasting.
[3] ML in Retail Sales Forecasting	XGBoost shows the best performance for structured sales data compared to other models.
[4] ML vs. Traditional Models for Sales Forecasting	ML models provide higher accuracy than ARIMA and Exponential Smoothing, but require higher computational power.
[5] ML for Vehicle Rental Demand Forecasting	Demand forecasting combined with customer segmentation improves fleet utilization and targeted marketing.
[6] ML for Used Car Price Prediction	XGBoost achieves over 91% accuracy in predicting used car prices.
[7] EV Charging Demand Forecasting	Hybrid ML models handle stochastic demand variations better than standalone probabilistic models.

[8] XGBoost for Passenger Car Sales Forecasting	XGBoost provides efficient and high-accuracy predictions with minimal computational overhead.
[9] ML for Large-Scale Retail Sales	XGBoost improves forecasting accuracy by 16.3% compared to traditional linear regression.
[10] ML for Firm-Level Auto Demand Forecasting	Hybrid ML models integrate firm-level and market-level data to improve forecasting accuracy.
[11] ML for EV Sales Forecasting	Combining GBRT and LSTM enhances prediction robustness for EV sales.
[12] ML for Used Car Price Prediction	Random Forest achieves higher prediction accuracy than Decision Tree models.
[13] ANN for Auto Sales Prediction	ANN models effectively capture complex auto sales trends.
[14] Hybrid Statistical & ML for Indian Auto Sales	A hybrid BPNN-AHW model achieves 97.46% prediction accuracy.
[15] Fuzzy Neural Networks for Indian Auto Sales	FBNN provides better accuracy than traditional multiple regression techniques.
[16] Contrastive ML for Promotional Sales Forecasting	GBDT improves interpretability in promotional sales forecasting.
[17] AHP-Based ML for Auto Sales Forecasting	Hybrid ML models combining RF, LR, and clustering improve forecasting accuracy.
[18] Deep Learning vs. Statistical Models for Auto Sales	LSTM outperforms ARIMA with 92% accuracy in handling nonlinear sales patterns.

6. Challenges

Although machine learning has improved the accuracy of sales forecasting, a number of issues still exist. Performance is impacted by aspects such as feature selection, data quality, interpretability of the model, and computing complexity.

Other challenges are brought about by privacy issues, real-time adaptation, and the unpredictable behaviour of the external market. Improving the accuracy and usefulness of ML-based forecasting models requires addressing these constraints.

- Data Quality & Availability: Incomplete, inconsistent, and imbalanced datasets impact model accuracy.
- Feature Selection: Identifying key factors like seasonality, pricing, and consumer behavior is complex.
- Model Performance Variability: No universal model fits all scenarios; deep learning requires extensive tuning.
- Computational Complexity: High processing power is needed for large datasets and deep learning models.
- Interpretability Issues: Many ML models act as “black boxes,” reducing transparency and trust.
- Unpredictable Market Factors: Economic shifts, competitor actions, and evolving consumer preferences affect sales.
- Ethical & Privacy Concerns: Regulations like GDPR/CCPA and biased models pose risks.
- Real-Time Forecasting Limitations: Models require continuous learning to adapt to changing trends.

Despite the use of Machine Learning (ML) improving sales forecasting, issues, including data quality, interpretability of the model, and real-time flexibility, need to be resolved. In order to improve forecasting accuracy and corporate decision-making, future research should concentrate on explainable AI, adaptive learning, and hybrid models.

7. Future Scope

Future studies should investigate hybrid models such as RF-ARIMA and ARIMA-LSTM to improve forecasting accuracy. Predictions will be more trustworthy if Explainable AI (XAI) methods like SHAP and LIME are used to increase model transparency. Adapting to changing market conditions can be facilitated by real-time forecasting that makes use of online learning and reinforcement learning. Prediction accuracy can be increased by including outside variables such as competitive price, consumer sentiment, and economic data. Efficiency can be increased by automating model selection and optimization with autoML frameworks. Cultured deep learning models such as Transformers and Graph Neural Networks (GNNs) provide better long-term forecasting potential. AI-driven sales forecasting will become equitable and trustworthy once biases are addressed and GDPR and CCPA compliance are guaranteed. Dynamic pricing, inventory control, and targeted marketing can all be improved using customized forecasting models.

8. Conclusion

By identifying intricate patterns in sales data, this study demonstrates the superiority of machine learning-based sales forecasting over conventional techniques. Prediction accuracy is increased by models like Random Forest, XGBoost, and LSTM; nonetheless, issues with data quality, model interpretability, and real-time adaptability still exist. Upcoming developments can further improve forecasting reliability in deep learning architectures, AutoML, Explainable AI (XAI), and hybrid models. ML-driven forecasting can be improved to become a more accurate and business-focused tool for decision-making by including outside variables and refining models.

References

[1] Emerson Martins, and Napoleon Verardi Galegale, "Sales Forecasting Using Machine Learning Algorithms," *Journal of Management and Secretarial Research*, vol. 14, no. 7, pp. 11294-11308, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos, "The M4 Competition: Results, Findings, Conclusion and Way Forward," *International Journal of Forecasting*, vol. 34, no. 4, pp. 802-808, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[3] D. Snehal Bhosale, Ravikant Suryawanshi, and Sandeep Musale, "Comparative Analysis of use of Machine Learning Algorithm for Prediction of Sales," *Journal of Electrical Systems*, vol. 20, no. 3, pp. 851-863, 2024. [Google Scholar] [Publisher Link]

[4] Md Kamal Ahmed et al., "Forecasting Sales Trends using Time Series Analysis : A Comparative Study of Traditional and Machine Learning Models," *Membrane Technology*, vol. 2025, no. 1, pp. 668-682, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[5] Maimuna I. Abdullahi, Gilbert I.O. Aimufua, and Umar A. Muhammad, "Application of Sales Forecasting Model Based on Machine Learning Algorithms," *Proceedings of the 28th iSTEAMS Intertertiary Multidisciplinary Conference*, pp. 205-216, 2021. [CrossRef] [Publisher Link]

[6] Rashmi Amardeep et al., "Machine Learning-Based Sales Forecasting and Inventory Optimization," *International Journal of Science, Engineering and Technology*, vol. 12, no. 3, pp. 1-9, 2024. [Publisher Link]

[7] Pooja Ghude et al., "Sales Forecast Prediction & Recommendation System," *Journal of Emerging Technologies and Innovative Research*, vol. 11, no. 6, pp. 61-68, 2024. [Publisher Link]

[8] Jil P. Bhatti, Janak H. Maru, and Devangi R. Paneri, "A Review on Sales Forecasting Using Machine Learning Techniques," *International Journal of Innovative Research in Technology*, vol. 11, no. 9, pp. 2761-2764, 2025. [Publisher Link]

[9] D. Preethi, "An Analysis of Machine Learning Algorithms to Predict Sales," *International Journal of Science and Research (IJSR)*, vol. 11, no. 6, pp. 462-466, 2022. [Google Scholar]

[10] Obumeneme Ukandu, and Olamide Kalesanwo, "Leveraging Artificial Intelligence for Customer Segmentation and Demand Forecasting in the Car Rental Industry," *Asian Journal of Research in Computer Science*, vol. 18, no. 4, pp. 452-472, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[11] Fahad Rahman Amik et al., "Application of Machine Learning Techniques to Predict the Price of Pre-Owned Cars in Bangladesh," *Information*, vol. 12, no. 12, pp. 1-24, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Mamunur Rashid, Tarek Elfouly, and Nan Chen, "A Comprehensive Survey of Electric Vehicle Charging Demand Forecasting Techniques," *IEEE Open Journal of Vehicular Technology*, vol. 5, pp. 1348-1373, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Zhenchang Xia et al., "ForeXGBoost: Passenger Car Sales Prediction Based on XGBoost," *Distributed and Parallel Databases*, vol. 38, no. 3, pp. 713-738, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Xie Dairu, and Zhang Shilong, "Machine Learning Model for Sales Forecasting by Using XGBoost," *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, Guangzhou, China, pp. 480-483, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[15] Sehoon Kim, "Innovating Knowledge and Information for a Firm-Level Automobile Demand Forecast System: A Machine Learning Perspective," *Journal of Innovation & Knowledge*, vol. 8, no. 2, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[16] Rishav Dev Mishra et al., "Hybrid LSTM-GBRT Based Machine Learning Technique Implementation for Electric Vehicle Sales Prediction Analysis," *International Journal of Sustainable Engineering*, vol. 17, no. 1, pp. 84-99, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Purwa Hasan Putra et al., "Random Forest and Decision Tree Algorithms for Car Price Prediction," *Journal of Mathematics and Natural Sciences LLDikti Region 1 (JUMPA)*, vol. 4, no. 1, pp. 81-89, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] V. Sreenivas et al., "Automotative Sales Prognostication via Neural Network: Predicting Purchase Power of Customer to Purchase a Car," *International Journal of Engineering Science and Advanced Technology (IJESAT)*, vol. 24, no. 05, pp. 308-312, 2024. [Publisher Link]

[19] Kayalvizhi Subramanian et al., "An Integrated Model for Forecasting Indian Automobile," *International Journal on Advanced Science, Engineering and Information Technology (IJASEIT)*, vol. 10, no. 6, pp. 2593-2598, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[20] Rashmi Sharma, and Ashok K. Sinha, "Sales Forecast of an Automobile Industry," *International Journal of Computer Applications*, vol. 53, no. 12, pp. 25-28, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[21] Carlos Aguilar-Palacios, Sergio Muñoz-Romero, and José luis Rojo-Álvarez, "Cold-Start Promotional Sales Forecasting through Gradient Boosted-Based Contrastive Explanations," *IEEE Access*, vol. 8, pp. 137574-137586, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[22] K. Madhuvanthi et al., "Car Sales Prediction Using Machine Learning Algorithmns," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 5, pp. 1039-1050, 2019. [Google Scholar] [Publisher Link]

[23] Sharath Kariya Shetty, and Rajesh Buktar, "A Comparative Study of Automobile Sales Forecasting with ARIMA, SARIMA and Deep Learning LSTM Model," *International Journal of Advanced Operations Management*, vol. 14, no. 4, pp. 366-387, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[24] Minshun Zhang, "Decoding Consumer Behavior in the Used Car Market : A Machine Learning Approach to Key Decision Factors," *2024 2nd International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2024)*, vol. 70, pp. 1-6, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[25] Abdul Subhahan Shaik, Nazeer Shaik, and C. Krishna Priya, "Predictive Modeling in Remote Sensing Using Machine Learning Algorithms," *International Journal of Current Science Research and Review*, vol. 7, no. 6, pp. 4116-4123, 2024. [CrossRef] [Google Scholar] [Publisher Link]