As per UGC guidelines an electronic bar code is provided to secure your paper

International Journal for Modern Trends in Science and Technology
Volume 12, Issue 05, pages 08-13
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol12issue05.html
DOI: https://doi.org/10.5281/zenodo.19613524

Machine Learning-Based Sales Forecasting Model with Automated Data Processing Using Pandas and Visualization

A. Venkata Ramana, B. Leela Kalyani, B. Mani Kumar, Ch. Ramyasri, Ch. Pranavi, Dr. S. Krishna Rao

Department of Computer Science and Engineering (AI & ML), Sir C R Reddy College of Engineering, Eluru, Andhra Pradesh, India

To Cite this Article
A. Venkata Ramana, B. Leela Kalyani, B. Mani Kumar, Ch. Ramyasri, Ch. Pranavi & Dr. S. Krishna Rao (2026). Machine Learning-Based Sales Forecasting Model with Automated Data Processing Using Pandas and Visualization. International Journal for Modern Trends in Science and Technology, 12(05), 08-13. https://doi.org/10.5281/zenodo.19613524

Article Info
Received: 28 March 2026; Revised: 24 April 2026; Accepted: 26 April 2026.

KEYWORDS

Electric Vehicles,
Deep Learning,
Machine Learning,
Time Series Forecasting.

ABSTRACT

The global shift towards sustainable transportation has significantly accelerated the demand for Electric Vehicles (EVs), creating a need for accurate forecasting and insightful market analysis. This project presents a comprehensive machine learning and deep learning-based forecasting system for global EV sales, leveraging both historical data and advanced predictive algorithms. Utilizing the International Energy Agency (IEA) dataset containing regional EV sales across multiple years, this study conducts exploratory data analysis, trend visualization, growth assessment, and market share distribution across key global regions. A central component of this work is the development and deployment of a Long Short-Term Memory (LSTM) neural network model to forecast EV sales for the period 2024–2030. The LSTM model is trained on globally aggregated historical sales data after appropriate normalization using MinMaxScaler to handle the temporal patterns in sales data effectively To provide interactive insights, a dynamic and visually appealing Streamlit-based web application is developed. This dashboard allows users to select regions, analyze year-over-year growth, view historical sales trends, explore regional market share distributions, and visualize both actual and forecasted sales trajectories. Plotly-powered interactive charts enhance user engagement, while custom CSS styling ensures a professional and modern user interface. The system successfully combines data preprocessing, machine learning model deployment, real-time analytics, and web-based visualization, providing stakeholders, policymakers, and business leaders with an end-to-end solution for strategic decision-making in the rapidly evolving EV market.

1. INTRODUCTION

The primary objective of this project is to develop a robust and intelligent forecasting system for predicting Electric Vehicle (EV) sales across different regions of the world. The proposed system leverages advanced Machine Learning (ML) and Deep Learning (DL) techniques to analyze historical sales data, identify underlying patterns, and generate accurate future predictions for the period 2024 to 2030. By utilizing data-driven approaches, the system aims to improve forecasting accuracy and support informed decision-making in the rapidly evolving EV market.

A key component of this work is the implementation of Long Short-Term Memory (LSTM) networks, a specialized type of deep learning model designed for time-series forecasting. LSTM models are capable of capturing long-term dependencies and complex nonlinear relationships in sequential data, making them highly suitable for analysing EV sales trends. Their ability to retain relevant historical information while filtering out noise significantly enhances prediction performance compared to traditional statistical methods

In addition to predictive modelling, the system is deployed through a modern and interactive Streamlit-based dashboard. This dashboard provides a user-friendly interface that enables users to explore historical trends, compare regional performance, and visualize forecast results effectively. The integration of visualization tools ensures that complex analytical outputs are presented in an intuitive and accessible manner, making the system suitable for a wide range of users including analysts, policymakers, investors, and industry professionals.

The significance of this project lies in its contribution to the global transition toward sustainable mobility. As electric vehicles play a crucial role in reducing carbon emissions and dependence on fossil fuels, accurate forecasting becomes essential for planning infrastructure, optimizing production, and designing effective policies. By providing reliable insights into future EV sales patterns, the proposed system supports strategic decision-making and contributes to the sustainable growth of the automotive sector.

The key objectives of this study include performing an in-depth analysis of historical EV sales data, providing region-wise insights through advanced visualization techniques, forecasting future sales using an LSTM-based deep learning model, and developing an interactive dashboard that enhances data interpretation and usability. Through the integration of machine learning, deep learning, and visualization, this project presents a comprehensive solution for EV sales forecasting.

2. LITERATURE REVIEW

1. Liang et al. proposed an improved method to predict product demand across multiple stores by using both product similarity and store-related information. This approach increases prediction accuracy compared to traditional methods.

2. Sleem et al. enhanced the Bass model to better predict product sales patterns over time. Their model accurately captures the growth and peak stages of product adoption.

3. Geertsema and Lu studied the use of financial data and market data for prediction. They found that market data gives quick insights, while financial data provides long-term understanding. Combining both improves prediction accuracy and customer segmentation.

4. Makridakis et al. compared different forecasting models such as ARIMA and regression techniques. They concluded that no single model works best in all situations.

5. Petropoulos et al. used multiple time-based data sources to improve forecasting accuracy. By combining product, location, and time data, better predictions can be achieved.

6. Snyder et al. focused on forecasting irregular or low-demand data. Their model performs well when sales are not continuous or predictable. It helps businesses identify occasional buyers and regular customers.

7. Petersen et al. introduced a structured method to organize and review research studies. This approach helps identify patterns, trends, and research gaps. It is useful for building strong and systematic literature reviews.

8. Swaminathan and Venkitasubramony reviewed forecasting methods used in the fashion industry. They found that machine learning models perform better than traditional methods.

9. Pinciroli et al. developed a method to analyze software development approaches systematically. This method can also be applied to customer behavior analysis.

10. Li et al. proposed a model that balances general patterns and unique product behaviors. This improves the accuracy of sales predictions. It also supports better personalized customer segmentation.

11. Omar et al. used shopping basket data to forecast demand across different sales channels. Their model analyzes customer purchase patterns effectively.

12. Wang et al. applied location-based and time-based data for demand forecasting. Their model provides accurate region-specific predictions. It helps businesses plan based on local customer behavior.

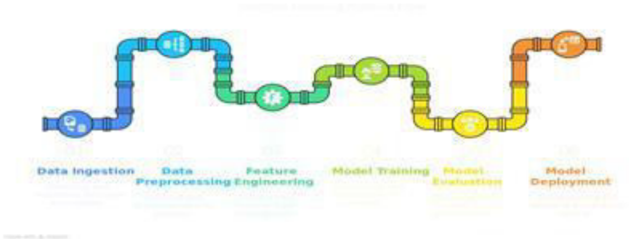
13. Tillmann et al. developed a simple and transparent forecasting model using public data. Their approach focuses on easy implementation and reproducibility. It is useful for demand prediction and planning.

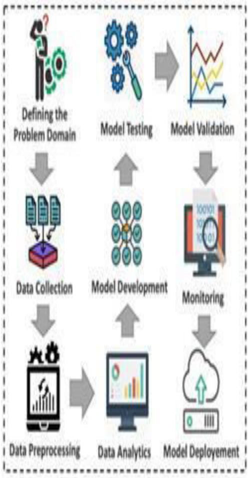
3. METHODOLOGY

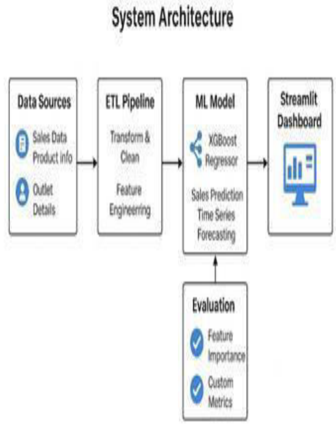
The proposed system for Electric Vehicle (EV) sales forecasting follows a structured approach that integrates data preprocessing, exploratory analysis, deep learning modelling, and visualization. The methodology is designed to ensure accurate predictions while maintaining interpretability and usability through an interactive dashboard.

3.1 System Overview

The overall workflow of the system consists of multiple stages, starting from data collection to final visualization and deployment.







3.2 Data Collection

The dataset used in this study contains historical EV sales data across different regions and time periods. The data includes important features such as:

- Year-wise sales records
- Region/country-wise distribution
- Growth trends over time

Sample DataSet:

This dataset serves as the foundation for training the forecasting model.

2	region	category	parameter	mode	powertrain	year	unit	value
3	Australia	Historical	EV sales	Cars	BEV	2011	Vehicles	49
4	Australia	Historical	EV stock sl	Cars	EV	2011	percent	0.00039
5	Australia	Historical	EV sales sl	Cars	EV	2011	percent	0.0065
6	Australia	Historical	EV sales	Cars	BEV	2011	Vehicles	49
7	Australia	Historical	EV stock	Cars	BEV	2011	Vehicles	49
8	Australia	Historical	EV stock	Cars	BEV	2012	Vehicles	220
9	Australia	Historical	EV stock	Cars	PHEV	2012	Vehicles	80
10	Australia	Historical	EV sales	Cars	PHEV	2012	Vehicles	80
11	Australia	Historical	EV sales sl	Cars	EV	2012	percent	0.03
12	Australia	Historical	EV stock sl	Cars	EV	2012	percent	0.0024
13	Australia	Historical	EV sales	Cars	BEV	2012	Vehicles	170
14	Australia	Historical	EV sales	Cars	BEV	2013	Vehicles	190
15	Australia	Historical	EV stock sl	Cars	EV	2013	percent	0.0046

3.3 Data Preprocessing

Before feeding the data into the model, several preprocessing steps are performed to ensure data quality and consistency:

- Handling Missing Values: Any missing or incomplete data points are cleaned or filled.
- Normalization: Data is scaled to a uniform range to improve model performance.
- Time-Series Formatting: The dataset is converted into sequential format suitable for LSTM input.

3.4 Exploratory Data Analysis (EDA)

Exploratory Data Analysis is conducted to understand patterns and trends in EV sales data. This includes:

- Visualization of yearly growth trends
- Region-wise comparison of EV adoption

3.5 LSTM-Based Forecasting Model

The core of the system is the Long Short-Term Memory (LSTM) model, a type of Recurrent Neural Network (RNN) designed for sequential data.

Working of LSTM Model:

- Input Gate: Decides which new information to store
- Forget Gate: Removes irrelevant information
- Output Gate: Produces the final prediction

The LSTM model captures long-term dependencies in EV sales data, making it highly effective for time-series forecasting.

3.6 Model Training and Testing

The dataset is divided into training and testing sets:

- Training Set: Used to train the LSTM model
- Testing Set: Used to evaluate prediction accuracy

3.7 Forecasting Future Sales

After training, the model is used to predict EV sales for future years (2024–2030). The predictions are generated based on historical patterns learned by the model.

3.8 Visualization and Dashboard (Streamlit)

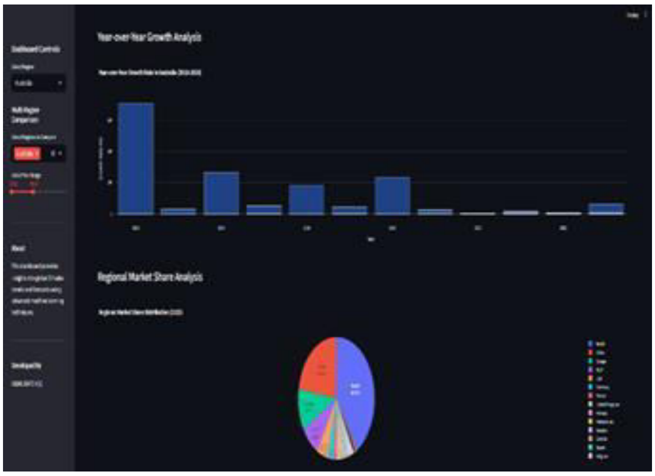
The final step involves presenting the results through a Streamlit dashboard:

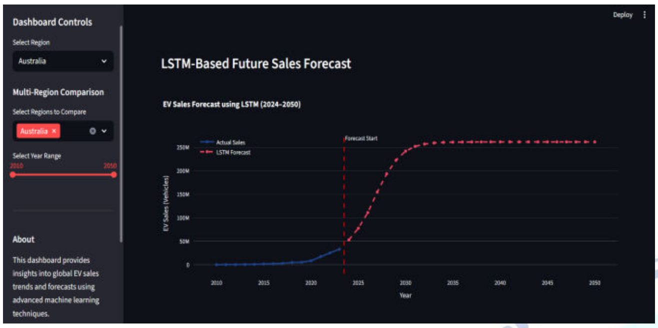
- Interactive graphs for sales trends
- Region-wise comparisons
- Future forecast visualization

This dashboard allows users to easily interpret results and make informed decisions.

4. RESULTS & OUTPUTS

The forecasting system analyzes historical EV sales data and generates predictions from 2024 to 2030 using an LSTM-based deep learning model. The results are presented through an interactive dashboard that enables users to explore trends across different regions, with validation performed using both visual and quantitative methods. Historical analysis shows strong growth in EV adoption, particularly in China, Europe, and North America, with China leading in total sales, while Year-over-Year (YoY) analysis highlights rapid growth during subsidy periods, slower growth in saturated markets, and steady increases in emerging regions. The forecasting results indicate a continuous upward trend, with 2024 predictions closely matching 2023 data and global EV sales expected to exceed 17 million by 2030. Regional market share analysis shows China dominating the market, followed by Europe, while other regions exhibit strong future potential. The dashboard includes key features such as latest, average, and total sales metrics, along with interactive charts and region-based customization. A comparison between actual and forecasted values confirms consistent growth trends and model reliability. Overall, the results demonstrate that EV adoption is rapidly increasing worldwide, and the combination of LSTM forecasting and interactive visualization provides an effective tool for understanding and analyzing future EV market trends.





5. CONCLUSION

This project focused on developing an intelligent end-to-end forecasting system for global Electric Vehicle (EV) sales by combining Machine Learning and Deep Learning techniques, particularly using LSTM neural networks. The system integrates interactive visualizations and a user-friendly Streamlit dashboard to provide accurate and dynamic forecasting across multiple regions. It successfully overcomes limitations of traditional models by capturing complex, non-linear patterns in historical data and includes advanced features such as region-wise analysis, year-over-year growth evaluation, and real-time updates. The outcomes include a functional time-series forecasting model capable of predicting EV sales from 2024 to 2050, along with an interactive web-based dashboard that enhances user experience through clean design and customizable features. However, certain limitations exist, such as reliance only on historical sales data without considering external factors, manual data updates, and lack of uncertainty estimation. Key takeaways highlight the effectiveness of LSTM in time-series forecasting, the importance of data preprocessing and model selection, and the value of combining analytics with intuitive visualization. Overall, this project demonstrates how modern data science and visualization tools can support informed decision-making, and with future improvements like automated data pipelines and integration of real-time data sources, the system can become even more powerful for EV market analysis and strategic planning.

6. FUTURE SCOPE

The project can be enhanced by integrating real-time data to improve prediction accuracy and making the system dynamic. Advanced models like GRU and Transformer can be used for better forecasting performance. It can also be extended to other domains such as retail and e-commerce. Additionally, deploying the system on cloud platforms with automated model retraining and adding Explainable AI features will make it more scalable, reliable, and user-friendly.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Liang, M.; Yang, L.; Li, K.; Zhai, H. Improved collaborative filtering for cross-store demand forecasting. Comput. Ind. Eng. 2024, 190, 110067. [CrossRef]

[2] Sleem, A.A.; Alromema, M.; Abdel-Aal, M.A.M. Improved bass model using sales proportional average for one condition of mono peak curves. arXiv 2024, arXiv:2403.08993.

[3] Geertsema, P.; Lu, H. Return Predictability: Accounting versus Market Information. 2024. Available online: https://ssrn.com/abstract=4725107 (accessed on 1 May 2024).

[4] Makridakis, S.; Andersen, A.; Carbone, R.; Fildes, R.; Hibon, M.; Lewandowski, R.; Newton, J.; Parzen, E.; Winkler, R. Accuracy of forecasting: An empirical investigation. J. R. Stat. Soc. Ser. A (General) 1983, 142, 97–145. [CrossRef]

[5] Petropoulos, F.; Makridakis, S.; Assimakopoulos, V.; Nikolopoulos, K. Forecasting with multivariate temporal aggregation: The case of promotional modelling. Int. J. Prod. Econ. 2018, 204, 161–171.

[6] Snyder, R.D.; Ord, J.K.; Beaumont, A. Forecasting the intermittent demand for slow moving inventories: A modelling approach. Int. J. Forecast. 2012, 28, 485–496. [CrossRef]

[7] Petersen, K.; Feldt, R.; Mujtaba, S.; Mattsson, M. Systematic mapping studies in software engineering. In Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE), BCS Learning & Development, Bari, Italy, 26–27 June 2008; pp. 1–10.

[8] Swaminathan, K.; Venkitasubramony, R. Demand forecasting for fashion products: A systematic review. Int. J. Forecast. 2024, 40, 247–267. [CrossRef]

[9] Pinciroli, F.; Justo, J.L.B.; Zeligueta, L.; Pma, M. Systematic Mapping Protocol Coverage of Aspect-oriented Methodologies

for the Early Phases of the Software Development Life Cycle. arXiv 2017, arXiv:1702.02653. 41

[10] Li, D.; Li, X.; Gu, F.; Pan, Z.; Chen, D.; Madden, A. A Universality-Distinction Mechanism-Based Multi-Step Sales Forecasting for Sales Prediction and Inventory Optimization. Systems 2023, 11, 311. [CrossRef]