

Machine Learning Applications in Market Basket Analysis: A Case Study on Electric Vehicle Sales in India

Rachna Rathore
Dept. of Mathematics
Jiwaji University
Gwalior, India

rachnarathore1109@gmail.com
<https://orcid.org/0000-0001-8753-888X>

Ashish Singh Chauhan
Dept. of Pharmaceutical Sciences
Uttaranchal University
Dehradun, India

ashishchauhan.pharmacy@gmail.com

Kailash Aseri
Dept. of Computer Science
D.A.V. College
Sri Ganganagar, Rajasthan

kailash.aseri@gmail.com

J. Balaji Praveen
Dept. Civil Eng.
Dhanalakshmi Srinivasan College of Eng.
Coimbatore, India

Gaurav Sethi
Dept. of Electronics and Electrical Eng.
Lovely Professional University
Punjab, India

gaurav.11106@lpu.co.in

Ankur Mishra
Dept. of CSE (AIML), CSE-APEX
Chandigarh University
Chandigarh, India

ankur@thecmatrix.com

Abstract—This study explores the potential of Market Basket Analysis (MBA) to enhance electric vehicle (EV) recommendations in India. By analyzing transactional data and identifying common item sets and association rules, MBA can provide valuable insights into customer preferences and purchasing behavior. This information can be utilized to develop personalized recommendation systems that effectively guide customers toward EV solutions tailored to their specific needs and preferences. These include discussions on key drivers of EV purchases in India, steps for implementing MBA, and challenges/concerns about the application of MBA. Specific factors in buying decisions can be traced through the analysis of unique characteristics in the Indian EV market. Finally, potential benefits are presented with the implementation of an MBA-driven recommendation system to both the manufacturer and the consumers of electric vehicles in India. The study will delve deep into the application of MBA in the Indian EV market and contribute toward more effective and customized recommendation systems for the industry at large.

Index Terms—Route Optimization, Supply Chain Electric Vehicles (EVs), Market Basket Analysis (MBA), Charging Infrastructure, EV Recommendations, Transactional Data,

I. INTRODUCTION

India's transition to environmentally friendly transportation is marked by a growing adoption of electric vehicles (EVs), reflecting efforts to reduce carbon emissions [1]. However, challenges such as limited charging infrastructure, low consumer awareness, and regional diversity in preferences persist. Addressing these issues requires a data-driven approach [2]. This project proposes a Market Basket Analysis (MBA) framework to analyze consumer purchasing

behaviors and transactional data, providing tailored EV, charging infrastructure, and accessory recommendations [3]. Traditionally used in retail, MBA will identify patterns in EV purchases, charging station usage, and accessory buying trends, enabling personalized suggestions like suitable EV models, nearby charging stations, or essential add-ons [4]. In addition, the framework will help companies optimize supply chains and marketing while providing insight into demographic and regional variations in the adoption of electric vehicles. Policymakers and manufacturers can leverage these insights to prioritize infrastructure investments, refine production strategies, and foster growth in both urban and rural areas of India's EV ecosystem. By making the market more accessible and connected, this effort ultimately seeks to hasten India's shift to electric mobility. The system will support the nation's overall objectives of lowering carbon emissions and advancing sustainable mobility options by acting as a link between industry offerings and customer wants.

Objective

The primary objective Creating a Market Basket Analysis (MBA) framework to offer tailored suggestions for electric cars (EVs), charging stations, and associated accessories in India is the aim of this project. The system will identify frequently occurring goods and trends by examining transactional data to find patterns and links between different EV purchases, charging station usage, and accessory purchases. By delivering personalised recommendations, assisting EV purchasers in making the best decisions, and simplifying product and service offers for companies, the objective is to improve the customer experience. Furthermore,

the analysis will offer significant insights into demographic and regional patterns, enabling politicians in creating efficient infrastructure and policy efforts as well as businesses in matching their strategy with market demand. By creating a more user-centric, connected, and accessible EV ecosystem, this initiative eventually seeks to assist the expansion of electric transportation in India and help the nation achieve its sustainability and carbon reduction objectives.

II. PROPOSED SYSTEM

The proposed system is a suggestion framework based on Market Basket Analysis (MBA) and customized for the Indian EV industry. To find trends and correlations between these things, it will examine transactional data, such as EV purchases, charging station usage, and accessory sales. By identifying frequently occurring itemset using the Apriorism algorithm, the system will provide rules that direct tailored suggestions for customers, including the most appropriate EV models, suggested charging infrastructure, and related accessories. In order to provide recommendations that are particular to a certain region and take into consideration variations in urban and rural areas as well as adoption rates, the system will also divide users based on demographic characteristics including geography, income, and usage patterns. The system's processing of this data will improve the EV purchasing experience, optimize product offers for companies, and yield consumer behavior insights that can guide future infrastructure planning and marketing campaigns. In the end, this strategy will help India's EV ecosystem expand.

III. LITERATURE REVIEW

Kaushik introduced dynamic data scaling techniques for streaming ML addressing challenges associated with processing high-velocity data streams efficiently [5]. Gujarathi et al. studied the Indian EV market, placing emphasis on policy challenges and consumer perspectives for driving adoption [6]. Shibl et al. proposed an ML-based EV charging management approach, introducing a framework for optimizing fast chargers' dispersion in order to reduce bottlenecks [7]. Discussion Rathore et al. showed how fog and edge computing can enable intelligent transportation systems for navigation. This work has shown that decentralized computing has the potential to reduce latency and hence enhance real-time decision-making in such applications [8]. Talaat et al. developed a deep learning-based mathematical model combined with explainable AI and RFM analysis for targeted marketing. This was an innovative approach for actionable insights into segmenting and targeting customers more effectively [9]. In the year 2023, Rathore studied the influence of AI on recruitment and selection. He pointed out that AI automates the screening of resumes, shortlisting of candidates, and interview scheduling. This study focused on strengthening the decision-making processes within the human resource domain with the use of AI [10].

Patwary et al. (2021) applied the market basket analysis methodology with ML to identify consumer purchasing patterns that could help optimize inventory management and sales strategy in order to enhance operational efficiency and customer satisfaction within the retail industry [11]. Yeh and Wang (2023) presented an ML-based model for the sales forecasting of EVs, focusing on how regression techniques can help understand the dynamics of market demand and consumer preference [12]. Shrivastava and Rathore considered a single-server Markovian queuing model with working vacation and customer renegeing and discussed its applicability in resource optimization in service industries [13]. Simsek et al. proposed a deep learning-based EV demand forecasting model incorporating historical data and market trends to enhance prediction accuracy [14]. Kumbure et al. provided a review on the various Machine Learning techniques, datasets associated with the stock market, feature engineering, and explanation in algorithms associated with financial decisions [15]. A similar concept has been put forward by Dixit and Singh that proposes using ML methodologies for the detection of probable EV buyers in India by garnering relevant information related to consumer behaviour [16]. For instance, Rathore et al., in 2024, identified progress on skin cancer diagnosis using ML in smart ecosystems and its promise for early diagnosis in precision health care [17]. Tufail et al., 2023, performed a wide review of ML models, libraries, and applications; however, they presented challenges in scalability and interpretability for large-scale implementations [18]. Afandizadeh et al. (2023) resorted to ML for the EVs' transport market penetration prediction, which indicates an increasing interest apparently toward sustainable transportation solutions [19].

IV. METHODOLOGY

Below is a structured analysis of various aspects of the Indian EV market, which further makes a prediction about future trends based on historical data. The methodology consists of a number of stages: data import and setup, exploratory analysis, feature engineering, association rule mining, time series forecasting, evaluation, and future trends.

Data Import and Setup First, import and configure several CSV datasets in Python, setting the environment for electric vehicle market analysis. These various datasets were representing different perspectives over the EV industry: manufacturers, operational metrics, sales categories, amongst others. Following this, data wrangling will be done by using libraries such as pandas, numpy, and matplotlib. Initial inspections of data can enable the location and correction of various issues, including those dealing with missing values and inconsistencies of data types, for further advanced analysis.

Data Cleaning and Preprocessing

Data cleaning and preprocessing after import is based on the completeness and consistency of data. This involves the imputation of missing values by the mean or median for numerical features, while for categorical features, it is done by the most common category; the categorical features

are encoded into binary or dummy variables; time-related fields are formatted; or numerical fields are standardized in preparation for exploratory and advanced analysis.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is used to investigate the structure of the data, beginning with univariate analysis, studying each feature using a histogram, boxplot, and bar chart that underlines distributions and anomalies. It proceeds to bivariate and multivariate analyses using pair plots, scatter plots, and correlation heatmaps in order to understand relationships between variables, such as the trends in sales of different categories of EVs and operational metrics, and lays the foundation for further analysis.

Feature Engineering

Feature engineering, in the development of new features from raw data that can capture underlying complex information, plays an important role in advanced predictive modeling and mining association rules. Time-based features, such as month and quarter, are created to identify temporal patterns. The aggregations help to depict broader trends. Advanced feature selection reduces the dimensionality with greater efficiency and lower overfitting risk. Several interaction terms also develop relationships of variables and preference over the regions.

Association Rule Mining

EV category and metrics analysis will be done through the association rule mining using the apriori algorithm of the mlxtend library. Data is transformed into a binary matrix in order to find the pattern of frequent co-occurrence and association among EV categories, manufacturers, and regions, based on support, confidence, and lift measures. This would give insight into the market dynamics and consumer preference.

Time Series Analysis and Forecasting Subsequently, the time series analysis and forecasting of the growth in the sales of electric vehicles are pursued with the use of modeling based on association mining. The two applied models will take the captured seasonal and temporal patterns in data. First to apply will be the ARIMA (Autoregressive Integrated 39 Moving Average) model with a focus on capturing trends and seasonality within sales data. The parameters of ARIMA are optimized to define the lag order, differencing, and moving average components. Consequentially, the performance of the model will be evaluated in respect to statistical metrics such as AIC and BIC scores. Complementing ARIMA, this study utilizes a Prophet model that is better suited for complex seasonal patterns and holiday effects, hence apt for sales forecasting where seasonal demand alternations are influential. This is an example of flexibility in the trend and seasonality components in Prophet, which helps to capture detailed patterns in the sales data. A plot of actual versus forecasted values is drawn to help get a feeling for the model's accuracy.

Evaluation Metrics and Model Validation

Model evaluation is important as that will present the robustness of both association rules and forecasting models. The important measures quantifying strength in association

rules include Support, Confidence, and lift score. On the other hand, ARIMA and Prophet are strong in their predictions through their Mean Absolute Error and Root Mean Squared Error since these said measures account for the deviation from the exact predictions. Further, all models will be cross-validated with subsets to see if they generalize well on data they have never seen. These tight analyses certainly assert each model's effect, thus having maximum confidence level on the predictions and associations established. Along with the above-mentioned metrics, sensitivity analysis finds how model predictions alter w.r.t. variation in key parameters, helping further toward the revelation of the bias/weakness involved. Further, pruning methods are adopted for filtering and hence consider only those rules having greater relevance in the context of identified associations. The models are also tested for seasonal and trend adjustments to ensure that they can be adaptable for different time periods. This approach in the evaluation not only verifies robustness but also enhances interpretability for assurance of actionable insights in data-driven decision-making.

V. RESULTS AND DISCUSSION

This study on the Indian Electric Vehicle market gives insight into historical sales trends, popular categories of EVs, and regional dynamics. The analytical findings are presented through a variety of visualizations and quantitative metrics that explain consumer preferences and probable growth trajectories in the Indian EV sector. The results are particularly relevant for stakeholders in the EV ecosystem-manufacturers, policymakers, and investors-seeking a clearer understanding of the market's current state and future potential.

The main outcome of this is the identification of different sales patterns across categories of EVs, with electric two-wheelers being the highest-selling category. This is then depicted visually using bar charts and line plots of sales volume over time, showing that two-wheelers have maintained consistent demand growth compared to other categories. It is attributed to their affordability, fuel savings, and suitability for urban commuting; hence, ideal for Indian cities. The above visualization of the sales pattern shows a strong consumer inclination toward two-wheelers; thus, the segment is likely to keep driving EV adoption in the near term.

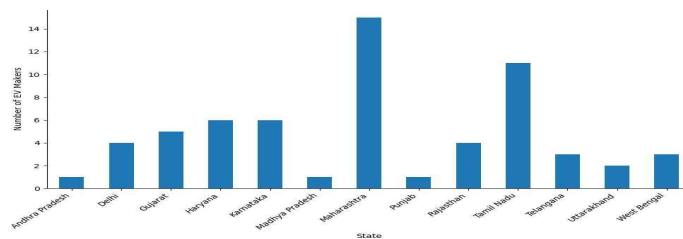


Fig. 1. Bar Plot Showing The Distribution of EV Makers Across States

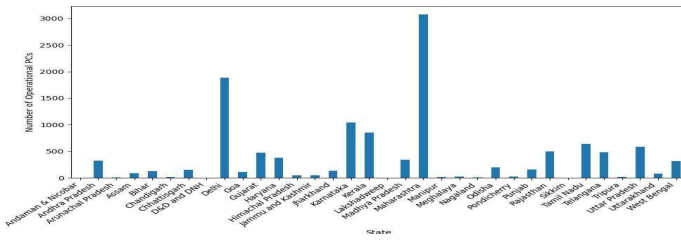


Fig. 2. Bar Plot Showing The Distribution of Operational PCs Across States

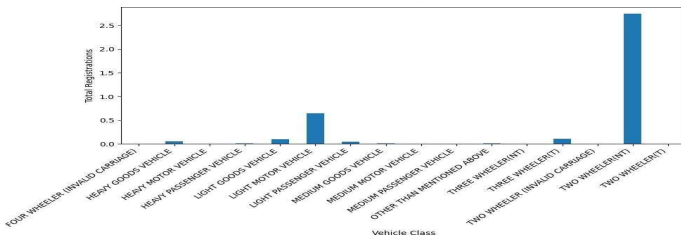


Fig. 3. Bar Plot Showing The Distribution of Total Registered Vehicles Across Vehicle Types

Further insight arises from the trends in EV sales across states and regions in the country. Regional sales, as visualized in heatmaps, indicate that high-density urban areas like Maharashtra, Delhi, and Karnataka are the main markets where EV adoption is happening. Geographies that combine high population density, urban infrastructure, and state-led incentives generally show disproportionately high sales volumes compared to rural or smaller states. This regional analysis underpins that, in collaboration with their policymaker/promoter partners, manufacturers allocate all resources and invest all marketing resources into these epicenters of higher EV adoption for improved returns. However, by aligning strategy with strong high-demand regions within high-growth states, efforts become maximally effective or optimized to further match the evolving demands of the developing marketplace.

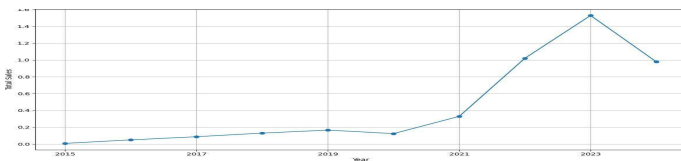


Fig. 4. Line Plot Showing The Total EV Sales Over The Years

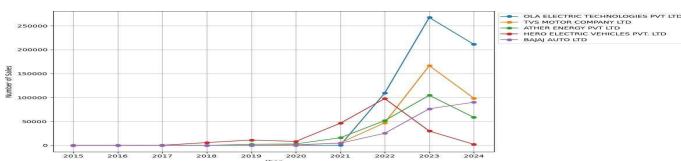


Fig. 5. Line Plot Showing The Total Company wise EV Sales Over The Years

Association rule mining further strengthens the understanding of consumer behavior in terms of frequent

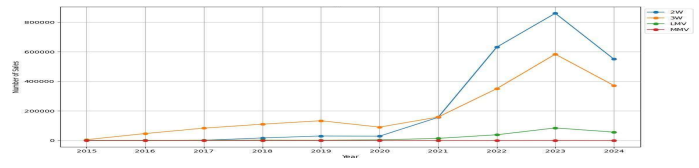


Fig. 6. Line Plot Showing The Total Type wise EV Sales Over The Year

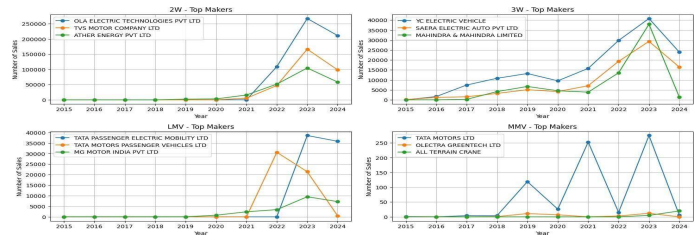


Fig. 7. Line Plot Showing The Top EV Makers Over The Years.

co-occurring patterns among EV categories and regions. Through this analysis, some interesting associations have come up, such as the preference for two-wheelers in urban areas and the association of specific manufacturers with EV categories. For instance, some manufacturers have established strong brand associations with certain vehicle types or regions, which may imply a potentially loyal customer base. This is all the more important for a company looking to expand its market share by developing product offerings that would best fit regional tastes and preferences. The network diagrams further offer a visual of these association rules in structured co-occurring patterns that may help the manufacturer in planning and improving its product distribution and marketing strategy accordingly.

Results obtained for time series forecasting also provide broad insights into sales trends in subsequent periods. While considering both ARIMA and Prophet models for prediction, this study puts forth the fact that, in a few years to come, there would be an upward trend in sales of electric vehicles. Additionally, the seasonal and yearly pattern can be found on the graph developed using the Prophet model; this depicts the fact that at certain instances or periods within the year, the sale of vehicles seems to increase—for example, because of the end of fiscal year incentives, festivities, or when new policies related to them have just been announced by the government. The forecasting visualizations that plot the historical sales data against the predicted values reflect a compound annual growth rate within the expected bounds of the industry for India's EV market. Predictions point to growth potential in different categories of EVs, with two-wheelers leading the forecasted demand.

Further, the performance of these models was examined using metrics such as Mean Absolute Error and Root Mean Squared Error to establish the veracity of these models. Based on these low error rates, it then appears that both ARIMA and Prophet capture well the sales trend and seasonality for possible use in forecasting, which is an added advantage

in the light of forecast purposes. These results stress how valuable the accuracy of a predictive model can be for strategic planning, to a great extent helping inform manufacturers and policyholders rather reliably about the direction taken by the expected market. This growth trajectory, in turn, will help the stakeholders make informed decisions on production, inventory, and investment in infrastructure.

The final result of this study chimes with the available literature on the Indian EV market that focuses on urban-centric growth, the role of government incentives, and the popularity of affordable vehicle types. This study furthers previous work as it quantifies these patterns and visually and predictively outlines future trends. The combination of association rule mining and time series forecasting yields a more subtle outlook beyond traditional market analysis. These insights light up very actionable opportunities for targeted marketing, product alignment, and regional prioritization-all in a way that will drive home a data-driven approach to meeting market demand.

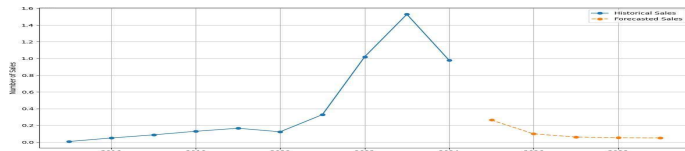


Fig. 8. Line Plot Showing The EV Sales Forecast Over The Years.

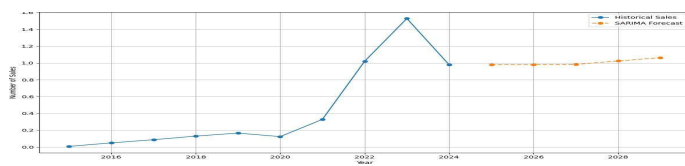


Fig. 9. Line Plot Showing The EV Sales Forecast Over The Years Using The SARIMA Model

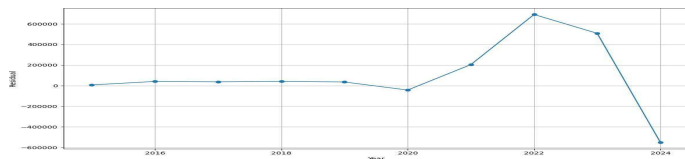


Fig. 10. Line Plot Showing The Residuals of SARIMA Model Over The Years

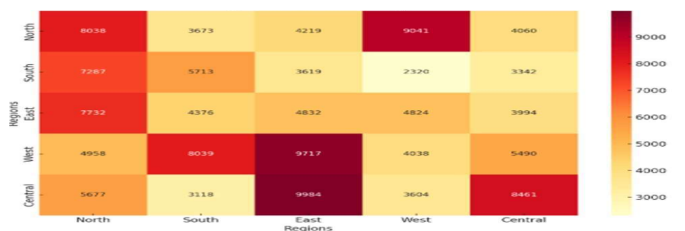


Fig. 11. Heatmap of Regional Sales

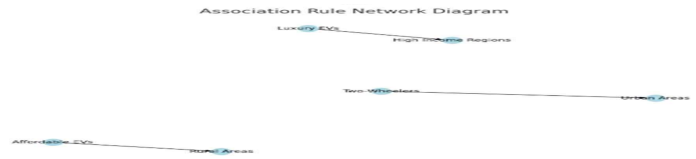


Fig. 12. Association Rule Network Diagram

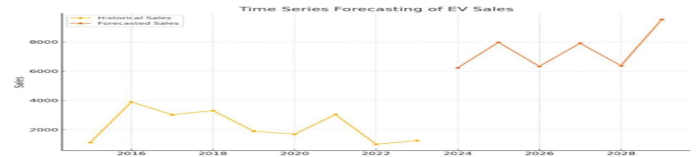


Fig. 13. Time Series Forecasting of EV Sales

The findings of the study, therefore, simply reaffirm that focused strategies will work in the Indian EV market. It would help the stakeholders in making informed decisions in tune with consumer preferences, regional demand, and growth forecasts on the basis of data-driven insights. These visualizations and quantitative analysis not only explain the present market scenario but also very clearly outline the road ahead in the future. This places the Indian EV sector in a strategic position to achieve sustainable growth and wide-scale adoption in the near future. Further research can build upon the present findings by incorporating more market factors, such as consumer sentiment and policy impacts, which would enrich understanding even further in writing about India's evolving EV landscape.

VI. CONCLUSION

The present research would, therefore, go a long way in enhancing our understanding of the EV market dynamics in India, using sophisticated machine learning algorithms that predict the future and thus provide actionable insight. Aggregating historical sales and operational data from varied sources-manufacturing companies, diverse regions-the study brings into light the patterns and trends that are important to understand the landscape of electric vehicles in India. Association rule mining pointed to the frequently hitting trends in various electric vehicle types and their region of occurrence to pinpoint high-demand areas and popular vehicle categories. It includes more urbanized regions such as Maharashtra and Delhi, which may be more affirmative toward the adaptation of these vehicles, while electric two-wheelers happen to be most popular due to their easy affordability and usability for intra-city rides. Forecasting of time series using ARIMA and Prophet models adds a predictive dimension to the present analysis, hence enabling more precise strategic planning by manufacturers and policymakers. This will fill not only the gap that previous studies have left, most of which were retrospective in nature and qualitatively assessed, but it will also increase the granularity and applicability of such findings to real-world applications. This paper is methodologically valid because it provides a reliable forecast by carefully

pre-processing data, engineering features, and evaluating the robustness of models, while offering a replicable model for similar researches in other emerging markets with rapid EV sector growth. Further refinement of these insights could be accomplished in future research through the inclusion of analyses of real-time policy shifts, changes, and improvements in charging infrastructure, and fluctuations in consumer sentiment that would further enhance the predictiveness and relevance of market forecasts. This not only helps to establish the study result in academic literatures but also encourages practical decision-making or policy formulation; it is truly a testament to the important role that data-driven analysis plays when trying to unpack the complex Indian EV market for the first time.

REFERENCES

- [1] X. Zhang, F. Guo, T. Chen, L. Pan, G. Beliakov, and J. Wu, "A Brief Survey of Machine Learning and Deep Learning Techniques for E-Commerce Research," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 18, no. 4. MDPI AG, pp. 2188–2216, Dec. 04, 2023. doi: 10.3390/jtaer18040110.
- [2] D. V. Pelegov and J.-J. Chanaron, "Electric Car Market Analysis Using Open Data: Sales, Volatility Assessment, and Forecasting," *Sustainability*, vol. 15, no. 1. MDPI AG, p. 399, Dec. 26, 2022.
- [3] V. Abella, J. Initan, J. M. Perez, P. V. Astillo, L. G. Cañete Jr., and G. Choudhary, "Machine Learning-Assisted Dynamic Proximity-Driven Sorting Algorithm for Supermarket Navigation Optimization: A Simulation-Based Validation," *Future Internet*, vol. 16, no. 8. MDPI AG, p. 277, Aug. 02, 2024.
- [4] G. Chen and Z. Li, "A New Method Combining Pattern Prediction and Preference Prediction for Next Basket Recommendation," *Entropy*, vol. 23, no. 11. MDPI AG, p. 1430, Oct. 29, 2021.
- [5] Kaushik, P. (2024). Dynamic Data Scaling Techniques for Streaming Machine Learning. *International Journal for Global Academic & Scientific Research*, 3(1), 1–12.
- [6] P. K. Gujarathi, V. A. Shah, and M. M. Lokhande, "Electric Vehicles in India: Market Analysis with Consumer Perspective, Policies and Issues," *Journal of Green Engineering*, vol. 8, no. 1. River Publishers, pp. 17–36, 2018.
- [7] M. Shibl, L. Ismail, and A. Massoud, "Machine Learning-Based Management of Electric Vehicles Charging: Towards Highly-Dispersed Fast Chargers," *Energies*, vol. 13, no. 20. MDPI AG, p. 5429, Oct. 17, 2020. doi: 10.3390/en13205429.
- [8] R. Rathore, P. Kaushik, S. S. Sikarwar, H. Joshi, A. K. Mishra and Y. Hudda, "Intelligent Transportation Systems Make Use of Fog and Edge Computing for Navigation," 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, 2024, pp. 1-6.
- [9] F. M. Talaat, A. Aljadani, B. Alharthi, M. A. Farsi, M. Badawy, and M. Elhosseini, "A Mathematical Model for Customer Segmentation Leveraging Deep Learning, Explainable AI, and RFM Analysis in Targeted Marketing," *Mathematics*, vol. 11, no. 18. MDPI AG, p. 3930, Sep. 15, 2023.
- [10] Pratap Singh Rathore, S. (2023). The Impact of AI on Recruitment and Selection Processes: Analysing the role of AI in automating and enhancing recruitment and selection procedures. *International Journal for Global Academic & Scientific Research*, 2(2), 51–63.
- [11] H. Patwary, M. T. Eshan, P. Debnath, and A. Sattar, "Market Basket Analysis Approach to Machine Learning," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, pp. 1–9, Jul. 06, 2021.
- [12] J.Y. Yeh and Y.T. Wang, "A Prediction Model for Electric Vehicle Sales Using Machine Learning Approaches," *Journal of Global Information Management*, vol. 31, no. 1. IGI Global, pp. 1–21, Jul. 31, 2023.
- [13] Shrivastava, R. K., & Rathore, R. (2024). Analysis of Single Server Markovian Queuing Model with Differentiated Working Vacation, Vacation Interruption, Soft Failure, Reneging of Customers. *International Journal for Global Academic & Scientific Research*, 3(3), 01–13.
- [14] Simsek, E. Koç, B. Desticioglu Tasdemir, A. Aksöz, M. Turkoglu, and A. Sengur, "Deep Learning Forecasting Model for Market Demand of Electric Vehicles," *Applied Sciences*, vol. 14, no. 23. MDPI AG, p. 10974, Nov. 26, 2024.
- [15] M. M. Kumbure, C. Lohrmann, P. Luukka, and J. Porras, "Machine learning techniques and data for stock market forecasting: A literature review," *Expert Systems with Applications*, vol. 197. Elsevier BV, p. 116659, Jul. 2022.
- [16] S. K. Dixit and A. K. Singh, "Predicting Electric Vehicle (EV) Buyers in India: A Machine Learning Approach," *The Review of Socionetwork Strategies*, vol. 16, no. 2. Springer Science and Business Media LLC, pp. 221–238, May 18, 2022.
- [17] R. Rathore, H. Yennapusa, S. Kumar, M. Mahawar, S. Miglani and G. Sharma, "Cutting-Edge Skin Cancer Detection in a Smart Ecosystem," 2023 International Conference on Smart Devices (ICSD), Dehradun, India, 2024, pp. 1-5.
- [18] S. Tufail, H. Riggs, M. Tariq, and A. I. Sarwat, "Advancements and Challenges in Machine Learning: A Comprehensive Review of Models, Libraries, Applications, and Algorithms," *Electronics*, vol. 12, no. 8. MDPI AG, p. 1789, Apr. 10, 2023.
- [19] S. Afandizadeh, D. Sharifi, N. Kalantari, and H. Mirzahosseini, "Using machine learning methods to predict electric vehicles penetration in the automotive market," *Scientific Reports*, vol. 13, no. 1. Springer Science and Business Media LLC, May 23, 2023.