International Journal of AI, Big Data, Computational and Management Studies
Noble Scholar Research Group | Volume 7, Issue 3, PP 1-6, 2026
ISSN: 3050-9416 | https://doi.org/10.63282/3050-9416.IJAIBDCMS-V7I3P101

Original Article

Machine Learning Framework for Electric Vehicle Customer Acquisition in the Automotive Industry

Vaibhav Tummalapalli
Atlanta, GA, USA.

Received On: 21/05/2026 ***Revised On: 15/06/2026*** ***Accepted On: 24/06/2026*** ***Published On: 01/07/2026***

Abstract**:** *The rapid growth of the EV market presents a major opportunity for automotive companies to expand their customer base and drive sustainable revenue. This paper outlines the development of a data-driven customer acquisition strategy for an automotive client—their first such initiative. The strategy leveraged historical EV sales trends to identify two key segments: existing brand customers and conquest customers. Initially focusing on existing customers, we built a scalable framework for future expansion. A machine-learning-based propensity model was developed using demographic and lifestyle attributes, sales and service history, charging infrastructure, state incentives, and fuel price trends. Key predictors included purchase recency, financial health, infrastructure availability, and technological affinity. Implemented end-to-end, from data preparation to model deployment. This paper explores the methodology, insights, challenges, and the broader impact of data science in advancing EV adoption*

Keywords: *Machine Learning, Data-Driven Marketing, Luxury Automotive, Customer Acquisition, Electric Vehicles (EV).*

1. Introduction

The global automotive industry is at the forefront of a paradigm shift, driven by the increasing adoption of Electric Vehicles (EVs). As governments push for stricter emission regulations and consumers demand more sustainable transportation options, EVs have emerged as a critical focus area for automakers. According to market forecasts, EV adoption is expected to grow exponentially, fueled by advancements in battery technology, expanding charging infrastructure, and substantial government incentives [1][3]. In this competitive and rapidly evolving landscape, automotive brands face the dual challenge of capturing market share while meeting the unique needs and preferences of EV buyers.

For many automotive companies, transitioning to EV sales requires rethinking traditional customer acquisition strategies. Unlike traditional gasoline vehicles, the decision to purchase an EV is influenced by a complex interplay of factors, including regional infrastructure, financial incentives, and lifestyle preferences. Understanding these nuances is key to successfully targeting potential EV buyers and achieving marketing efficiency.

This paper discusses the development of a data-driven customer acquisition strategy for EVs, implemented for a leading automotive client. The project aimed to capitalize on the growing EV market by identifying high-potential buyers and optimizing the client's marketing efforts to maximize return on investment (ROI). By analyzing historical sales trends, segmenting target audiences into existing and conquest customers, and leveraging advanced machine learning models, the strategy not only provided actionable insights but also established a scalable framework for future campaigns.

Through this initiative, we demonstrated how data science can serve as a strategic enabler for the automotive industry, driving innovation and measurable business impact in the rapidly growing EV market.

2. Challenges in EV Targeting

The increasing adoption of Electric Vehicles (EVs) presents a unique opportunity for automakers to capture market share in a rapidly evolving segment. However, this opportunity is accompanied by significant challenges in identifying and acquiring customers for EVs. Unlike traditional gasoline-powered vehicles, purchasing an EV involves a highly complex decision-making process influenced by factors such as financial incentives, regional charging infrastructure, and individual lifestyle preferences [2].

2.1. Why a Customer Acquisition Strategy for EVs Was Necessary

Traditional customer acquisition strategies, designed for internal combustion engine (ICE) vehicles, are often insufficient for targeting EV buyers. EV adoption is influenced by a distinct set of factors, including environmental awareness, the availability of government subsidies, and perceived cost-benefit trade-offs [4][7]. Automotive brands cannot rely solely on existing methods to identify potential buyers; instead, they require a tailored, data-driven approach that considers the unique dynamics of the EV market.

Furthermore, as the market becomes increasingly competitive, early movers have the opportunity to build strong brand loyalty among EV buyers. For this reason, automakers must proactively understand their target audience and refine

their marketing efforts to maximize ROI, ensuring they remain ahead of competitors in the race for EV adoption.

2.2. Challenges in Identifying and Targeting EV Buyers

- Diverse Buyer Personas: EV buyers do not fit a single demographic or psychographic profile. They range from environmentally conscious consumers to tech-savvy early adopters and cost-conscious commuters. Targeting these diverse groups requires advanced analytics.
- Regional Disparities: The availability of charging infrastructure and financial incentives varies widely by region, directly impacting the attractiveness of EVs for potential buyers. This regional variability complicates the identification of high-potential markets.
- Limited Historical Data: EVs represent a relatively new segment, meaning historical sales data is sparse or inconsistent, making it difficult to build predictive models using traditional techniques.
- Behavioral and Psychological Barriers: Range anxiety, concerns about battery longevity, and a lack of familiarity with EV technology create psychological barriers that further complicate customer acquisition efforts.
- Balancing Existing and Conquest Customers: Automakers must strategically balance efforts to retain existing customers transitioning to EVs while also targeting conquest customers who may be loyal to competing brands or unfamiliar with EVs.

These challenges highlight the necessity for a comprehensive, data-driven customer acquisition strategy tailored to the unique characteristics of the EV market. By addressing these issues, automakers can not only increase EV sales but also build a foundation for long-term growth in a sustainable automotive future.

3. Analytical Framework for Customer Repurchase

To ensure precise targeting in the customer acquisition strategy, I developed an analytical framework that incorporated a wide range of data sources for analysis. The framework aimed to identify individuals likely to be in the market for an Electric Vehicle (EV) by testing the predictive power of various factors influencing EV adoption. This process involved developing a propensity model trained to differentiate between EV buyers and non-buyers, while carefully selecting only the most significant predictors for the final model.

The analysis considered the following data sources:

- Demographics and Lifestyle Attributes: Variables such as income, education, and environmental consciousness were analyzed for their ability to capture EV adoption propensity. This data provided a 360-degree view of each prospect, offering comprehensive insights through 1,500 attributes per individual. This dataset covered critical dimensions such as demographics (age, household composition, ethnic background etc.), lifestyle factors (interests in books, gardening, travel, sports etc.), and financial health (income, credit, debt, asset information etc.). Additionally, it included market activity indicators (transaction behaviors, spending trends) and automotive data (vehicle ownership details)
- Historical Sales and Service Transactions: Patterns in past vehicle purchases, service interactions, and spending behavior proved to be highly predictive.
- Charging Infrastructure Data: Proximity to EV charging stations was hypothesized to play a key role and was validated as a significant factor influencing adoption [1][3].
- State-Level Incentives and Gasoline Prices: These macroeconomic variables were included in the initial analysis to assess their relevance but were ultimately found to have limited predictive power in this specific context.

Only the most predictive variables, based on their statistical significance and business relevance, were included in the final model. This rigorous feature selection process ensured that the model focused on key drivers of EV adoption while minimizing complexity. By integrating these insights, the framework provided actionable recommendations for targeting high-potential customers and maximizing the campaign's return on investment.

3.1. Data Structure & Machine Learning Framework

The development of the machine learning framework for predicting Electric Vehicle (EV) repurchase required a carefully designed approach to structuring data and formulating the problem. The framework aimed to effectively leverage temporal customer behavior data and seasonality patterns to build a predictive model.

To create a robust foundation for analysis and model development, the first step was organizing the data into cohorts. Cohort-based structuring [14] was chosen to capture time-specific customer behaviors and seasonality effects, which are critical in understanding EV purchase patterns.

Cohort Definition: Each cohort consisted of customers observed at a specific time point, and their behaviors were tracked across a predefined performance window (PW) of 8 months.

Outcome Labeling for model development:

- Customers who purchased an EV within the 8-month performance window were flagged as 1 (buyer).
- Customers who did not purchase were labeled as 0 (non-buyer).

Rationale for Performance Window: Various performance windows (e.g., 8 months, 10 months) were tested to evaluate their sufficiency for modeling. The 8-month window was selected as it demonstrated the highest balance between response frequency and data sufficiency, while the 10-month window was deemed too long

Contextual Factors for EV Purchase Timelines:

- Learning Curve: Buyers often take time to understand EV benefits, technology, and cost considerations.
- Range Anxiety: Concerns about driving range and charging infrastructure availability extend the decision-making timeline.
- Incentives: Variations in the availability of tax credits and subsidies influence the purchase window.

Features Representing Temporal Behaviors: Cohorts allowed the creation of time-specific features to capture patterns in customer behavior, including:

- Recency: Time since the last service or purchase.
- Frequency: Number of transactions within a specified period.
- Trend Indicators: Changes in sales & service transactions over time.

3.2. Evaluating Metrics for the model.

When evaluating models for customer targeting, metrics like Lift and Cumulative Capture Rate are more effective than traditional metrics like accuracy or confusion matrices. This is because these models prioritize ranking prospects by their likelihood to convert, rather than just classifying them as buyers or non-buyers.

Deciles: The records are ranked by predicted likelihood and divided into 10 equal groups. The baseline for comparison is a random targeting approach, where lift is always 1, and the capture rate follows a linear path.

Lift: Measures how much more likely prospects chosen by the model are to respond compared to random targeting. A lift of 1 represents random targeting, and higher lift values in the top deciles indicate better model performance. The lift in each decile is calculated as follows.

$$\text{Lift} = \frac{\text{Response Rate of Decile}}{\text{Overall Response Rate (Baseline)}}$$

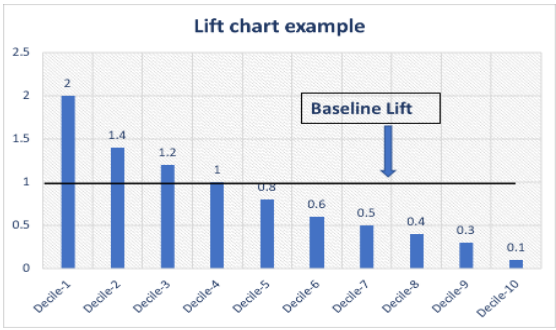

Fig 1: Example Lift Chart

Here is an example of a Lift Chart that illustrates how the model's lift is distributed across deciles. The lift in each decile represents how much more likely the model's predictions are to capture buyers compared to random selection. The x-axis represents the decile, and the y-axis represents the lift in each decile.

In this example, decile 1 (top-scoring prospects) shows a lift of 2, meaning that the prospects in this decile are 2 times more likely to respond than a random sample of the same size. As you move to lower deciles, the lift decreases, which indicates that lower-scoring prospects are progressively less likely to respond. The baseline lift (represented by the black line at 1) represents the random targeting approach, where the response rate is equal to the overall average.

Capture Rate: Shows the cumulative percentage of total buyers captured by the model in each decile. Higher capture rates in early deciles reflect the model's ability to focus on high-potential buyers.

$$CR = \frac{Number\ of\ Buyers\ captured\ by\ the\ model}{\text{Total number of buyers}}$$

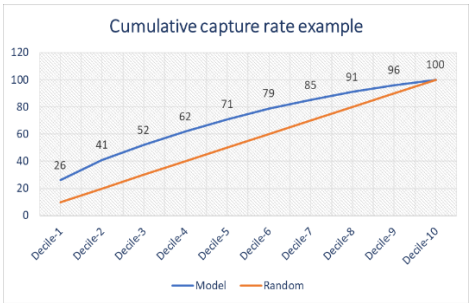

Fig 2: Example Cumulative Capture Rate Chart

Here is an example of a Cumulative Capture Rate chart that illustrates the model's ability to identify and capture buyers compared to a random selection approach. The capture rate shows the cumulative percentage of buyers identified by the model across each decile. The x-axis represents the decile, and y-axis represents the cumulative capture rate. The blue line represents the model's performance, while the orange line reflects random targeting.

In this example, Decile 1 captures 26% of the total buyers, whereas random targeting captures only 10% in the same decile. As we move through the deciles, the model consistently captures a higher percentage of buyers compared to random selection, indicating its effectiveness in concentrating high-potential leads in the top deciles.

3.3. Data preparation, processing & ML Model development

- Sampling: The first step in the process is sampling the data. Despite having access to millions of data points, computational resource constraints often necessitate sampling for model development. Stratified sampling by cohort is employed to ensure that the proportions of observations and event rates in the sample remain consistent with the overall population. This approach maintains the integrity of the dataset while making it computationally manageable [7].

3.3.1. Data Formatting and Feature Creation

- After sampling, the raw data from sales, service, and demographic sources undergoes formatting. This includes:
- Renaming and standardizing field names.
- Dropping irrelevant fields for modeling purposes.
- Converting fields to the appropriate format for analysis.
- Aggregating categorical variables with excessive levels into broader categories based on business logic.

Feature Creation: Customer historical data is aggregated to generate features across dimensions like recency, frequency, and monetary value [8]. Examples of features include:

- Purchase history (items purchased, spending amount, and purchase frequency).
- Service transactions (last service date, types of services availed, and cost).
- Behavioral attributes (mileage driven and service intervals).

3.3.2. Data cleaning and Pre-processing

1. Data Partitioning: The dataset is split into training and validation subsets. The partitioning ratio depends on dataset size and computational resources, ranging from 90:10 for large datasets to 60:40 or 70:30 for smaller datasets [13].
2. Data Cleaning
   - Column Elimination: Fields with low variance, high missing percentages, or single values are dropped as they provide little to no predictive power.
   - Missing Value Thresholds: Columns with missing percentages exceeding predefined thresholds (e.g., 40%-50%) are removed, while others are treated with imputation techniques.
   - Categorical Feature Treatment: Encoding methods, such as ordinal encoding, one-hot encoding, or grouping rare categories based on event rates, are applied depending on the variable type.

Outlier Detection and Imputation Outlier Treatment:

- Gaussian-distributed variables are treated using standard deviation (STD) methods.
- Non-Gaussian variables are treated using Median Absolute Deviation (MAD) or Interquartile Range (IQR).
- For business-specific fields (e.g., payment data), manual thresholds are applied.[11]

Imputation Techniques:

- Simple imputation (mean/median/mode) is used for low missing percentages (<10%).
- Advanced methods like KNN imputation or Multivariate Imputation by Chained Equations (MICE) are applied for higher missing percentages, ensuring data completeness without bias [9][10].

Variable Selection**:** To reduce the dataset's dimensionality, Statistical methods like IV, Chi-Square, and ANOVA are used for initial filtering. Machine learning methods such as Recursive Feature Elimination (RFE) with tree-based models (Decision Trees, Random Forest, Gradient Boosting Machines) provide a robust variable selection process. Redundant variables are further identified and removed using Variance Inflation Factor (VIF) or correlation analysis [12][13].

- Feature Transformations: Transformations are applied to optimize model input, including log, square root, power transforms, or optimal binning, based on chi-square values relative to the target variable. These transformations are particularly beneficial for linear models like logistic regression.
- Modeling: The modeling process begins with a baseline model (e.g., logistic regression or decision tree) to establish a reference. Subsequently, advanced models, including tree-based algorithms and neural networks, are evaluated [13].
- Model Evaluation and Comparison: The models are assessed using metrics such as KS Statistic, Lift and Capture Rate. The model with the highest performance metrics is chosen as the champion model.
- Interpretability: Once a champion model is selected, its interpretability is assessed using tools like SHAP values to understand feature contributions to predictions. Partial Dependence (PD) plots to visualize relationships between input variables and the target.
- Back Testing: To validate model generalization, a portion of the dataset (The latest cohort population) is set aside at the beginning of the process for back testing. This unseen dataset ensures that the model performs consistently across training, validation, and test subsets [13]
- Deployment & Monitoring: Once the champion model was finalized, it was deployed into the production environment to support the customer acquisition strategy. Continuous monitoring was implemented to track the model’s performance over time, with key metrics such as conversion rates, Lift, and Cumulative Capture Rate being reviewed periodically.
- This approach allows for timely adjustments, ensuring the model adapts to changes in consumer behavior. If necessary, the model is retrained to maintain accuracy and relevance, ensuring the ongoing success of the campaign [13].

4. Model Results

4.1. Training & Validation results

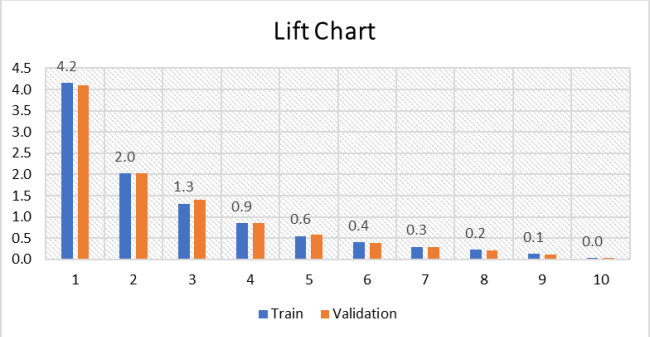

Fig 3: Model Lift – Train & Validation

Lift Analysis:

- Top Decile: Achieves a lift of 4.2, capturing 42% of responders, indicating strong predictive accuracy.
- Top 3 Deciles: Capture 75% of total responders, demonstrating effective prioritization.
- Model Stability: Consistent lift values between training and validation datasets confirm reliability and generalizability [13].

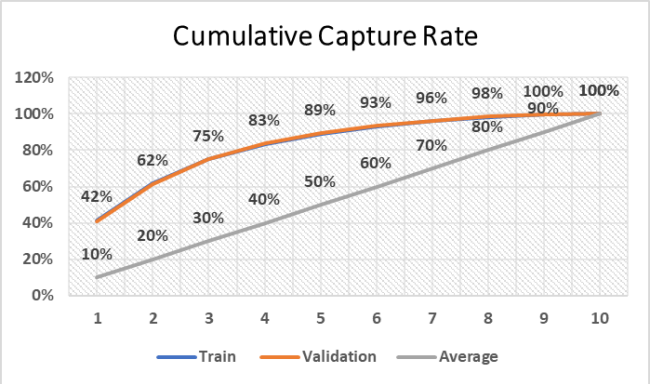

Fig 4: Model Cumulative Capture Rate – Train & Validation

Capture Rate Analysis:

- Top Decile: Captures 24% of responders, outperforming random selection.
- Top 3 Deciles: Cover 56% of total responders, balancing precision and coverage.
- Consistency: Similar capture rates for training and validation highlight robust model performance.

Key Takeaways:

- The model effectively targets high-propensity customers, optimizing marketing efforts.
- Stability across training and validation ensures scalability and reliability for future cohorts.
- Focused campaigns targeting the top deciles will yield maximum impact.

4.2. Back test results

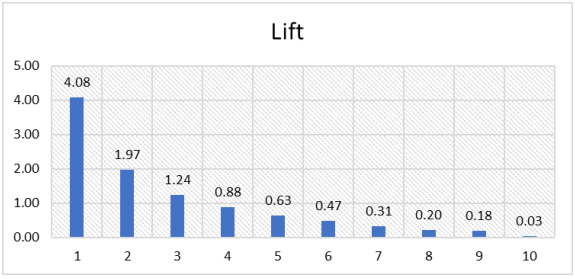

Fig 5: Lift – Back test results

The model demonstrates strong performance on back-test data, maintaining trends consistent with training and validation datasets. The top 3 deciles successfully capture 56% of total responders, highlighting its effectiveness in prioritizing high-propensity customers**.**

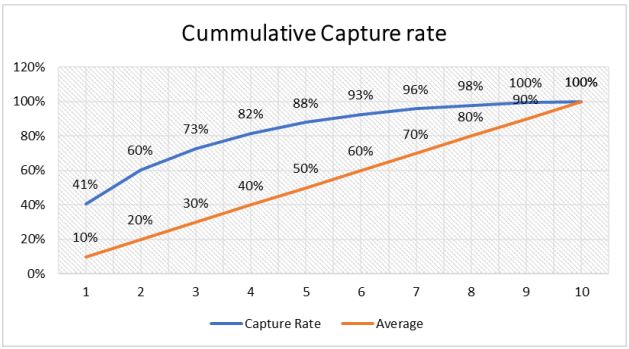

Fig 6: Cumulative Capture Rate – Back Test Results

5. Insights from the model

The predictive model developed for this study identified key variables influencing the likelihood of customers purchasing electric vehicles (EVs). These insights provide actionable takeaways for designing targeted marketing strategies and improving customer engagement. The top variables and their implications are as follows:

- Historical Purchase Frequency and Recency: The model highlighted that individuals who recently purchased a vehicle (within the past 2-3 years) show a higher propensity for considering an EV. This behavior is likely driven by their inclination to remain current with the latest advancements in automotive technology and their desire to own newer models.
- Financial Health: Financial stability emerged as a critical factor, as EVs are generally more expensive compared to conventional vehicles. Customers with high net worth, income, educational attainment, and credit scores demonstrated a higher likelihood of EV adoption [1]. These findings emphasize the importance of targeting financially secure demographics for EV marketing efforts.
- Infrastructure Accessibility: The availability of charging infrastructure was another significant determinant. Customers residing in states with extensive and accessible charging networks exhibited a higher probability of purchasing EVs [6].

This finding underscores the role of infrastructure in supporting EV adoption and suggests that marketing efforts could be concentrated in regions with well-developed charging ecosystems.

- Interest in Autonomous Vehicles and Technology Adoption: Individuals with a demonstrated interest in autonomous vehicle technology and a general predisposition towards embracing new technologies also showed a higher propensity for EV adoption [5]. Although not all EVs offer autonomous features, the association between EVs and cutting-edge technology likely drives this interest.

6. Conclusion

This study highlights the power of predictive modeling in understanding EV purchase behavior. Key factors such as purchase recency, financial stability, gender preferences, infrastructure accessibility, and interest in technology were identified as critical drivers. These insights provide a strategic foundation for targeting high-propensity customers and optimizing marketing efforts. By leveraging data-driven approaches, luxury auto brands can effectively position themselves in the evolving EV market and drive customer adoption

References

1. S. Afandizadeh, D. Sharifi, N. Kalantari, and H. Mirzahossein, "Using machine learning methods to predict electric vehicle penetration in the automotive market," *Scientific Reports*, vol. 13, p. 8345, 2023. [Online]. Available: https://www.nature.com/articles/s41598-023-35366-3
2. 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, 2023. [Online]. Available: https://www.igi-global.com/article/a-prediction-model-for-electric-vehicle-sales-using-machine-learning-approaches/327277
3. Z. Li, H. Fan, and S. Dong, "Electric Vehicle Sales Forecasting Model Considering Green Premium: A Chinese Market-based Perspective," *arXiv preprint*, arXiv:2302.13893, 2023. [Online]. Available: https://arxiv.org/abs/2302.13893
4. Yeğin, Tuğba & Ikram, Muhammad. (2022). Analysis of Consumers' Electric Vehicle Purchase Intentions: An Expansion of the Theory of Planned Behavior. Sustainability. 14. 10.3390/su141912091.
5. Lilhore, Aakash & Prasad, Kavita & Agarwal, Vivek. (2023). Machine Learning-based Electric Vehicle User Behavior Prediction. 1-6. 10.1109/GlobConHT56829.2023.10087780.
6. N. Arechiga, F. Chen, R. Iliev, and A. Molnar, "Understanding and Shifting Preferences for Battery Electric Vehicles," *arXiv preprint*, arXiv:2202.08963, 2022. [Online]. Available: https://arxiv.org/abs/2202.08963
7. Tummalapalli Vaibhav. (2025). Stratified sampling in Cohort-based data for Machine learning Model development. International Scientific Journal of Engineering and Management. 04. 1-8. 10.55041/ISJEM03377
8. V. Tummalapalli, "Feature Engineering for Building Machine Learning Models in Automotive Industry," International Scientific Journal of Engineering and Management, vol. 4, no. 8, pp. 1–9, 2025. doi: 10.55041/ISJEM04985.
9. V. Tummalapalli, "Comprehensive study of data imputation techniques for machine learning models," International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences, vol. 13, no. 4, 2025, doi: 10.37082/IJIRMPS.v13.i4.232674
10. V. Tummalapalli, "Understanding distance metrics in KNN imputation: Theoretical insights and applications," Journal of Mathematical & Computer Applications, vol. 4, no. 4, pp. 1–4, 2025. doi: 10.47363/JMCA/2025(4)208.
11. Vaibhav Tummalapalli. (2025). Outlier Detection & Treatment for Machine Learning Models. International Journal of Innovative Research and Creative Technology, 11(3), 1–8. https://doi.org/10.5281/zenodo.16500050
12. V.Tummalapalli and K. Konakalla, "Statistical Techniques for Feature Selection in Machine Learning Models," International Journal for Innovative Research in Multidisciplinary Pursuit and Studies (IJIRMPS), vol. 13, no. 3, pp. 1-8, 2025, doi: 10.37082/IJIRMPS.v13.i3.232566
13. V. Tummalapalli, "Machine learning pipeline for automotive propensity models," International Journal of Core Engineering & Management, vol. 8, no. 3, 2025, ISSN 2348-9510
14. https://ijcem.in/wp-content/uploads/MACHINE-LEARNING-PIPELINE-FOR-AUTOMOTIVE-PROPENSITY-MODELS.pdf
15. Tummalapalli, V. (2026). Cohort-Based Segmentation Framework for Machine Learning: Structuring Temporal Data for Enhanced Feature Engineering. International Journal of Intelligent Data and Machine Learning, 3(03), 05-17. https://doi.org/10.55640/ijidml-v03i03-02
16. Veershetty, G. (2026). Automated Root Cause Analysis in SAP Landscapes Using Large Language Models and Operational Telemetry. International Journal of Emerging Trends in Computer Science and Information Technology, 7(1), 186-191. https://doi.org/10.63282/3050-9246.IJETCSIT-V7I1P127
17. Shashank, A. (2025). AI-Enhanced ETL Processes: Leveraging Artificial Intelligence for Optimized Data Integration Systems. Journal Of Multidisciplinary, 5(8), 219-225.
18. Kaur, M., Bonkra, A., Verma, R., Khanna, N., Maken, P., & Sunkara, S. K. (2025). Comparative study of traditional and hybrid models in short-term financial forecasting using machine learning. In Innovations in Computing (pp. 13-18). CRC Press.