



# Predictive Modeling for Identifying Potential EV Buyers Using Demographic and Behavioral Data

Dr. V. Kanimozhi<sup>1</sup> and S. Kaamesh Kumar<sup>2</sup>

[<sup>1</sup>] Associate Professor, Department of Management Sciences, Hindusthan college of engineering and technology, Coimbatore, Tamil Nadu.

[<sup>2</sup>] Student, Department of Management Sciences, Hindusthan college of engineering and technology, Coimbatore, Tamil Nadu.

## How to Cite this Article:

Kumar, S. K. (2026). Predictive Modeling for Identifying Potential EV Buyers Using Demographic and Behavioral Data. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(05). <https://doi.org/10.55041/ijcope.v2i4.1047>

## License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i4.1047>

## ABSTRACT

The automotive industry's transition to electric vehicles (EVs) requires moving beyond early adopters to capture the pragmatic "Early Majority"—a task at which traditional demographic-based marketing has proven inefficient. This study develops a predictive machine learning model integrating demographic, behavioral, and psychographic data to identify potential EV buyers. Primary survey data (n=251) was collected and used to train four classification models, with Random Forest emerging as the best performer (accuracy: 61%, AUC-ROC: 0.513). Feature importance analysis revealed that occupation, current vehicle type, EV feature importance, carbon footprint importance, and perceived barriers were stronger predictors than income or age. Cost (39.8%), charging infrastructure (25.1%), and range anxiety (23.5%) were identified as the top adoption barriers, while range (40.6%) and charging speed (25.5%) were the most valued features. Critical awareness gaps were uncovered: 61.8% were unfamiliar with government incentives, and 63.4% found charging stations inaccessible. The model enables precise targeting for marketers and provides policymakers with data-driven insights for infrastructure and incentive planning. Despite modest predictive power and low recall, this study demonstrates that integrating psychographic and behavioral factors offers a more accurate foundation for EV adoption modeling than traditional demographic profiling alone.

**Keywords:** Electric Vehicle adoption, predictive modeling, Random Forest, consumer behavior, Early Majority, psychographic segmentation



## INTRODUCTION

The automotive industry is shifting from Internal Combustion Engine vehicles to Electric Vehicles (EVs). While early adopters have embraced EVs for technology and environmental reasons, mass adoption requires winning over the "Early Majority"—pragmatic, risk-averse consumers focused on cost, reliability, and convenience. Traditional marketing strategies relying on broad demographics (income, age) are failing to identify *why* someone might buy an EV, resulting in wasted budgets and missed opportunities.

In India, EVs comprise less than 1% of vehicle sales, though the government aims for 30% EV penetration in private cars by 2030. Current research remains fragmented, using data types without strong behavioral theory. This study develops a predictive machine learning model integrating demographic, behavioral, and psychographic data to identify potential EV buyers, determine key purchase predictors, and uncover adoption barriers—enabling efficient targeting and accelerating sustainable transportation

## STATEMENT OF THE PROBLEM

Automotive marketing is failing to keep pace with the EV transition. Traditional strategies targeting broad demographics (income, age) are inefficient for the complex EV purchase decision, which involves environmental values, charging anxiety, and total cost of ownership. This leads to four critical issues: high marketing costs wasted on unmotivated buyers, inability to identify the pragmatic "Early Majority," ineffective generic messaging, and poor strategic planning for inventory and infrastructure. The industry lacks a precise, data-driven method to identify which mainstream consumers are actively considering an EV.

## NEED OF THE STUDY

This study is needed because traditional demographic-based marketing is failing to reach the pragmatic "Early Majority" consumers essential for mass EV adoption. Current strategies waste budgets by targeting the wrong people while missing potential buyers, as demographics reveal *who* a person is but not *why* they would buy an EV. Existing research remains fragmented, using single data types without strong behavioral theory, and crucially fails to classify the large "neutral" consumer segment. Additionally, policymakers lack data-driven insights for optimal charging infrastructure and incentive design. Therefore, this study develops a predictive model integrating demographic, behavioral, and psychographic data to accurately identify potential EV buyers, reduce marketing waste, fill research gaps, and accelerate sustainable transportation.

## OBJECTIVES

- To develop a predictive model that identifies potential EV buyers by analyzing demographic, behavioral, environmental, social, and infrastructural factors.
- To determine the strongest predictors of EV purchase intention, including age, income, commute distance, charging accessibility, and social circle EV ownership.
- To identify the most significant barriers (such as cost, range anxiety, and charging infrastructure) that differentiate likely buyers from unlikely buyers for targeted interventions.

## SCOPE OF THE STUDY

This study focuses on developing a predictive machine learning model to identify potential electric vehicle (EV) buyers by integrating demographic, behavioral, and psychographic factors within the context of the Indian automotive market. The research targets adult vehicle owners aged 18 years and above who hold a valid driver's license and have decision-making authority regarding vehicle purchases. The study analyzes demographic factors (age, income, education, occupation), behavioral factors (current vehicle type, commute distance, fuel efficiency importance), psychographic factors (climate concern, technology comfort, EV perception), social factors (social circle EV ownership), and infrastructural factors (charging accessibility, government incentive



familiarity). Using survey data from 251 respondents, four classification models—Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting—are trained and evaluated. The study excludes real-time economic variables (fuel prices, interest rates), actual purchase behavior tracking, commercial fleet buyers, and comprehensive regional variations within India, focusing instead on individual consumer targeting for marketing applications.

## RESEARCH METHODOLOGY

### Research Design

The research follows a quantitative, predictive modeling approach comprising four phases. The first phase involves Exploratory Data Analysis and Preprocessing, which includes cleaning and integrating demographic, behavioral, and psychographic data. The second phase is Model Development and Training, where Logistic Regression, Decision Tree, Random Forest, and XGBoost models are trained. The third phase consists of Model Evaluation and Interpretation, where models are tested, compared, and interpreted using SHAP analysis. The final phase is Validation and Application, which generates "EV Propensity Scores" for targeted marketing campaigns.

### Sampling Design

The target population comprises adult individuals aged 18 years and above who currently own or lease a vehicle, hold a valid driver's license, and have decision-making authority regarding vehicle purchases. The sampling frame was constructed using online consumer panels, social media automotive groups, personal and professional networks, and community organization email lists. A convenience sampling technique was employed due to time and resource constraints. A total of 251 valid responses were collected, exceeding the minimum requirement for a 95% confidence level with approximately 6.7% margin of error. To minimize bias, the survey link was shared across multiple platforms including Facebook, WhatsApp, LinkedIn, and email with no incentives offered, and early respondents were compared with late respondents with no significant differences found.

### Data Collection

Google Forms linked to Google Sheets was used as the primary data collection tool for real-time data capture. The Google Form was designed with logical grouping of questions, required fields to prevent incomplete submissions, skip logic for relevance, and an embedded attention check question to identify random responders. The form was tested on multiple devices before distribution. The link was distributed through email invitations, social media platforms (Facebook and LinkedIn), WhatsApp automotive groups, and community forums. Responses were monitored daily, and data quality control measures included validation rules for numerical ranges and exclusion of responses failing the attention check. Upon completion, data were exported as Excel and CSV files and stored on password-protected local computers with backup on encrypted USB drives.

### Data Analysis Methodology

Pandas and NumPy were used for data manipulation, Matplotlib and Seaborn for visualizations, and Scikit-learn for machine learning models. The target variable was created by coding "Very likely" and "Likely" responses as 1 (potential buyer) and "Neutral," "Unlikely," and "Very unlikely" as 0 (non-buyer). Feature engineering produced new features including numeric income, numeric commute distance, high income indicator, long commute indicator, high climate concern indicator, and social circle EV ownership indicator. The dataset was split into 70% training and 30% testing using stratified sampling, and Standard Scaler was applied for feature standardization. Four classification models—Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting—were trained and compared. Random Forest was selected as the best performing model based on accuracy and AUC-ROC scores, with feature importance analysis revealing top predictors including income, commute distance, and climate concern.



## FINDINGS

**Model Performance:** The Random Forest classifier emerged as the best-performing model, achieving the highest accuracy of 61% and the most stable cross-validation score of 0.6057 across five folds. However, the AUC-ROC score of 0.513 indicates modest predictive power, only slightly better than random guessing. The model demonstrated high specificity, correctly identifying 173 true non-buyers with only 13 false positives, but low recall, capturing only 10 out of 114 actual buyers (9%).

**Top Predictors:** Feature importance analysis revealed that occupation (0.055) was the strongest predictor of EV purchase intention, followed by current vehicle type (0.054), EV feature importance (0.053), carbon footprint importance (0.051), and perceived EV barriers (0.049). Other notable predictors included tech comfort level, trusted information source, EV perception, preferred price range, and age group.

**Buyer Profile:** Potential EV buyers had significantly higher median household income (100,000 vs. 62,500 for non-buyers) and longer commute distances. Among current electric vehicle owners, 31.8% were classified as likely buyers, followed by 24.2% of gasoline owners and 23.4% of hybrid owners. Additionally, 42.1% of buyers reported being "very" or "extremely concerned" about climate change, compared to lower levels among non-buyers. The 36-45 age group showed a slightly higher buyer proportion.

**Key Barriers:** Cost (39.8%) was identified as the single most significant barrier to EV adoption, followed by charging infrastructure (25.1%) and range anxiety (23.5%). Technology reliability (6.0%) and other factors (5.2%) were cited by very few respondents.

**Desired Features:** Range (40.6%) was the most valued EV feature, followed by charging speed (25.5%) and price (18.3%). Performance (11.6%) and environmental impact (3.6%) were least important to respondents.

**Infrastructure & Awareness Gaps:** A large majority (63.4%) of respondents found charging stations inaccessible or very inaccessible. Nearly two-thirds (61.8%) were unfamiliar with government incentives for EV adoption. Independent reviews (37.1%) were the most trusted source of EV information, while social media (10.4%) and word of mouth (9.6%) were the least trusted.

**Preferred Price Range:** The most preferred EV price range was 20,000 – 35,000 (33.935,001-50,000 (26.3%) and under 20,000 (25.565,000 (4.4%).

**Purchase Intention Distribution:** The largest group of respondents was neutral (23.5%) regarding EV purchase likelihood, followed closely by unlikely (23.1%), likely (20.7%), very likely (19.9%), and very unlikely (12.4%). Overall, 59% leaned toward uncertainty or unwillingness compared to 40.6% likely or very likely to purchase.

## SUGGESTIONS

### For Marketers:

Shift from demographic to psychographic targeting. Stop relying solely on broad demographics like income or age. Instead, use the identified key predictors—occupation, current vehicle type, feature importance, and perceived barriers—to create precise, lookalike audiences for targeted digital campaigns.

Focus marketing spend on independent reviews and government agency channels. Avoid heavy investment in social media or word-of-mouth campaigns for initial EV education, as these are the least trusted sources.

Emphasize real-world range and charging speed over environmental benefits in advertising, as practical features far outweigh environmental impact in purchase decisions for the Pragmatic Majority.



### **For Policymakers:**

Launch an aggressive, multi-channel awareness campaign using trusted sources (government websites and independent media) to educate consumers about existing EV incentives, as current unfamiliarity is extremely high at 62%.

Direct funds to significantly expand public charging accessibility, particularly in areas where respondents report "inaccessible" or "very inaccessible" stations, since charging infrastructure is a top-three barrier for nearly one-quarter of consumers.

Design income-tiered subsidies to make EVs affordable for the 20,000 –35,000 preferred price range, which was the most popular choice among 33.9% of respondents.

### **For Automakers:**

Develop and market EV models in the 20,000 –35,000 price bracket to capture the largest segment of potential buyers.

Prioritize improvements in range and charging speed over performance or environmental features, as these are the most valued attributes by consumers.

Create referral programs or ambassador initiatives that leverage existing EV owners within social circles, as social influence is a notable predictor of purchase intention.

### **For Future Research:**

Address class imbalance using techniques like SMOTE (Synthetic Minority Over-sampling Technique) to improve the model's low recall (9%) and better capture actual potential buyers.

Incorporate real-time economic variables such as fuel prices, electricity tariffs, and interest rates to enhance predictive accuracy.

Track actual purchase behavior over time through longitudinal follow-up studies rather than relying solely on stated purchase intentions.

## **CONCLUSION**

predictive machine learning model to identify potential electric vehicle buyers by integrating demographic, behavioral, and psychographic data. The Random Forest classifier emerged as the best-performing model, achieving 61% accuracy, though the modest AUC-ROC score of 0.513 indicates that predictive power remains limited with the current feature set. Critically, the research demonstrated that psychographic and behavioral factors—particularly occupation, current vehicle type, EV feature importance, carbon footprint importance, and perceived barriers—are stronger predictors of purchase intention than traditional demographic variables like income and age. The analysis revealed that potential EV buyers tend to have higher median income (100,000 vs. 62,500), longer commutes, greater climate concern (42.1% very or extremely concerned), and are more likely to already own electric or hybrid vehicles. Cost (39.8%), charging infrastructure (25.1%), and range anxiety (23.5%) remain the most significant adoption barriers, while range (40.6%) and charging speed (25.5%) are the most valued features. A major awareness gap was uncovered, with 62% unfamiliar with government incentives and 63.4% finding charging stations inaccessible. Despite limitations including small sample size, convenience sampling, and low model recall (9%), this research provides a valuable framework for moving beyond ineffective demographic profiling. The findings enable marketers to target based on occupation, vehicle type, and feature preferences rather than broad demographics, while policymakers gain data-driven insights for infrastructure investment and incentive awareness campaigns. With future refinement using larger datasets, class



imbalance techniques like SMOTE, and real-time economic variables, this predictive modeling approach can evolve into a production-ready tool for accelerating EV adoption among the pragmatic mainstream market, ultimately contributing to sustainable transportation goals.

## **IMPLICATIONS**

### **Improve Model Performance with Advanced Techniques**

Future work should apply techniques such as SMOTE (Synthetic Minority Over-sampling Technique), balanced random forests, or anomaly detection frameworks to specifically address the severe class imbalance that caused low recall (only 9%). Collecting a larger, more balanced dataset with at least 1,000+ responses is critical for improving model sensitivity.

### **Incorporate Real-Time Economic Variables**

The study did not include dynamic economic factors such as petrol/diesel prices, electricity tariffs, inflation rates, or interest rates on auto loans. Future models must integrate these variables as they significantly influence vehicle purchase decisions, particularly for cost-sensitive consumer segments.

### **Track Actual Purchase Behavior**

Move beyond stated intentions by conducting longitudinal follow-up studies with survey respondents to track who actually buys an EV. This will enable development of "revealed preference" models that are more robust than "stated preference" models and reduce the intention-behavior gap.

### **Expand Geographical Coverage**

Test and validate the predictive model across different regions of India (urban vs. rural, tier-1 vs. tier-2 cities) and different countries to develop context-specific theoretical models. Regional variations in infrastructure, policy, and consumer attitudes must be captured.

### **Experiment with Advanced Algorithms**

Test more sophisticated algorithms including XGBoost with hyperparameter tuning, neural networks, and hybrid deep learning models (LSTM + Attention) as proposed by Shahriar Afandizadeh (2023) to capture non-linear and temporal dependencies in EV adoption patterns.

### **Develop Practical Marketing Tools**

Translate the final improved model into a production-ready marketing tool such as a CRM-integrated dashboard or lightweight web application that generates real-time "EV Propensity Scores" for individual leads, enabling automated, personalized customer outreach at scale.

### **Address the "Neutral Majority"**

Future research must specifically focus on classifying and understanding the large, pragmatic "neutral" consumer segment that was identified in this study (23.5% neutral, 59% uncertain/unwilling). This group is essential for mass adoption but remains poorly understood and under-studied.



## REFERENCES

- [1] ANDADELABARA, S.-V. O. (2025). How Predictable Is Electric Vehicle Adoption? Exploring the Broader Role of Renewables in Transportation Using a Data-Driven Approach. 22.
- [2] Cherchi, C. D. (2023). Electric vehicle forecasts: a review of models and methods including diffusion and substitution effects. *Electric vehicle forecasts: a review of models and methods including diffusion and substitution effects*, *Transport Reviews*, 27.
- [3] DEMIRCI, A. (2023). A Comprehensive Data Analysis of Electric Vehicle User Behaviors Toward Unlocking Vehicle-to-Grid Potential. *Bootstrap, charging behavior, distributed network, driving data, electric vehicle*, 17.
- [4] Jaman, O. (2025). A SYSTEMATIC LITERATURE REVIEW ON PREDICTIVE ANALYTICS FOR SUSTAINABLE CONSUMER BEHAVIOR. Master's thesis in Governance of Digitalization, 96.
- [5] Sanjib Kumar Shil, M. S. (2024). Forecasting Electric Vehicle Adoption in the USA Using Machine Learning Models. *Journal of Computer Science and Technology Studies*, 14.
- [6] Shahriar Afandizadeh, D. S. (2023). Using machine learning methods to predict electric vehicles penetration in the automotive market. *nature portfolio*, 16.
- [7] Sharma, K. S. (2020). Using extended theory of planned behaviour (TPB) to predict adoption intention of electric vehicles in India. *Automobile · Conventional vehicles (CVs) · Electric vehicles (EVs) · Environment · TPB · Sustainability*, 17.
- [8] Singh, S. K. (2022). Predicting Electric Vehicle (EV) Buyers in India: A Machine Learning Approach. *Electric vehicles · EV buyers in India · EV adoption · Factors affecting EV purchase · Machine learning model to predict EV buyers*, 18.
- [9] Trisha V. Ramadoss, J. H. (2025). Classifying electric vehicle adopters and forecasting progress to full adoption. *sustainable mobility and transport*, 15.
- [10] Zarazua de Rubens, G. (2019). Who will buy electric vehicles after early adopters? Using machine learning to identify the electric vehicle mainstream market. *Energy*, 32.