

Predicting Electric Vehicle Adoption Intention Using Machine Learning: An AI-Driven Framework for Charging Infrastructure Planning in Karnataka, India

Shivabeerappa M.¹, Priya K.², Ruchitha G.²

¹Department of Business Administration, MMK & SDM Mahila Maha Vidyalaya, University of Mysore, Mysuru, Karnataka, India

²Department of Master of Business Administration, GSSS Institute of Engineering and Technology for Women, Visvesvaraya Technological University, Mysuru, Karnataka, India

Corresponding Author: shivabeerappa@sdmmmkmysore.in

ABSTRACT

Forecasting electric-vehicle (EV) demand is important for charging-infrastructure planning, but aggregate sales models offer limited insight into which consumers are considering a purchase. The study's methodological contribution is a reproducible, leakage-controlled workflow that applies machine-learning classifiers to primary survey data to predict individual purchase consideration in a secondary-city context, as a complement to aggregate forecasts. A convenience and snowball sample of 175 respondents was drawn from Mysuru and surrounding districts of Karnataka, India; the analysis used the 162 non-EV owners (117 Yes/Maybe; 45 No). Logistic regression and random forest classifiers were compared using stratified five-fold cross-validation. Random forest achieved pooled out-of-fold accuracy of 0.784 (95% bootstrap CI: 0.722–0.846) and ROC-AUC of 0.843 (95% CI: 0.784–0.898), compared with 0.735 and 0.823 for logistic regression; repeated cross-validation gave slightly lower but consistent estimates. Attitude-related measures ranked highest, but their importance is predictive, not causal. Because the sample is small and non-probabilistic, the findings are a methodological proof of concept that requires external validation, not a tool ready for planning use.

Keywords: adoption intention; charging infrastructure; electric vehicles; machine learning; random forest

1. INTRODUCTION

Electric vehicle (EV) adoption in India is accelerating: total EV sales rose from 1.68 million units in FY 2023–24 to 1.96 million units in FY 2024–25, an increase of about 17% [10], and the sector is projected to represent a \$200 billion opportunity by the early 2030s [23]. Karnataka was the first Indian state to notify a dedicated EV and Energy Storage Policy in 2017 and has since built one of the country's larger public charging networks [16][35]. Translating this policy momentum into efficiently sited, adequately scaled charging infrastructure, however, depends on accurately forecasting where and among whom EV demand will actually materialize.

Artificial intelligence and machine learning (ML) are increasingly used for exactly this purpose. Deep learning architectures combining convolutional and recurrent layers have been used to forecast aggregate EV market demand across vehicle categories with higher accuracy than conventional regression methods [28], and hybrid long short-term memory (LSTM) models have been applied to predict EV penetration into the broader automotive market from historical sales and registration time series [1]. These approaches are powerful for macro-level, aggregate forecasting, but they are largely agnostic to why individual consumers do or do not intend to adopt an EV, information that is directly useful for deciding where to prioritize charging infrastructure, dealership outreach, and awareness campaigns at a finer geographic and demographic grain.

This study complements that macro-forecasting literature with a micro-level, survey-based machine learning approach. Using primary data from 175 respondents in Karnataka, it trains and evaluates supervised classifiers that predict individual-level EV purchase consideration from a small set of demographic and perceptual features, and asks which of those features actually drive predictive performance. The resulting model and its feature-importance profile are intended as a lightweight, interpretable input that planners could combine with aggregate time-series demand forecasts of the kind reported in the recent deep-learning literature [1][28]; once externally validated, such a combination could inform charging-infrastructure planning.

1.1 Theoretical Foundations for Feature Selection

The feature set was informed by established technology- and green-product-adoption theories. The Theory of Planned Behavior (TPB) proposes that attitude, subjective norms, and perceived behavioral control shape behavioral intention [2][15][27][33][34]. The Technology Acceptance Model emphasizes perceived usefulness and ease of use [11], the Unified Theory of Acceptance and Use of Technology incorporates social influence and facilitating conditions [31], and Diffusion of Innovations highlights relative advantage and observability [26]. These theories informed the selection of attitude, perceived subsidy-awareness, and purchase-attribute measures. The present dataset, however, does not contain a complete operationalization of every construct in these theories; the TPB-related composite should therefore be interpreted as a survey-derived attitudinal index rather than a validated full TPB scale.

1.2 Determinants of EV Purchase Intention: Global and Indian Evidence

Prior EV-adoption studies identify environmental concern, perceived value, risk, incentives, affordability, driving range, and charging access as correlates of purchase intention [3][8][13][14][20][27][29]. Indian and Bengaluru-focused research has examined attitude, perceived behavioral control, and infrastructure constraints [8][15][18][32]–[34]. Much of this literature uses explanatory statistical models or aggregate demand data, while the present study tests whether a compact set of survey-based consumer attributes can classify individual purchase consideration in a secondary-city context. The contribution is therefore localized and methodological: it evaluates a micro-level classification workflow and its potential complementarity with macro-level forecasts, rather than claiming a new general theory of EV adoption.

1.3 Machine Learning Applied to Adoption Survey Data

Machine learning has increasingly been applied to survey data on technology adoption, usually alongside or after conventional behavioral models. In an early example, Chong [9] combined structural equation modelling with a neural network to predict mobile-commerce adoption from questionnaire constructs, showing that the nonlinear model captured relationships missed by the linear one. In EV research, Bas et al. [4] applied support vector machines, neural networks, gradient boosting, and random forest variants to United States survey data combining socioeconomic, attitudinal, social-network, and vehicle-attribute information to classify potential EV purchasers, and Bas et al. [5] used interpretable machine learning to show how attitudinal and ridesourcing factors shape EV adoption. Ong et al. [24] used random forest and deep learning on 1,048 Filipino drivers, sampled by convenience and snowball methods, to predict hybrid-car purchase intention from theory-of-planned-behavior and UTAUT2 constructs. Chen et al. [7] used random forest to predict EV purchase intention from customer evaluations and personal information in China, and Naseri et al. [21][22] applied ensemble learners with interpretation techniques to large Canadian survey and choice-experiment datasets of 2,062 and 1,077 respondents.

These studies establish that survey-based classifiers can identify likely adopters, but they differ from the present study in three ways. Most draw on larger samples in high-income or East Asian markets; several report very high accuracies (for example, 94% to 99% in [7][24]) without uncertainty intervals or comparison with a simple benchmark; and none that we located models non-owners in an Indian secondary city. The present study therefore contributes a small-sample workflow that controls target leakage, fits all preprocessing inside cross-validation folds, compares a nonlinear learner with a regularized linear benchmark, and reports bootstrap intervals, repeated cross-validation, and an explicit check for overfitting.

1.4 Machine Learning for EV Demand Forecasting and Infrastructure Planning

Machine-learning studies have used time-series and deep-learning methods to forecast aggregate EV sales, market penetration, or demand [1][28]. Such approaches address market-level volume and temporal patterns, whereas survey-based classifiers address differences in stated purchase consideration across respondents. This study compares logistic regression, used as a regularized linear benchmark [17], with a random forest [6], implemented through scikit-learn [25]. Unlike aggregate forecasting, the present analysis offers a respondent-level predictive profile; however, it does not estimate actual station utilization or establish that any specific infrastructure or outreach intervention will change adoption.

1.5 Karnataka's EV Policy Context

Karnataka's EV policy and infrastructure context makes consumer-level evidence relevant to planning [16][19][35]. Nevertheless, evidence from a geographically concentrated, non-probability sample cannot establish statewide adoption patterns. The specific gap addressed here is the limited evidence from localized survey-based classification of EV purchase consideration in Mysuru and nearby districts, alongside an explicit comparison with a simple linear benchmark. This is a proof-of-concept contribution, not a representative estimate for Karnataka.

1.6 Research Objectives

The study has four objectives, each linked to an analytical procedure: (1) profile respondents' demographic and perceptual characteristics using descriptive frequencies and percentages; (2) construct theory-informed survey features and train logistic regression and random forest classifiers using a documented preprocessing pipeline; (3) compare classifier performance using stratified five-fold cross-validation, fold-level metrics, pooled out-of-fold confusion matrices, and bootstrap confidence intervals; and (4) examine model-based feature importance, test the robustness of the results to overfitting and to the low-reliability belief index, and discuss cautious, potential applications to infrastructure and outreach planning. Figure 1 presents the conceptual decision-support framework.

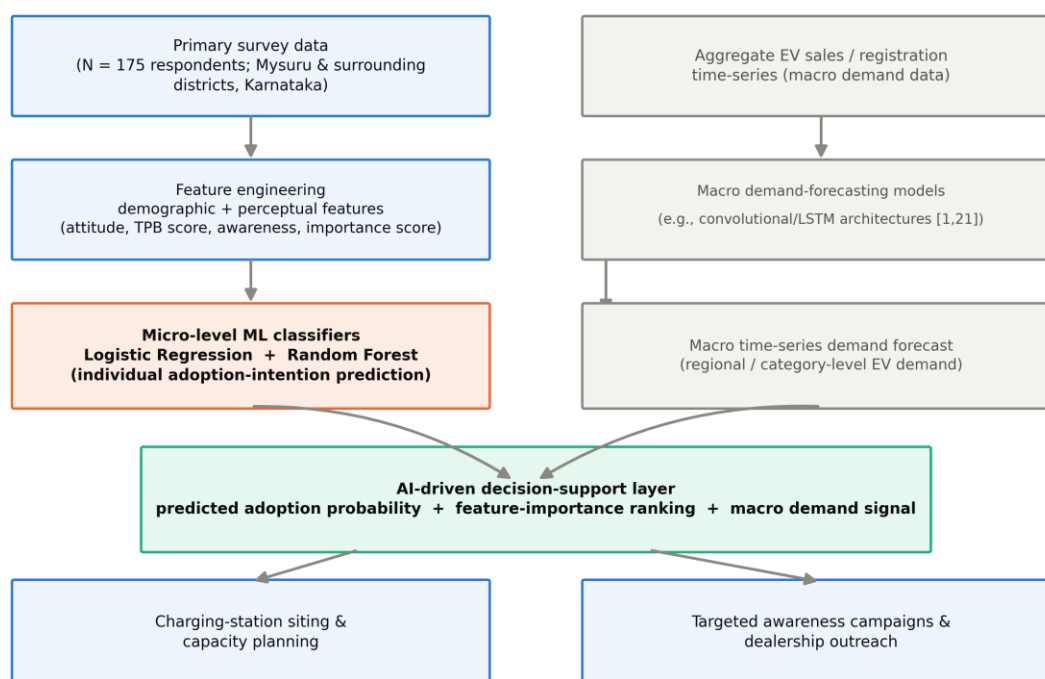


Figure 1. Conceptual (not yet validated) AI-driven framework linking survey-based, individual-level EV adoption prediction with macro demand forecasting to support charging-infrastructure and outreach planning.

2. METHODS

2.1 Research Design and Respondents

The study used a cross-sectional survey design. A structured online questionnaire was administered to 175 respondents in Mysuru and surrounding districts (Mandya, Hassan, Kodagu, Bengaluru, and Chamarajanagar), using convenience and snowball sampling. The questionnaire's purchase-consideration item was explicitly directed to respondents who did not own an EV. Four respondents had no recorded outcome; of the 171 with a response, nine were current EV owners and were excluded because the item was not designed for them. The primary modelling cohort therefore comprised 162 non-EV owners (117 Yes/Maybe = 1; 45 No = 0; majority-class accuracy = $117/162 = 72.2\%$). Figure 2 summarizes the complete data-to-evaluation pipeline.

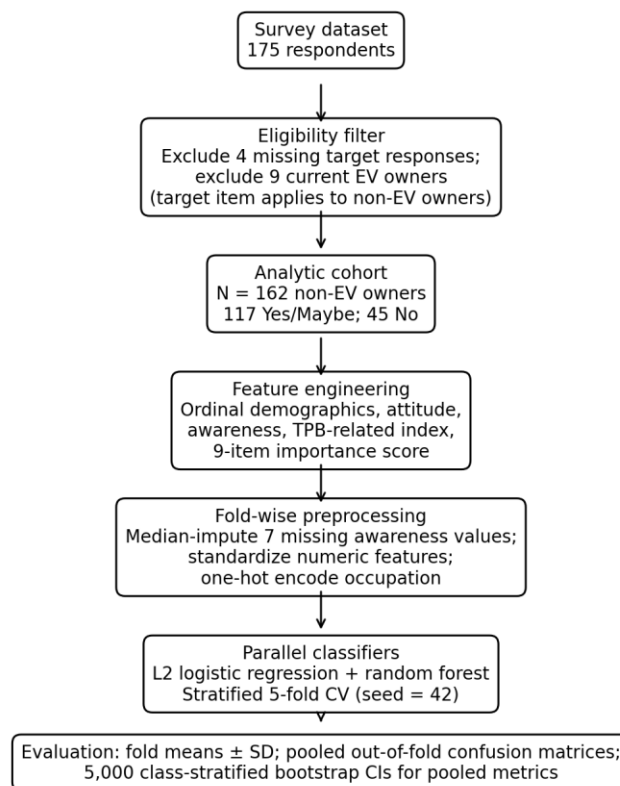


Figure 2. Analysis pipeline, including the non-EV-owner eligibility filter, fold-wise imputation, and out-of-fold evaluation.

2.2 Instrument, Validity, and Reliability

The Google Form had three sections: an introduction describing the study (Section 1); general information on age, gender, educational qualification, occupation, annual income, and district (Section 2); and perceptions of electric vehicles, items Q1–Q20 (Section 3). The Section 3 items used in this study were: EV ownership (Q3, “Is your current vehicle an electric vehicle (EV)?” Yes/No), used as the eligibility filter; purchase consideration (Q5, “If you do not own an electric vehicle, have you ever considered purchasing one?” Yes/No/Maybe), the outcome; overall opinion (Q6, “What is your overall opinion towards electric vehicles?” five points from Very negative to Very positive), referred to below as overall attitude; eight Likert statements on EV perceptions (Q9, 1 = Strongly Disagree to 5 = Strongly Agree); nine purchase-factor importance ratings (Q11, 1 = Not important at all to 5 = Very important); and awareness (Q19, “How would you rate the current level of awareness among the public regarding government subsidies for Electric vehicles?”, with the options Very Low, Low, Moderate, High, Very High). Q19 therefore measures respondents' perception of public awareness of government EV subsidies, not their own knowledge of EVs; the variable is labelled “perceived subsidy awareness” in the results. The full questionnaire as administered is provided as Supplementary File S1.

Content validity was supported by drawing the perception and purchase-factor items from constructs in established adoption theories and prior EV studies [2][11][15][26][27][31]. Internal consistency was assessed using Cronbach's α among the 162 analysed respondents. The nine-item Q11 purchase-factor importance score showed high reliability ($\alpha = 0.94$). The eight Q9 statements showed low internal consistency after the two negatively worded items were reverse-coded ($\alpha = 0.38$). Item analysis explains why: the two reverse-coded statements, on purchase cost and battery-life concern, correlated negatively with the rest of

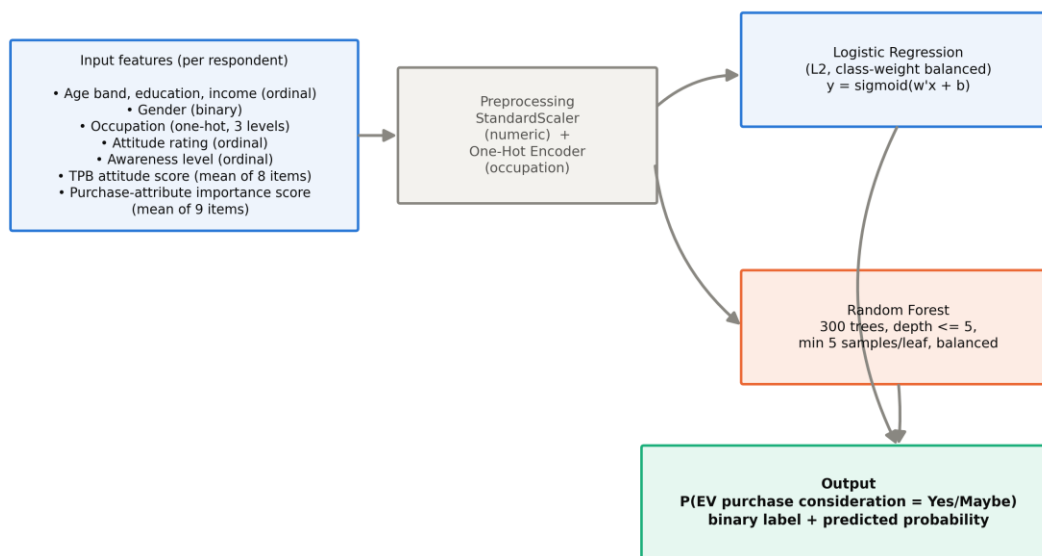
the index (corrected item–total $r = -0.18$ and -0.28), whereas the six positively worded statements correlated positively ($r = 0.20$ to 0.46). Respondents who saw EVs as environmentally beneficial or well performing therefore did not necessarily see them as cheap or free of battery risk. The statements capture distinct beliefs (environmental benefit, cost, charging availability, range, maintenance, performance, incentives, and battery life) rather than a single unidimensional construct. Such a set of distinct but jointly relevant beliefs is better treated as a formative index, for which internal consistency is not an appropriate validity criterion [12]. The Q9 mean is therefore used as a composite belief index for prediction, not as a reliable latent scale, and Section 3.3 tests whether the results depend on it.

2.3 Feature Engineering

Features were derived from the respondent-level dataset. Ordinal codes were assigned to age, education, income, overall attitude, and awareness; gender was binary-coded; occupation was grouped as Private Employee, Not Employed, or Other. The TPB-related attitudinal index was calculated as the mean of the eight Q9 Likert items. To align direction, the negatively worded purchase-cost and battery-concern items were reverse-coded ($6 - \text{response}$) before averaging; higher scores thus indicate more favorable perceptions. The purchase-attribute importance score was the mean of the nine Q11 ratings. The perceived subsidy-awareness item (Q19) was ordinal-coded (Very Low = 1 to Very High = 5); seven of the 162 eligible respondents did not answer it. These missing values were imputed with the ordinal-scale median (3, Moderate), a robust central-value choice that retained eligible cases; imputation was fitted within each training fold using a pipeline, preventing validation-fold information from informing imputation. Current EV ownership and vehicle-type variables were excluded to reduce near-definitional target leakage.

2.4 Model Specification

Two classifiers were fitted in scikit-learn pipelines [25]. Logistic regression used L2 regularization and balanced class weights as an interpretable linear benchmark [17]. The random forest used 300 trees, maximum depth 5, minimum 5 samples per leaf, and balanced class weights [6]. These hyperparameters were prespecified as conservative settings rather than selected through an exhaustive search; no claim of optimal tuning is made. Numeric features were median-imputed where needed and standardized; occupation was one-hot encoded. The same preprocessing and feature set were used for both models (Figure 3). A fixed random seed (42) was used for reproducibility.



Both branches share the same standardized input feature vector; predictions are compared under stratified 5-fold cross-validation (Fig. 2).

Figure 3. Model architecture: a shared preprocessing pipeline (standardization and one-hot encoding) feeds parallel logistic regression and random forest classifiers evaluated on the same feature set.

2.5 Evaluation Protocol

Performance was evaluated using shuffled, stratified five-fold cross-validation (random_state = 42) on the 162 non-EV-owner cases. Preprocessing, including awareness imputation and scaling, was fitted within each training fold. Fold-level accuracy, balanced accuracy, precision, recall, F1-score, and ROC-AUC are reported as means and standard deviations. Pooled out-of-fold predictions were used to construct confusion matrices and calculate respondent-level performance estimates. Approximate 95% confidence intervals for pooled out-of-fold metrics were obtained using 5,000 class-stratified bootstrap resamples; these intervals condition on the cross-validated predictions and do not capture all model-training or external-sample uncertainty. Feature importances and logistic-regression odds ratios from a full-data refit are descriptive and should not be interpreted causally.

Because cross-validation on small samples can give optimistic or unstable estimates [30], three further checks were run. First, the cross-validation was repeated 20 times with different random partitions (100 train-validation splits), and training and validation ROC-AUC were compared to gauge overfitting. Second, a permutation test (200 random relabellings of the outcome) assessed whether random-forest performance exceeded chance. Third, the model was re-estimated with the Q9 composite replaced by its eight items, with the composite removed, and with overall attitude removed. To limit overfitting, no hyperparameter tuning or data-driven feature selection was performed, and the random forest's depth and leaf size were capped.

2.6 Ethical Considerations

Participation was voluntary, and respondents were informed of the study's purpose on the first page of the online form before responding. E-mail addresses collected by the form were removed before analysis. Details are given in the Ethics Statement.

3. RESULTS AND DISCUSSION

3.1 Sample Profile

The overall sample was concentrated in Mysuru district (76.6%), with 75.4% aged 26–45 years, 75.4% holding postgraduate education or higher, and 65.1% reporting income below ₹5 lakh. Ten respondents (5.7%) reported current EV ownership. Since the purchase-consideration item was directed to non-EV owners, the primary model used 162 non-EV owners with complete outcomes: 117 (72.2%) answered Yes/Maybe and 45 (27.8%) answered No.

3.2 Model Performance

Across stratified five-fold validation of the 162 non-EV-owner cases, random forest had mean accuracy of 0.784 (SD = 0.050) and ROC-AUC of 0.848 (SD = 0.072), while logistic regression had mean accuracy of 0.735 (SD = 0.073) and ROC-AUC of 0.826 (SD = 0.044) (Table 1; Figure 4). The pooled out-of-fold random-forest confusion matrix was TN = 39, FP = 6, FN = 29, TP = 88 (Table 2); pooled accuracy was 0.784 (95% bootstrap CI: 0.722–0.846), balanced accuracy 0.809 (0.744–0.870), F1-score 0.834 (0.780–0.884), and ROC-AUC 0.843 (0.784–0.898). For logistic regression, the pooled matrix was TN = 34, FP = 11, FN = 32, TP = 85; accuracy was 0.735 (95% CI: 0.667–0.802) and ROC-AUC was 0.823 (95% CI: 0.757–0.883). Random forest had higher point estimates on several metrics, but uncertainty and the localized sample warrant cautious interpretation.

Table 1. Cross-validated performance of the two classifiers among non-EV owners (N = 162)

Metric	Logistic Regression	Random Forest
Accuracy	0.735 (0.073)	0.784 (0.050)
Balanced accuracy	0.741 (0.077)	0.810 (0.052)
Precision	0.885 (0.049)	0.939 (0.044)
Recall	0.727 (0.080)	0.753 (0.074)
F1-score	0.797 (0.063)	0.833 (0.045)
ROC-AUC	0.826 (0.044)	0.848 (0.072)

Note. Stratified five-fold cross-validation, N = 162 non-EV owners. Majority-class baseline accuracy = 0.722. Values are fold means with standard deviations in parentheses.

Table 2. Pooled out-of-fold confusion matrices among non-EV owners (N = 162)

Model	True negative	False positive	False negative	True positive
Logistic regression	34	11	32	85
Random forest	39	6	29	88

Note. Class 1 = Yes/Maybe; class 0 = No. Values summarize pooled held-out predictions from stratified five-fold cross-validation.

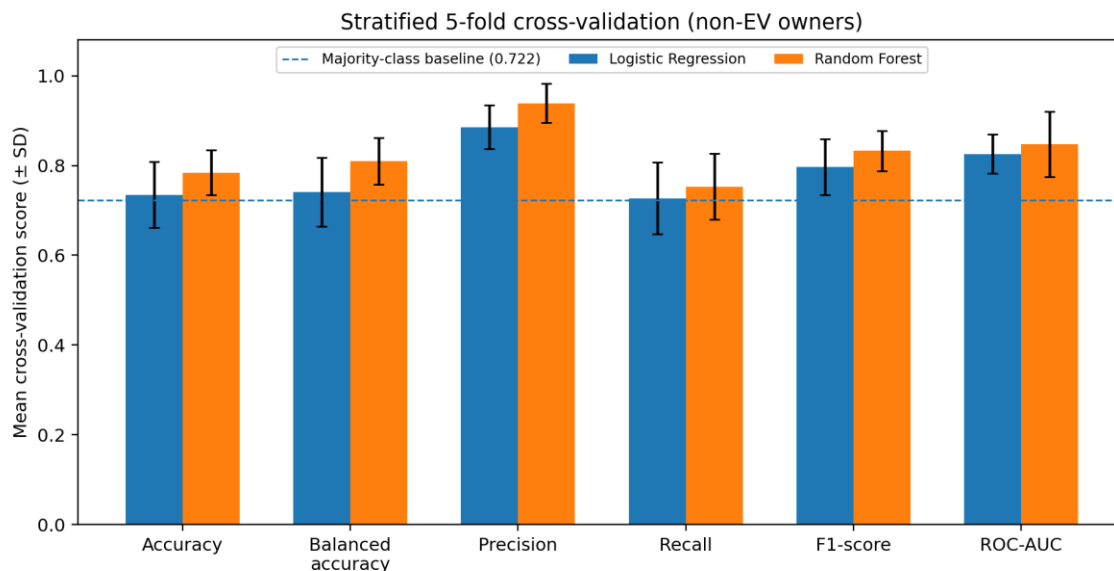


Figure 4. Fold-level cross-validation performance among non-EV owners; error bars show ± 1 SD across five folds. The dashed line marks the majority-class accuracy (0.722).

Sensitivity analysis: excluding awareness yielded random-forest mean accuracy of 0.772 (SD = 0.053) and ROC-AUC of 0.844 (SD = 0.068). Restricting the model to 155 complete-awareness cases yielded accuracy of 0.781 (SD = 0.043) and ROC-AUC of 0.853 (SD = 0.010). These results suggest that performance was not solely attributable to median imputation of the seven missing awareness values, although the complete-case analysis is smaller and its fold-level AUC variability should be interpreted cautiously.

3.3 Robustness, Overfitting, and Sensitivity to the Q9 Index

Across 20 repetitions of stratified five-fold cross-validation, the main random-forest specification averaged accuracy of 0.751 (SD = 0.068) and ROC-AUC of 0.837 (SD = 0.064); logistic regression averaged 0.738 (SD = 0.073) and 0.811 (SD = 0.069) (Table 4). These estimates are slightly lower than the single-partition results in Table 1, which therefore lie toward the favorable end of the plausible range; the level of performance and the ordering of the two models are unchanged. The random forest's mean training ROC-AUC was 0.950, against 0.837 on held-out folds. This gap of about 0.11 shows that, despite the depth and leaf-size limits, the forest fits the training data noticeably better than new cases; logistic regression showed a smaller gap (0.862 versus 0.811). A permutation test nevertheless placed random-forest performance well above chance: ROC-AUC was 0.849, against a mean of 0.504 for randomly relabelled outcomes (95th percentile 0.601; $p = 0.005$).

The results did not depend on the low-reliability Q9 index. Removing it left ROC-AUC at 0.818 (random forest) and 0.814 (logistic regression). Entering its eight items separately raised random-forest ROC-AUC to 0.883 but slightly lowered logistic-regression ROC-AUC (0.804). The item-level model was not adopted as the primary specification because it was examined after the main analysis, and choosing a specification by its performance would bias the estimates upward [30]; with 16 features it also showed a larger training-validation gap (0.977 versus 0.883). By contrast, removing overall attitude reduced ROC-AUC to 0.739 (random forest) and 0.705 (logistic regression), confirming that the single overall-opinion item carries most of the predictive signal.

Table 4. Robustness and sensitivity analyses: 20 × repeated stratified five-fold cross-validation (N = 162)

Specification	Model	Accuracy	ROC-AUC	Training ROC-AUC
A. Main (Q9 composite)	LR	0.738 (0.073)	0.811 (0.069)	0.862
	RF	0.751 (0.068)	0.837 (0.064)	0.950
B. Q9 items separately	LR	0.740 (0.065)	0.804 (0.068)	0.898
	RF	0.816 (0.065)	0.883 (0.061)	0.977
C. Q9 composite removed	LR	0.762 (0.071)	0.814 (0.069)	0.861
	RF	0.766 (0.055)	0.818 (0.061)	0.934
D. Overall attitude removed	LR	0.625 (0.078)	0.705 (0.085)	0.780
	RF	0.696 (0.071)	0.739 (0.082)	0.944

Note. Values are means (SD) over 100 validation folds; training ROC-AUC is the mean over the corresponding training folds. LR = logistic regression; RF = random forest. All variants use the same preprocessing and hyperparameters as the main analysis.

3.4 Feature Importance

Table 3 reports the full-data refit results. Random-forest impurity importances ranked overall attitude (0.411), the TPB-related attitude index (0.174), and purchase-attribute importance (0.147) highest (Figure 5). These values are model-dependent rankings, not estimates of causal effects. Overall attitude and the multi-item attitude index partly measure overlapping perceptions; their combined prominence may reflect shared information, and impurity-based importance can distribute or inflate importance among correlated predictors. Logistic-regression odds ratios are also conditional associations within this fitted model and should not be read as causal effects or standalone policy levers.

Table 3. Full-data refit: random-forest impurity importance and logistic-regression odds ratios (N = 162)

Feature	RF importance	LR odds ratio
Attitude (overall rating)	0.411	3.76
TPB attitude score	0.174	1.42
Purchase-attribute importance score	0.147	1.39
Age	0.079	0.46
Income	0.065	1.10
Perceived subsidy awareness (Q19)	0.043	1.18
Gender (female)	0.032	1.35
Education	0.022	1.12
Occupation (Other)	0.014	2.61
Occupation (Private Employee)	0.009	0.75
Occupation (Not Employed)	0.004	0.51

Note. RF = random forest; LR = logistic regression. Numeric-feature odds ratios correspond to a one-standard-deviation increase. Occupation is represented by one-hot indicator columns; with all categories encoded, their exponentiated coefficients should be interpreted cautiously and not as direct pairwise contrasts. All estimates are descriptive conditional associations, not causal effects.

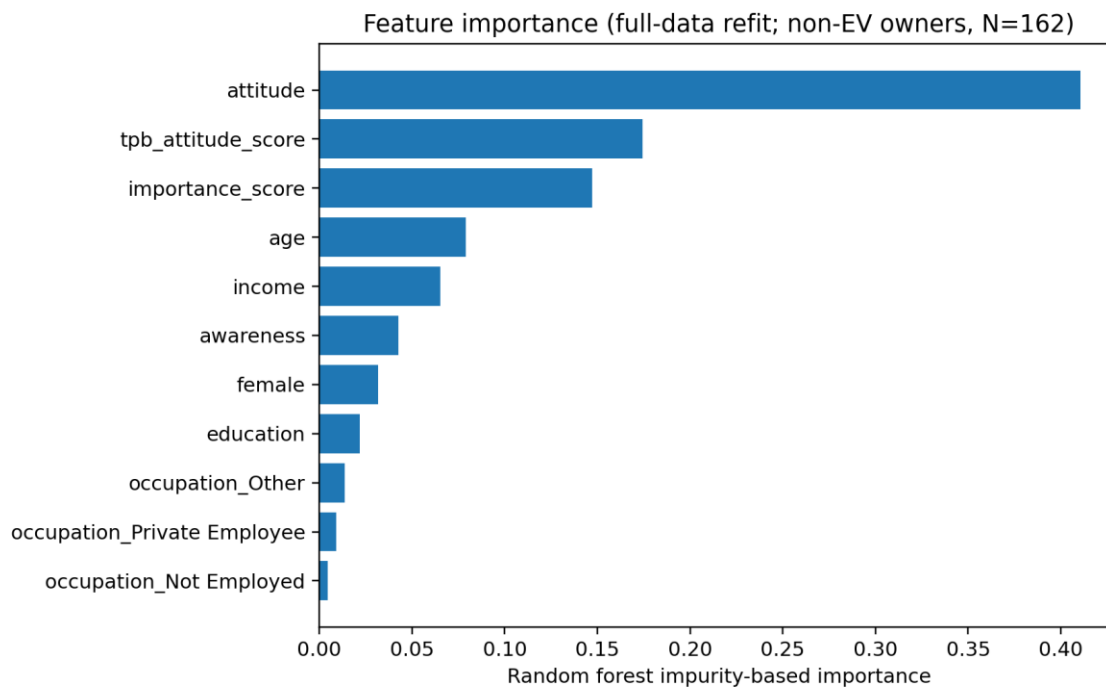


Figure 5. Random-forest impurity-based feature importance from the full-data refit among non-EV owners (N = 162). Importance values are descriptive and may be affected by correlated predictors. In the figure, “awareness” denotes perceived public awareness of government EV subsidies (Q19).

3.5 Discussion

The reanalysis indicates that the selected survey features contain some predictive information about stated EV purchase consideration among non-EV owners, although performance is preliminary. In pooled out-of-fold predictions, random forest reached 0.784 accuracy and 0.843 ROC-AUC; its confusion matrix shows 88 true positives, 39 true negatives, six false positives, and 29 false negatives. The model therefore identifies many respondents considering an EV, but it also misclassifies a meaningful share of Yes/Maybe respondents as No. Attitude variables rank highly, consistent with TPB-based EV studies that identify attitude as a central predictor of intention [15][18][27], but overall attitude and the TPB-related composite overlap conceptually and were derived from related self-reported items. Their importance indicates predictive association in this sample, not causal influence or evidence that changing attitudes will necessarily change purchasing behavior.

The false negatives matter for planning. At the default 0.5 threshold, the random forest missed 29 of the 117 Yes/Maybe respondents (24.8%). If predicted probabilities were aggregated by locality to help prioritize charging sites or outreach, areas with many such respondents would be under-ranked: their latent demand would be invisible to the model, and outreach would bypass consumers who are in fact open to an EV. Because limited charging access is itself a barrier to adoption [8][29], this error is arguably costlier than a false positive, which mainly wastes outreach effort. The threshold can be lowered to trade one error for the other. Across the repeated cross-validation runs, lowering it from 0.5 to 0.4 cut the average number of missed considerers from 31.5 to 16.6 (sensitivity 0.73 to 0.86), but roughly doubled false positives from 9.9 to 20.9 (specificity 0.78 to 0.54). Any planning use would therefore need a threshold chosen by the relative cost of missed demand and wasted outreach, and should work with aggregated probabilities rather than individual yes/no labels.

Aggregate demand models [1][28] and individual-level survey classifiers answer different questions. A classifier of this kind could, after external validation, complement regional forecasts by describing stated consideration among surveyed consumers. It cannot by itself determine charging-station locations, forecast actual charging demand, or demonstrate that subsidy-awareness campaigns or incentives will cause purchases. Because the sample is non-probabilistic and concentrated in Mysuru, the current findings should not be generalized to all Karnataka consumers or used to rank districts.

The random forest's point estimates exceeded those of logistic regression on several metrics, but the difference should not be treated as definitive evidence of a universally superior algorithm. The models were evaluated on one small dataset, and the confidence intervals are approximate. Repeated cross-validation (Section 3.3) confirmed the ordering of the two models but also showed that the random forest overfits the training data to a degree, so its advantage over the simpler benchmark should be regarded as modest.

3.6 Limitations

This study is limited by its modest analytic sample ($N = 162$ non-EV owners), geographic concentration in Mysuru, and convenience/snowball recruitment, which restrict representativeness and statewide generalization; the findings cannot be assumed to hold for other cities or for rural Karnataka. Four respondents lacked a recorded target; nine current EV owners with recorded responses were excluded because the questionnaire's purchase-consideration item was directed to non-EV owners. The outcome is self-reported purchase consideration, not observed purchasing or subsequent EV adoption; responses may be affected by social-desirability, recall, and hypothetical-intention biases. The dataset includes seven missing awareness values, and median imputation was performed within training folds; sensitivity checks comparing median-imputed awareness, complete cases ($N = 155$), and exclusion of awareness showed broadly similar random-forest performance. The overlapping attitude measures, and the low internal consistency of the Q9 belief index, complicate feature-importance interpretation. Although cross-validation and bootstrap intervals provide internal performance estimates, no independent or temporal external validation was available. With 162 cases, the random forest showed a training-validation ROC-AUC gap of about 0.11, and repeated cross-validation gave slightly lower accuracy than the single partition; performance on new respondents may therefore be lower than reported here. Class imbalance, possible unmeasured confounding, and limited hyperparameter search further constrain inference.

4. CONCLUSIONS/IMPLICATIONS OF THE STUDY

This study evaluated logistic regression and random forest classifiers for predicting self-reported EV purchase consideration among non-EV owners in a localized survey dataset. After excluding four missing outcomes and nine current EV owners for whom the questionnaire's target item was not intended, the primary analysis included 162 respondents. Random forest achieved pooled out-of-fold accuracy of 0.784 (95% CI: 0.722–0.846) and ROC-AUC of 0.843 (0.784–0.898). Attitude-related measures ranked highly, but they overlap and should be interpreted as predictive signals rather than causal drivers. The results are preliminary and limited by non-probability sampling, missing awareness values, and the absence of external validation.

Robustness checks support this reading. Repeated cross-validation gave slightly lower but consistent estimates (random-forest ROC-AUC 0.837), a permutation test placed performance well above chance, and the results held when the low-reliability Q9 belief index was removed or disaggregated. They also showed that the random forest overfits the training data to a degree, and that a single overall-opinion item carries most of the predictive signal. In relation to the four objectives, the study profiled a localized sample, built a documented and leakage-controlled classification pipeline, showed that a constrained random forest and

a regularized logistic regression both predict stated purchase consideration substantially better than chance, and identified attitude as the dominant predictor.

The novelty of the work lies in shifting the unit of analysis from aggregate market volume to the individual consumer in a secondary-city context. The implication for infrastructure planning is that a validated consumer-level classifier could serve as a complement to aggregate forecasts, not as a direct estimate of actual adoption or station demand, and it is not ready for state-level planning use. The study predicts stated purchase consideration among non-EV owners—not actual adoption, charging-station utilization, or response to an intervention—and does not establish that subsidy-awareness campaigns will raise adoption.

5. RECOMMENDATIONS

For policy and practice. The findings suggest a possible research workflow rather than a planning tool that is ready for state-level use. State EV cells, utilities, or researchers could test short attitude and subsidy-awareness measures alongside representative consumer surveys and administrative demand data. Only after external validation, calibration, fairness checks, and spatial linkage should predicted probabilities be considered as one input to charging-site or outreach planning, combined with registration, travel, grid-capacity, utilization, and spatial-accessibility data and evaluated prospectively.

For prospective evaluation. Planners or researchers wishing to test the framework could proceed in five steps. (1) Pilot: administer the short instrument to a fresh probability sample in one or two districts and record each respondent's predicted probability before any intervention. (2) Follow-up: after 12 to 24 months, re-contact respondents, or use anonymized local Vahan registration counts, to compare predictions with actual purchases and assess out-of-sample discrimination and calibration. (3) Spatial check: aggregate predicted probabilities to ward or taluk level and compare them with subsequent EV registrations and charger utilization. (4) Intervention test: where outreach is planned, roll it out in a staggered or randomized order across comparable localities so that any effect on adoption can be estimated. (5) Decision rule: pre-specify the performance needed before planning use, for example an out-of-sample ROC-AUC of at least 0.75 with acceptable calibration, and re-estimate the model periodically as the market matures.

For future research. Future work should use larger probability-based multi-district samples, validate models on independent datasets, preregister model selection, apply nested cross-validation if tuning is added, develop and validate a multi-item belief scale to replace the Q9 index, assess calibration and subgroup performance, and integrate validated consumer-level predictions with spatial and aggregate demand models.

ETHICS STATEMENT

This study involved a voluntary online survey of adult respondents. The form collected respondents' e-mail addresses; these were removed before analysis, and the analysed dataset contains no identifying information. No sensitive personal data or biometric data were collected. All participants were informed of the purpose of the study at the point of data collection and provided informed consent before responding, and were free to withdraw at any time. Under the applicable institutional policy, minimal-risk survey research of this kind is exempt from full ethics-committee review.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

DATA AVAILABILITY STATEMENT

The anonymized respondent-level dataset and the analysis scripts (feature engineering, model training, and cross-validation) that support the findings of this study are available from the corresponding author upon reasonable request. The data are not deposited in a public repository because participants consented to the use of their responses for this study only. The administered questionnaire is provided as Supplementary File S1. All software used is open source and freely available: analyses were conducted in Python with the scikit-learn library [25].

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design, data collection, analysis, and manuscript preparation. Detailed CRediT role assignments are provided on the separate title page. All authors read and approved the final version of the manuscript.

REFERENCES

1. Afandizadeh, S., Sharifi, D., Kalantari, N., & Mirzahosseini, H. (2023). Using machine learning methods to predict electric vehicles penetration in the automotive market. *Scientific Reports*, 13, Article 8345. <https://doi.org/10.1038/s41598-023-35366-3>
2. Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
3. Balasubramanian, N., Dhalmahapatra, P. K., & Pragma, M. S. (2024). Sustainable transportation in developing countries: Uncovering factors influencing electric vehicle purchase intention in India. *Transportation Planning and Technology*, 47(7), 1042–1067. <https://doi.org/10.1080/03081060.2024.2313138>
4. Bas, J., Cirillo, C., & Cherchi, E. (2021). Classification of potential electric vehicle purchasers: A machine learning approach. *Technological Forecasting and Social Change*, 168, Article 120759. <https://doi.org/10.1016/j.techfore.2021.120759>
5. Bas, J., Zou, Z., & Cirillo, C. (2023). An interpretable machine learning approach to understanding the impacts of attitudinal and ridesourcing factors on electric vehicle adoption. *Transportation Letters*, 15(1), 30–41. <https://doi.org/10.1080/19427867.2021.2009098>
6. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
7. Chen, J., Zhang, Z., Zhao, C., Zhang, S., Guo, W., Lu, C., & Sun, X. (2024). Research on the purchase intention of electric vehicles based on customer evaluation and personal information. *World Electric Vehicle Journal*, 15(1), Article 9. <https://doi.org/10.3390/wevj15010009>
8. Chhikara, R., Garg, R., Chhabra, S., Karnatak, U., & Agrawal, G. (2021). Factors affecting adoption of electric vehicles in India: An exploratory study. *Transportation Research Part D: Transport and Environment*, 100, Article 103084. <https://doi.org/10.1016/j.trd.2021.103084>
9. Chong, A. Y.-L. (2013). A two-staged SEM-neural network approach for understanding and predicting the determinants of m-commerce adoption. *Expert Systems with Applications*, 40(4), 1240–1247. <https://doi.org/10.1016/j.eswa.2012.08.067>
10. Dalvi, A. (2025, April 1). EV sales grow 17% to 1.96 million in FY2025, 2- and 3Ws, cars and SUVs hit new highs. *Autocar Professional*. <https://www.autocarpro.in/analysis-sales/ev-sales-in-india-rise-17-to-196-million-in-fy2025-2-and-3ws-cars-and-suvs-hit-new-highs-125656>
11. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
12. Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, 38(2), 269–277. <https://doi.org/10.1509/jmkr.38.2.269.18845>
13. Dutta, B., & Hwang, H. G. (2021). Consumers' purchase intentions of green electric vehicles: The influence of consumers' technological and environmental considerations. *Sustainability*, 13(21), Article 12025. <https://doi.org/10.3390/su132112025>

14. Featherman, M., Jia, S. J., Califf, C. B., & Hajli, N. (2021). The impact of new technologies on consumers' beliefs: Reducing the perceived risks of electric vehicle adoption. *Technological Forecasting and Social Change*, 169, Article 120847. <https://doi.org/10.1016/j.techfore.2021.120847>
15. Gayathiri, B., & Ahamed, S. B. I. (2025). Assessing consumer perceptions of electric vehicles in India through the lens of the theory of planned behavior. *Journal of the Institution of Engineers (India): Series A*, 106(1), 155–174. <https://doi.org/10.1007/s40030-024-00863-9>
16. Government of Karnataka. (2017). *Karnataka electric vehicle & energy storage policy 2017*. Commerce and Industries Department. <https://wordpress.missionstartupkarnataka.org/wp-content/uploads/2022/12/Karnataka-Electric-Vehicle-and-Energy-Storage-Policy-2017-1.pdf>
17. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley. <https://doi.org/10.1002/9781118548387>
18. Khurana, A., Kumar, V. V. R., & Sidhpuria, M. (2020). A study on the adoption of electric vehicles in India: The mediating role of attitude. *Vision*, 24(1), 23–34. <https://doi.org/10.1177/0972262919875548>
19. KPMG in India & Confederation of Indian Industry. (2020). *Shifting gears: The evolving electric vehicle landscape in India*. KPMG. <https://assets.kpmg.com/content/dam/kpmgsites/in/pdf/2020/10/electric-vehicle-mobility-ev-adoption.pdf>
20. Liu, X., Sun, X., Zheng, H., & Huang, D. (2021). Do policy incentives drive electric vehicle adoption? Evidence from China. *Transportation Research Part A: Policy and Practice*, 150, 49–62. <https://doi.org/10.1016/j.tra.2021.05.013>
21. Naseri, H., Waygood, E. O. D., Patterson, Z., & Wang, B. (2024). Who is more likely to buy electric vehicles? *Transport Policy*, 155, 15–28. <https://doi.org/10.1016/j.tranpol.2024.06.013>
22. Naseri, H., Waygood, E. O. D., Patterson, Z., & Wang, B. (2026). Which variables influence electric vehicle adoption? *Transportation*, 53(3), 1141–1178. <https://doi.org/10.1007/s11116-024-10525-1>
23. NITI Aayog. (2025). *Unlocking a \$200 billion opportunity: Electric vehicles in India*. Government of India. <https://niti.gov.in/sites/default/files/2025-08/Electric-Vehicles-WEB-LOW-Report.pdf>
24. Ong, A. K. S., Cordova, L. N. Z., Longanilla, F. A. B., Caprecho, N. L., Javier, R. A. V., Borres, R. D., & German, J. D. (2023). Purchasing intentions analysis of hybrid cars using random forest classifier and deep learning. *World Electric Vehicle Journal*, 14(8), Article 227. <https://doi.org/10.3390/wevj14080227>
25. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://www.jmlr.org/papers/v12/pedregosa11a.html>
26. Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.
27. Shalender, K., & Sharma, N. (2021). Using extended theory of planned behaviour (TPB) to predict adoption intention of electric vehicles in India. *Environment, Development and Sustainability*, 23(1), 665–681. <https://doi.org/10.1007/s10668-020-00602-7>
28. Simsek, A. I., Koç, E., Desticioglu Tasdemir, B., Aksöz, A., Turkoglu, M., & Sengur, A. (2024). Deep learning forecasting model for market demand of electric vehicles. *Applied Sciences*, 14(23), Article 10974. <https://doi.org/10.3390/app142310974>
29. Singh, V., Singh, V., & Vaibhav, S. (2020). A review and simple meta-analysis of factors influencing adoption of electric vehicles. *Transportation Research Part D: Transport and Environment*, 86, Article 102436. <https://doi.org/10.1016/j.trd.2020.102436>
30. Vabalas, A., Gowen, E., Poliakov, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLOS ONE*, 14(11), Article e0224365. <https://doi.org/10.1371/journal.pone.0224365>
31. Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
32. Verma, M., Verma, A., & Khan, M. (2020). Factors influencing the adoption of electric vehicles in Bengaluru. *Transportation in Developing Economies*, 6(2), Article 17. <https://doi.org/10.1007/s40890-020-0100-x>
33. Wang, S., Fan, J., Zhao, D., Yang, S., & Fu, Y. (2016). Predicting consumers' intention to adopt hybrid electric vehicles: Using an extended version of the theory of planned behavior model. *Transportation*, 43(1), 123–143. <https://doi.org/10.1007/s11116-014-9567-9>
34. Yeğin, T., & Ikram, M. (2022). Analysis of consumers' electric vehicle purchase intentions: An expansion of the theory of planned behavior. *Sustainability*, 14(19), Article 12091. <https://doi.org/10.3390/su141912091>
35. Zehra, F. (2025). Electric vehicle adoption and infrastructure development in Karnataka: Trends, challenges, and policy implications. *International Education and Research Journal*, 11(9). <https://doi.org/10.5281/zenodo.17322377>