Electric Vehicle Adoption Trends and Consumer Sentiment Analysis

*Note: Sub-titles are not captured in Xplore and should not be used

Name1
Department
University of North Texas

Name2
Department
University of North Texas

Name4
Department
University of North Texas

Name4
Department
University of North Texas

***Abstract*—This study investigates the adoption of electric vehicles (EVs) using a dual-faceted approach: prediction of sales and consumer sentiment analysis. In more detail, the quantitative methods are forecast models like the Bass Diffusion Model and its derivatives (Generalized and Uncertain Bass Models) that estimate the tendency of adoption relying on historical records of sales. Consumer sentiment is the use of machine learning and deep learning models to assess the emotions of consumers. This has been evidenced in ev adoption results that demonstrate the continuous increase in growth patterns occasioned by government incentives, market forces, and shifting consumer choice. Charging infrastructure and product affordability are thus vital issues to address in order to sustain the level of growth.**

I. Introduction

Introduction The transport sector is going through a transition to electric vehicles (EVs) on account of environmental improvement and curbing greenhouse gas emissions. The use of EV"s has been embraced globally due to the lowered Co2 emissions and improved energy efficiency that it takes to the fight against climate change. Governments globally have understood this potential, they have applied incentives like subsidies and tax credits for the same. For instance, overreliance of the electrical vehicle or firm policy support has seen many nations achieve rapid sales growth over the last decade.

Despite , there are still challenges in the way of introducing the technology to mainstream use, such as, firstly, high initial costs; secondly, insufficient number of recharging stations; and, thirdly, consumer acceptance about its stability in the long-term usage. Such challenges are even more apparent in emerging markets where other accompanying structures and frameworks are comparatively less developed. These concerns, and the drivers for and against adoption, are critical to the long-term sustainment of growth in the EV market.

To overcome these challenges this study will use two converging research approaches. First, it uses sophisticated accuracy forecasts models, such as the Bass Diffusion Model, which is used to explain the patterns of innovation and imitation, hence, used to adopt new technologies. Second, it uses the machine learning sentiment analysis to analyze and interpret the perception and concern that consumers have for EV"s. Employing both analytical methods, this research seeks to offer recommendations to policymakers and manufacturers that will improve their understanding of the changing nature of the EV market.

Identify applicable funding agency here. If none, delete this.

II. Review of Literature

This research integrates insights from various studies focusing on:

A. Diffusion Models: Bass Diffusion and Generalized Bass Models Predicting EV Adoption

Global Ratios of battery electric vehicles When undertaking comparisons of the ratios of different countries in adoption of battery electric vehicles, this research has noted that all the conditions present exponential growth. If one bases his/her estimates on modeling, then there is a propensity for system-wide adoption that appears significantly faster than the potential extraordinary analysis of economics have predicted. [3]

Prediction of PEV Adoption with Agent-Based Parameterized Bass Network Diffusion Model: This research develops an agent-based parameterized Bass network diffusion model for EV population data, achieving high estimation accuracy for EV adoption in both temporal and geographical scopes. [2]

B. Sentiment Analysis: Machine Learning Models and Deep Learning Techniques (e.g., LSTM) Analyzing Textual Data

Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018.This survey provides an overview of deep learning methods in sentiment analysis, discussing various models and their effectiveness in capturing sentiment from text data. [4]

But then, as we have seen, the baseline approach using traditional powered machine learning models is still useful for sentiment classification while the LSTMs and Transformers add a new dimension to the domain of deep learning. [4] [5]

III. Objectives of the Study

The purpose of this paper is therefore to present current and potential future trends in the growth of electric vehicles and analyze the consumers' response to bring opportunities for manufacturers, policy makers, and other interested parties. Through the application of predictive modeling and implementing sentiment analysis, the research aims at

A. To Forecast EV Adoption Trends:

Estimate electric vehicles adoption using Bass Diffusion Model, Generalized Bass Model (GBM) and Uncertain Bass Model. It is important to note that these models are applied frequently in studying the patterns of technology adoption . [2] [3]

Incorporate external factors, such as governmental policies, incentives, and technological advancements, to enhance forecasting accuracy, as demonstrated in policy-driven adoption models .

B. To Analyze Consumer Sentiments Using Machine Learning:

Employ machine learning techniques, including Logistic Regression and Random Forest, and advanced deep learning methods, such as LSTM, to classify consumer sentiments into positive, neutral, or negative categories. These approaches are effective for handling textual data and capturing sentiment nuances. [6]

Identify key factors influencing consumer behavior, such as affordability, charging infrastructure, and performance, through sentiment analysis of online reviews and social media content . [7]

C. Quantitative and Qualitative Insights

Including forecasting models and sentiment analysis to provide a comprehensive understanding of the EV market. This hybrid approach enables stakeholders to address both quantitative adoption metrics and qualitative consumer feedback .

Offer actionable recommendations for policymakers and manufacturers to address consumer concerns and drive EV adoption, following frameworks that integrate data analytics with market insights .

IV. Data Collection

The study depends on two distinct datasets: EV sales data and consumer review data

A. EV Sales Data

EV sales data and consumer reviews were used as the data sources for the analysis, with each providing unique information for the study. The EV sales data was collected from the Atlas EV Hub that tapped into both, national level data for 16 states in the USA as well as the state level data concerning the years between 2010 and 2024. Original datasets of these variables were downloaded, combined to form a single data set and data duplication was also removed. Some of these columns are standardized, namely: Vehicle Name, Year, and Sales, which is number of annual EV registrations per model. Data cleaning entails treatments such as handling null values, normalizing vehicle naming, and removing invalid years encoded in the data set. The data was then converted to a time series format for trend-analysis and modeling necessary for Bass Diffusion model and its derivatives [8].

B. Consumer Reviews Dataset

The consumer reviews dataset was collected from Kaggle which contains the reviews on different models of EV. The types of fields in the dataset include Review Text, Ratings and Vehicle Model. In data preparation some operations like cleaning the text from symbols, URLs and stop words, as well as tokenization were performed. Sentiment, numerical values that gives a clue regarding the positive or negative nature of sentence or document was derived using machine learning extractors from text mining framework, while the word vectors for traditional and deep learning model was extracted using TF-IDF and Word2Vec respectively. This was used for sentiment analysis where playing field of Logistic Regression/Random Forest LSTM was prepared to categorize consumer feedback sentiments as positive, neutral, or negative. [6], [9]. Using the above datasets, an EV adoption trend analysis was conducted alongside with a proposed consumer sentiment analysis to facilitate decision making for adoption by the key stakeholders.

V. System Architecture and Design

One is called the sales forecasting pipeline and the other is called the consumer sentiment pipeline Each one is running in parallel and independently from the other but give the same kind of information different in that they are looking at the market share of EVs from the forecast angle as well as the consumer opinion angle

The system architecture is divided into five distinct layers:

1. **Data Ingestion Layer:** Derived from the sales of EV and the feedback elicited from the customers.

2. **Data Processing Layer:** Surfaces structured sales details from the input data as well as raw review textual data.

3. **Analytical Core:** Adopts trend and opinion prediction algorithms.

4. **Visualization and Insights Delivery:** Creates business analysis and visualization outputs.

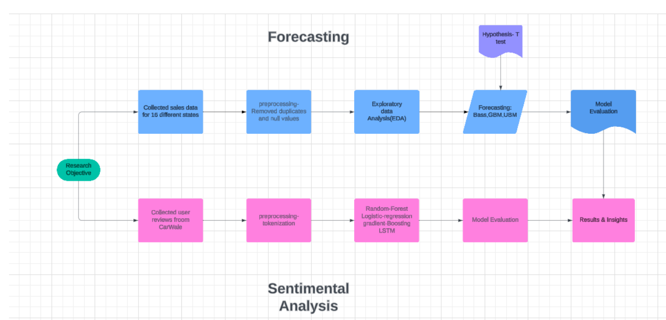

Fig. 1. System Architecture Flowchart for Electric Vehicle Adoption Trends and Consumer sentiment Analysis

A. Data Ingestion Layer

The data ingestion layer handles the acquisition of two datasets: Sales data for EVs and experience of consumers. This EV sales dataset has fields such as; the name of the car, year of registration, and total sales made. Consumer reviews are saved in another CSV file which contains textual reviews, ratings, and associated reviews given by the consumer [10]. The current data files are loaded using Python's pandas library for the purpose of providing proper formatting and validation for the completeness of the data. [10]

B. Data Processing Layer

The layer of data processing is able to process the datasets that are to be analyzed.Pre-processing for sales data is to remove null values, standardise vehicle name to reduce its variance and convert it in the structure appropriate for time series data for forecasting. Yearly registration count is aggregated in a manner that will allow for trend identification. In the consumer reviews dataset textual data preprocessing, that includes the removal of special characters, URLS and unnecessary space are performed. Tokenisation is then used to separate the words from the text with feature extraction using either TF, IDF or word2Vec. Such a step makes sure that both datasets have a proper structure that can match their desired analyses. [11]

C. Analysis

The analytical core is divided into two independent pipelines: one for sales forecasting and the other for sentiment analysis. The pipeline includes models such as the Bass Diffusion Model, Generalized Bass Model GBM, Uncertain Bass Model, and others for future EV sales forecast. Model parameters are optimized using Python's scipy.optimize.curve fit to achieve accurate forecasts. The sentiment analysis pipeline utilizes machine learning algorithms such as Logistic Regression and Random Forest for text classification, alongside Long Short-Term Memory LSTM networks for capturing sequential patterns in consumer reviews.For the testing of models for forecasting statistical validity of parameters we employ p-value analysis and visualization comparisons while for sentiment analysis we use a measure such as accuracy, precision and recall, F1 score. [12] [7]

D. Visualization and Insights Delivery

This layer then produces the implementation of the analysis in forms of display and understandable documents. For sales forecasting, time-series plots illustrate historical trends and model predictions, enabling comparisons between models like Bass Diffusion, Generalized Bass, and Uncertain Bass. Sentiment analysis results are summarized using pie charts, bar graphs, and word clouds to show sentiment distributions and key consumer concerns, such as "charging infrastructure" or "battery life. Such visuals are made using tools like Matplotlib and Seaborn and compiled into useful reports and presentations for readers and market leaders respectively.

VI. Exploratory data analysis (EDA) and Hypotheses for the Study

EDA means paying attention to the overall architecture of the datasets and important features before introducing complex models. For the sales dataset, time-series plots were created to observe trends in EV registrations over the years. These plots demonstrated an increase from the early years up to the later years to depict the increase expansion of EVs in the market. Brand specific analysis showed that few brands and models of vehicles were dominating the industry.

For the customer review dataset, a distribution of sentiment ratings was analyzed to identify patterns in consumer feedback. For the positive, neutral, and negative sentiments, bar charts were used, and for the various topics discussed such as "charging infrastructure" and "battery performance." Word clouds were used to visualize frequently mentioned words in reviews, offering an overview of recurring themes in consumer opinions. This exploratory phase helped identify trends and areas of focus for forecasting and sentiment modeling.

A. Hypotheses

The hypotheses for sales are therefore formulated for testing the performance of the Bass Diffusion Model and then comparing the result with the Generalized Bass Model (GBM). First, the parameter significance hypothesis analyses the Bass Model's parameters distributions (p, q, m) being able to make prediction regarding EV adoption tendencies. The first set of hypothesis includes the null hypothesis, H_0 which asserts that p = 0, q = 0 and m = 0 while the hypothesis against it is H_A which postulates that one or more of the parameters p, q and or m is significantly different from zero.Secondly, the model comparison hypothesis assesses the models fit using AIC and BIC. The null hypothesis (H_0) being that the performance of the Bass model is no better, or even poorer than GBM, the alternative hypothesis (H_A) being that it outperforms. Therefore according to the shown results we can confirm the superiority of the Bass Model as has a lower AIC equal to 477.49 as compared to the GBM's AIC equal 499.87 and BIC equal to 480.90 as compared to the GBM's BIC equal 507.25.

The hypotheses of the sentiment analysis are meant to test the effectiveness of traditional machine learning algorithms and state-of-art deep learning algorithms in effectiveness of consumer sentiments classification. The model performance hypothesis tests whether deep learning models, such as LSTM, outperform traditional models like Logistic Regression and Random Forest. The null hypothesis (H_0) is that there is no major disparity in performance between the deep learning models and the other models while the factorial hypothesis (H_A) is that the deep learning models' F_1 scores are at least 0.03 greater than the other set of models. In particular, the sentiment distribution hypothesis examines if the consumer sentiments are well distributed among the positive, neutral and the negative sentiments. The null hypothesis (h0) assumes an even distribution, while the alternative hypothesis (H_A) expects a skew, with positive sentiments comprising at least 50

VII. Data Visualization and Results

A. Sales Data Trends

Visualization: Time-series plot that illustrates history of EV sales and predicts the future.
Insights: This is the level of growth in the later years where ev market presents exponetial shift, The EV diffusion can be estimated by the Bass Diffusion Model and the Generalized Bass Model. Among these, there are Government policies and the outcomes of market saturation effects. Some of these are Government policies and effects of market saturation effects.

B. Market Leaders and Consumer Preferences

Visualization: The variations are shown here through bar charts and pie charts:

- Mainstream EV manufacturers (with Tesla in the forefront).
- Market share by brand and segment.

Insights: Tesla dominates consumer preferences, followed by growing competition. Consumer satisfaction correlates with performance, eco-friendliness, and comfort.

C. Sentiment Distribution

Visualization: Bar charts showing the distribution of the sentiment ratings:

- Positive, Neutral, and Negative sentiments from customer reviews.

Insights:

- Positive sentiment is focused on performance with reference to the environmental issues.
- The major concern of negative sentiment revolves around cost and perhaps the lack of adequate charging stations.

D. Feature Correlation Matrices

Visualization: Heatmaps of correlation matrices:

- Between sales data variables (e.g., vehicle count vs. year).
- Sub-categories of sentiment ratings (e.g., Comfort, Performance, Value for Money).

Insights: Significant features reveal that performance and comfort influence the satisfaction of consumers most significantly.

VIII. Results

A. Sales Forecasting

Models Used:

- **Bass Diffusion Model:** Predicts steady, organic growth but lacks uncertainty considerations, aligning well with historical data.
- **Generalized Bass Model:** Explains market coverage and explain external factors. Growth is expected to start growing at high rates but will then level off after some time.
- **Uncertain Bass Model:** Contains variability for cases such as policy change, ideal for decision making under conditions of uncertainty.

Metrics: Comparison of models using AIC and BIC; historical vs. predicted data visualizations.
Outcome: Generalized Bass Model provides the most reliable forecasts for market planning.

B. Sentiment Analysis

Models Used:

- **Traditional:** Logistic Regression, Random Forest, Gradient Boosting.
- **Deep Learning:** LSTM.

Metrics: Accuracy, Precision, Recall, F1-score.
Outcome:

- Traditional models provide robust baselines for structured data.
- LSTM excels in capturing complex patterns in textual data, achieving higher accuracy in sentiment classification.

C. Key Findings

Positive Sentiment Drivers:

- Performance, Comfort, Eco-friendliness.

Negative Sentiment Drivers:

- Cost, Charging Infrastructure.

Market Strategy: Insights from sentiment and forecasting models suggest focusing on improving infrastructure and reducing costs to accelerate adoption.

D. Hypothesis Testing

Sales Forecasting Hypothesis: Rejection of null hypothesis confirms significant contributions of model parameters (P, Q, M) to the Bass model.
Sentiment Analysis Hypothesis: LSTM significantly outperforms traditional models, proving advanced models' advantage in sentiment understanding.

IX. Conclusion

This project combines sales forecasting with sentiment analysis to provide better insights into the EV market.

A. Sentiment Analysis

- **Positive Feedback:** It is clear that performance is highly appreciated by the consumers, and this can be implemented.
- **Negative Feedback:** Charging infrastructure and costs continue to be considered as one of the significant challenges to adoption.
- **Model-Specific Insights:** Popular models like Tesla outperform in consumer satisfaction, while others face challenges related to reliability or pricing.

B. Sales Forecasting

- **Bass Diffusion Model:** Points at precise average growth rate increase but misses uncertainty considerations.
- **Generalized Bass Model (GBM):** Effectively captures market saturation dynamics.
- **Uncertain Bass Model:** It explains as to why it has fluctuations which provide accurate forecast under volatile and uncertain environments.

To boost EV adoption, addressing concerns such as charging infrastructure and pricing is crucial. Combining consumer sentiment insights with advanced forecasting provides a strategic roadmap for manufacturers and policymakers to drive growth and satisfaction in the evolving EV market.

References

[1] S. Singhal, Y. Bano, and O. Singh, "Demand forecasting of technological products considering re-purchasing and dynamic pricing: A modified diffusion model," International Journal of System Assurance Engineering and Management, 2024

[2] Prediction of PEV Adoption with Agent-Based Parameterized Bass Network Diffusion Model Yuhao Yuan, Yihua Zhou, Zhounan Lin, Kai Jin

[3] ung F, Schröder M, Timme M (2023) Exponential adoption of battery electric cars. PLoS ONE 18(12): e0295692.

[4] Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018

[5] Sahoo, C., Wankhade, M. Singh, B.K. Sentiment analysis using deep learning techniques: a comprehensive review. Int J Multimed Info Retr 12, 41 (2023)

[6] S. Kumar and S. Gawade, "Sentiment Analysis of opinions about Electric Vehicles Using Twitter data," 2023 IEEE 7th Conference on Information and Communication Technology (CICT), Jabalpur, India, 2023, pp. 1-5, doi: 10.1109/CICT59886.2023.10455654.

[7] Jia, C.; Ding, C.; Chen, W. Research on the Diffusion Model of Electric Vehicle Quantity Considering Individual Choice. Energies 2023, 16, 5423.

[8] Patil, Amruta Bisen, Mahak Lokachari, Prakash Sai. (2023). Diffusion of Multigenerational Technologies in the Indian Machine Tool Industry: Bass Model. 10.46254/AU02.20230122.

[9] Sharma, Hemlata Din, Faiz Ogunleye, Bayode. (2024). Electric Vehicle Sentiment Analysis Using Large Language Models. Analytics. 3. 425 - 438. 10.3390/analytics3040023.

[10] elegov, D.V.; Chanaron, J.-J. Electric Car Market Analysis Using Open Data: Sales, Volatility Assessment, and Forecasting. Sustainability 2023, 15, 399.

[11] Usmani, M.; Memon, Z.A.; Zulfiqar, A.; Qureshi, R. Preptimize: Automation of Time Series Data Preprocessing and Forecasting. Algorithms 2024, 17, 332.

[12] Gudari, Sai , G, Dr. Vadivu. (2023). Comparative Study of Logistic Regression and LSTM for Sentiment Classification Across Diverse Textual Dataset. 10.13140/RG.2.2.24661.60641.