Leveraging Machine Learning and Natural Language Processing for Electric Vehicle Market Analysis

Sharvari Subhedar[1]
Student, Department of Computer Engineering
K J Somaiya School of Engineering
Somaiya Vidyavihar University
Mumbai, India

Prof. Shweta Dhawan Chachra[2]
Professor, Department of Computer Engineering
K J Somaiya School of Engineering
Somaiya Vidyavihar University
Mumbai, India

ABSTRACT

The adoption of Electric Vehicles (EVs) in India is majorly influenced by consumer perceptions regarding performance and reliability. Traditional market analysis methods often fail to capture real-time consumer sentiment, creating a gap in understanding EV market dynamics. This research leverages Machine Learning (ML) and Natural Language Processing (NLP) techniques to analyse consumer sentiment and predict the success of EV product launches in the Indian market. Consumer reviews from platforms like Bikewale, Carwale, and Cardekho were collected and pre-processed for sentiment analysis. ML models including Logistic Regression, Ridge Classifier, Random Forest, AdaBoost, and LightGBM were employed for sentiment classification. Logistic Regression achieved an accuracy of 71.03% and F1-score of 0.6609, while Ridge Classifier achieved 72.89% accuracy and 0.7066 F1-score. Random Forest performed significantly better with 96.73% accuracy and 0.9603 F1-score, and AdaBoost improved on this with 98.13% accuracy and 0.9733 F1-score. LightGBM outperformed all other models with an accuracy of 99.53% and F1-score of 0.9972. These results were bench-marked against standard studies to validate effectiveness. The findings provide valuable insights for EV stakeholders by identifying consumer concerns and preferences, ultimately supporting data-driven strategies to enhance EV adoption. Future work will explore real-time sentiment tracking and the development of a interactive dashboard.

Keywords: Electric Vehicles (EVs), Machine Learning, Natural Language Processing, Sentiment Analysis, Indian Market Analysis, LightGBM

1. INTRODUCTION

The worldwide shift toward sustainable transportation has greatly accelerated the uptake of Electric Vehicles (EVs). As concerns over carbon emissions, air pollution, and fossil fuel dependency grow, EVs are increasingly being promoted as a practical substitute to traditional internal combustion engine vehicles. Governments worldwide, including India, are implementing policies and incentives to accelerate EV adoption. However, despite technological advancements in battery efficiency, driving range, and cost reduction, consumer skepticism and market uncertainties remain key challenges.

The Indian EV market presents unique dynamics influenced by factors such as infrastructure limitations, high initial costs, and consumer concerns regarding reliability and performance. Understanding consumer sentiment and market trends is essential for manufacturers, marketers, and policymakers to address these challenges effectively. While conventional market research relies on structured data like sales figures and industry reports, a vast amount of valuable consumer feedback exists in the form of unstructured text, including online reviews and social media discussions. Analyzing this unstructured data can provide deeper insights into consumer preferences, pain points, and expectations [19].

This study leverages Machine Learning (ML) and Natural Language Processing (NLP) techniques to analyze consumer sentiment and predict the success of EV product launches in the Indian market. By utilizing datasets from platforms such as Bikewale, Carwale, and Cardekho, this research aims to extract actionable insights into the factors driving EV adoption. Sentiment analysis is employed to gauge public perception, while ML models are used to predict market trends and evaluate the success potential of newly launched EV models [20]

The key objectives of this research are:

- To perform sentiment analysis on consumer reviews of two-wheeler and four-wheeler EVs using NLP techniques.
- To apply ML models for predicting the success of EV launches.
- To compare the effectiveness of different ML algorithms in analyzing consumer sentiment and predicting market trends.

By bridging the gap between consumer perceptions and market strategies, this research aims to assist EV manufacturers, marketers, and policymakers in making informed decisions to enhance EV adoption in India. The findings will contribute to optimizing product development, refining marketing strategies, and improving consumer engagement in the evolving EV market.

2. LITERATURE REVIEW

Understanding electric vehicle (EV) adoption requires an interdisciplinary approach, integrating machine learning (ML), natural language processing (NLP), and economic modeling. This section summarizes key studies that contribute to EV market analysis.

2.1 Predictive Modeling of EV Market Penetration

Afandizadeh, S., D. Sharifi, and N. Kalantari et al. 2023 [1] utilized Long Short-Term Memory (LSTM) networks and Support Vector Machines (SVM) to forecast EV adoption, demonstrating that economic indicators such as GDP and technological readiness significantly impact market penetration.

2.2 Consumer Perceived Risk in EV Adoption

Shu, T., Z. Wang, and L. Lin et al. 2022 [2] applied NLP to analyze consumer concerns, identifying key risk factors such as technology maturity, charging infrastructure, and time constraints. The study highlighted the role of sentiment analysis in capturing hidden consumer anxieties.

2.3 Forecasting EV Market Trends in India

Mishra, R. D., S. K. Dash, and S. Chudjuarjeen et al. 2024 [3] developed LSTM-based models that proved highly effective in predicting Indian EV sales, particularly in the 2-wheeler and 3-wheeler types vehicles. Support Vector Regression (SVR) and Linear Regression provided comparative insights, emphasizing the importance of data-driven forecasting.

2.4 Price Forecasting and Market Dynamics

"EV Sales Price Forecasting Using Machine Learning" [5] 2024 demonstrated that XGBoost outperformed traditional regression models in predicting EV pricing trends by incorporating manufacturing costs, technological advancements, and government policies. This research provided a comprehensive framework for understanding EV market economics.

2.5 Sentiment Analysis Using Social Media Data

Kumar, S. and S. Gawade 2023 [6] conducted Twitter-based sentiment analysis using the BERT model, which revealed consumer preferences, with environmental benefits driving positive sentiment, while concerns about cost and battery life tempered enthusiasm.

2.6 Federated Learning for Privacy-Preserving Analysis

Thiruneelakandan, A. and A. Umamageswari 2023 [7] employed federated learning to enable collaborative EV market analysis while preserving data privacy. This approach allowed the integration of data from manufacturers, charging networks, and policymakers, creating a holistic view of sales trends.

2.7 Dynamic Topic Modeling for EV Market Trends

Lee, S. and M. Park 2021 [8] used YouTube data and dynamic topic modeling to track evolving consumer discussions on EVs, projecting a fourfold market growth from USD 287.36 billion in 2021 to USD 1,318.22 billion by 2028.

Key Insights and Future Directions :

The reviewed studies underscore the importance of ML and NLP in understanding EV adoption. Key takeaways include:

- The effectiveness of advanced predictive algorithms in market analysis.
- The role of sentiment analysis in understanding consumer behavior.
- The necessity of multi-dimensional approaches for accurate forecasting.

Future research should focus on integrating diverse data sources, refining predictive models, and exploring cultural variations in EV adoption to support sustainable transportation transitions.

3. METHODOLOGY

3.1 Overview

This study adopts a data-driven methodology combining Natural Language Processing (NLP) and Machine Learning (ML) to analyze consumer sentiment toward electric vehicles (EVs) in India. Unstructured review data was collected from three major automotive platforms—Bikewale, Carwale, and Cardekho—and processed using NLP techniques such as text cleaning, lemmatization, and sentiment scoring via VADER. The derived sentiment scores and vehicle attributes were used as inputs to various ML models, including Logistic Regression, Random Forest, AdaBoost, and LightGBM. These models were then evaluated to classify sentiment and predict EV market success. This approach enables a comprehensive understanding of consumer perception and supports informed decision-making for manufacturers and policymakers.

The data processing and model development were implemented in Python within the Google Colaboratory (Colab) environment, which provided the computational resources necessary for efficient execution and reproducibility of the analysis.

3.2 Data and Variables

Data Source

The datasets used in this research were obtained from Kaggle, containing user-generated reviews from platforms such as Bikewale, Carwale, and Cardekho. These platforms host extensive reviews on electric two-wheelers and four-wheelers available in the Indian market.

2-Wheeler Dataset

Table 1 presents the structure of the dataset used in this study, including 14 columns with varying levels of completeness. The dataset contains 844 entries, with key features such as 'rating', 'Visual Appeal', and 'Comfort' being partially filled.

Table 1: *Summary of attributes in the 2-wheeler electric (Bikewale) vehicle dataset*

Column Name	Non-Null Count	Data Type
Review	797	object
Used it for	844	object
Owned for	844	object
Ridden for	668	object
Ratings	844	int64
Visual Appeal	739	float64
Reliability	716	float64
Performance	345	float64
Service Experience	703	float64
Extra Features	185	float64
Comfort	530	float64
Maintenance cost	180	float64
Value for Money	390	float64
Model Name	844	object

4-Wheeler Datasets

Table 2 summarizes the structure of the first four-wheeler electric vehicle dataset, which primarily includes compact and sedan models. It contains 129 entries with attributes evaluating both qualitative (e.g., review, condition) and quantitative features (e.g., performance, comfort, fuel economy). These fields are used to assess user satisfaction and contribute to sentiment classification and predictive modeling.

Table 2: *Summary of attributes in the 4-wheeler electric vehicle(Carwale) dataset*

Column Name	Non-Null Count	Data Type
Review	129	object
Exterior	129	float64
Comfort	129	float64
Performance	129	float64
Fuel Economy	129	float64
Value for Money	129	float64
Condition	129	object
Driven	129	object
Rating	129	float64
Model Name	129	object

Table 3 outlines the structure of the second four-wheeler electric vehicle dataset, consisting of 140 entries related to premium models and user-expressed features. This dataset emphasizes open-text reviews and highlighted vehicle attributes, offering a qualitative perspective for deep sentiment analysis. It was particularly useful for extracting specific aspects (e.g., battery, design, comfort) mentioned by users in their feedback.

Table 3: *Summary of attributes in the 4-wheeler electric vehicle(Cardekho) dataset*

Column Name	Non-Null Count	Data Type
Review	140	object
Rating	140	float64
Attributes Mentioned	140	object
Model	140	object

Variables

Independent Variables (Predictors)

These variables influence consumer sentiment and market trends:

- Vehicle Attributes: Visual appeal, reliability, performance, comfort, maintenance cost, extra features, etc.
- Consumer Experience: Usage duration, driven kilometers, service experience.
- Economic Factors: value for money.
- Sentiment Scores: NLP-derived polarity scores from customer reviews.

Dependent Variable (Target Variable)

- Sentiment Classification: Categorized as positive, neutral, or negative, determined using sentiment analysis techniques.
- Market Success Prediction: Probability of an EV model's success based on user feedback and key influencing factors.

3.3 Natural Language Processing for Market Analysis

Natural Language Processing (NLP) techniques were employed to extract meaningful insights from unstructured text data such as consumer reviews. An overview of the full NLP pipeline is shown in **Fig 1**. It illustrates the flow from raw review data collection to final feature and label preparation for model training.

Text Preprocessing

Text preprocessing is a foundational step in Natural Language Processing (NLP) that prepares raw textual data for analysis by standardizing and simplifying the content. The following techniques were applied to clean and normalize the electric vehicle review data:

- **Cleaning:** Special characters, numerals, and punctuation marks were removed using regular expressions. All text was converted to lowercase to ensure consistency. *Example*: "Excellent performance!!! :)" → "excellent performance"
- **Stopword Removal:** Commonly used words such as "the," "and," "is," etc., that do not contribute significant semantic meaning, were removed using NLTK's predefined stopword list. *Example*: "the vehicle is very fast" → "vehicle fast"
- **Lemmatization:** Each word was reduced to its base or dictionary form using WordNetLemmatizer. This helps in grouping similar words (e.g., "drives," "driving," "driven") under a common lemma. *Example*: "driving" → "drive"

Feature Extraction

To transform cleaned textual data into a machine-readable format, the following techniques were used:

- **TF-IDF Vectorization:** The TfidfVectorizer was used to convert text into numerical feature vectors. It captures both term frequency and inverse document frequency, thereby emphasizing unique terms in each review. *Example*: In a corpus, the term "battery" appearing frequently in one document but rarely in others would get a high weight.
- **Sentiment Analysis using VADER:** The Valence Aware Dictionary and sEntiment Reasoner (VADER) tool from the NLTK library was used to compute sentiment polarity scores. VADER is particularly effective for social media text and short product reviews because it considers lexical features and also applies heuristic rules based on:
  1. Capitalization: Enhances emphasis (e.g., "AMAZING" is more positive than "amazing").
  2. Punctuation: Increases intensity (e.g., "great!!" is more intense than "great").
  3. Negation: Reverses sentiment polarity (e.g., "not good" becomes negative).
  4. Degree modifiers: Adjust intensity (e.g., "very bad" is worse than "bad").
  5. Each review is passed to SentimentIntensityAnalyzer().polarity_scores() which returns a dictionary with the following structure:

```
{
  'negative': 0.0,
  'neutral': 0.3,
  'positive': 0.7,
  'compound': 0.8316
}
```

The compound score (ranging from -1 to +1) was extracted to represent the overall sentiment of the review and stored in the sentiment_score column.

Example:
Review: "The EV drives smoothly and charges quickly!"
Compound Score: 0.84 → Positive Sentiment

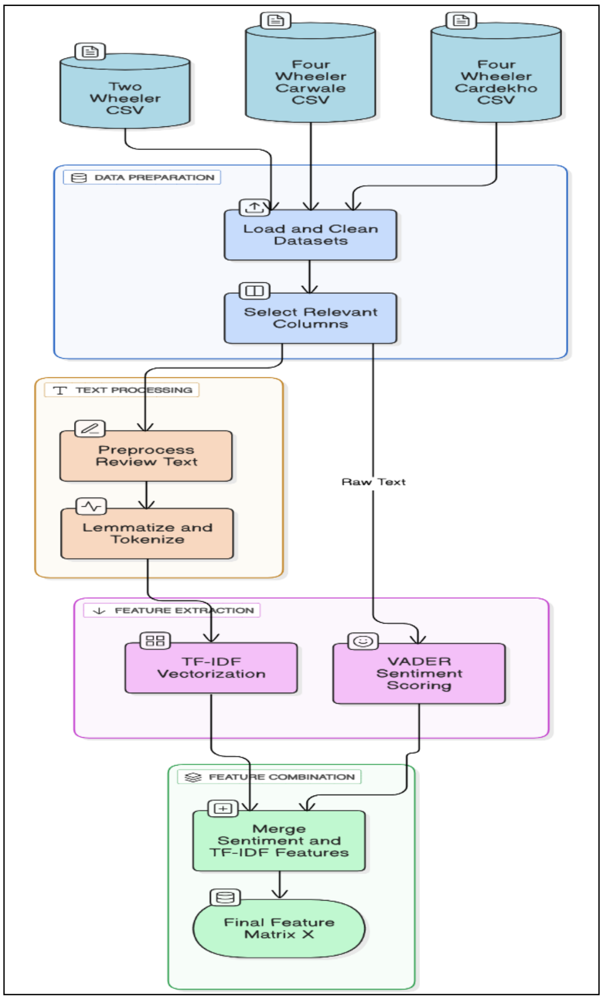

Fig. 1: *NLP Pipeline for Electric Vehicle Review Sentiment Analysis*

3.4 Feature Engineering

Key features were derived from both structured and unstructured components of the datasets:

- Ratings & Reviews: Sentiment scores were calculated from textual feedback.
- Vehicle Attributes: Included features such as performance, comfort, maintenance cost, and value for money.
- Sentiment Scores: Generated using NLP tools like VADER to quantify consumer opinions.

- Label Encoding: Sentiment labels (e.g., *positive*, *neutral*, *negative*) were converted into numerical values using *LabelEncoder*, making them suitable for supervised learning.

Sentiment Analysis using VADER

To extract sentiment-related insights from user-generated reviews, the **VADER (Valence Aware Dictionary and sEntiment Reasoner)** sentiment analysis tool was employed. VADER calculates sentiment intensity based on a predefined lexicon of sentiment-laden words and accounts for contextual cues such as punctuation, capitalization, negation, and degree modifiers.

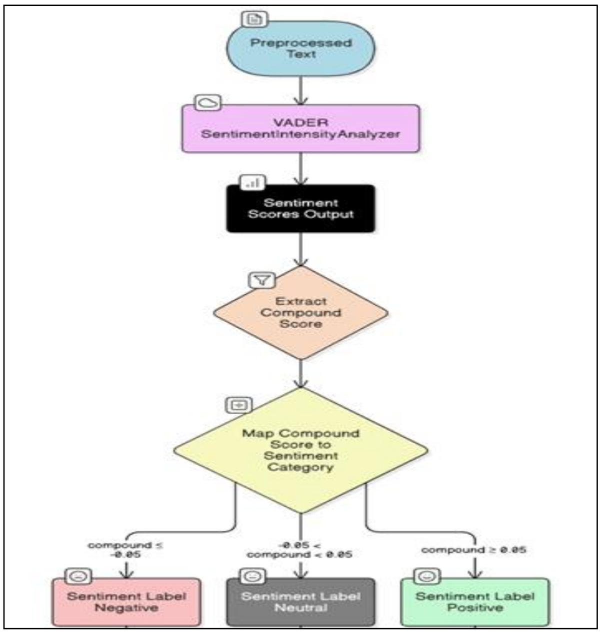

Fig.2: *VADER-Based Sentiment Scoring and Rating Mapping Workflow*

When applied to pre-processed textual reviews, VADER returns a dictionary of four sentiment scores:

- **pos**: Proportion of positive sentiment in the text
- **neu**: Proportion of neutral sentiment
- **neg**: Proportion of negative sentiment
- **compound**: A normalized, aggregated sentiment score ranging from -1 (most negative) to +1 (most positive)

The **compound score** is a weighted sum of the valence scores of each word in the text, adjusted according to syntactic and grammatical rules. It serves as a comprehensive measure of the overall sentiment of a sentence or document.

To categorize reviews into discrete sentiment classes, the following thresholds were applied to the compound score:

- Compound ≤ -0.05 → **Negative Sentiment**
- -0.05 < Compound < 0.05 → **Neutral Sentiment**
- Compound ≥ 0.05 → **Positive Sentiment**

Each review was then labelled accordingly, and sentiment labels were subsequently mapped to numerical rating categories to align with the 5-point feedback system commonly used in consumer evaluations:

- **Negative sentiment** → Ratings **1–2**
- **Neutral sentiment** → Rating **3**
- **Positive sentiment** → Ratings **4–5**

This VADER-based sentiment scoring and mapping process ensures consistency between qualitative feedback and quantitative ratings, as illustrated in Fig. 2 thereby improving the reliability of downstream machine learning models trained to predict or classify user sentiment.

Integration in the Pipeline

As illustrated in Fig. 1, data from three separate CSV files were merged into a unified DataFrame. The complete NLP pipeline was applied as follows:

1. Reviews were cleaned, tokenized, and lemmatized.
2. Features were extracted using TF-IDF vectorization and VADER () sentiment scoring.
3. These text-derived features were then combined with structured vehicle attributes.
4. Sentiment labels were encoded and used to train classification models.

This comprehensive feature matrix X and the corresponding label vector y served as the input to multiple classification models including Random Forest, Logistic Regression, and LightGBM.

3.5 Predictive Modeling

As illustrated in Fig. 3, the machine learning model training pipeline follows a systematic approach to sentiment-based prediction. After preparing the feature matrix X and target vector y, the data undergoes a structured modeling workflow starting with a train-test split, followed by feature selection to retain the most informative variables.

Predictive modeling was conducted to estimate market sentiment and evaluate the potential success of EV products using the following machine learning techniques:

Regression Algorithms

- Logistic Regression: Used for binary sentiment classification and market success prediction.
- Ridge Regression: Applied for continuous variable prediction with regularization to prevent overfitting.

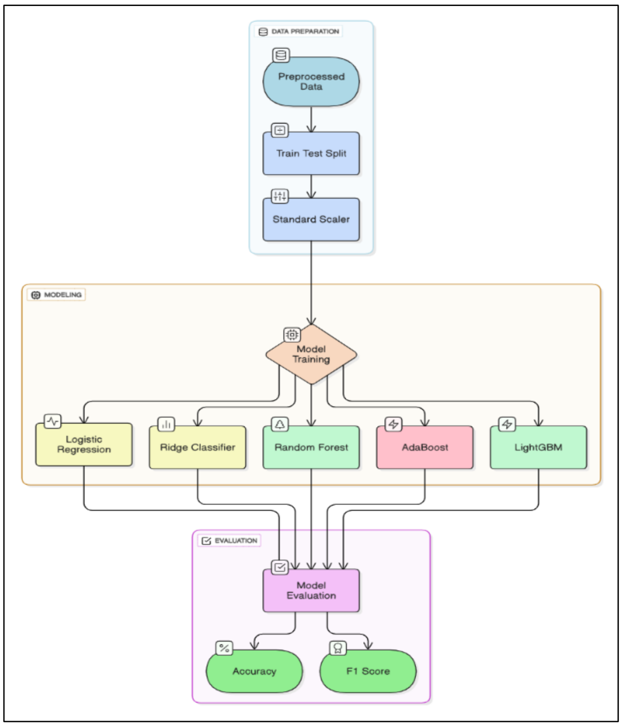

Fig. 3: *Machine Learning Model Training Pipeline*

Ensemble Methods

- Random Forest: Utilized multiple decision trees to enhance classification accuracy.
- AdaBoost: Focused on misclassified instances to improve the performance of weak learners.

Gradient Boosting Framework

- LightGBM: Employed for its high speed and efficiency, and it delivered the best accuracy in sentiment classification among all models tested.

4. RESULTS

The Machine Learning models were evaluated for sentiment classification using key performance metrics such as Accuracy, F1-Score, Mean Squared Error (MSE), and R^2 score. The results indicate the effectiveness of each model in predicting sentiment with varying degrees of accuracy and robustness. Below is a summary of their performance :

	Model Name	Accuracy	F1-Score	MSE	R²
0	Random Forest	0.967290	0.960327	0.574766	0.950897
1	Logistic Regression	0.710280	0.660929	4.000000	0.658275
2	AdaBoost	0.981308	0.973298	0.168224	0.985628
3	LightGBM	0.995327	0.997196	0.168224	0.985628
4	Ridge Classifier	0.728972	0.706598	3.434579	0.706580

Fig. 4: *Evaluation Metrics of ML Models for EV Sentiment Analysis*

- Logistic Regression served as the baseline model, achieving an accuracy of 71.03% and an F1-score of 0.6609. Despite its simplicity, it provided a reasonable benchmark for comparison.
- Random Forest Classifier demonstrated strong predictive power with an accuracy of 96.73% and an F1-score of 0.9603, showcasing its ability to handle complex patterns effectively.
- AdaBoost further improved upon the Random Forest model, achieving an accuracy of 98.13% and an F1-score of 0.9733, indicating its strength in boosting weak learners to enhance performance.
- Ridge Classifier, incorporating regularization, achieved an accuracy of 72.89% with an F1-score of 0.7066, offering moderate performance while addressing overfitting concerns.
- LightGBM outperformed all other models with the highest accuracy of 99.53% and an exceptional F1-score of 0.9972. Its efficiency in handling large datasets and feature extraction made it the most effective model for sentiment classification.

Overall, LightGBM emerged as the best-performing model, followed closely by AdaBoost and Random Forest, while Logistic Regression and Ridge Classifier provided baseline comparisons. The low MSE and high R^2 scores for LightGBM and AdaBoost further validate their reliability in classification tasks.

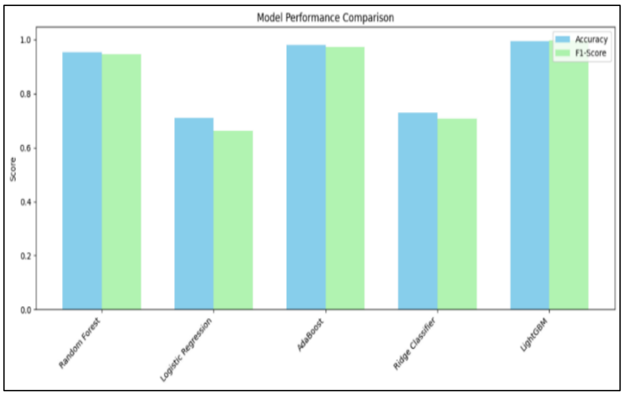

Fig. 5 : *Comparison of Accuracy and F1-Score Across ML Models for Sentiment Analysis*

Predictions Using the Best Performing Model

The Fig. 6 presents example outputs generated by the LightGBM model, which demonstrated the highest accuracy among the tested machine learning models. The model was trained on a variety of natural language processing (NLP) features extracted from customer reviews and their corresponding ratings.

Initially, the textual reviews were analysed using sentiment analysis techniques, including VADER, to obtain compound sentiment scores. These scores were then mapped onto a 1–5 rating scale to align with the original review ratings. Multiple machine learning models were subsequently trained on these NLP-derived features, with LightGBM outperforming others in both accuracy and F1 score.

The predictions shown in the Fig. 6 illustrate the model's ability to capture sentiment polarity and generate predicted ratings consistent with both the compound sentiment scores and the review context. This demonstrates the model's effectiveness for automated sentiment-based rating prediction.

```
Review: Absolutely love this bike! It's smooth and fuel efficient.
Predicted Rating : 5.0
Compound Score: 0.822
Mapped VADER Score (1-5): 4.64
Sentiment Label: Positive
---
Review: Worst experience ever. The engine failed in just two months.
Predicted Rating : 1.0
Compound Score: -0.813
Mapped VADER Score (1-5): 1.37
Sentiment Label: Negative
---
Review: It's okay, not great but not too bad either.
Predicted Rating : 4.5
Compound Score: 0.473
Mapped VADER Score (1-5): 3.95
Sentiment Label: Positive
---
Review: It's an average experience
Predicted Rating : 2.0
Compound Score: 0.000
Mapped VADER Score (1-5): 3.0
Sentiment Label: Neutral
```

Fig. 6 : *Predictions of Sentiment and Ratings from the LightGBM Model*

To further validate our model performance, we compared our results with findings from recent literature on electric vehicle forecasting and sentiment analysis.

Afandizadeh, S., Sharifi, D., Kalantari, N., et al. [1] proposed a hybrid machine learning model incorporating social media sentiment (via the VADER method), Google Trends, and vehicle registration data for forecasting EV adoption. Their model achieved a Mean Absolute Error (MAE) of 3.5% (Afandizadeh et al. 2023). Although this approach was regression-based and targeted at predicting market penetration, it confirmed the predictive strength of sentiment-informed features, aligning with our use of unstructured text reviews for market insight.

Mishra, R. D., Dash, S. K., Chudjuarjeen, S., et al. [4] examined forecasting methods for EV market trends using models such as Support Vector Regression and Long Short-Term Memory (LSTM) networks. While exact performance metrics were not reported, LSTM was highlighted as the best-performing model for time-series forecasting tasks. In contrast, our work targets sentiment classification, where LightGBM yielded an F1-score of 0.9972 and an accuracy of 99.53%, surpassing previously reported performance in related contexts (Mishra et al. 2024).

Warpe, V. S., Buchkul, S. D., et al. [21]explored various models, including Logistic Regression, SVM, Decision Trees, and LightGBM, for EV sales price forecasting. Although specific metrics were not disclosed, they emphasized the importance of F-score and sensitivity in model evaluation—metrics that are especially crucial in our use case, given the class imbalance in sentiment data. Our LightGBM model outperformed baseline classifiers with the highest F1-score and minimal error, confirming its suitability for consumer sentiment analysis (Warpe et al. 2024).

These comparisons indicate that our ensemble models—particularly LightGBM and AdaBoost—not only align with current research trends but also achieve superior results in EV-related prediction tasks using unstructured consumer data.

5. CONCLUSION

This study demonstrates the effectiveness of machine learning and natural language processing techniques in analyzing consumer sentiment and predicting market trends in the Indian electric vehicle (EV) industry. By leveraging datasets from Bikewale, Carwale, and Cardekho, the research uncovers key factors influencing EV adoption, such as performance, reliability, cost, and consumer perception. The analysis highlights that two-wheeler and four-wheeler EV reviews exhibit distinct sentiment trends, with a general inclination toward positive feedback.

Among the machine learning models tested, LightGBM emerged as the best-performing algorithm, achieving the highest accuracy in sentiment classification. Its ability to handle large datasets efficiently while prioritizing complex samples makes it a superior choice for predictive modeling in EV market analysis. The insights derived from this study can assist EV manufacturers, marketers, and policymakers in designing strategies to enhance adoption rates and optimize product offerings based on consumer sentiment.

To make these findings more accessible and actionable, a real-time dashboard can be developed using Streamlit. This dashboard can provide interactive visualizations, allowing users to predict sentiment scores for new consumer reviews and recommend top-performing EV models based on sentiment analysis. By integrating machine learning predictions into an intuitive interface, stakeholders can gain real-time insights into consumer perceptions and market dynamics, enabling data-driven decision-making for sustainable growth in the EV sector.

Future work can explore the integration of real-time data sources such as social media sentiment, government policy updates, and emerging market trends to further refine predictive accuracy. Additionally, advancements in explainable AI techniques can enhance trust in model-driven recommendations, fostering greater adoption of data-driven methodologies in the EV industry.

6. REFERENCES

[1] Afandizadeh, S., Sharifi, D., Kalantari, N., et al. 2023. "Using machine learning methods to predict electric vehicles penetration in the automotive market." Scientific Reports 13 (May). Available: https://www.nature.com/articles/s41598-023-35366-3.

[2] Shu, T., Wang, Z., Lin, L., et al. 2022. "Customer Perceived Risk Measurement with NLP Method in

Electric Vehicles Consumption Market: Empirical Study from China." Energies 15, no. 5: 1637 (February). Available: https://www.mdpi.com/1996-1073/15/5/1637.

[3] Cao, C. 2022. "Research on Marketing Strategy of Electric Vehicle Based on Bayesian Processing and Natural Language Analysis." In Proceedings of SDPIT 2022 Conference (July). Available: https://drpress.org/ojs/index.php/HSET/article/view/850.

[4] Mishra, R. D., Dash, S. K., Chudjuarjeen, S., et al. 2024. "Forecasting EV Market Trends in India: A Deep Learning Approach for Two/Three-Wheelers." In Proceedings of 2024 iEECON Conference (March). Available: https://ieeexplore.ieee.org/document/10537867.

[5] "EV Sales Price Forecasting Using Machine Learning." 2024. In Proceedings of ICSCSS Conference (July). Available: https://ieeexplore.ieee.org/document/10625450.

[6] Kumar, S., and Gawade, S. 2023. "Sentiment Analysis of Opinions About Electric Vehicles Using Twitter Data." In Proceedings of 2023 CICT Conference (December). Available: https://ieeexplore.ieee.org/document/10455654.

[7] Thiruneelakandan, A., and Umamageswari, A. 2023. "Federated Learning Approach for Analyzing Electric Vehicle Sales in the Indian Automobile Market." In Proceedings of 2023 RMKMATE Conference (November). Available: https://ieeexplore.ieee.org/document/10369875.

[8] Lee, S., and Park, M. 2021. "Understanding EV Market Trend: Using Time Series Dynamic Topic Modeling with YouTube Data." In Proceedings of 2021 IEEE International Conference on Big Data (December). Available: https://ieeexplore.ieee.org/document/9671884.

[9] Araiza, J. A. G., Luna, S., Santiago, I., et al. 2024. "Perceptions of Electric Vehicle Adoption through Natural Language Processing and Machine Learning." In Proceedings of 2024 IEEE SysCon (April). Available: https://ieeexplore.ieee.org/document/10553625.

[10] Sathyan, S., Peedikayil, J. J., P. V., R., et al. 2023. "Two-Layered Machine Learning Approach for Sentiment Analysis of Tweets Related to Electric Vehicles." In Proceedings of 2023 ICIET Conference (July). Available: https://ieeexplore.ieee.org/document/10220717.

[11] Bau, R. T. R. L., Hermila, A., and Latief, M. 2024. "Electric Vehicle Adoption in Indonesia: Insights from Sentiment Analysis of Major News Portals." In Proceedings of 2024 ICoDSA Conference (July). Available: https://ieeexplore.ieee.org/document/10651692.

[12] Elouariaghli, F. N., Kozderka, S. M., Quaranta, T. G., et al. 2023. "A Machine Learning-Based Method for Parametric Environmental Impact Model for Electric Vehicles." Resources, Conservation and Recycling 193. Available: https://www.sciencedirect.com/science/article/pii/S0959652624017566.

[13] Shri, R. M., and Raj, S. N. V. 2022. "Machine Learning Models for Predicting Customer Willingness to Buy Electric Vehicles." In Proceedings of 2022 Springer LNDECT Conference (September). Available: https://link.springer.com/chapter/10.1007/978-981-19-3015-7_30.

[14] Miconi, F., and Dimitri, G. M. 2023. "A Machine Learning Approach to Analyze and Predict the Electric Cars Scenario: The Italian Case." PLoS ONE 18 (January). Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0279040.

[15] Yeh, J., and Wang, Y. 2023. "A Prediction Model for Electric Vehicle Sales Using Machine Learning Approaches." Journal of Global Information Management 31, no. 1. Available: https://www.igi-global.com/gateway/article/327277.

[16] Dixit, S. K., and Singh, A. K. 2022. "Predicting Electric Vehicle (EV) Buyers in India: A Machine Learning Approach." Review of Socionetwork Strategies 16 (May). Available: https://link.springer.com/article/10.1007/s12626-022-00109-9.

[17] Tripathy, N., Hota, S., Satapathy, P., et al. 2023. "An Empirical Analysis of Electric Vehicle in Urban Transportation Market Using Deep-Learning Techniques." In Proceedings of 2023 IEEE SEFET Conference (August). Available: https://ieeexplore.ieee.org/document/10245669.

[18] Prstkrish, D. "EV Cars User Reviews India." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/deadprstkrish/ev-cars-user-reviews-india.

[19] Jena, Rabindra. 2019. “An Empirical Case Study on Indian Consumers' Sentiment Towards Electric Vehicles: A Big Data Analytics Approach.” Industrial Marketing Management 86 (May): 210–220. https://doi.org/10.1016/j.indmarman.2019.12.012.

[20] Myneni, M. B., Akkineni, H., Mai, C. K., and Boppana, S. 2024. “Analytical Framework to Understand Electric Vehicle Adoption by Leveraging Sentiment Analysis.” *Journal of Mobile Multimedia* 20, no. 05: 1067–1088. https://doi.org/10.13052/jmm1550-4646.2054.

[21] Warpe, V. S., Buchkul, S. D., et al. 2024. “EV Sales Price Forecasting Using Machine Learning.” *2024 2nd International Conference on Sustainable Mobility*. https://ieeexplore.ieee.org/document/10625450