Sentiment Analysis of Electric Vehicles

Work Done- Steps involved in analysing sentiment towards electric vehicles (EVs) in India:

➢ Data Collection

1. **Target Websites:** Identify the top websites in India that focus on reviews related to EVs. Consider including a mix of general automotive sites and those specializing in EVs.
2. **Web Scraping:** Utilize web scraping technique s to extract reviews and relevant content from these websites. Used Octoparse, a web scraping tool that allows us to extract data from websites.

Websites:

- **India:**
  - Bikewale.com
  - Bikedekho.com
  - Cardekho.com
  - Carwale.com

➢ Data Preprocessing

1. **Cleaning:** Remove irrelevant information like HTML tags, punctuation, and stop words (common words like "the," "a").
2. **Language Processing:** Depending on the website language (English or the local language), apply Natural Language Processing (NLP) techniques lemmatization to reduce words to their root form.
3. **Sentiment Labelling:** Annotate a small portion of the data (reviews) manually with sentiment labels (positive, negative, neutral) for training a machine learning model.

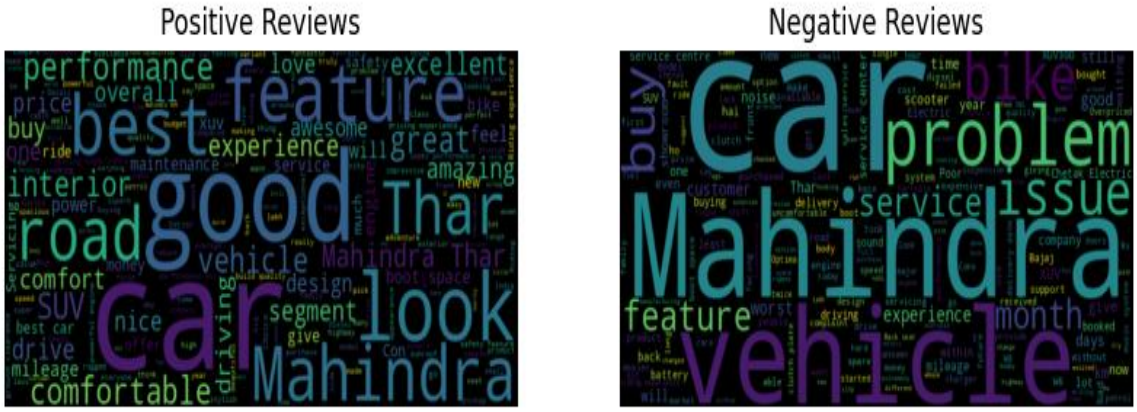
➢ Data Analysis

Dataset consists of 2107 reviews after data cleaning. The structure of the dataset from different websites is as follows:

Websites	Reviews
Carwale.com	618
Bikedekho.com	875
Bikewale.com	43
Cardekho.com	571
Total	2107

This is the dataset with 6 features:

	Title	Review	Review_by	Time	URL	Sentiment
0	Very very Delayed Delivery of Car	This is a safety car. \nBut the only issue whi...	Ranjan Kumar Meher	3 years ago	https://www.carwale.com/mahindra-cars/xuv300/u...	negative
1	Disturbing noise from brake and suspension	After 10 days of purchase.. Sound started comi...	Jumpe Maro	4 years ago	https://www.carwale.com/mahindra-cars/xuv300/u...	positive
2	XUV300 has Clutch Issues.	After driving a few 100 kilometres I found the...	Risabh	4 years ago	https://www.carwale.com/mahindra-cars/xuv300/u...	negative
3	Amazing machine with a great value for money	I bought this car 3 months back after doing 6 ...	Anurag sharma	2 years ago	https://www.carwale.com/mahindra-cars/xuv300/u...	positive
4	Noise issue in xuv300 do not buy	This car has noise from front axle on bumpy ro...	pulkit chauhan	4 years ago	https://www.carwale.com/mahindra-cars/xuv300/u...	negative



Word Cloud by Sentiment

Top 5 Positive Reviews:

Best

Comfortable

Good

Look

Excellent

Top 5 Negative Reviews:

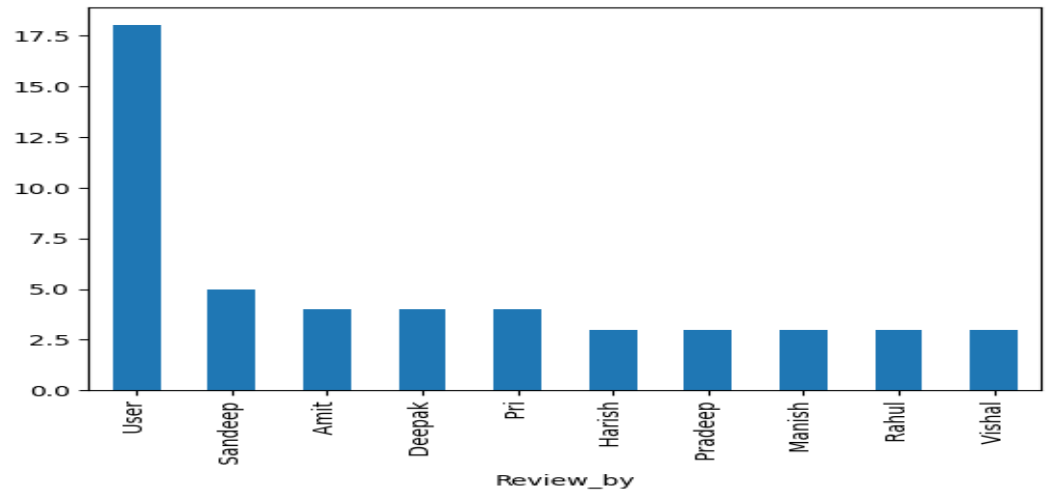
Buy - Cost

Service

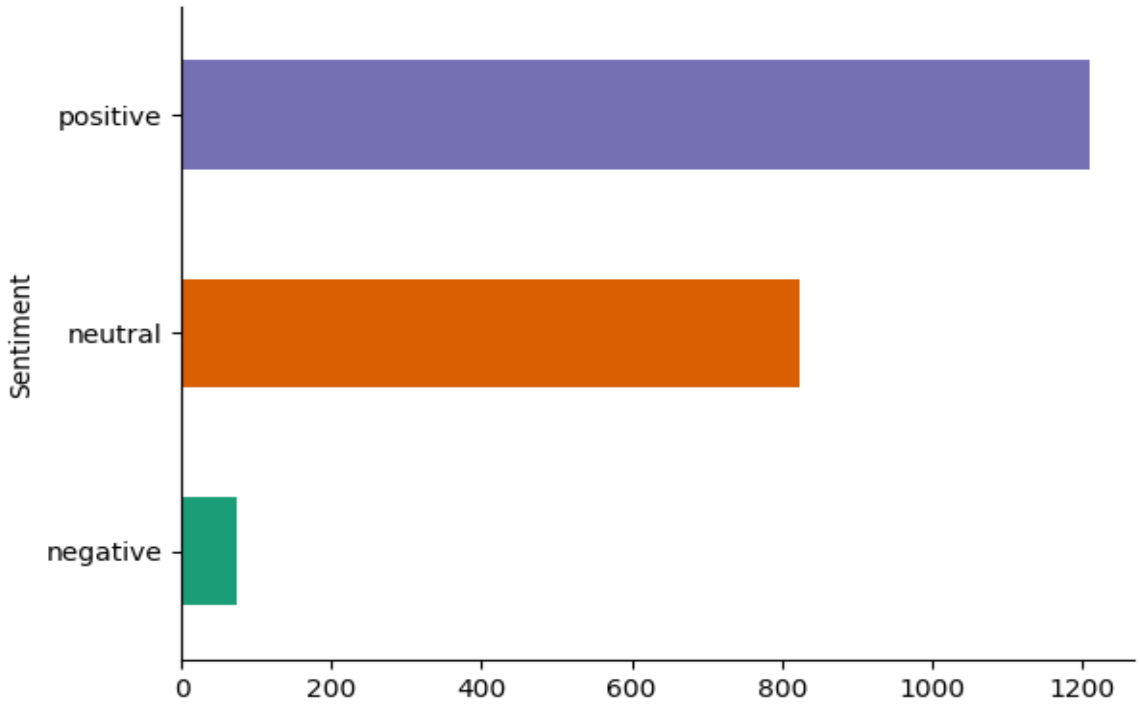
Feature

Issue - Charging

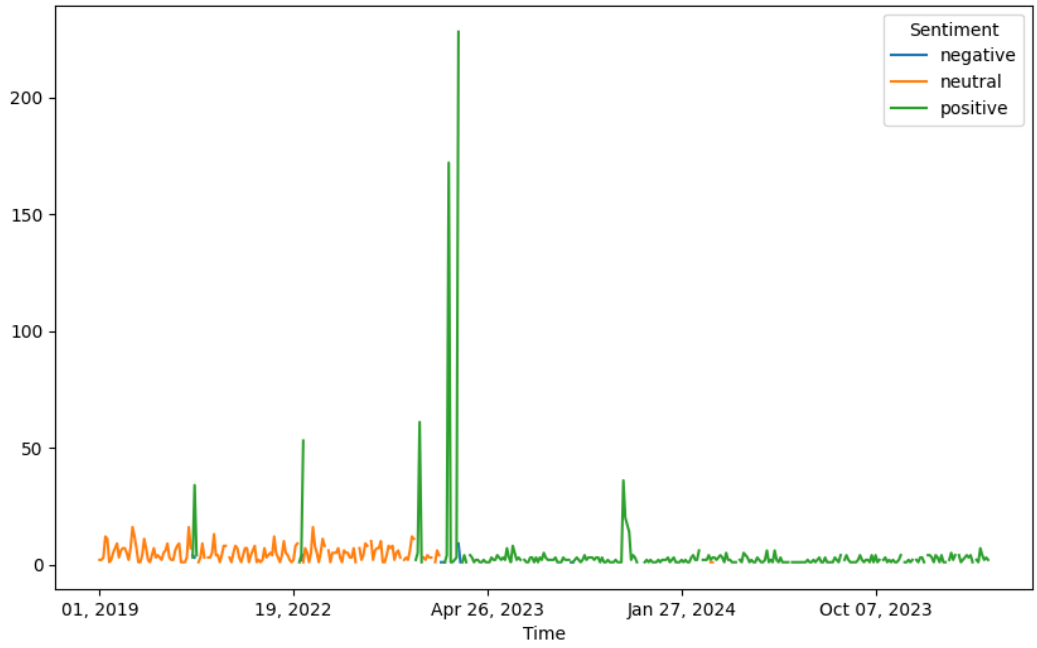
Problem



Top Positive Reviewers



Sentiment



Sentiment over Time

➢ Data Preprocessing for NLP

Used one hot encoding and the word2vec for data preprocessing and the feature engineering.

1. **One-hot Encoding:** Used to represent categorical data numerically.
2. **Word2Vec:** Used to represent words as dense vectors in a continuous vector space.

➢ Machine Learning/Deep Learning

1. **Model Selection:**.
   - **Machine Learning:** Used Naive Bayes.
   - **Deep Learning:** Used Long Short-Term Memory (LSTM), SimpleRNN, Bidirectional LSTM, CNN and CNN-LSTM.
2. **Training:** Train the chosen model using the labelled data.
3. **Evaluation:** Evaluate the model's performance on a separate validation dataset to ensure its accuracy. evaluate their performance on the testing data using metrics like accuracy, precision, recall, and F1-score.

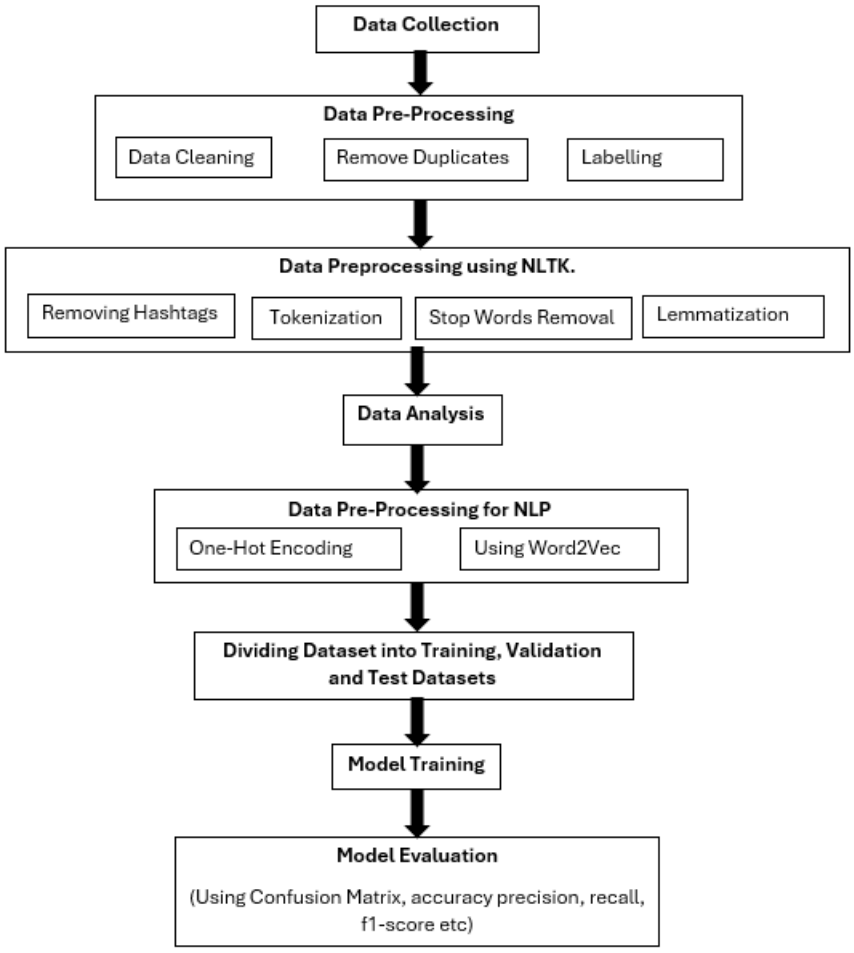


Figure: Methodology

➢ Analysis and Insights

1. **Sentiment Analysis:** Apply the trained model to analyze the sentiment of the scraped reviews from country.
2. **Insights:** Identify key themes and trends in the reviews. This could include:
    - Common concerns about EVs (e.g., range anxiety, charging infrastructure)
    - Positive aspects of EVs (e.g., environmental benefits, performance)

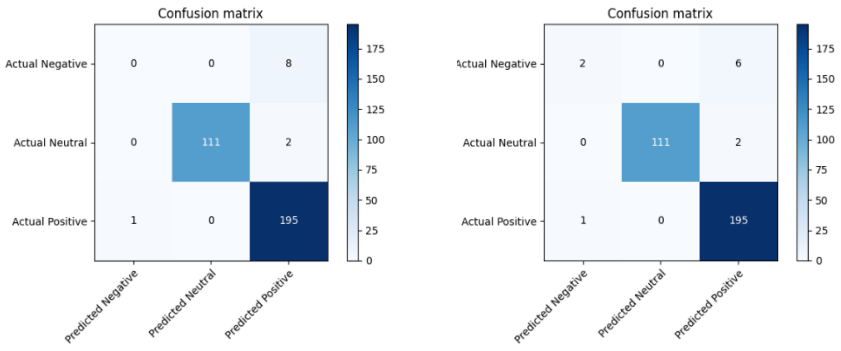
Results:

Table: Accuracies of different algorithms:

Algorithm used	Accuracy
LSTM	96.84%
Naïve Bayes	72.56%
Bidirectional LSTM	95.58%
SimpleRNN	96.52%
CNN	97.16%
CNN-LSTM	96.85%

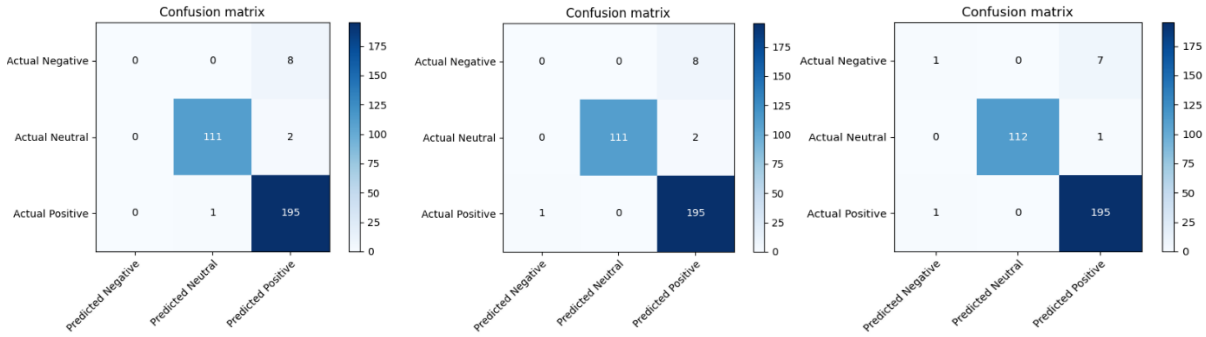
Table: Performance report for Different Algorithms:

Algorithm Used	Precision	Recall	F1-Score
LSTM	0.960	0.968	0.961
Naïve Bayes	0.942	0.725	0.803
Bidirectional LSTM	0.960	0.955	0.957
SimpleRNN	0.940	0.970	0.950
CNN	0.970	0.970	0.970
CNN-LSTM	0.960	0.970	0.960



(a) LSTM

(b) Bidirectional LSTM



(c) SimpleRNN (d) CNN (e) CNN-LSTM

Conclusion:

- We've received 1200 positive reviews on EVs in India, indicating a predominantly favorable sentiment among users.

- Sentiment analysis reveals a notable shift over time. Initially, in 2019, there was some negativity surrounding EVs, but as time progressed, this sentiment diminished significantly, transitioning into overwhelmingly positive reviews by 2023.

- Among various deep learning models compared, both CNN and CNN-LSTM models exhibited the highest accuracies at 97.16%. Given their strong performance on our dataset, utilizing deep learning for analysis appears advantageous.

- In the positive reviews, we observed mentions of eco-friendliness, aesthetics, and overall quality, indicating a strong emphasis on environmental consciousness and satisfaction with the product's appearance and performance.

  Conversely, negative reviews primarily highlighted concerns regarding cost and the adequacy of charging infrastructure, underscoring challenges related to affordability and accessibility in the EV market.

- **Policymakers** can leverage this analysis to shape policies promoting EV adoption. Insights into consumer concerns can guide efforts addressing challenges such as charging infrastructure development and battery range limitations.

- **Companies** in the EV industry stand to benefit from understanding consumer sentiment. This knowledge enables manufacturers to tailor their products and marketing strategies to better meet customer needs and preferences across different countries.