Twitter data Sentiment Analysis on Electric Vehicles

Abstract

We intend to explore Twitter data related to tweets about electric vehicles and electrical mobility by using sentiment analysis. The analysis has been carried out to capture the sentiment among the people about the change in the traditional transportation system, from fossil fuels to electrical mobility.
1 Lakh tweets have been extracted for roughly a year using the Social network scrape (Snscrape) python package. The analysis of these tweets has been caried out using AFINN package.
The excerpts from the analysis have been visualised to further draw insights using matplotlib and pandas visualisation packages.

Introduction:

Twitter is a popular social networking site where users from all over the world express their opinions. Twitter generates over 500 million tweets every day, around 8 TB of data (Shreya Ahuja, 2017).

When we evaluate the data that Twitter generates, we may use opinion mining to extract useful information, which makes the data very beneficial. Twitter data may clearly display thoughts on any news item, new product launch, or specific trend.

Sentiment analysis is the technique of analytically identifying and categorising opinions stated in a piece of text, particularly to determine if the writer has a positive, negative, or neutral attitude toward a specific subject, product, etc. The main objective of sentiment analysis, commonly referred to as opinion mining, is to discover subjectivity, subject, emotion, and opinion in a natural text.

Using one or more electric motors to power the vehicle for transportation is called electric mobility. Currently, this form of transportation provides options for both short-distance, light-weight journeys (using bicycles, scooters, and electric motorbikes) and long-distance, heavy-weight ones with electric public transport vehicles.

Electric mobility has several benefits, one of which is that it enhances people's quality of life by not generating harmful pollutants. Additionally, because electric vehicles lack combustion engines, they don't emit as much greenhouse gases, helping in the fight against the consequences of climate change.

Though it is a way forward in the sustainable transportation industry, the infrastructure and the customer awareness about the EVs is still in the early stage, when compared to its potential. So it serves as a great topic to know what the people really feel about this shift.

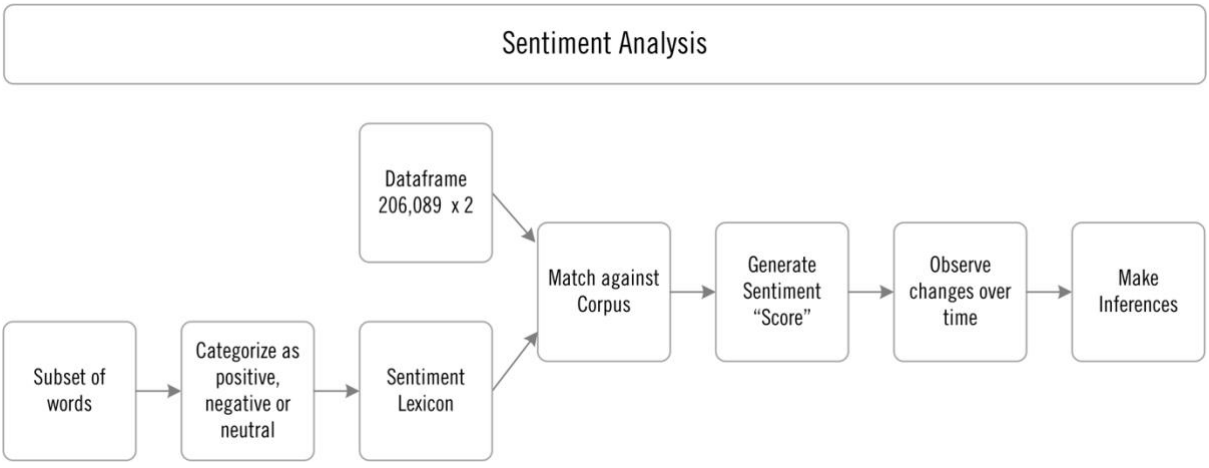
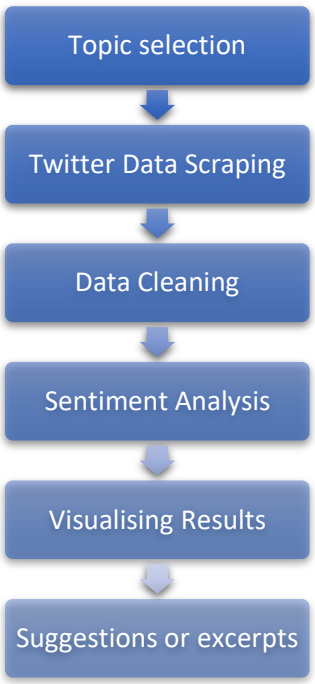


Figure 1: Roadmap of sentiment analysis

Methodology:



Topic selection:

We started off with selecting the hashtags for the topic Electrical Mobility (#electricalvehicles, #EV, #electricalmobility).

Twitter Data Scraping:

We used the **Snscrape** python package to scrape the data from twitter. Snscrape allows you to scrape basic information such as a user's profile, tweet content, source, and so on.

Its advantages are that there are no limits to the number of tweets you can retrieve or the window of tweets (that is, the date range of tweets). So Snscrape allows you to retrieve old data.

Using Snscrape we extracted 1 Lakh tweets with the hashtags #electricalvehicles, #EV and #electricalmobility from 1st January 2022 to 1st November.

Data Cleaning:

```
def preprocess_tweet(sen):
    '''Cleans text data up, leaving only 2 or more char long non-stepwords composed of A-Z & a-z only
    in lowercase'''

    sentence = sen.lower()

    # Remove RT
    sentence = re.sub('RT @\w+: '," ", sentence)

    # Remove special characters
    sentence = re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t])|(\w+:\/\/\S+)"," ", sentence)

    # Single character removal
    sentence = re.sub(r"\s+[a-zA-Z]\s+", ' ', sentence)  # When we remove apostrophe from the word "Mark's", the apostrophe is replaced by

    # Remove multiple spaces
    sentence = re.sub(r'\s+', ' ', sentence)  # Next, we remove all the single characters and replace it by a space which creates multiple

    return sentence
```

The raw tweets scraped are cleaned by removing the retweets, special characters, multiple spaces and stop words to analyse as precisely as possible.

Sentiment Analysis:

- Sentiment analysis is done using an Unsupervised Learning technique, based on the AFINN lexicon.
- The afinn object contains a method, called score(), which receives a sentence as input and returns a score as output. The score may be either positive, negative or neutral.
- We calculate the score of a book, simply by summing all the scores of all the sentence of that book.
- We define three variables positive, negative and neutral, which store respectively the sum of all the positive, negative and neutral scores of all the sentences of a book.
- If the score > 0, it is considered ‘positive’. If score < 0, then ‘negative’. Else is considered to be ‘neutral’.

Results:

	Total	Percentage
neutral	49609	49.61
positive	39861	39.86
negative	10533	10.53

We could see the count of the tweets classified into neutral, pose and negative classes.

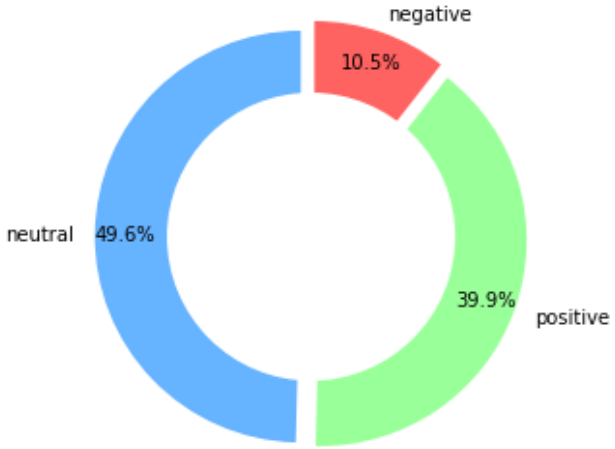


Figure 2: Donut chart showing distribution of tweets

The positive tweets are 39.9% whereas the negative tweets are only 10.5%. This shows that the majority of the people are pro EV and are optimistic & encouraging the use of electric vehicles.

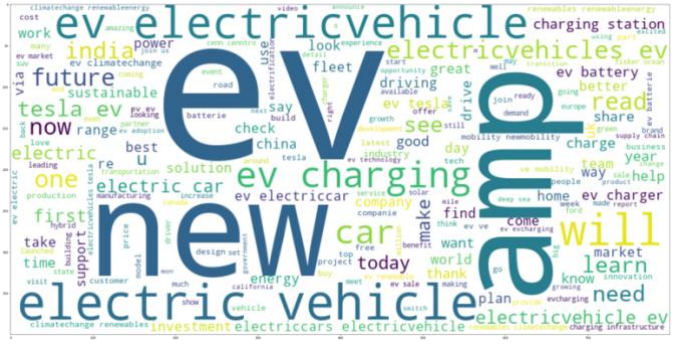


Figure 3: Positive tweets wordcloud

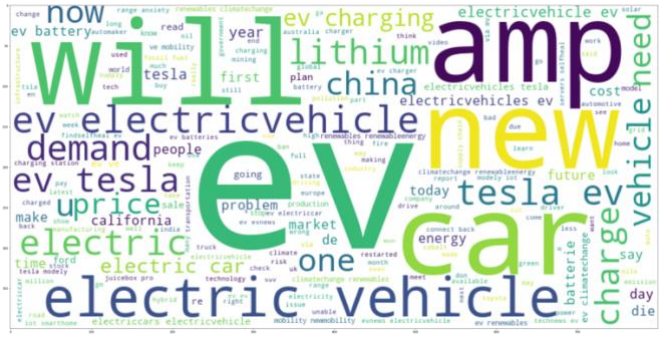


Figure 4: Negative tweets wordcloud

The above word cloud shows the words used mostly in positive and negative tweets.

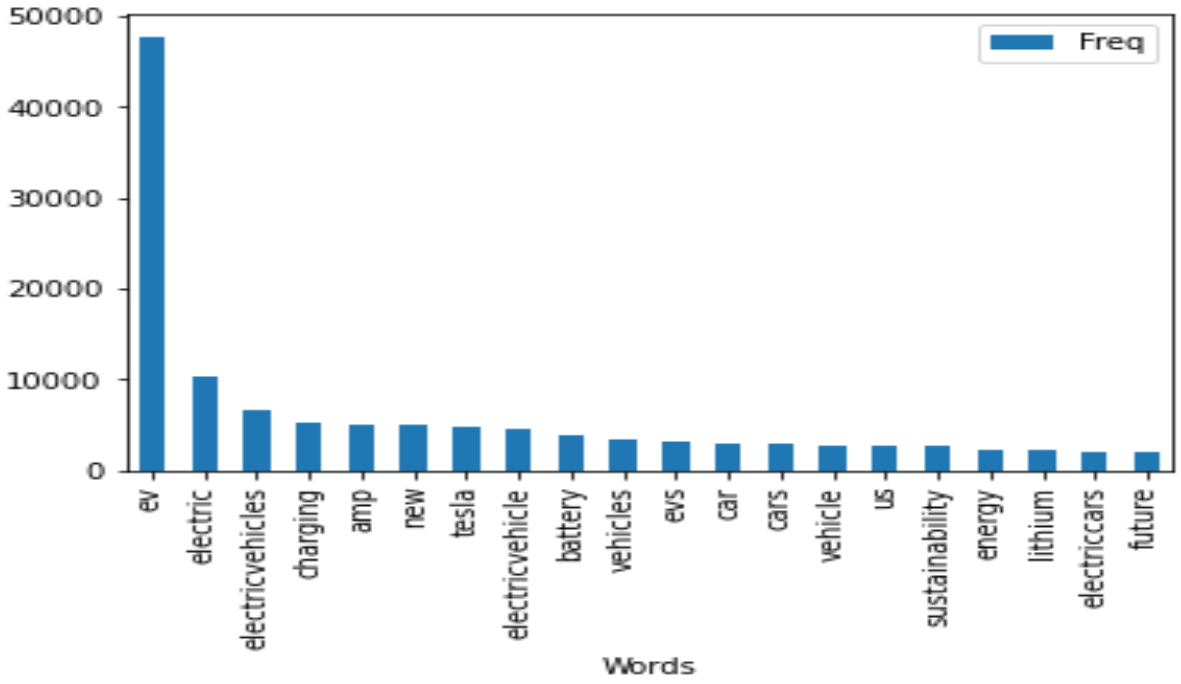


Figure 5: Top 20 word frequency in positive tweets

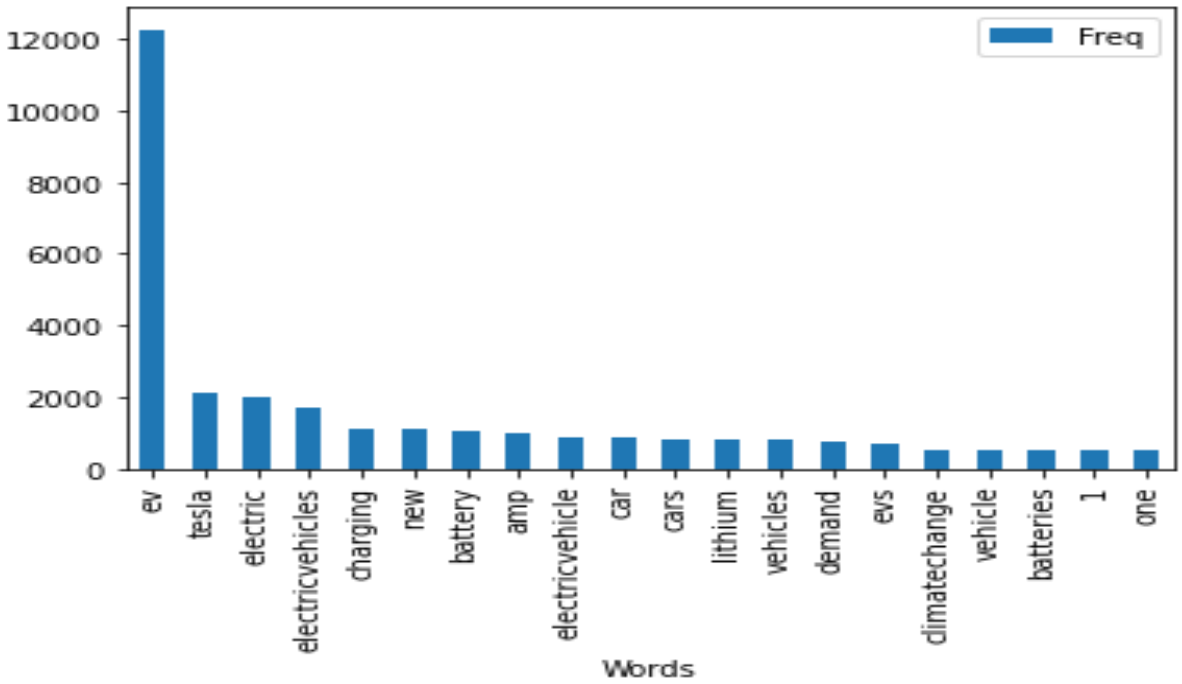


Figure 6: Top 20 word frequency in positive tweets

Bibliography

Shreya Ahuja, G. D. (2017). Clustering and Sentiment Analysis on Twitter Data. *2nd International Conference on Telecommunication and Networks (TEL-NET 2017)*.

(n.d.). Retrieved from https://medium.com/codex/making-wordcloud-of-tweets-using-python-ca114b7a4ef4

(n.d.). Retrieved from https://bookdown.org/psonkin18/berkshire/sentiment.html

(n.d.). Retrieved from https://medium.com/@kvnamipara/a-better-visualisation-of-pie-charts-by-matplotlib-935b7667d77f