Initial Plan

Cardiff University School of Computer Science and Informatics

ANALYSING THE DISCOURSE RELATED TO ELECTRICAL VEHICLES ON SOCIAL MEDIA

CM3203 – One Semester Individual Project – 40 Credits

Author: Anthos Kountouris

Supervisor: Dr Padraig Corcoran

Project Description

By 2035, the UK government has committed to ban the sale of new vehicles that are solely powered by petrol and diesel and transition to electrical vehicles [1]. Many people support this goal and believe that it will result in a more sustainable world. On the other hand, other people have concerns about this transition. These concerns include not being able to buy new electrical vehicles, but also not being able to charge them.

This project focuses on analysing the discourse related to electrical vehicles on social media and specifically Twitter. Twitter is a social media platform in which users are able to post and interact with messages up to 280 characters, known as "tweets" [2]. The first stage of the project will be the collection of the data from Twitter. The programming language I will be using is Python because it is a language that I have some background on from previous years of the University and it has many libraries and tools for Natural Language Processing. For the purposes of the analysis, I will apply a number of NLP analysis techniques including sentiment analysis and topic modelling. I will use the Twitter API as it gives the ability to collect tweets in real-time, based on keywords, user ids and locations. The tweets that will be collected must have been posted in the UK and must be related to electrical vehicles. For the storage of the data, in a format that can be used later for analysis, I will be using a Microsoft Excel file (xlsx).

After collecting the data, I will need to prepare the data using pre-processing techniques. Social media data is unstructured and needs to be cleaned before using it to train a Sentiment Analysis model. In other words, the dataset needs to be cleaned from irrelevant information like special characters, extra spaces and emojis.

Then, I will need to define the sentiment of each tweet in the created dataset in order to proceed to the analysis. This will be done by using a Natural Language Processing analysis technique, a sentiment analysis. The sentiment analysis will identify people's opinions and classify them into positive, neutral or negative. From this, I will determine what percentage of people support or have concerns about the transition to electric vehicles.

Having the sentiment of the tweets, I will then use another Natural Language Processing technique, the topic modelling, to determine the aspects of the transition that people are concerned about. Topic modelling is a type of statistical model which can be used to extract abstract topic for documents. I will identify the topics that make people have a positive or a negative opinion about the transition.

The final part of the project is the analysis where, based on the findings from the sentiment analysis and the topic modelling, I will gain insights into what people feel about the transition to electric vehicles and what their concerns are.

Ethical Considerations

As I will be gathering data and information from an online platform, ethical approval needs to be sought. I have taken the Research Integrity Training course offered by our University which is a requirement that need to be met in order to apply and get ethical approval. Once I submit the ethical review form, I will not be permitted to gather data from the online platform until my application gets approved by the Ethics Committee of the University. Only the text content, the location and the date of the tweets will need to be stored for my project.

Project Aims and Objectives

The main aims of the project are:

1. Retrieving the data related to the transition to Electrical Vehicles from Twitter
2. Creating the dataset of the tweets
3. Perform sentiment analysis and topic modelling of tweets
4. Identifying the aspects that people are concerned about
5. Visualising and analysing the findings to gain insights

There are key objectives that must be met in order to achieve the aforementioned aims and ensure that the project is successful. Careful research must be done in order to find the best possible tools that will help in achieving the main aims of the project.

Key Objectives:

- Determine a set of hashtags or keywords which will be used for the retrieval of the relevant tweets.
  - Selecting terms and words that will create relevant Twitter streams will need both research and personal knowledge.

- A dataset of tweets gathered using the Twitter API that falls into the keywords and terms given.
  - A Twitter developer account and a Python library to stream the tweets are required for the retrieval of the tweets. There are several libraries in python for this functionality, so research will be required.

- The dataset of the tweets must be cleaned before applying any analysis models
  - Will require Natural Language Processing tools to achieve this. There are several Python libraries for NLP, so research will be required to find the most suitable one for the project.

- A sentiment analysis model will produce the sentiment scores for a negative, neutral and positive emotion for each tweet.
  - Python has many different libraries and tools which work with different sentiment analysis models; therefore, research will be required to choose the one which is most suitable for my project.

- Topic modelling algorithm will be used to identify people's concerns
  - Research will be required to find the most suitable libraries in Python for Natural Language Processing and the most suitable algorithm to use for topic modelling in order to identify the concerns that people have about the transition to electric vehicles.

- Visualising the sentiment and people concerns

- Research will be required to find the best suitable tools in Python or other software to visualise the findings.

- Analysing the findings
  - Analyse the findings using the visuals in order to gain insights.

Work Plan

In the Work Plan, I have not included the weeks of Easter Break even though I will be still working on the project. I will try to follow the work plan as closely as possible.

- Task to be completed

🎖 Milestone

Week 1 – 31/1/2022

- Research:
  - How Twitter API works and what information can be retrieved
  - What pre-processing techniques of text are available
  - What machine learning models are available
  - What data visualisation tools are available
- Create a Twitter developer account to get the API keys.
- Complete Ethics Review form
- Work on the Initial Plan

Week 2 – 7/2/2022

🎖 Submit Initial Plan
- Continue research about how to retrieve the desirable data from Twitter and find the best keywords.
- Continue research about NLP tools and conclude to which is best to use.
- Continue research about sentiment analysis models and conclude to which is most suitable to use.
- Continue research about topic modelling techniques and conclude to which is most suitable to use.
- Continue research about tools for visualisation and analysis to gain insights.

Week 3 – 14/2/2022

- Having a clear idea on how to retrieve the desirable tweets using the Twitter API.
- Start writing a Python program to stream tweets using the Twitter API.

🎖 The Python program for the streaming of the tweets should be finished.

Week 4 – 21/2/2022

🎖 Ethical application should have been approved at this stage.

- Start saving tweets with the correct criteria being met. E.g., uses the hashtag #ElectricVehicles and was tweeted in the United Kingdom.
- Start writing a Python program to clean the data that are being saved.

Week 5 – 28/2/2022

- Python program that cleans the dataset should be finished.
- Continue streaming and saving tweets.
- Begin writing a Sentiment Analysis model in Python to classify the sentiment of each tweet.

Week 6 – 7/3/2022

- Sentiment Analysis model in python should be finished.
- Run the dataset of tweets collected through the sentiment analysis model and create another dataset including the content of the tweets, the locations the tweets were posted to and the predicted sentiment.
- Continue streaming and saving tweets.

Week 7 – 14/3/2022

- Using the dataset with the sentiments, visualisation and analysis should be done to analyse the sentiment of the tweets and derive the percentages of people that support the transition to electric vehicles (positive sentiment) and people that have concerns (negative sentiment).
- The analysis of the sentiment of the tweets should be finished.

Week 8 – 21/3/2022

- Begin writing a Topic Modelling program in python.
- Finish the Topic Modelling program in python.
- Using the dataset containing the sentiments of the tweets apply the Topic Modelling technique.

Week 9 – 28/3/2022

- Visualisation and analysis of the findings to derive insights about the reasons why people support the transition and what are the people's concerns.
- Begin writing the final report.

Week 10 – 25/4/2022

- Continue writing up the final report.

Week 11 – 2/5/2022

- Continue writing up the final report.
- Complete a draft of the final report and send it to the supervisor.

- Make amendments on the final report based on the supervisor's feedback.

Week 12 – 9/5/2022

- Finish final report

- Submit Final report

References

[1] **https://researchbriefings.files.parliament.uk/documents/CBP-7480/CBP-7480.pdf**

[2] **https://en.wikipedia.org/wiki/Twitter#Tweets**