Ann Lian, Yuhan Ma, Junling Shao

Project Proposal: Sentiment Analysis of Electric Vehicles

Backgrounds: Growing Popularity of EVs and Relevated Policy

Concerns related to climate change and air quality are the primary drivers behind the shift toward cleaner and more sustainable transportation options. EVs produce zero tailpipe emissions, making them a more environmentally friendly choice compared to traditional vehicles. Additional reasons contributing to the growing popularity of EVs include lower operation costs, improved performance and advancements in battery technology.

The federal government has historically offered various incentives to encourage EV adoption. Some of the primary federal policies include federal tax credits for producers and buyers, as well as funding for charging infrastructure. In addition to the federal government, state governments also play a significant role in shaping EV policy, and there is substantial variation across the United States. Key state-level policies and incentives include state tax credits, HOV lane access, and zero-emission vehicle mandates.

Our Research Goals

This study aims to analyze and interpret the sentiment surrounding EVs on social media platforms. Specifically, the objectives are to gain insights into the dynamics of public opinion related to EVs and EV policies, identify the factors influencing both positive and negative sentiment, and assess how these sentiments might impact policies and regulations. The project will address the following key questions:

1. What are the predominant sentiments expressed on social media platforms regarding Electric Vehicles?
2. What are the primary factors driving positive and negative sentiment towards EVs?
3. What is the public sentiment and perception regarding policies and regulations related to Electric Vehicles?
4. What valuable insights can be derived from sentiment analysis to inform the development of future EV policies and strategies?

Data Collection and Preprocessing

YouTube is one of the most popular platforms today, hosting millions of videos and serving billions of users. Most conversations on YouTube, particularly comments on videos, are publicly available and easily accessible through YouTube's Application Programming Interface (API). Therefore, YouTube was chosen as the data source for this study.

The Python programming language, particularly its libraries for data manipulation and analysis (Pandas, NumPy), web scraping, and API interaction (Requests, BeautifulSoup, Selenium), will be used for data collection and preprocessing. We will make use of the YouTube API to extract video metadata and comments related to specific keywords, such as "EVs" and "EV tax credit."

Text Mining Approaches – Sentiment Analysis

The primary objective of our sentiment analysis is to identify and extract sentiment from the publicly available YouTube comments related to electric vehicles (EVs). By applying Sentiment Analysis (SA), we aim to analyze subjective information, including opinions and emotions expressed in the text data.

There are two primary methods for conducting SA: lexicon-based and corpus-based. Given the extensive and diverse nature of our YouTube comment dataset, which includes varying languages and sentiment expressions, the corpus-based sentiment analysis method will be the most suitable choice.

There are numerous available Lexicon, and we want to use the NRC Emotion Lexicon (EmoLex) due to its unique ability to classify emotions into eight fundamental categories: anger, fear, sadness, anticipation, disgust, joy, surprise, and trust. Additionally, EmoLex allows us to identify positive and negative sentiments in the data, making it an ideal choice for our study.

To perform sentiment analysis, we will use Python's Natural Language Toolkit (NLTK), which integrates with the NRC EmoLex. Prior to sentiment analysis, we will preprocess the text data using NLTK, which involves tokenization (segmenting the text into words or phrases), converting text to lowercase, and removing punctuation and stopwords. These preprocessing steps are crucial for preparing the text data for sentiment analysis. By applying NLTK and EmoLex, we will extract and analyze the sentiments expressed in YouTube comments related to electric vehicles, providing valuable insights into public opinion on this topic.

Definition of "Success Project"

Through the mining of social media data from the YouTube comments section, we aim to discern the relationship between the release of Electric Vehicle (EV) policies and the public sentiment towards the EV. A project can be deemed successful if it accomplishes the following objectives:

1. Selection of Highly Representative Keywords: The chosen keywords, which describe the public opinion in the EV industry, should precisely encapsulate the public's attitude. After the release of different policies, these keywords should exhibit shifts in public opinion. Additionally, the selection of keywords and their synonyms should be as comprehensive as possible to ensure the completeness of the data set.
2. Accuracy in the Selection of Data Scraping Objects: The selection of YouTube videos for analysis should correspond temporally with different policies and be highly relevant to the central topic. The content of the videos should be as unbiased and neutral as possible, ensuring that the data used for analysis is accurate and unbiased. The video publishers should be authorized and provide truthful information.
3. Accurate Identification of Significant Sentiments: After analyzing various EV policies, it is essential first to have an accurate anticipation of their impact on public sentiment. Subsequently, this anticipation can be validated through data scraping.
4. Meaningful and Constructive Analysis Results from Sentiment Analysis: Based on the analysis results, we hope to propose meaningful policy suggestions.

The ideal outcome is observing noticeable sentiment changes after the release of different policies, with certain temporal characteristics such as a sudden shift from negative to positive, a gradual deepening of positive attitudes, or fluctuations between positivity and negativity. Through these changes, we can identify corresponding variations in the policies, thereby offering policy recommendations or making predictions about developmental trends.

Our analysis was the anticipation of a more favorable public sentiment towards the Electric Vehicle (EV) industry, given the global shift towards sustainability and green technologies. However, the study might unveil a surprising prevalence of negative sentiments. This unexpected trend of skepticism could potentially be attributed to some accidents, like EV malfunctions, or other possible factors.