**Final Project**

**Data Design & Representation**

***Sentiment Analysis of TESLA Brand from***

***Reddit Community***

**Group Member:**

**Hema Sudarsan Rao**

**Florentina Vania Santosa**

**Jiateng Chen**

1. **Title:**

Sentiment Analysis for Tesla Brand from Reddit Community

1. **Project Summary**

For our final project here, our team decided to research public opinion and sentiment about Tesla as a brand within the Reddit community. We employed web scraping techniques to collect data “teslamotors” subreddit, we try to analyze the insights into the brand’s perception. The goal here is to uncover any patterns, trends, and general sentiment, providing valuable knowledge to stakeholders interested in understanding the brand's online reputation.

1. **Background & Business Questions**

We chose Reddit as our course to scrape public opinion for several reasons:

* It is a diverse and engaged community. With the amount of users it has, this diversity would allow brands to find a wide range of opinions, discussions, and sentiments related to their product or services.
* Source of candid feedback. We believe that even though some users might just write some short comments, some provide in-depth conversation and offer comprehensive public sentiment.

In an era where people rely on the opinion of others to make decisions, online forums like Reddit might have significant influence on the brand image and therefore understanding what the public says toward their products is crucial to almost every company. Our project seeks to answer business questions:

* How does the Reddit community perceive Tesla as a brand?
* What are the most common topics of discussion related to Tesla?

1. **Data Sources & Description of the Web-Scraping Procedures**

Our source for the web scraping is the Reddit website using “teslamotors” as the subreddit.

Web-scraping Procedure:

1. Setting up PRAW:

We started by setting up PRAW, a library that we use in python to help provide a simple interface accessing Reddit’s API.

1. Setting up Reddit Account to obtain API credentials
2. Accessing Reddit Data:

When we have our API, then we can start accessing Reddit specifically targeting the “teslamotors” subreddit, focusing on the top 719 hot posts.

1. Extracting Unique Titles:

We try to avoid duplicates and focus on the unique titles, we iterate through these posts and come up with unique titles into a set.

1. Sentiment Analysis:

We examined using SentimentIntensityAnalyzer to score each title we scrape whether it is negative, neutral, or positive sentiment. Libraries such as NLTK and TextBlob are used to categorize the posts.

1. Common Topic Analysis:

We use LDA(Latent Dirichlet Allocation) to extract common topics. LDA is a widely used method to cluster and analyze text information. Compared with the world cloud and other methods, its advantage is to automatically discover topic patterns without human labeling.

1. **Discussion on Insights**

Several insights uncovered during our analysis:

* Neural Sentiment:

Based on our analysis, we found more general sentiment toward Tesla. The "teslamotors" subreddit has a high neutral score (0.878), indicating that most comments there are factual, educational, or neutral in tone. This may suggest that rather than an excessive amount of good or negative feedback, the community shares a great deal of information, news, and unbiased debates regarding Tesla and its products.

* More Positive than Negative:

The positive score (0.076) is higher than the negative score (0.047), although both scores are relatively low. This indicates that, of the sentiment-bearing posts, there are more posts with a positive sentiment towards Tesla than negative. This could suggest that the community's overall perception of Tesla is cautiously optimistic or generally favorable, albeit with some reservations or criticisms.

* Low Compound Score:

By summing up the positive, negative, and neutral ratings, the compound score (0.051) indicates the public overall sentiment. The fact that the compound score is low closer to 0 confirms the finding that the subreddit's conversation on Tesla is mainly general, but slightly optimistic.

* Common topics:
* Topic 1: Keywords: cybertruck, model, as, now, it, car, back, are, from, that This topic seems to be about the Cybertruck and Tesla's model. It might discuss the current status or availability of the Cybertruck.
* Topic 2: Keywords: model, cybertruck, no, performance, at, highland, video, this, now, us. This topic appears to focus on the Tesla models’ performance, features, and events like highland testing or video releases.
* Topic 3: Keywords: fsd, cybertruck, model, update, as, beta, out, has, after, berlin This topic seems to talk around updates, especially related to the Full Self-Driving (FSD) feature, as well as developments concerning the Cybertruck, possibly mentioning beta releases and events in Berlin.
* Topic 4: Keywords: cybertruck, model, at, berlin, it, giga, spotted, over, range, that This topic looks similar to topic 3.
* Topic 5: Keywords: cybertruck, model, at, elon, musk, nacs, range, will, from, fsd This topic could relate to the Cybertruck and Tesla's model, potentially discussing announcements or statements made by Elon Musk, as well as mentions of events like NACS (North American Cybertruck Show) and updates on range and FSD.

