Seneca POLYTECHNIC

PROJECT PROPOSAL

BAN200: Text Mining & Sentiment Analysis
Instructor: Professor William Dick
Group Members:
Arpit Trivedi | Ryan Lobo | Samarth Chugh | Kareena Dhamija
Dhairya Gada | Danish Memon

Project Title:

Unfiltered Reactions: Sentiment and Topic Mining of YouTube Comments on Tesla Cybertruck Launch

Dataset Description and Source:

MKBHD Cybertruck Video

The total numbers of the received user generated videos on the given video which is posted by the most-followed tech reviewer internationally, Marques Brownlee (MKBHD), exceed 5000. The video under consideration can be connected to the review of Tesla Cybertruck one of the most controversial and radical products. The attributes are the data in the following:

- **Comment Text:** All the user produced content.
- **Published Date:** The day the remark has been published.
- **Username:** common name of writer of legation.
- **Likes:** Count of the likes of the remark.
- **Last Updated**: Mark of the end of final interaction/editing.

This is pure, unrefined, uncodified mass taste and this is where one has best place to work on consumer reaction, market sentiments and what gets the heart to feel so that the design and innovation can build up on reactions.

Business Problem / Research Question:

Tesla thrives largely due to its brand image, particularly in the case of such a radical thing as the Cybertruck that violated the established principles of automotive design. Although some conventional forms of customer feedback (e.g., surveys, reviews) are designed, the comments on YouTube are unarranged and very emotionally content voluminous.

Research questions are explored in this project:

- What does the YouTube commenters think of the Cybertruck?
- Which are the most repeated themes or topics talked about by the viewers?

- Is there any other sentiment or pattern of opinion on extensively liked comments?
- Are there any suggestions that can be made on the marketing, product communication, or improvement in the design of Tesla?

Planned Methods and Tools:

1. Text Preprocessing:
   - Lower case-uppercase Write small
   - Remove punctuations, stop words and even emojis

2. Sentiment Analysis:
   - Polarity score: Vader (adjusted to social media) TextBlob
   - Make the remark something. Slow, fast or Mute
   - The overall favored things distribution in the sentiment world in the case with the element of weight and without it
   - Pattern or vocal identification of tones to roar and determine irony or sorrowless

3. Experimental analysis of Keyword Extraction or Topic Modeling:
   - LDA (Latent Dirichlet Allocation) and BERTopic are linked with the clustering of themes
   - Make word clouds handically manageable
   - Determine 5-10 topics of discussion (e.g. design, price, durability, innovation)

4. Competent scope- Date/Time Analysis and Engagement Analysis:
   - To get information about the best time to undertake the commenting activity, it is important to compare the spacing of the comments with time and tabulate when there is a trend of activity.
   - The attitude of the early and late commenter
   - Examine the relationship between like of comments and the level of the feelings

5. Reporting and visualisation:
    - Interactive plot; either Plotly, Seaborn or Tableau
    - There are sentiments distribution dashboard, topical topics and level of user interactions to recent topics
    - Word cloud, bar chart or heat map of the time driven comment analysis

6. Tools:
    - Python Jupyter Notebooks
    - Libraries: Pandas, NLTK spaCy, TextBlob, Vader, Gensim, matplotlib, seaborn and wordcloud

Expected Deliverables:

- Cleaned dataset/Analysis dataset
- The outcomes of the sentiment classification as well as visual distribution
- Labeled cluster output of topic modelling
- Trend analysis and business insights/recommendation to Tesla
- An orderly and succinct presentation/report detailing methods, results and implication

Business Value & Uniqueness:

This is because the project will take the advantage of the voice the voice that people have on the web and the project will provide Tesla (or other firms of its caliber) an understanding of what people do towards them and how they see them and what people say on how people see and treat them about doing things. Nothing is censored and hence, there is a special access to uncensored words of the tech-savvy viewers and crowds. It is possible to use the findings to educate:

- Response plans in social media
- Marketing tone and content the tone and the content of marketing
- Emphasis to be placed on features of the products in the future upgrades
- Waiting to be hailed critically or to appear to become viral