Graduate Certificate in Artificial Intelligence with Machine Learning
AIGC 5002 - Machine Learning and Deep Learning
Fall 2023

Project Proposal

Project Title: YouTube Comments Sentiment Analysis for Content Creators
Team Members: Bhavesh Waghela, Siri Srujana Karnala

[Section 1: Project Overview]

- YouTube Comments Sentiment Analysis for Content Creators
- **Brief Description**:

  This project aims to provide a tool for YouTube content creators to analyze the sentiment of comments on their videos. Content creators often get overwhelmed with comments, and manually searching through them can be time-consuming. This project intends to help content creators gain a better understanding of the comments they receive, allowing them to reflect upon their content and make data-driven decisions to improve the quality of their videos. Creators can quickly identify and prioritize positive, negative, or neutral comments by using sentiment analysis. This tool will assist creators in better understanding audience feedback and interacting with their audience in a more targeted and efficient manner.

[Section 2: Problem Statement]

- **Problem Background:**

  YouTube has evolved into a powerhouse platform for content creators. With millions of videos uploaded daily, the comment section has become a dynamic space for viewer interaction and feedback. Managing the sheer volume of comment effectively has become a challenge for content creators. Understanding the sentiment behind these comments provides creators with a perspective on how their content is received. By automating sentiment analysis creators can streamline this process allowing them to focus more on content creation and strategic engagement.

- **Problem Definition**:

  The problem at hand is the manual and time-consuming process of analyzing comments on YouTube videos for meaningful insights. Content creators receive a diverse range of comments, which include both positive and negative feedback. This mix makes it challenging to identify trends or common issues that need addressing. The emotional toll of sorting through negative comments can also be significant.

- **Objectives**:

  - Develop a sentiment analysis tool to automatically categorize comments as positive, negative, or neutral.
  - Provide content creators with an efficient means of processing feedback without being overwhelmed.
  - Enable data-driven decision-making by extracting valuable insights from audience comments.
  - Improve the content creation experience by facilitating more focused interaction with viewers.
  - Implement measures to protect the privacy and security of both the content creators and the commenters.

[Section 3: Dataset Description]

If you still don't have a dataset, you can describe the dataset you want to find and use for the project.

- **Dataset Overview**:
  The dataset will be collected from the comment sections of YouTube videos, specifically focusing on content related to a particular YouTube channel of the content creator.

  The dataset will contain the following columns:

  - Comment_ID: Unique identifier for each comment.
  - Author_Name: Name of the commenter.
  - Comment_Text: The actual text of the comment.
  - Like_Count: Number of likes on the comment.
  - Reply_Count: Number of replies to the comment.
  - Video_ID: Unique identifier for the video.
  - Video_Title: Title of the associated video.
  - Published_Date: Date when the comment was posted.
  - Sentiment_Label: Sentiment label assigned through sentiment analysis (Positive, Negative, Neutral).

- Data Collection:
  The data collection process will be done from YouTube API using Google Developers Console. Comments will be extracted from each video ID and it will be stored into a .csv file, we would use those files in python to do some data processing before applying sentiment analysis on those comments.

[Section 4: Proposed Methodology]
Can be from models that we'll learn in coming lectures (Can also be changed later after with permission)

- Algorithm/Model Selection:

  Machine Learning Algorithm like VEDAR will be used to identify the sentiments for the comments attached to the video.
  - VDER is specifically designed for sentiment analysis and performs well, especially with social media text.
  - It provides a polarity score (compound score) that indicates the overall sentiment (positive, negative, or neutral).

[Section 5: Expected Outcomes]

- Desired Results:

The expected results for the project will be Automated Sentiment Analysis Dashboard categorizing comments into positive, negative, or neutral sentiments. This will significantly reduce the manual effort required for content creators to analyze viewer feedback. It would benefit the content creators by gaining valuable insights into audience sentiment, allowing them to understand how their content is perceived.
They will be able to identify trends, popular topics, and areas for improvement.

[Section 6: Timeline and Milestones]

- Milestones:

Milestone 1: Data Collection and Preprocessing (Deadline: November 10, 2023)
Task 1: Gather a diverse dataset of YouTube comments for a particular channel using YouTube API.
Task 2: Preprocess the data to clean and format comments for analysis.
Task 3: Explore and understand the dataset's characteristics and distributions.

Milestone 2: Model Selection and Development (Deadline: November 20, 2023)
Task 1: Evaluate and select the most suitable for sentiment analysis model for YouTube comments (VADER, TextBlob).
Task 2: Implement the chosen model and integrate it into the analysis process.
Task 3: Fine tune and validate the model using a subset of the data.

Milestone 3: Initial Sentiment Analysis (Deadline: November 30, 2023)
Task 1: Conduct initial sentiment analysis on a portion of the dataset to assess model performance.
Task 2: Evaluate the accuracy, precision, and recall of the sentiment predictions.
Task 3: Identify any areas for improvement or fine-tuning of the model.

Milestone 4: Full-Scale Sentiment Analysis (Deadline: December 10, 2023)
Task 1: Apply the sentiment analysis model to the entire dataset of YouTube comments that has been captured with the API.
Task 2: Generate sentiment labels (positive, negative, neutral) for each comment.
Task 3: Validate the results and ensure they align with expected outcomes.

Milestone 5: Results Interpretation and Reporting (Deadline: December 15, 2023)

Task 1: Analyze the sentiment distribution and trends across the dataset.
Task 2: Summarize key findings and insights gained from the sentiment analysis.
Task 3: Prepare a detailed project report highlighting the methodology, results, and potential applications of the sentiment analysis.

[Section 7: Ethical Considerations]

- Ethical Concerns:

There are a few ethical considerations that can be taken for the dataset that is been used for sentiment analysis.

- Privacy and Consent: The data that is processed for sentiment can be ensured that it is compliance with privacy regulations.
- Bias and Fairness: Identify the potential biases in the data, as well as in the sentiment analysis model. Ensuring that the model does not disproportionately favor or discriminate against certain groups or perspectives.
- Transparency and Explainable: We would ensure that we provide transparency in how the sentiment analysis is performed. Including documentation of the models used, feature selection, and any data preprocessing steps. Also ensuring that the model's decisions can be explained to stakeholders.
- Data Security: Safeguard the collected data from unauthorized access or breaches. Implement access controls, and other security measures to protect sensitive information.

- Citations:

Project Idea: https://www.stratascratch.com/blog/19-data-science-project-ideas-for-beginners/

Project 7: YouTube APIs with Python
https://www.youtube.com/watch?v=fklHBWow8vE&ab_channel=StrataScratch

https://github.com/Strata-Scratch/api-youtube/blob/main/README.md

Dataset Description:
https://github.com/hellotinah/youtube_sentiment_analysis/blob/main/tinas_comments.csv

Proposed Methodology:
https://www.youtube.com/watch?v=kHOVWiZKpHM&ab_channel=TinaHuang

Milestone: https://chat.openai.com/

Ethics: https://ts2.space/en/the-ethics-of-ai-and-sentiment-analysis-navigating-privacy-and-bias-concerns/