#### **Youtube Comment Sentiment Analysis**

#### **Objective**

The primary goal of this project is to conduct sentiment analysis on YouTube comments for the video "OneRepublic - I Ain't Worried (From Top Gun: Maverick)". We aim to extract meaningful insights from the comments by classifying them into positive, negative, and neutral sentiments using various machine learning models.

#### **Dataset**

The dataset consists of comments fetched from the YouTube video using the YouTube Data API. After preprocessing, the dataset contains a balanced number of comments for each sentiment class.

#### **Data Preprocessing**

Several preprocessing steps were applied to the dataset using the SpaCy framework and NLTK library:

1. Normalization: Convert all text to lowercase.
2. Removal of punctuation and numerical characters.
3. Elimination of stop words.
4. Lemmatization: Reduce words to their base form.
5. Handling class imbalance: Upsampling the minority classes to ensure a balanced dataset.

#### **Feature Engineering**

* Text Vectorization: Text data was transformed into numerical format using Count Vectorization.
* Sentiment Analysis: The SentimentIntensityAnalyzer from the NLTK library was used to assign initial sentiment scores (positive, negative, neutral, compound) to each comment.

#### **Model Selection and Training**

Two machine learning models were trained using the Count Vectorization technique:

1. Naive Bayes Classifier:
   * Achieved an accuracy of 63.52%.
2. SVM Classifier:
   * Achieved an accuracy of 94.92%.

#### **Conclusion**

SVM outperformed Naive Bayes, achieving an accuracy of 94.92%, making it suitable for this sentiment analysis task..

### **Future Work**

Future work could involve exploring more advanced models such as:

* Deep Learning Models: Utilizing LSTM, or Transformer-based models for improved accuracy.
* Word Embeddings: Using pre-trained word embeddings like Word2Vec, GloVe, or BERT to capture semantic relationships.
* Hyperparameter Tuning: Further tuning the hyperparameters of the SVM model and other potential models to enhance performance.
* Web Application: Developing a web application where users can input a YouTube link and receive sentiment analysis for each comment, categorizing them into positive, negative, and neutral sentiments.

This project highlights the potential of machine learning in understanding and categorizing viewer sentiments, providing valuable insights for content creators and marketers.