Entity-level Sentiment Analysis on YouTube Comments

Overview

This project focuses on performing entity-level sentiment analysis on YouTube comments using natural language processing (NLP) techniques. The objective is to gain insights into the sentiment distribution across different entities mentioned in the comments, such as video content or creators, to better understand audience reception and engagement on YouTube.

Project Structure

The project structure is organized into several key components:

Data Preprocessing: Raw YouTube comments data is cleaned and prepared for analysis. This involves tasks such as text normalization, tokenization, and noise removal from the comments.

Model Development: Two approaches are explored for sentiment analysis: one using TfidfVectorizer and another using LSTM with TensorFlow. Models are trained on preprocessed data to classify comments' sentiment towards specific entities.

Evaluation: The performance of the trained models is evaluated using metrics such as accuracy, precision, recall, and F1-score. Confusion matrices are also generated to visualize the model's performance.

Visualization and Analysis: six visualizations are explored to get meaningful insights into the analysis.

Documentation & Reporting: Findings, insights, and recommendations are summarized in a report, and the project is documented in a power point.