



SENTIMENTAL ANALYSIS

On YouTube Comments

ABSTRACT

This project focuses on sentiment analysis of YouTube comments using VADER sentiment intensity analyser and Gaussian Naive Bayes classifier, providing insights into user sentiments and their implications for content creators and brand managers.

Spandan Ghosh

Introduction:

Sentiment analysis, also known as opinion mining, is a technique used to determine the sentiment expressed in a piece of text. It involves analyzing the emotions, attitudes, and opinions conveyed by individuals in their written or spoken language. Sentiment analysis has gained significant importance in various fields, including marketing, customer feedback analysis, social media monitoring, and brand reputation management.

This report aims to provide a detailed description of the process of sentiment analysis using a dataset of YouTube comments. The report will cover the data description, methodology, and conclude with the findings and implications of the analysis.

Data Description:

The dataset used for sentiment analysis consists of YouTube comments related to a specific video. The comments were scraped using the YouTube Comment Scraper Python library. The dataset contains 180 rows and one column, which includes the comments made by users.

The data preprocessing phase involved removing unnecessary columns, such as UserLink, Likes, Time, and user. The remaining column, "Comment," contains the text of the comments. The data was then labeled based on sentiment using the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analysis tool. The sentiment labels include positive, negative, and neutral.

Methodology:

1. Data Transformation:

- Libraries required: numpy, pandas, matplotlib.pyplot, os, sklearn.preprocessing, sklearn.utils, sklearn.feature_extraction.text, nltk.sentiment.vader, nltk.tokenize, nltk.stem, nltk.corpus, string, re.
- The data was read from a CSV file and unnecessary columns were dropped.
- Sentiment analysis was performed using the VADER sentiment intensity analyzer.
- The sentiment scores were calculated for each comment, and sentiment labels were assigned based on the scores.

2. Data Preprocessing:

- Libraries required: nltk.corpus, nltk.stem, nltk.tokenize, string, re.
- Text processing techniques were applied to clean the comments.
- The text was converted to lowercase, punctuation and special characters were removed, and stopwords were eliminated.
- Stemming and lemmatization were performed to reduce words to their base form.

3. Balancing Data:

- The dataset was balanced by up sampling the minority classes (negative and neutral) using the resampling technique.
- The unsampled data was concatenated with the positive sentiment data.

4. Feature Extraction:

- The CountVectorizer from `sklearn.feature_extraction.text` was used to convert the text data into numerical features.
- The maximum number of features was set to 1500.

5. Machine Learning Model:

- The Gaussian Naive Bayes classifier from `sklearn.naive_bayes` was used for sentiment classification.
- The dataset was split into training and testing sets.
- The classifier was trained on the training set and evaluated on the testing set.

Conclusion:

In this report, we performed sentiment analysis on a dataset of YouTube comments using the VADER sentiment intensity analyser and the Gaussian Naive Bayes classifier. The analysis involved data transformation, data preprocessing, balancing the data, feature extraction, and model training.

The results of the sentiment analysis showed that the classifier achieved an accuracy of 91.77% in predicting the sentiment of YouTube comments. The confusion matrix revealed that the classifier performed well in classifying positive and neutral sentiments but had some difficulty in classifying negative sentiments.

The findings of this analysis can be valuable for understanding the sentiment of YouTube users towards a particular video. This information can be used by content creators, marketers, and brand managers to gain insights into user opinions and make informed decisions.

In conclusion, sentiment analysis is a powerful technique that can provide valuable insights into user sentiments. By analysing the sentiment expressed in text data, organizations can better understand customer opinions, improve their products or services, and enhance customer satisfaction.