# Qualitative Sentiment Analysis of YouTube Contents based on User Reviews

Sentiment analysis is a branch of Natural Language Processing that involves computational techniques to identify and extract subjective information from text data. YouTube, a content platform with thousands of uploads, has a problem with effective sentiment analysis to get user opinions and customize advertising and content. This study focuses on sentiment analysis in the digital content market, specifically in YouTube reviews. Employing the solution with 'Encoder-Decoder with Attention', a specialized deep learning approach, the authors assessed nearly 700,000 comments from 8,000 YouTube channels, categorizing them as positive, negative, or neutral for training the model.

The model is trained on a dataset of user comments from YouTube videos, which are labeled as positive, negative, or neutral which is then used to predict the sentiment of new comments. The Encoder-decoder-based Attention model consists of two main components: an Encoder and a Decoder. The encoder takes the input sequence of words and converts it into a fixed-length vector representation. The decoder then takes this vector representation and generates the output sequence of words. The attention mechanism allows the model to focus on the most relevant parts of the input sequence when generating the output sequence improving the accuracy of the model.

The proposed model also uses a squeeze-and-excitation attention layer. This layer is used to improve the performance of the model by allowing it to focus on the most important features of the input sequence. It is designed to learn the importance of each feature in the input sequence and then use this information to weight the features accordingly. The results show that the proposed model outperforms other state-of-the-art models in terms of accuracy, precision, recall, and F1 score.

The result can be used to assess user opinions and customize advertisements and future content, accordingly, monitor the reputation of a YouTube channel, and identify potential issues or problems with a YouTube video, etc. The study also suggests that the proposed model can be extended to other social media platforms such as Facebook, Twitter, and Instagram.

Research Paper Citations:

https://ieeexplore.ieee.org/document/10141517

# Sentiment Analysis of YouTube Comments Using Naïve Bayes Support Vector Machine (NBSVM)

The study of sentiment analysis on YouTube video comments, with an emphasis on the extraction and processing of textual data to assess the sentiment included in the comments. The primary goal is to categorize comments as positive or negative to assist YouTube users in determining the significance of video material based on user opinions. The comments made by users have a significant impact on the reputation of both the material and the channel. Understanding the sentiment expressed in these comments is critical for content makers to successfully personalize their videos.

The author applies a Naive Bayes - Support Vector Machine (NBSVM) technique with binary classification to attain this purpose. It says combination of these two classification methods was chosen for its high performance in dealing with varying amounts of text data. Support Vector Machine does best with larger datasets or full-length documents, but Naive Bayes excels with tiny data samples and document snippets.

The authors have tested the model on YouTube comments within the education category, specifically on the "Kok Bisa?" YouTube channel, known for its high number of subscribers and views. Data collection involved web scraping and employed Natural Language Processing to ensure data accuracy. The methodology includes text preprocessing stages such as case folding, tokenization, filtering, and stemming. Once the data is ready, Multinomial Naive Bayes calculates the probability of words appearing in positive and negative comments. This information, along with log-count ratios, is used to generate feature vectors for further processing by Support Vector Machine, which determines the optimal hyperplane for classifying data.

In conclusion, their study shows that use of text preprocessing techniques and combining Naïve Bayes and SVM gave optimal results with more accuracy of the classification. The performance of the classification model was evaluated on different data scales (60:40, 70:30, and 80:20), and the 70% training data with 30% testing data configuration yielded the best results (91% precision, 83% recall and 87% f1 score).

Research Paper Citations:

https://ieeexplore.ieee.org/document/8920923