Azrul Zulhilmi Ahmad Rosli
P153478

Sentiment Analysis

Sentiment Analysis: A Review of YouTube Comments on Short Chinese Drama

Introduction

YouTube is one of the most popular platforms where people can watch a movie, drama and others videos related especially with short drama that are concise yet emotionally impactful. The comments left by audience is reflecting audience perception which can offer some insight for the public sentiments. This project using the short chinse drama on YouTube with title "Rural girl was forced to marry a poor boy,didn't expect he was CEO and fell in love with her!" which have duration at 1:46:42 and was uploaded by KIKI SHORTDRAMA.

This dataset has been collected using "Comment Exporter for Youtube" extension from Google with version 0.0.4.3. The dataset will exported in csv format. The three main attributes in this dataset are author, content, and link directing to user's Youtube account. This analysis only use content to do a sentiment analysis. The objective of this study is classifying overall sentiment along common emotional expressions alongside identification of most frequent positive and negative words. Additionally, emojis and informal comments accompanied by stories add charm to the dataset making it ideal for lexicon-based sentiment classification. Figure 1 below show the word cloud for this text data.
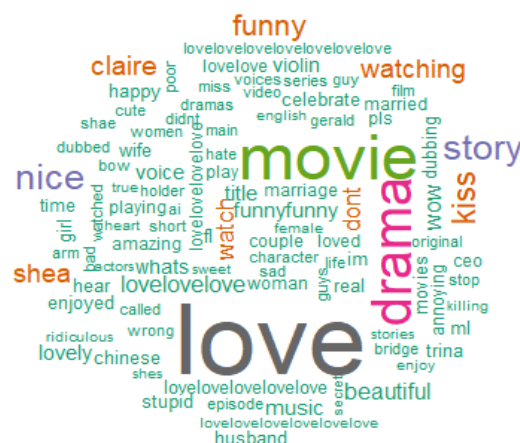

Figure 1   Wordcloud for YouTube review data

Sentiment Analysis

Methodology

Dataset are imported and separated using semicolon. Then, preprocessing was done by replacing commonly used emoji with the corresponding sentiment words while the rest of emoji will be remove using Unicode ranges that covers emoticons, pictographs, transport symbols and flags. For example, change "😂" to "funny". Then, remove numbers and punctuation and convert to lowercase to standardize the input. There might have some missing value after this process so row with missing values in Content column will be filtered out.

Next, tokenized the words and remove the stop word in English to focus on meaningful terms. Sentiment analysis was conducted using four lexicons which are Bings, AFINN, NRC and syuzhet. The Bing lexicon provide binary classification of words as either positive or negative while AFFIN lexicon assigns a numeric score to each words. Meanwhile, NRC lexicon used classify words into eight basic emotion such as joy and others. In additions syuzhet lexicon was used to estimate the overall emotional progression in the reviews. This comprehensive approach allowed for an in-depth interpretation of the public's response to the drama, combining both sentiment polarity and emotional tone.

Analysis and Discussion

First of all, find the top 10 frequent words using in this dataset to understand the general review. Figure 2 shows top 10 most frequent words.
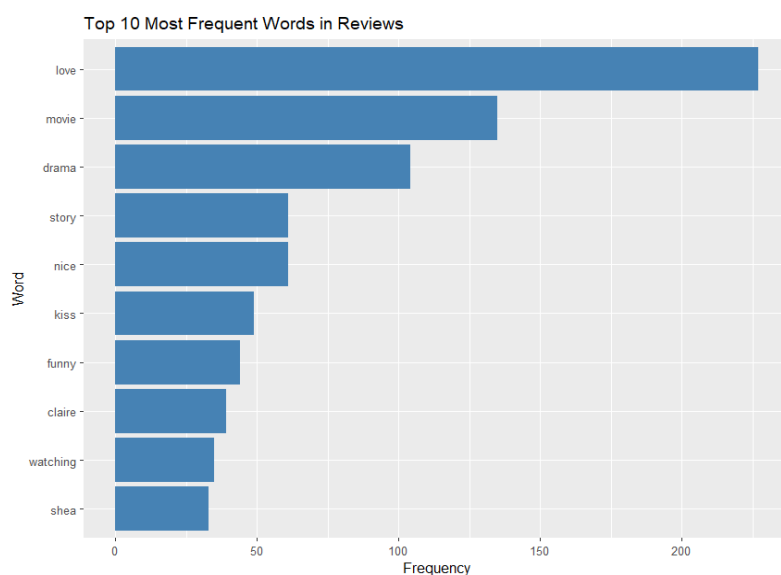

Figure 2   Top 10 most frequent words in reviews.

Azrul Zulhilmi Ahmad Rosli
P153478

Sentiment Analysis

The word 'love' dominates with 227 mentions in the comment shows a strong emotional connection and positive sentiment. The word 'movie' has frequently appeared due to viewers referring to the movies itself and it similarly with words 'drama' and 'story'. The character names like 'Claire' and 'Shear' also rank at top which show how the reviewers love their acting and character. Then, the term 'kiss' may refer to a memorable scene that triggered a strong reactions. Overall, all the top frequent words have positive emotions.

The first lexicon used in this sentiment analysis is Bing lexicon. Bing lexicon is performs categorical sentiment classification where each word is labelled as either "positive" or "negative". When aggregating the total occurrence of each label, positive words appeared is 609 times while negative words appeared is 320. This shows the audience reaction to the YouTube short drama were favourable. Figure 3 below shows top sentiment words in YouTube drama reviews based on negative and positive sentiment by using Bing lexicon.
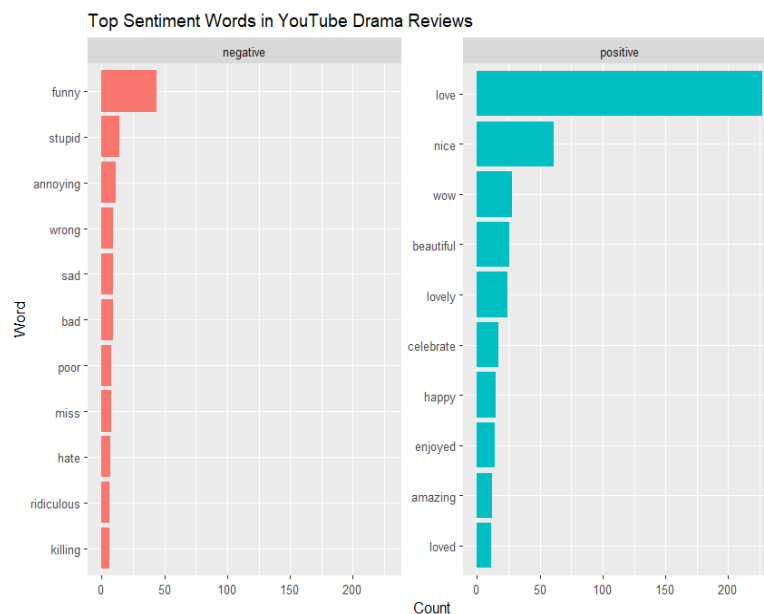

Figure 3   Top sentiment words in Youtube Drama Reviews

Based on Figure 3, the most frequent words on the comment are 'love' which is a positive sentiment with 227 counts. This is show most of the audience admire with the drama might be because of the plot, actors and others. Others positive term such as 'nice', 'beautiful', and 'lovely' is supporting the drama generally favourable by the audience. On the other hand, term like 'stupid', 'annoying' and 'bad' also appear but having a lower frequency around 15 and below. This also shows some of the audiences doesn't think the drama is good. However,

the word 'funny' was classified under negative by the Bing lexicon is reflecting a misalignment with the casual or humorous context common in YouTube comments. Overall, this analysis show the audience or viewer was overwhelmingly positive.

Next, AFFIN sentiment analysis shows the mean score of this review are 1.55 which indicate the reviews carry a moderate positive sentiment because AFFIN score is from -5 to +5 which form strong negative to strong positive. This might be from the effect of term like 'love' which have higher count for positive sentiment. Then, standard deviation here is 2.26 suggesting a notable variation in the sentiment across different reviews where some might have stronger expression of satisfaction or criticism. This variability shows the balanced dataset where most of reviewer express enjoyment when watching the drama. Figure 4 shows the emotion classification in YouTube reviews using NRC lexicon.
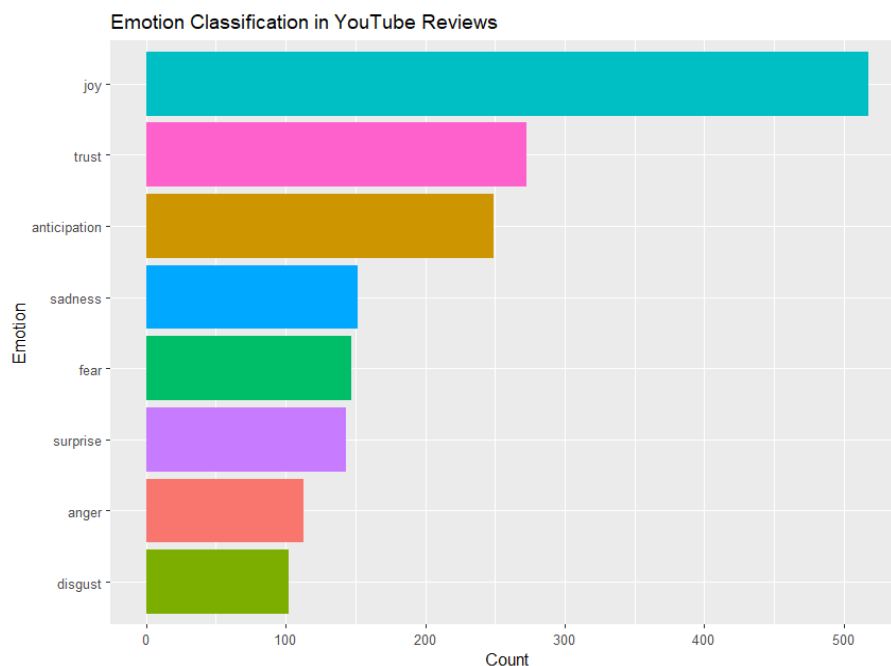

Figure 4   Emotion classification in YouTube reviews

The NRC classification shows 'joy' is the most dominant emotion with 518 mention in the drama review. It followed by 'trust' with 273 times and 'anticipation' with 249 times which shows viewers generally responded to the drama with happiness, emotional, connection and eagerness if drama is good. However negative emotions like 'sadness', 'fear', 'anger' and 'disgust' also appear to show the criticism on the drama. This emotional mix reflects the dramatic and engaging nature of the content where emotional highs are balanced with moments

of tension and conflict typical in short dramas. Figure 5 below shows Syuzhet sentiment analysis.

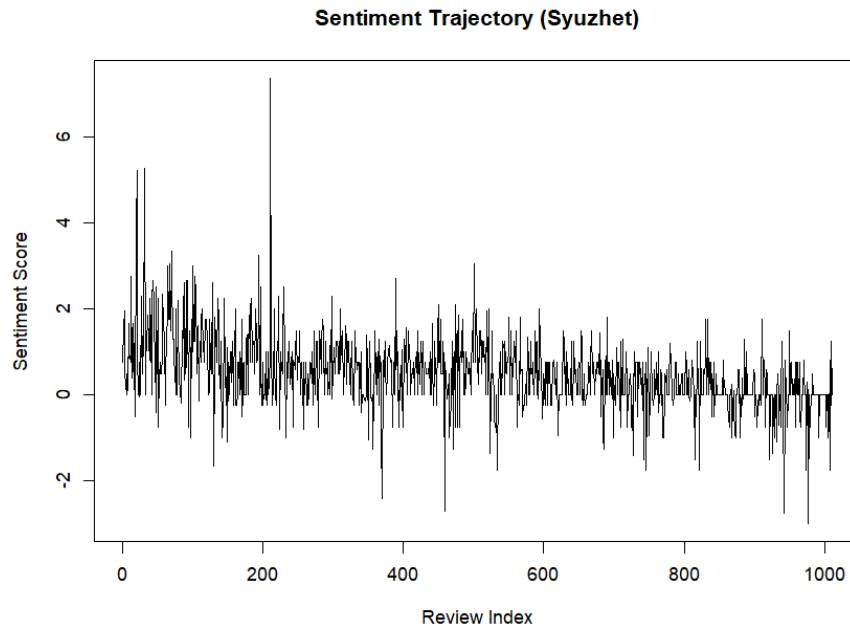
Figure 5   Syuzhet sentiment analysis

This Syuzhet sentiment analysis have a mean score with 049 which indicate a slightly positive overall sentiments in the reviews. The sentiment trajectory shows emotional tone across all comments which for 1010 reviews. The plotting shows clear fluctuation of emotions where some review show very high positive sentiment and others are neutral or mildly negative. The graph above also shows that earlier viewers expressed more positive sentiment compared to those who commented later. This shows majority of reviewer react positively to the drama and some might have mixed opinion about the drama.

Conclusion

In conclusion, sentiment analysis shows that reviewers are response positively to the short Chinese drama where most of them expressed joy, trust and anticipation with 'love' can be emoji love being the most frequent word. AFINN and Syuzhet scores confirmed a generally positive tone, while NRC revealed emotional engagement, proving the drama resonated well with its audience.