### **Project Report**

Title: Sentiment Analysis of Movie trailers based on youtube comments

Team: Raptors

**Introduction**

Analysis of the video sharing platform, YouTube reveals a huge amount of community feedback through various comments posted on these videos. One can get insights such as expectations of the audience or the reviews about a movie through these comments. However, It is very difficult to read all the comments for every video. Therefore, we propose a web based application where we scrap the comments on trending movie trailers on YouTube as a data source. We then perform analysis on this data and try to predict how a movie will be received based on these comments. This will be helpful in giving reviews about new movies and the expectations of the audience before the movie releases.

### **Original Contribution**

We are conducting an in-depth analysis on YouTube comments to shed light on different aspects of comments for videos which are mainly movie trailers and trying to come up with a sentiment review score of these movies. The community feedback is being used to determine how the movie would be received by the audience based on millions of comments posted on these videos. So far, there hasn’t been any such application to predict a movie’s reception.

### **System Architecture**

### 

### **Technologies used**

1. Frontend: Interactive Dynamic Web Page, CSS, JavaScript
2. Dataset: Json files
3. Sentiment Analysis: Python NLTK (Natural Language Processing Tool)

### **Technical Details**

Python Version 3.4 was used to implement and import NLTK toolkit. NLTK is a toolkit which is used for NLP( Natural Language Processing) and perform sentiment analysis.

**nltk.setiment\_Intensity\_analyzer**- Library to implement and import Sentiment Analysis Task using NLTK features based on classifier and objects.

**sid.polarity\_scores**-A code to analyze a piece of text and classify the sentences under positive, negative and neutral polarity of sentiments. The output for the above code can be ‘pos’, denoting positive. The training data here is an array. of sentences with corresponding class types – positive (pos) or negative (neg) to train the classifier.

There's also something called compound, which is a feature allowing us to label our result in a more flexible way. It ranges from -1 (Extremely Negative) to 1 (Extremely Positive). For our purpose we consider statement as neutral when compound is equal to 0.

**emoji\_pattern = re.compile-** A regular expression specifies a set of strings that matches it; the functions in this module let you check if a particular string matches a given regular expression and can filter out unrecognized words//emoticons/symbols & pictograph.

**json.load/json.dump**- JavaScript Object Notation is used to import/export data from python to internal/external files. The data in json file is separated by format like (', ', ': ';’). Character e*ncoding of* **UTF-8** is used to include the support for all the latest characters and symbols used in json data file.

The UI consists of HTML, CSS and JavaScript. The JavaScript makes use of ajax to get the json data. This data is then fed into pie chart representation. A modal is designed through CSS and javascript which pops up when a movie poster is clicked. This modal contains the movie trailer and the pie chart describing the sentiment score of the comments analyzed. Amcharts js library is used to design the pie chart.

### **Related Work**

Opinion mining and work on sentiment analysis and assigning opinion values such as positive negative or neutral on document and tweets has been done before. However, in this application we are analyzing comments and trying to provide a generalized opinion rather with the help of predicting the sentiments of the comments. We are trying to use the comments in a novel context to find out how the movie might be received after release by comments posted on the movie trailers.

### **Team Members and Responsibilities**

|  |  |
| --- | --- |
| Name | Responsibility |
| Dhanashree Joshi | YouTube comments extraction API exploration |
| Tusharika Mishra | Third party comments extraction API |
| Hiten Nirmal | Python NLTK sentiment analysis |
| Gaurav Agarwal | Evaluating sentiment analysis tools; UI design. |

### **Conclusion and Future work**

After observing the output on the webpage of sentiment, the spectators could make sense out of it and judge the movie based on the trailer’s sentiment score. Continuing this work with improved sentiment analysis by training the model and adding more features like text mining, semantic analysis to get a summary of the comments will give a more detailed and clear idea to the audience.

### **References**

Stefan Siersdorfer, Sergiu Chelaru, Wolfgang Nejdl,Jose San Pedro. 2010. How useful are your comments- Analyzing and Predicting Youtube Comments and Comment Rating. In World Wide Web Conference Committee(IW3C2)

Stefan Siersdorfer, Aliaksei Severyn, Alessandro Moschitti, Olga Uryupina, Barbara Plank,Katja Filippova.2014. Opinion Mining on YouTube. In Association for Computational Linguistics(ACL)

Shravan I.V, (2016, December 1). Analyzing Sentiments with NLTK. Retrieved from http://www.opensourceforu.com