**Project Proposal**

**Generic Sentiment Analysis for Movie Reviews**

**Team Members:**

Harshini Reddy Neerugutti

Sai Sreya Sri Pitta

Varsha Reddy GummyReddy

Venkata Jayanth Mandava

Github Link:

<https://github.com/Harshinireddyy19/NLP-Project-group9>

Goals and Objectives:

The sentiment analysis of a movie will rate how positive or negative a movie is, which gives us the overall rating of a movie. Hence, the process of recognizing if a movie review is positive or negative can be made automated as the machine learns while training the data.

Motivation:

Whenever we try to buy a product online (from amazon or any other shopping website), we usually see the reviews of that product and buy it if it has good reviews. Some reviews may be good or bad and other reviews could be neutral and reading all the reviews to buy a single product is neither efficient nor time saving process. Therefore, using the Sentiment Analysis, it is possible for us to extract positive, negative and neutral words.

Similar to product reviews, Movie Reviews can help us in determining whether a particular movie is watchable or not.

Objectives:

The objectives of this idea are:

* To extract the opinions of the reviewers and classify the texts as positive and negative from various perspectives.
* To analyze the given reviews and sort out the movies to understand what their overall reaction to the movie was i.e., if they liked or hated the movie.

Significances:

Sentiment analysis is a method of determining the emotional tone behind a group of words that is used for understanding the opinions, attitudes and emotions expressed in an online mention.

* Here, we will be classifying the movie reviews based on whether the person liked the movie or not. This is useful when the director/creator of the movie wants to measure the overall performance of the movie using the reviews.
* The model that we are generating here can also be used for creating a recommender system that provides recommendations for the viewers to watch the movie or not based on their previous reviews.
* Another application can be to be find out a group of similar viewers with similar taste of the movies.

Features:

* As we know, the reviews are given in word representations, and it is necessary for us to transform these words to numerical versions for a machine/model to understand. For doing this, we will be using different mechanisms such as “Bag of Words”, “Word2Vec” and other methods.
* The movie reviews contain enough information like the time, location, genres, characters etc., these features can be extracted and could be used for finding out movies with similar features.
* We have three types of emotion polarities based on the breadth of the text i.e., document level, sentence level and entity/aspect level. Ex:”I enjoy the xphone features, but the battery life is terrible”. Here, the emotion is positive if the xphone is considered and at the same time, it’s negative when it comes to the battery life of the phone. Therefore, instead of considering the entire sentence, it is important for us to select only those opinions that are very relevant to a particular feature(like the batterylife or the features) and classify based on them.

Visualization:

The data we have is the raw data that is simply downloaded from the sources without changing or updating it. This data may have null values, missing values or erroneous special characters which may not be helpful in predicting the results and may give wrong results if they are present. In order to increase the efficiency and accuracy of the model we have created, it is very important to preprocess the data. The basic steps involved in this model for preprocessing are:

1. Removing the HTML tags
2. Removing the special characters
3. Removing StopWords
4. Performing Stemming

The basic step of sentiment analysis is given below:

Diagram

Description automatically generated

References:

1. Datasets for movie reviews: <https://www.cs.cornell.edu/people/pabo/movie-review-data/>
2. About sentiment analysis: [IRJET-V6I9188.pdf](https://www.irjet.net/archives/V6/i9/IRJET-V6I9188.pdf)
3. [(PDF) Sentiment Analysis-An Objective View | J4R - Journal for Research - Academia.edu](https://www.academia.edu/25824227/Sentiment_Analysis_An_Objective_View)