**SENTIMENTAL ANALYSIS ON IMDB REVIEW DATA**

**Title of Contents**

* Introduction
* Project timeline
* Project Description
* References

**Introduction**

Sentiment analysis, a highly recommended technique in Natural language processing (NLP) that helps the user to understand the emotion behind the given text, the text can be either a paragraph, sentence or a document and this technique classifies text as positive, negative, or neutral in terms of sentiment. This technique helps in analyzing social media posts and product reviews, evaluating customer feedback, understanding public opinion regarding an article, identifying customer satisfaction or market trends. It’s one of the useful tools for business, researchers, and organizations to understand and respond to public sentiment effectively.

The given IMDB dataset [1] has 25,000 data points for each training and test data. The data consists of both positive and negative reviews which will be further analyzed and identified by the deep learning models such as LSTM [3], RNN [2], TF-IDF [2][5], Word2Vec [6], whereas these models predict whether the review that is given a positive or a negative.

**Project Timeline**

**Idea discussion:**

We have reviewed the research papers form IEEE and other websites and learned about Natural language processing and its applications to work on the sentimental analysis.

**Data collection and analysis:**

Once we have chosen NLP as the project domain we will go through the large movie review data set, analyze it and perform necessary transformations**.**

**Code Implementation:**

Understand the existing start of art models, analyze and evaluate them and develop code to improve the model performance.

**Final Presentation:**

Once code is developed and evaluated, we will summarize the proposed project and prepare the report and presentation slides for the final review.

**Project Description**

In this project, we considered IMDB dataset [1], which has 50 thousand reviews in total with 25 thousand each in training and testing with both positive and negative reviews. Firstly, we will be doing the data preprocessing where we will come over the following steps.

* Removal of HTML tags
* Remove non alphabetical characters
* Converting to lower case
* Stop words removal
* Tokenization
* Lemmatization

By doing all the above steps we are done with the data preprocessing.

Next, we will be using the Deep learning models to predict whether it is positive or negative review and see which model is performing best among the following models.

* LSTM (Long short-term memory networks) [3]
* RNN (Recurring neural networks) [2]
* TF-IDF (Term frequency and inverse document frequency) [2][5]
* Word2Vec (converts the data from word to vector) [6]

**References**

1. IMDB data set.

Link: <http://ai.stanford.edu/~amaas/data/sentiment/>

1. Prashuna,S.S.V.M.(2023) & Sai,S,S.(2023). Sentiment Analysis Of IMDB Movie Reviews, Thesis.

Link: https://www.diva-portal.org/smash/get/diva2:1779708/FULLTEXT02

1. Sayed,M.Q(2022).Sentiment Analysis of IMDb Movie Reviews Using Long Short-Term Memory, Research Gate, 978-1-7281-5467-1.

Link: https://www.researchgate.net/publication/346511493\_Sentiment\_Analysis\_of\_IMDb\_Movie\_Reviews\_Using\_Long\_Short-Term\_Memory

1. Abhimanyu,S., Chaitanya,K.& Necati.A.A(2020). Sentiment Analysis of IMDB Movie Reviews, GitHub.

Link: https://chaitanya1731.github.io/img/prj-1/report.pdf

1. Beatrice, J.,Minh,A,N.& Xavier,S(2019). IMDb Sentiment Analysis, GitHub.

Link: https://nminnie.github.io/pdf/Sentiment-analysis.pdf

1. Avi ,A(2022),Sentiment Analysis of IMDb Movie Reviews, IRJASET, 2321-9653

Link: https://www.ijraset.com/research-paper/sentiment-analysis-of-imdb-movie-reviews