**CIS 8694 ARTIFICIAL INTELLIGENCE**

**Project Group #6**

**SENTIMENT ANALYSIS FOR MOVIE REVIEWS**

**Anindita Sarmah** [**asarmah1@student.gsu.edu**](mailto:asarmah1@student.gsu.edu)

Panther ID: 002733978

**Ariba Surani** [**fariba1@student.gsu.edu**](mailto:fariba1@student.gsu.edu)

Panther ID- 002676489

**Mayuri Jadhav** [**mjadhav1@student.gsu.edu**](mailto:mjadhav1@student.gsu.edu)

Panther ID- 002732643

**Sanika Dhawale** [**sdhawale1@student.gsu.edu**](mailto:sdhawale1@student.gsu.edu)

Panther ID- 002733251

**Guided By: Prof. Yuan Long**

**Sentiment Analysis for Movie Review:**

**Project Summary:**

For our project we are developing a sentiment analysis model that can classify movie reviews into positive, negative, or neutral sentiment categories. This process involves collecting a dataset of labelled movie reviews, pre-processing the data to eliminate noise, selecting a suitable machine learning algorithm, training the model on a training set, and evaluating its performance using various metrics. The aim is to create a highly accurate model that can predict the sentiment of new movie reviews with confidence.

The technique of sentiment analysis is a part of natural language processing that aims to identify and extract the emotions and opinions conveyed in a given text, such as social media posts, news articles, or reviews. This process involves using machine learning algorithms to analyze the language and determine whether the sentiment expressed is positive, negative, or neutral. The process comprises several steps, including text pre-processing, feature extraction, and classification. Text pre-processing involves cleaning and preparing the text for analysis. Feature extraction is used to identify relevant elements such as words or phrases that help to determine the sentiment. Finally, classification is done using a machine learning algorithm to assign the sentiment category to the text. Sentiment analysis is used in various areas such as social media monitoring, market research, customer feedback analysis, and political analysis.

Sentiment analysis provides businesses with a means of comprehending customer sentiment, including their preferences, tastes, and aversions, by examining customer feedback, reviews, and social media posts. By analyzing the sentiment of their target audience, businesses can gain valuable insights into market trends, preferences, and sentiment, which can help them make informed decisions about product development, advertising, and marketing strategies. Furthermore, businesses can modify their strategies based on the insights gained from monitoring the sentiment of their competitors, such as identifying their strengths and weaknesses.

**Advantages of Movie review Sentimental analysis :**

Time and Cost Efficiency: Sentiment analysis of movie reviews can be done quickly and cost-effectively, especially when compared to traditional methods such as surveys and focus groups.

Objectivity: Sentiment analysis provides an objective assessment of the overall sentiment expressed in movie reviews, free from bias and subjectivity.

Insights into Audience Perception: By analyzing movie reviews, sentiment analysis can provide valuable insights into how audiences perceive and react to movies. This information can be used by filmmakers, production companies, and distributors to make informed decisions about future movie projects, marketing strategies, and release plans.

Improving Audience Engagement: Sentiment analysis can help movie makers and distributors understand their audiences better, enabling them to tailor their content and marketing efforts to appeal to their target audience.

Competitive Advantage: By monitoring and analyzing the sentiment of movie reviews, businesses can gain a competitive advantage over their rivals by making timely adjustments to their marketing strategies, products, and services.

**Dataset:** We used the data source from Kaggle <https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews>

**Data Model**: Neural Network -Keras

**Keras** is an open-source deep learning library written in Python that allows users to easily build, train, and deploy neural networks. It provides a user-friendly interface for defining and configuring various types of neural network architectures, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and multi-layer perceptrons (MLPs).

**Layers of our Neural Network models:**

1. One word embedding layer which creates word embeddings of length 16 from integer encoded review.

2. Second layer is global average pooling layer which is used to prevent overfitting by reducing the number of parameters.

3. Then a dense layer which has 16 hidden units and uses relu as activation function

4. The final layer is the output layer which uses sigmoid as activation function

**Code:**

model=keras.Sequential([keras.layers.Embedding(10000,16,input\_length=500),

keras.layers.GlobalAveragePooling1D(),

keras.layers.Dense(16,activation='relu'),

keras.layers.Dense(1,activation='sigmoid')])

**Overview of the Steps Involved:**  
1. Preprocess the data  
 2. Word Embedding  
 3. Defining the model architecture  
 4. Training the model  
 5. Testing the model

**Tools Used:**

1. Google Colab
2. Tableau (Data Visualization)
3. Python Libraries used-
   1. Keras from Tensorflow
   2. Numpy
   3. Pandas
   4. Pyplot from matplotlib

**Google Colab Link:** <https://colab.research.google.com/drive/1_BP0Mh895mbyMQw37HirXxWXpwLOzxof?usp=sharing>

**Video Recording:**

<https://studentgsu-my.sharepoint.com/:v:/g/personal/mjadhav1_student_gsu_edu/ESbl0R0uFgxOtbNgOyWxppQBlEg4W-152HcEZTS2MvSKvw>

**Conclusion:**

To conclude our dataset has been trained to predict the sentiment of movie reviews with an accuracy of 94%. This indicates that the algorithm is able to accurately classify 94 out of every 100 reviews correctly. However, when tested on a separate set of data, the accuracy of the algorithm drops to 88%. While this is still a relatively high level of accuracy, it suggests that there may be some limitations to the model's ability to generalize to new data. Therefore, it may be necessary to further refine the algorithm to improve its performance on previously unseen data.