**National University of Computing & Emerging Sciences**

**Karachi Campus**



**Sentimental Analysis of Movies**

*Project Report*

Course: Data structure

Group Members:

**Syeda Wara Batool 21k-3214**

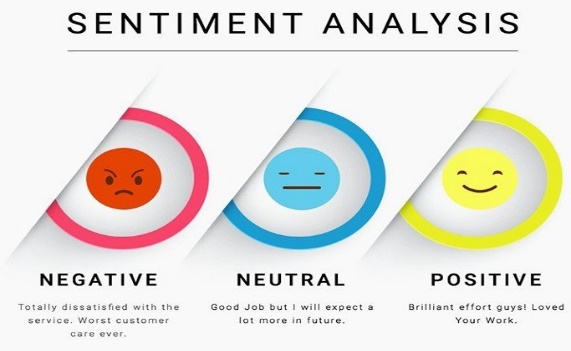
**Syed Hadi Raza 21k-3420**

**Humam Imtiaz 21k-3334**

*Submitted to:*

Sir Fatmi Ali

## **Objective-**

The objective of our project was to make sentimental analysis for movies reviews. Movie reviews help users decide if the movie is worth their time. A summary of all reviews for a movie can help users make this decision by not wasting their time reading all reviews. Movie-rating websites are often used by critics to post comments and rate movies which help viewers decide if the movie is worth watching. Sentiment analysis can determine the attitude of critics depending on their reviews. Sentiment analysis of a movie review can rate how positive or negative a movie review is and hence the overall rating for a movie. Therefore, the process of understanding if a review is positive or negative can be automated as the machine learns through training and testing the data. This project aims to rate reviews using two classifiers and compare which gives better and more accurate results.

## **Introduction**

Sentiment analysis is the interpretation and classification of emotions within text data using text analysis techniques. Sentiment analysis allows businesses to identify customer sentiment toward products, brands or services in online conversations and feedback. Sentiment analysis models focus on polarity (positive, negative, neutral) but also on feelings and emotions (angry, happy, sad, etc), and even on intentions (e.g. interested v. not interested). Sentiment analysis has become a hot topic and many big companies are investing their resources to predict the results for their businesses.

## **Data**

## 

The dataset is taken or download Kaggle. The first data set is Of size

27MB.Link for it is below:

<https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews>

The next data set is for stop word from kaggle of size 23 KB,link is below:

[https://www.kaggle.com/datasets/rtatman/stopword-lists-for-19-languages](%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20https:/www.kaggle.com/datasets/rtatman/stopword-lists-for-19-languages)

## **Abstract**

The working principle of sentiment analysis includes tokenization, word filteringand classifications. In tokenization, text needs to be segmented into units such as words/ numbers or punctuations.After preprocessing, we analyze the dataset by performing classification using Naïve Bayes, Vectors .Here, we determine the the review based on accuracy. Hence, We analyze and study the features that affect the scores of our review text and finally classify the movie as positive or negative. We have also implemented usage of:

1. Nodes
2. Linked list
3. Mapping

4. Tree

5. Vectors

6. Functions

## **Implementation**

Our project is implemented on Visual studios. We have added multiple libraries for example map and vector .

1. The number of reviews is stored in the linked list
2. Mapping is a type of fast key lookup data structure that offers a

flexible means of indexing into its individual elements. Maps provide an

alternative approach to searching. Mapping class

1. Stop word: Stop words are a set of commonly used words in a language. Examples of stop words in English are “a”, “the”, “is”, “are” and etc. Stop words are commonly used in Text Mining and Natural Language Processing (NLP) to eliminate words that are so commonly used that they carry very little useful information.

We have used a Functional called separation in which the separation of positive and negative words is being held and those positive and negative words are being saved in following linked list.

have used mapping a type of fast key lookup data structure that offers a

flexible means of indexing into its individual elements. Maps provide an

alternative approach to searching.

**RESULTS & EVALUATION**

1. Naive Bayes

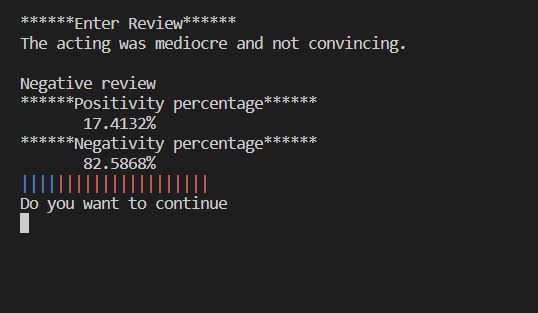
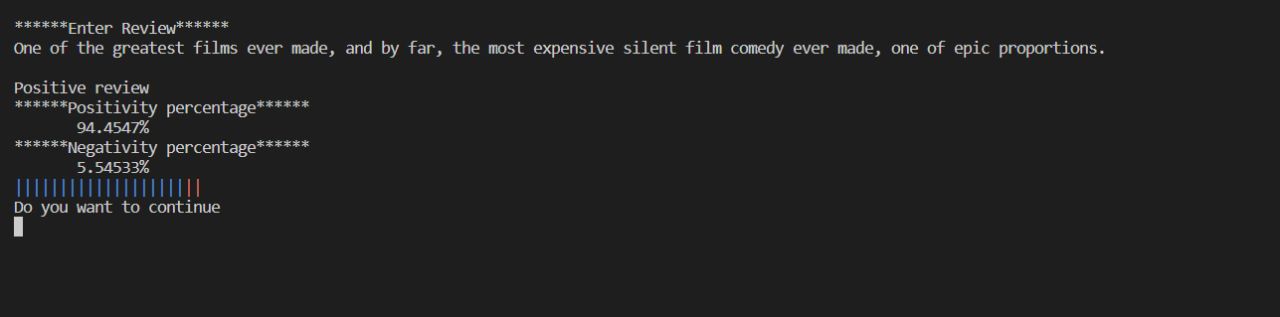
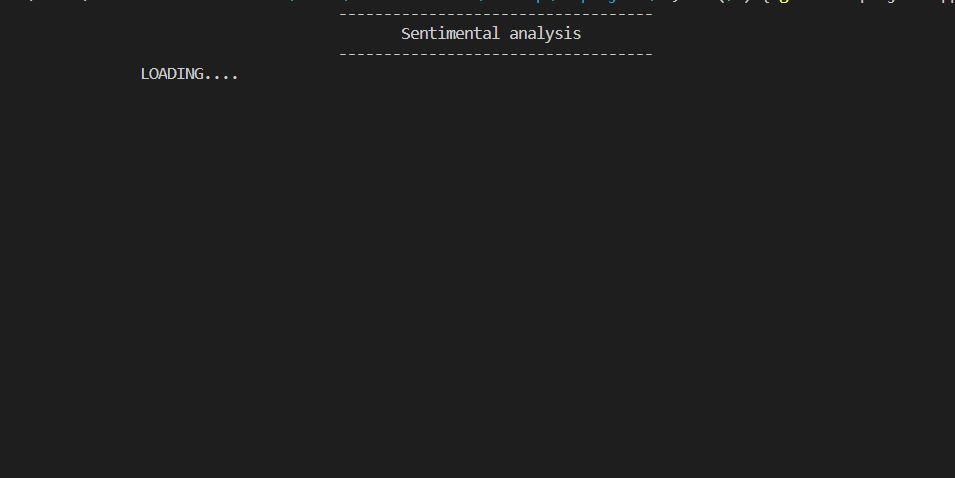
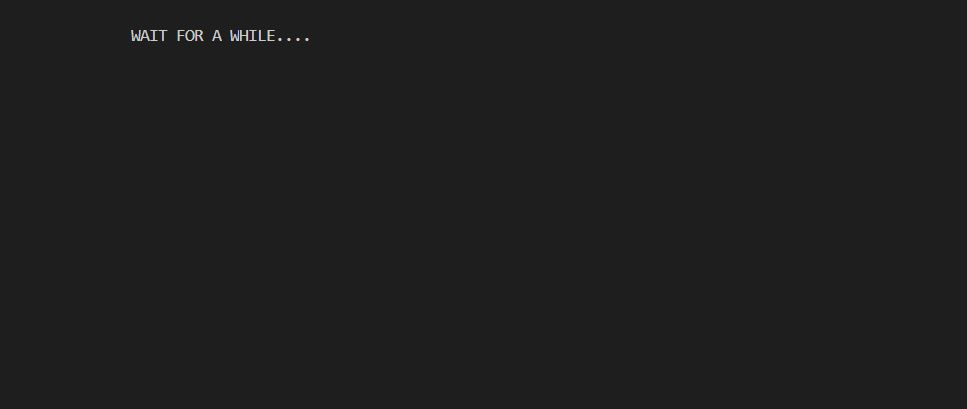
Naive Bayes works on the principle of probabilities and the Bayes rule given by:

P(c|d) =

It is a classification algorithm, primarily used for text classification involving high dimensional training data sets. Example spam filtering, sentiment analysis etc. This algorithm learns the probability of an object with certain features belonging to a particular class. It is a probabilistic classifier. This algorithm is called Naive Bayes because it makes a naive assumption that occurrence of certain features is independent of each other which in reality is not the case . A is called the proposition and B is called the evidence . P(A) is called prior probability of proposition and P(B) is called prior probability of evidence. P(A/B) is called the posterior. P(B/A) is called the likelihood.

P(A/B) = Probability of occurrence of event A, given event B has already occured P(A) = Probability of event A P(B) = Probability of event B P(B/A) = Probability of occurrence of event B, given event A has already occure.

**Output:**

****