



A Project Work at ATSS's Institute of Industrial & Computer Management & Research, Nigdi



Movies reviews Analysis using Machine Learning

SUBMITTED BY:

- **Rwittick Sadhu**
- **Chetan K. Choudhari**

UNDER THE GUIDANCE OF (MENTOR): Mr. Sanjay Mathpati (Department of MCA)

[Date of Submission: 5th March,2022]

TABLE OF CONTENTS

| Sr. No. | Chapter Name | Page No. |
|-------------------|--|-----------------|
| 1 | Chapter 1: Introduction | 2-6 |
| 1.1 | Introduction | 3 |
| 1.2 | Existing System and Need of System | 4 |
| 1.3 | Scope of work | 4 |
| 1.4 | Operating Environment-Hardware and Software | 5 |
| 1.5 | Technology Used | 6 |
| 2 | Chapter 2:Proposed System | 6-7 |
| 2.1 | Proposed System | 6 |
| 2.2 | Objectives of system | 6 |
| 2.3 | Scope of proposed system | 7 |
| 3 | Chapter 3. ANALYSIS & DESIGN | 7-16 |
| | 1. Use Case Diagram 2. FUNCTIONAL DECOMPOSITION DIAGRAM 3. Component Diagram 4. Deployment Diagram 5. Module Specifications | |
| 4 | User Interface Screens | 17 |
| 6 | Drawbacks and Limitations | 18 |
| | | |
| 7 | Proposed Enhancement | 18 |
| 8 | Bibliography | 18 |
| Annexure 1 | User Interface Screens | 19 |
| Annexure 2 | Sample Program Code | 19-21 |

Chapter 1. INTRODUCTION

The count of internet users is increasing day by day and with this, social media influences a lot to the people for their internet addiction. Due to this, people are shifted from print media to digital media. Instead of reading printed news papers, most of the users start using e-newspapers and their respective social media official pages to get update with latest news and all. Most of the users trusted a lot on online reviews of any product or services or even if any IMDb movie. So, the sentiments and emotions of users are hidden inside their reviews and they are expressing their thoughts through their reviews. New users or customer usually read previous customers reviews before buying or adapting products and services. They have also taken their decisions like to watch any movie or not from IMDb movie reviews. Hence, there is much more need of sentimental analysis of IMDb movie reviews. Through sentimental analysis, we can extract information like popularity of movie, audience positive and negative emotions analysis about movies. Using sentimental analysis, which age group is influenced by which character of movie that we can find out.

The sentiment analysis of IMDb movie review process extracted hidden emotions inside customer's comments and reviews by using the keywords which are used inside it. Sometimes, producer and directors want to show something else and their intentions behind it are somewhat different but when audience watch that movie their reactions are totally opposite to that has been expected by directors and producers. To analyze such kind of things, sentiment analysis of IMDb movie reviews plays a vital role. Sometimes, people like supportive actor and not lead role actor. That also can be found out through this process. Lot of times movie directors and producers do sentiment analysis on IMDb movie trailer's reviews. From that they understand pro and cons of movie and as per that they change or manipulate the scenes and dialogs of that movie. In the marketplace, the priority is always given to customer rather than seller. So the emotions of targeted customers are very much important for everyone. Many times, movie directors used to check Actors and Actresses previous IMDb movie reviews and if they are positive then only they offer their next movie to that actor or actress. Emotion is the measure factor in buying process of customer that's why in sentiment analysis we mostly focus on customer hidden intentions as they are positive only then customer will buy movie tickets.

Existing System and Need of System

About the Existing System:

Survivor usually enter manually all the details.

The Drawbacks of Existing system

Survivor usually provide a feedback form & enter manually all the details.

and analyze one by one.

- That is time consuming, hectic, error could prone data could be lost.
- Not owner-friendly environment.
- Accident easily occur either burning of papers or water damage.

Need of system

- New users or customer usually read previous customers reviews before buying or adapting products and services. They have also taken their decisions like to watch any movie or not from IMDb movie reviews. there is much more need of sentimental analysis of IMDb movie reviews Through sentimental analysis, we can extract information like popularity of movie, audience positive and negative emotions analysis about movies
- The aim of this project is to build an analysis model for Movies reviews which will allow us to categorize words based on their sentiments, that is whether they are positive, negative and also the magnitude of it.
- This project could show a path to reduce customer churn.

Scope of work:

- The sentiment analysis of IMDb movie review process extracted hidden emotions inside customer's comments and reviews by using the keywords which are used inside it.

- Movies reviews Analysis is a process of extracting opinions that have different polarities. By polarities, we mean positive, negative or neutral. It is also known as opinion mining and polarity detection.
- With the help of sentiment analysis, you can find out the nature of opinion that is reflected in documents, websites, social media feed, etc.
- Movies reviews Analysis is a type of classification where the data is classified into different classes. These classes can be binary in nature (positive or negative) or, they can have multiple classes (happy, sad, angry, etc.).

Operating Environment-Hardware and Software :

| Hardware Requirement | | |
|---|--|--|
| Processor | RAM | Free Space |
| Intel(R) Pentium(R) CPU N3700 @ 1.60 GHz | 4 GB or Above | 500MB or Above |
| Software Requirements | | |
| Operating System | Application/Software | |
| 64-bit Operating System, x64 based processor | Web Browser: Mozilla, Google Chrome,IE8, etc. | |
| Server-Side Requirements | | |
| Operating System | Processor | Storage Space |
| Win-7, Win-10, Linux or any other higher version | Intel core i3 or i5 | Ram 4GB or more and 5GB or More Free Space |

Technology Used

| | |
|--------------------|-------------------------------------|
| Front End Language | HTML, AJAX, JQUERY, JAVASCRIPT, CSS |
| Back End Language | Python Machine Learning |
| Library | |
| Tools | Jupiter, Visual Studio Code. |
| Framework | |

Chapter 2. PROPOSED SYSTEM

PROPOSED SYSTEM :

The count of internet users is increasing day by day and with this, social media influences a lot to the people for their internet addiction. Due to this, people are shifted from print media to digital media. Instead of reading printed news papers, most of the users start using e-newspapers and their respective social media official pages to get update with latest news and all. Most of the users trusted a lot on online reviews of any product or services or even if any IMDb movie. So, the sentiments and emotions of users are hidden inside their reviews and they are expressing their thoughts through their reviews.

. New users or customer usually read previous customers reviews before buying or adapting products and services. They have also taken their decisions like to watch any movie or not from IMDb movie reviews. Hence, there is much more need of sentimental analysis of IMDb movie reviews. Through sentimental analysis, we can extract information like popularity of movie, audience positive and negative emotions analysis about movies. Using sentimental analysis, which age group is influenced by which character of movie that we can find out.

Objectives of the Proposed System:

- New users or customer usually read previous customers reviews before buying or adapting products and services. They have also taken their decisions like to watch any movie or not from IMDb movie reviews. there is much more need of sentimental analysis of IMDb movie reviews Through sentimental analysis, we can extract information like popularity of movie, audience positive and negative emotions analysis about movies
- The aim of this project is to build an analysis model for Movies reviews which will allow us to categorize words based on their sentiments, that is whether they are positive, negative and also the magnitude of it.
- This project could show a path to reduce customer churn.

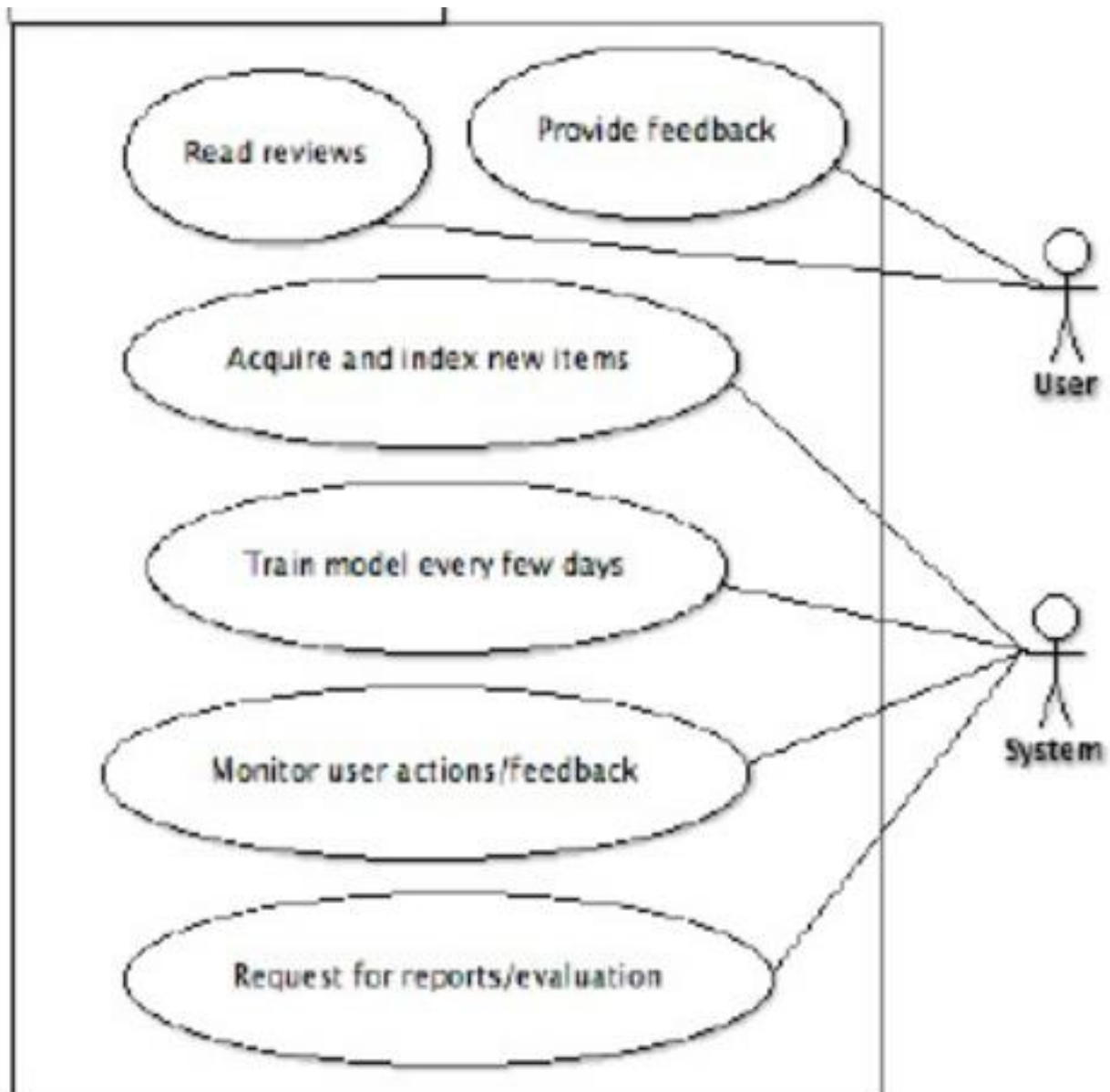
Scope of the proposed System:

The project is identified by the merits of the system offered to the user. The merits of this project are as follows: -

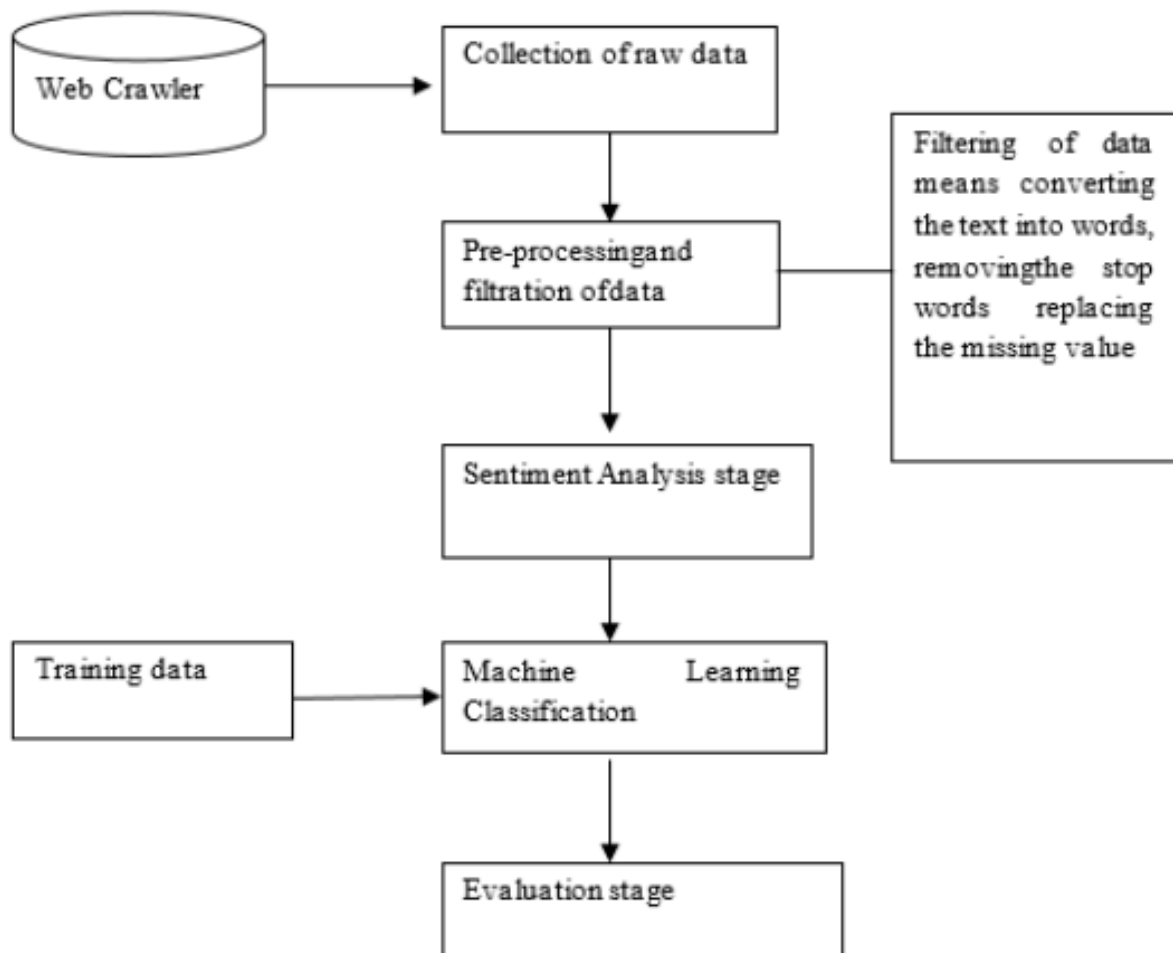
- The sentiment analysis of IMDb movie review process extracted hidden emotions inside customer's comments and reviews by using the keywords which are used inside it.
- Movies reviews Analysis is a process of extracting opinions that have different polarities. By polarities, we mean positive, negative or neutral. It is also known as opinion mining and polarity detection.
- With the help of sentiment analysis, you can find out the nature of opinion that is reflected in documents, websites, social media feed, etc.
- Movies reviews Analysis is a type of classification where the data is classified into different classes. These classes can be binary in nature (positive or negative) or, they can have multiple classes (happy, sad, angry, etc.).

Chapter 3. ANALYSIS & DESIGN

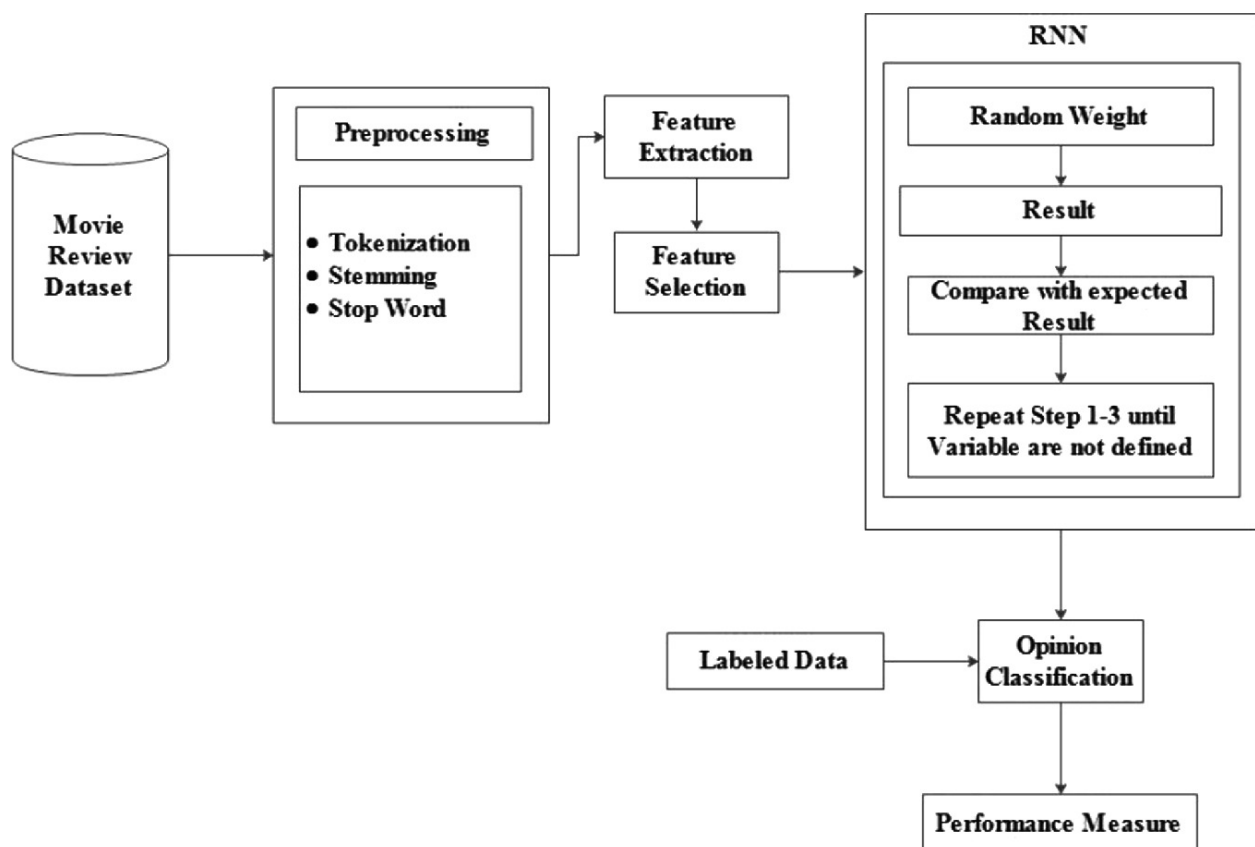
1. Use Case Diagram



2. FUNCTIONAL DECOMPOSITION DIAGRAM

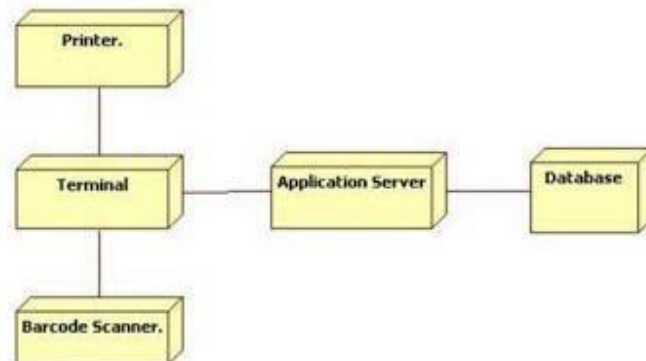


3. Component Diagram



4. Deployment Diagram

Deployment
Diagram



5. Module Specifications

- Defining the model

We have used one of the [Naive Bayes \(NB\)](#) classifier for defining the model. Specifically, we have used [MultinomialNB classifier](#).

Naive Bayes Model

This model applies Bayes theorem with a Naive assumption of no relationship between different features. According to Bayes theorem:

Posterior = likelihood * proposition/evidence

or

$$P(A|B) = P(B|A) * P(A)/P(B)$$

For ex: In a deck of playing cards, a card is chosen. What is the probability of a card being queen given the card is a face card?

This can be solved using Bayes theorem.

$$P(\text{Queen given Face card}) = P(\text{Queen}|\text{Face})$$

$$P(\text{Face given Queen}) = P(\text{Face}|\text{Queen}) = 1$$

$$P(\text{Queen}) = 4/52 = 1/13 \quad P(\text{Face}) = 3/13 \quad \text{From Bayes theorem:}$$

$$P(\text{Queen}|\text{Face}) = P(\text{Face}|\text{Queen}) P(\text{Queen})/P(\text{Face}) = 1/3$$

For an input with several variables:

$$P(y|x_1, x_2, \dots, x_n) = P(x_1, x_2, \dots, x_n|y) * P(y)/P(x_1, x_2, \dots, x_n)$$

with Naive Bayes we assume $x_1, x_2 \dots x_n$ are independent of each other, i.e:

$$P(x_1, x_2, \dots, x_n|y) = P(x_1|y) * P(x_2|y) \dots * P(x_n|y)$$

The assumption in distribution of $P(x_i|y)$ give rise to different NBM. For

example assuming Gaussian distribution will give rise to Gaussian Naive Bayes

(GNB) or multinomial distribution will give Multinomial Naive Bayes (MNB).

Naive Bayes Model works particularly well with text classification and spam filtering. Advantages of working with NB algorithm are:

Requires a small amount of training data to learn the parameters

Can be trained relatively fast compared to sophisticated models

The main disadvantage of NB Algorithm is:

It's a decent classifier but a bad estimator

It works well with discrete values but won't work with continuous values (can't be used in a regression)

The dilemma of NB Algorithm

A challenging question which can be asked regarding NB algorithm is:

although the conditional independence assumption in NB algorithm is hardly true in real life then how come NB Algorithm work so well as a classifier? I won't discuss the solution here, rather will direct you towards the resource which contains the solution.

The loss function for NB classification

NB classification uses a zero-one loss function. In this function error = number of incorrect classifications. Here accuracy of probability estimation is not taken into account by error function given that class with the highest probability is predicted right. For example, let's say there are two classes A and B, and different attributes (x_1, x_2, \dots, x_n) are given. $P(A|\text{all attributes}) = 0.95$ and $P(B|\text{all attributes}) = 0.05$ but NB might estimates $P(A|\text{all attributes}) = 0.7$ and

$P(B|\text{all attributes}) = 0.3$. Here although estimates are far from accurate but classification is correct.

- **Compiling the model**
we are using sklearn's modules and classes we just need to import the precompiled classes.

```
from sklearn.naive_bayes import MultinomialNB
```

- **Fitting the model**

In this step, we generate our model-fitting our dataset in the MultinomialNB. In order to look for the arguments which can be passed while fitting the model, it's advised to check the sklearn webpage of the module underuse

```
MNB = MultinomialNB()  
MNB.fit(X_train, Y_train)  
  
MultinomialNB(alpha=1.0, class_prior=None, fit_prior=True)
```

- **Evaluating the model**

Here we quantify the quality of our model. We have used metrics module from the sklearn library to evaluate the predictions

```
from sklearn import metrics
predicted = MNB.predict(X_test)
accuracy_score = metrics.accuracy_score(predicted, Y_test)
```

```
print(str('{:04.2f}'.format(accuracy_score*100))+ '%')
```

60.25%

- Making predictions with the model

```
from sklearn.naive_bayes import GaussianNB
GNB = GaussianNB()
GNB.fit(X_train.todense(), Y_train)
accuracy_score = metrics.accuracy_score(GNB.predict(X_test), Y_test)
print('GNB accuracy = ' + str('{:4.2f}'.format(accuracy_score*100))+ '%')
```

GNB accuracy = 47.53%

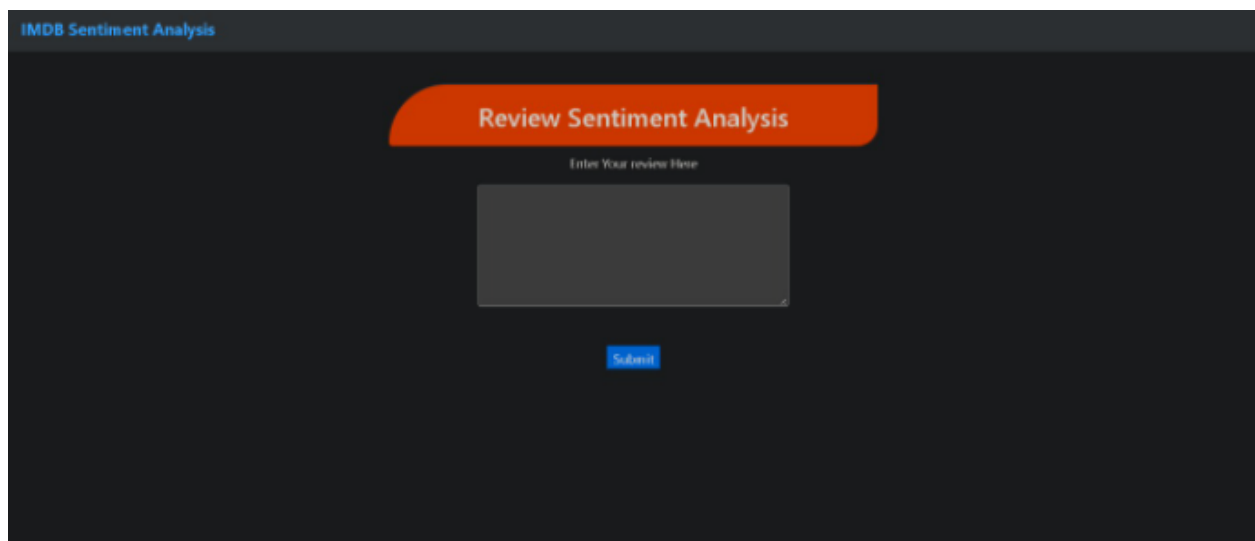
```
from sklearn.naive_bayes import BernoulliNB
BNB = BernoulliNB()
BNB.fit(X_train, Y_train)
accuracy_score_bnb = metrics.accuracy_score(BNB.predict(X_test), Y_test)
print('BNB accuracy = ' + str('{:4.2f}'.format(accuracy_score_bnb*100))+ '%')
```

BNB accuracy = 60.61%

6. User Interface Design

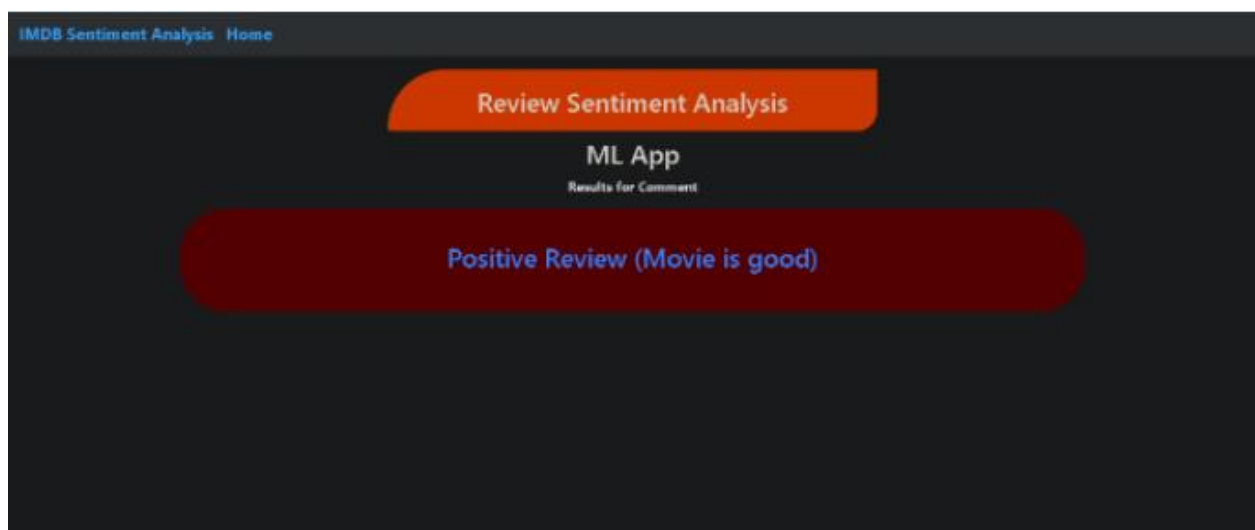
Annexure I:

1: Input Page



The screenshot shows a web application titled "IMDB Sentiment Analysis". It features a dark background with a prominent orange rounded rectangle containing the text "Review Sentiment Analysis". Below this, there is a smaller, lighter gray rounded rectangle with the placeholder text "Enter Your review Here". At the bottom center, there is a small blue button labeled "Submit".

2: Output page



The screenshot shows the output page of the "IMDB Sentiment Analysis" application. It has a dark background. At the top left, there is a navigation bar with the text "IMDB Sentiment Analysis Home". In the center, there is an orange rounded rectangle with the text "Review Sentiment Analysis". Below this, the text "ML App" is displayed, followed by "Results for Comment". A large, dark red rounded rectangle contains the text "Positive Review (Movie is good)" in a light blue font.

Chapter 4. USER MANUAL

Drawbacks and Limitations

- Train the model every few days..
- We will make prediction more accurate.

Proposed Enhancements

Bibliography

For ML

➤ <https://towardsdatascience.com/sentiment-analysis-introduction-to-naive-bayes-algorithm-96831d77ac91>

➤ <https://sites.pitt.edu/~naraehan/presentation/Movie%20Reviews%20sentiment%20analysis%20with%20Scikit-Learn.html>

➤ <https://acadpubl.eu/hub/2018-119-16/2/374.pdf>

➤ <https://www.wonderflow.ai/blog/sentiment-analysis-examples>

For MySQL

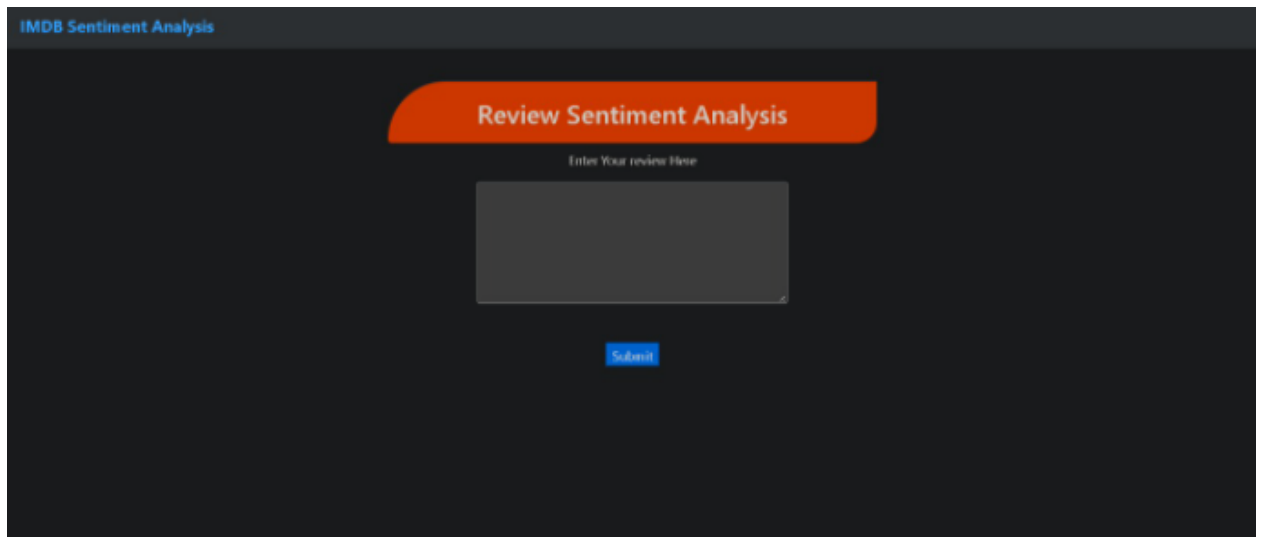
➤ <https://towardsdatascience.com/sentiment-analysis-a-how-to-guide-with-movie-reviews-9ae335e6bcb2>

For XAMPP

<https://www.apachefriends.org/download.html>

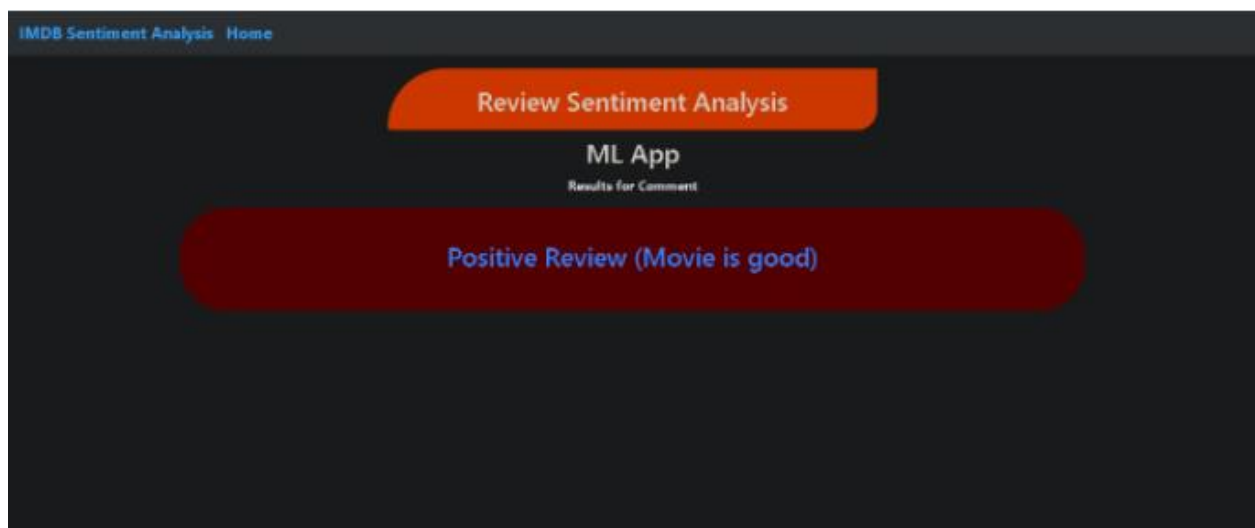
Annexure I:

1: Input Page



The screenshot shows a web application titled "IMDB Sentiment Analysis" in the top left corner. The main heading is "Review Sentiment Analysis" in white text on an orange rounded rectangle. Below this, there is a prompt "Enter Your review Here" in small white text. Under the prompt is a large, empty gray rectangular text input field. At the bottom center, there is a small blue button with the word "Submit" in white text.

2: Output page



The screenshot shows the output page of the "IMDB Sentiment Analysis" application. The top left corner has the text "IMDB Sentiment Analysis Home". The main heading is "Review Sentiment Analysis" in white text on an orange rounded rectangle. Below this, the text "ML App" is displayed in white, followed by "Results for Comment" in smaller white text. A large, dark red rounded rectangle contains the text "Positive Review (Movie is good)" in white.

Annexure II:

```
from flask import Flask,render_template,url_for,request

import numpy as np

import pickle

import pandas as pd

import flasgger
from flasgger
import Swagger
from sklearn.feature_extraction.text import CountVectorizer

from sklearn.naive_bayes import MultinomialNB

from sklearn.externals import *

app=Flask(__name__)

Swagger(app)

mnbc = pickle.load(open('Naive_Bayes_model_imdb.pkl','rb'))

countVect = pickle.load(open('countVect_imdb.pkl','rb'))

@app.route('/')

def home():

    return render_template('/home.html')

@app.route('/predict',methods=['POST'])

def predict():
```

```
if request.method == 'POST':  
  
    Reviews = request.form['Reviews']  
  
    data = [Reviews]  
  
    vect = countVect.transform(data).toarray()  
  
    my_prediction = mnbg.predict(vect)  
  
return render_template('/result.html',prediction = my_prediction)
```

```
if __name__ == '__main__':  
    app.run(debug=True)
```