Analysis of Amazon Cell Phone Reviews

A UG PHASE-1 PROJECT REPORT

Submitted to

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD

In partial fulfilment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE AND ENGINEERING

Submitted by

SHAHEDA NAAZ	**19UK1A0541**
NARAHARI SOWMYA REDDY	**19UK1A0544**
GUDLA PRIYA	**19UK1A0547**
MAKKALA PRASHANTH KUMAR	**19UK1A0558**

Under the esteemed guidance of
MS. M. SWETHA
(Assistant Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
VAAGDEVI ENGINEERING COLLEGE
(Affiliated to JNTUH, Hyderabad)
Bollikunta, Warangal – 506005
2019– 2023
I

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
VAAGDEVI ENGINEERING COLLEGE
BOLLIKUNTA, WARANGAL – 506005
2019 – 2023

CERTIFICATE OF COMPLETION
UG PROJECT PHASE-1

This is to certify that the UG Phase-1 Project entitled **"OTT PLATFORM ANALYSIS TOOL"** is being submitted by *SHAHEDA NAAZ (*H.NO:19UK1A0541)**, *N. SOWMYA REDDY* **(H.NO:19UK1A0544)**, *G. PRIYA* **(H.NO:19UK1A0547), M. PRASHANTH KUMAR (H.NO:19UK1A0558)** in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering** to **Jawaharlal Nehru Technological University Hyderabad** during the academic year **2022-23**, is a record of work carried out by them under the guidance and supervision.

Project Guide   **Head of the Department**
Ms. M. Swetha   **Dr. R. Naveen Kumar**
(Assistant Professor)   (Professor)

II

ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr. P. PRASAD RAO,** Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this UG Project Phase-1 in the institute.

We extend our heartfelt thanks to **Dr. R. NAVEEN KUMAR**, Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the UG Project Phase-1.

We express heartfelt thanks to Smart Bridge Educational Services Private Limited, for their constant supervision as well as for providing necessary information regarding the UG Project Phase-1 and for their support in completing the UG Project Phase-1.

We express heartfelt thanks to the guide, **Ms. M. Swetha,** Assistant professor, Department of CSE for her constant support and giving necessary guidance for completion of this UG Project Phase-1.

Finally, we express our sincere thanks and gratitude to my family members, friends for their encouragement and outpouring their knowledge and experience throughout the thesis.

SHAHEDA NAAZ   (19UK1A0520)
N. SOWMYA REDDY   (19UK1A0551)
G. PRIYA   (19UK1A0527)

M. PRASHANTH KUMAR (19UK1A0570)

ABSTRACT

Predicting the type of sentiment associated with the reviews provided by the users on a cell phone through Amazon website. Multinomial Naive Bayes model was trained for the prediction as the scenario involved text classification. This prediction would be useful for a new user who is trying to look for specific things before buying a new cell phone online by automatically identifying the prominent emotion of the review.

Amazon.com is a treasure trove of product reviews and their review system is accessible across all channels presenting reviews in an easy-to-use format. The product reviewer submits a rating on a scale of 1 to 5 and provides own viewpoint according to the whole experience.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Overview

In recent years E-Commerce has exploded everywhere in the world, and the majority of the population prefers to buy products through these websites. Consequently large amounts of data in the form of reviews are produced which helps prospective buyers choose the right product. Furthermore, these reviews contain opinionated contents which can be useful for the company to identify the areas which need to be enhanced.

However it is impractical for the user to read each and every review about the product. Moreover, reading only a few reviews may present a biased idea about the product. It is quite possible that some of the reviews lack credible sources, which the users have no means to differentiate. Besides the reviews and ratings provided little to assess the specific features of the product. Due to all the above constraints, the user is unable to make a fully informed decision about the product.

1.2 Purpose

Our project aims at building a model to predict the helpfulness of the review and the rating based on the review text. Corpus-based and knowledge-based methods can be used to determine the semantic similarity of review text. We used Natural language processing for the analysis of the sentiment (positive or a negative) of the given review.

Sentiment analysis can be used to extract customer reviews from different sources on the internet. This technique implements various algorithms to analyze the corpus of data and make sense out of it. This technique helps to identify the orientation of a sentence thereby recognizing the element of positivity or negativity in it. Automated opinion mining can be implemented through a machine learning based approach. Opinion mining uses natural language processing to extract the subjective information from the data.

The Sentiment Analysis is used to classify the polarity (positivity or negativity behind the opinion) of the reviews. There are three main categories in sentiment analysis: sentence-level, unit aspect-level and document level sentiment analysis.

2. LITERATURE SURVEY

2.1 Existing Problem (OR) Problem Statement:

When a person thinks of buying a product, his or her next immediate action would be to search for the product on the internet. Internet gives him/her a lot of choices based on brand, price, model, features, colors, quality, rating, discounts and many more. Introduction of new products, new fashion, new model, new brand, new business, new technology, new services, new marketing strategies happen daily. These may leave a consumer perplexed and confused when having to make a choice. When it's tough to make a choice, we tend to get the feedback from the people who have already bought and used those products. Customers register their review on online shopping sites.

Amazon is the 1st site in the list of the top 10 online shopping sites. Amazon gives overall ratings for each product based on all ratings from the reviewers of the product. But from that overall rating a person cannot necessarily conclude on quality of all the features of the product.

Getting feedback on major features of the product will help the manufacturer or seller to improve the quality and marketing of the product. The ultimate solution for these problems is to read through the text reviews to know specifically which feature of the product is lacking customer satisfaction. But a product may have thousands of reviews which would make this task tedious. So, we need a system to do that for us.
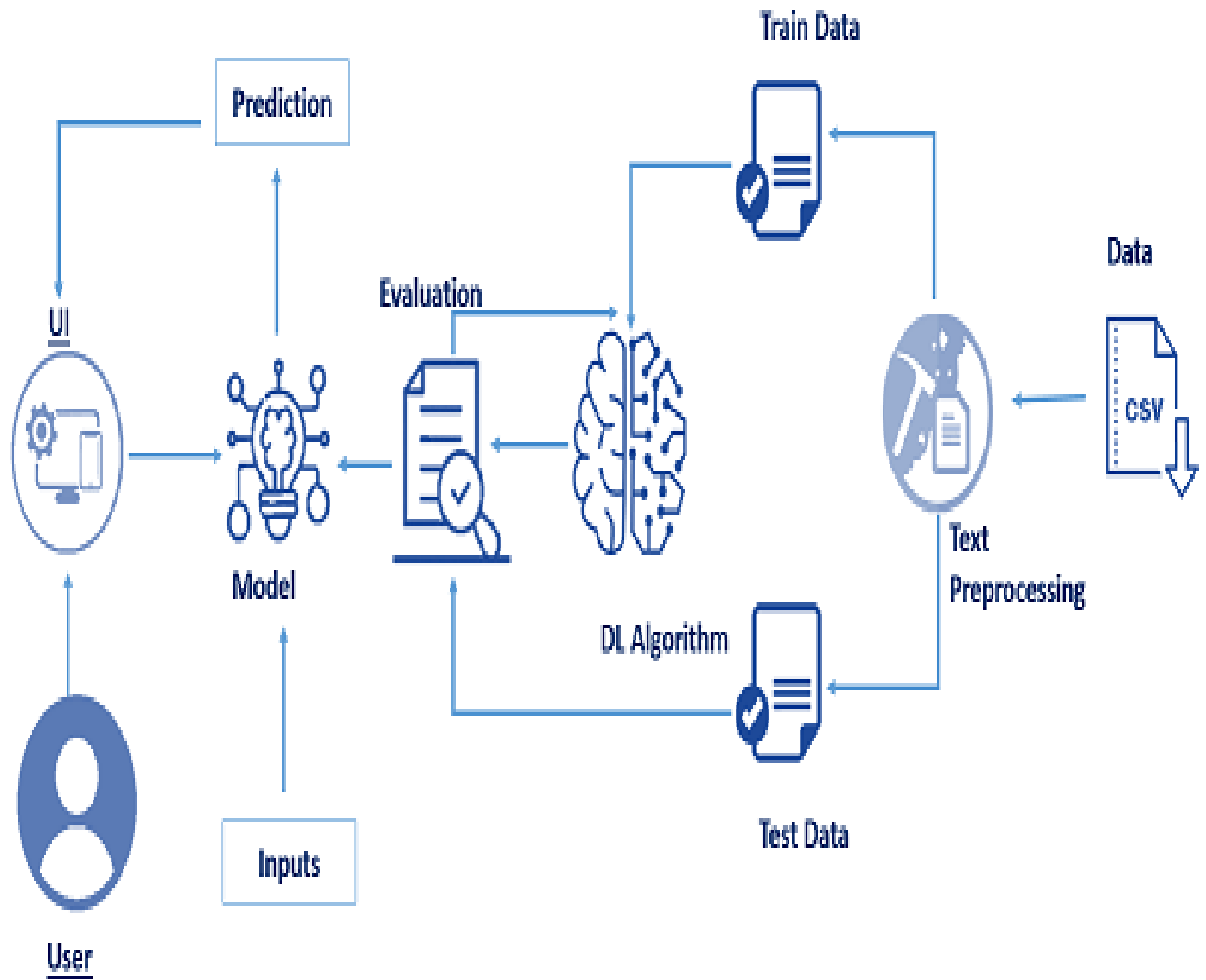
2.2 Proposed Solution

We started building a model to predict the helpfulness of the review and the rating based on the review text. Currently consumers who write opinion and experience online are increasing. If the consumer reads the whole review it can spend much time. But if it is read without some evaluation it will be biased. Sentiment classification aims to overcome this problem by automatically classifying user review by positive or negative opinion.

This project aims at building a model to predict the helpfulness of the review and therating based on the review text. Corpus-based and knowledge-based methods can be used to determine the semantic similarity of review text. We will be using Natural language processing to analyse the sentiment (positive or negative) of the given review. A simple web application is integrated to the model built.

3. THEORITICAL ANALYSIS

3.1 Block Diagram

Train Data

Prediction

Data

Evaluation

UI

CSV

Model

Text
Preprocessing

DL Algorithm

Inputs

Test Data

User

3.2 Hardware / Software Designing

The following is the Hardware required to complete this project:

● Internet connection to download and activate
● Administration access to install and run Anaconda Navigator
● Minimum 10GB free disk space
● Windows 8.1 or 10 (64-bit or 32-bit version) OR Cloud: Get started free, *Cloud account required.

Minimum System Requirements To run Office Excel 2013, your computer needs to meet the following minimum hardware requirements:

● 500 megahertz (MHz)
● 256 megabytes (MB) RAM
● 1.5 gigabytes (GB) available space
● 1024x768 or higher resolution monitor

The following are the software s required for the project:

● Google Colab/Jupyter Notebook
● Anaconda - Spyder
● Pycharm Community(For HTML files)

4. EXPERIMENTAL INVESTIGATIONS

1. **Data Collection:** The employed data set contains customer reviews about the Unlocked Mobile phones, which are taken from Amazon.com. The dataset is taken from "http://www.kaggle.com". This data is used to predict that what rating a customer would probably give, based on the sentiments hidden in their reviews.

2. **Preprocessing:** Preprocessing can increase the performance and accuracy of a classifier. The dataset is comprised of exclusively English reviews. Reviews contain information which are not clearly expressive and needed to be removed.

   a. **Remove Punctuation:** All punctuations are removed as they are not needed for analysis.

   b. **Remove Stop words:** Some words are called stop words. These pronouns, prepositions, conjunctions have no specific meanings. "I", "a", "an"," is", "are", "as", "at", "from", "in", "this", "on", "or", "to", "was", "what", "will", "with" etc. are examples of stop words, so these types of words have been removed. All text is converted into lower case.

   c. **Stemming:** It converts word into its grammatical root form. Stemming technique converts words like "teach", "teacher", " teaching", " teaches" to root word teach. M.F. Porter stemming algorithm is used for this task. It minimizes the feature set and increases the accuracy of classification.

3. **Model Building & Training:** We initialize the model and then we add input layer, hidden layer and output layer. We configured the learning process, trained and tested the model, and optimized the model. By training the model we achieved an accuracy of 95.7%. The model is then saved.

4. **Predicting:** The model is then loaded and various inputs are given to predict the responses. The predicted responses tell us how accurate our model is.

5. **HTML for Web UI:** A HTML file is made to accept review from the user and then model is deployed to display whether the review is positive or negative on the screen.

6. **Flask file:** A flask file is made to coordinate between the HTML script and the model. Once the file is run, it assigns a local host on the internet browser and displays its HTML file

5. FLOWCHART
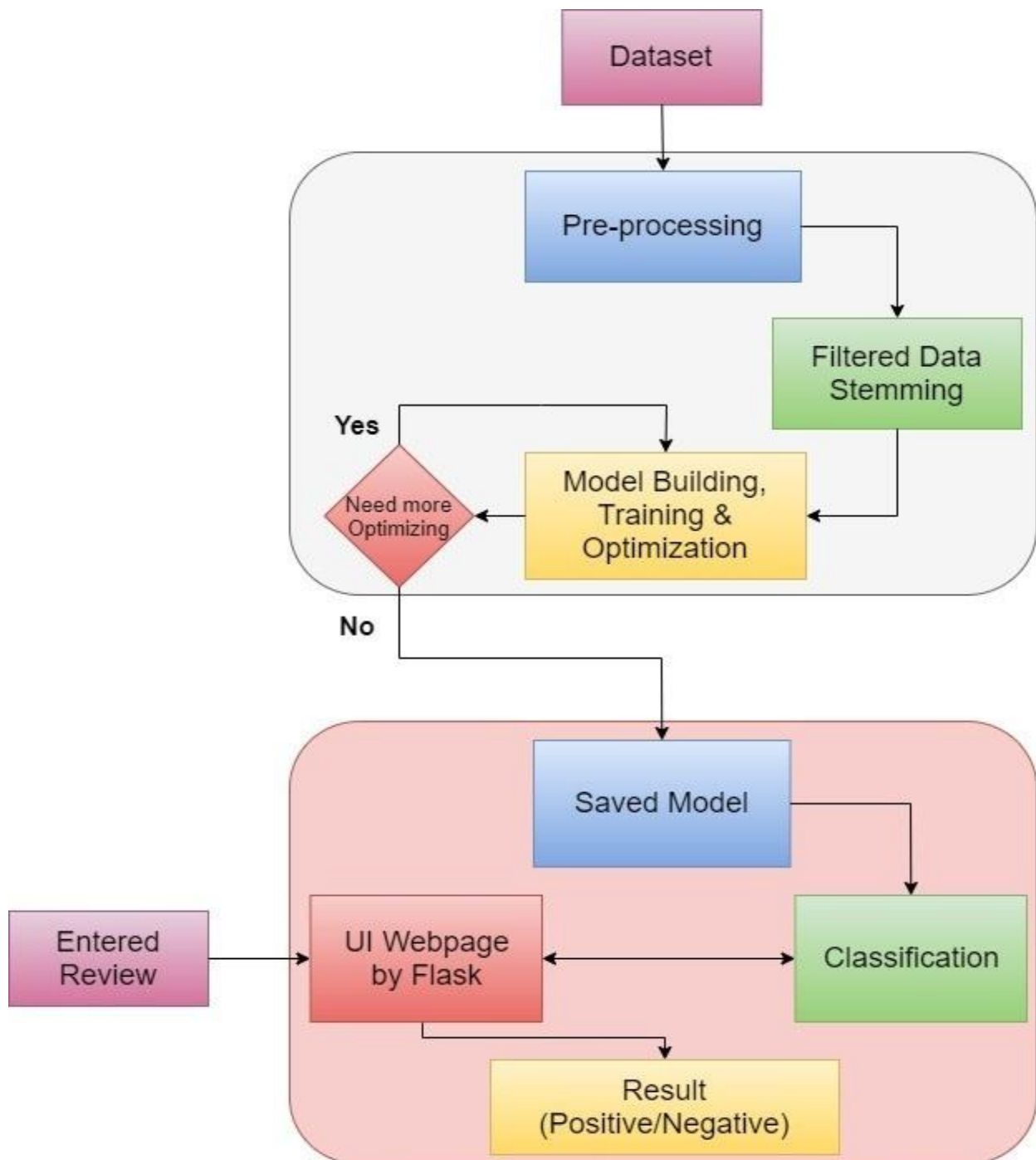
6. RESULT

The model and the flask file work very well together. We are able to distinguish a positive review from a negative one, which was our aim throughout the project. Our model is working perfectly and is predicting the correct sentiment from our reviews which was observed when we ran sample reviews through the model.
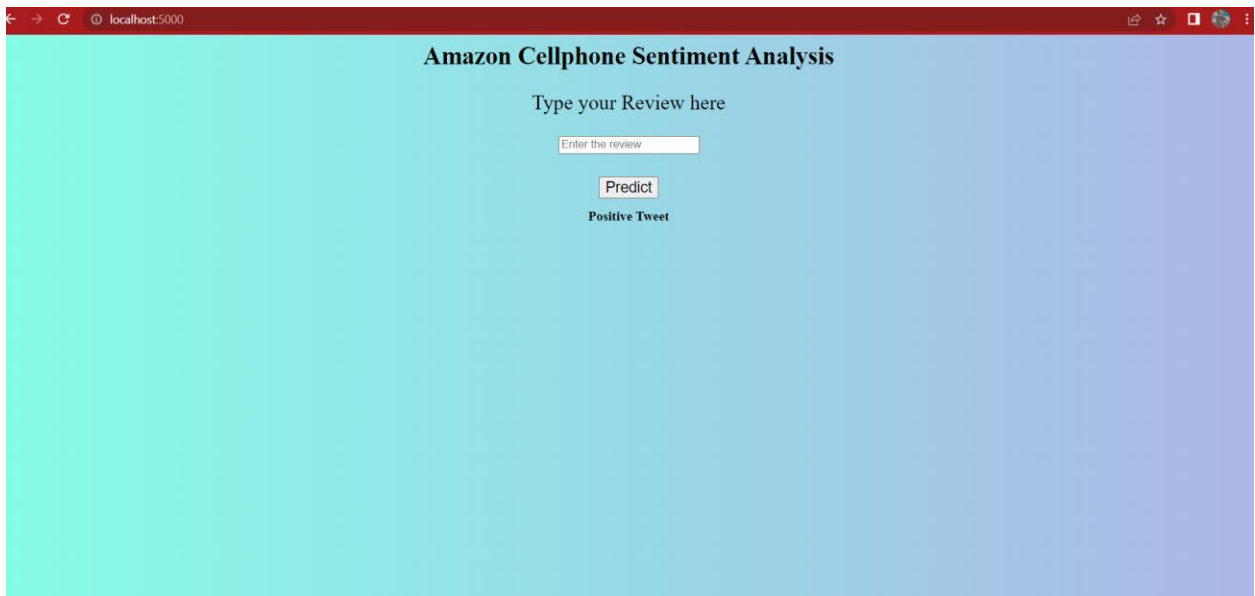
7. ADVANTAGES & DISADVANTAGES

7.1 ADVANTAGES :

➤ Increases confidence in new customers.

➤ Brings credibility to products and the company.

➤ Knowing which product works best.

➤ Helps a company to improve their products.

7.2 DISADVANTAGES :

➤ One negative review of a product or business can skew a potential customer's view of them.

➤ We need to keep reviews current and up to date. Otherwise they will seem out of date and irrelevant.

➤ Disgruntled customers have the freedom to say whatever they like. This could lead to malicious or damaging information being posted.

➤ Lack of touch or feel of products during online shopping is a drawback.

8. APPLICATIONS

1. Business
a. Consumers voice
b. Brand Reputation
c. Online advertising

2. Politic
a. Voting advise applications
b. Clarification of politicians' positions

3. Public Actions
a. Real-world events monitoring
b. Legal matters
c. Policy or government-regulation proposals
d. Intelligent transportation systems

Sentiment Analysis Examples

1. Brand Sentiment Analysis
Brand monitoring and reputation management is the most common use of sentiment analysis across different markets. It allows companies to track the perception of the brand by the customers, point out the specific details about the attitude, find patterns and trends and keep a close eye on the presentation by the influencers.

2. Social Media Monitoring
Sentiment analysis can be used to automate media monitoring process and the accompanying alert system, monitor mentions or reviews of the brand on different platforms and categorize urgency of mentions according to the relevant scoring.

3. Customer Support
Sentiment analysis can be used to give insight into customer's opinions regarding the product, intent Analysis for process automation and workflow management and customer prioritization.

9. CONCLUSION

This project was a great learning experience with interesting challenges. My passion for Data Science and Artificial Intelligence drove me to choose this topic and complete it in a good way. Among the different tasks involved in the system implementation data collection and pre-processing was the most time-consuming one. Sentiment Analysis is a challenging task.

The sentiment analysis is being implementing through deep learning techniques. Deep learning consists of numerous effective and popular models, these models are used to solve the variety of problems effectively. Different studies have been discussed in this review to provide a deep knowledge of the successful growing of deep learning applications in the field of sentiment analysis. Numerous problems have been resolved by having high accuracy of both fields of sentiment analysis and deep learning.

We have created a model which can predict the Amazon mobile phone reviews whether it is a positive or negative review. We also created a UI interface to co-ordinate between our model and the HTML web page via the python file.

10. FUTURE SCOPE

Future opinion-mining systems need broader and deeper common and commonsense knowledge bases. This will lead to a better understanding of natural language opinions and will more efficiently bridge the gap between multi-modal information and machine processable data. Blending scientific theories of emotion with the practical engineering goals of analyzing sentiments in natural language text will lead to more bio-inspired approaches to the design of intelligent opinion-mining systems capable of handling semantic knowledge, making analogies, learning new effective knowledge, and detecting, perceiving, and "feeling" emotions.

The major research scope areas in sentiment analysis are:

1. Spam Detection Sentiment Analysis
2. Sentiment Analysis on short Sentence like abbreviations
3. Improving sentiment word identification algorithm
4. Developing fully automatic analyzing tool
5. Effective Analysis of policy opinionated content
6. Successful handling of bi polar sentiments
7. Generation of highly content lexicon database

11. BIBLIOGRAPHY

□ Kaggle Dataset - https://www.kaggle.com/grikomsn/amazon-cell-phones-reviews

□ International Journal of Computer Applications, Volume 125 – No.3, September 2015
   ISSN : 0975 – 8887

□ International Journal of Advance Research in Science and Engineering, Volume 6 - Issue 7,
   July 2017
   ISSN(O) : 2319 - 8354
   ISSN(P) : 2319 - 8346

□ https://www.lexalytics.com/applications

□ Journal of Computer Science, Volume 12 - No.3, March
   2016 153 - 168

□ Prominent Feature Extraction for Sentiment Analysis by Basant Agarwal and Namita Mittal,
   Page 89-92, January 2016
   ISSN : 2509-5706

□ https://www.geeksforgeeks.org/html-tutorials/

□ https://www.w3schools.com/

□ https://smartbridge.teachable.com/courses/enrolled/846395

□ Fourth International Conference on Recent Trends in Computer Science & Engineering
   Sentiment Analysis:A Comparative Study On Different Approaches - Page 44-49 , 2016

□ https://google.com

APPENDIX :

a. amazon review main.ipynb (colab notebook)

https://colab.research.google.com/drive/1lRkC2AEVuCXkY3w1L-GCb6LkL4cbcAsb?usp=sharing

b. app.py (Flask code)

```python
from tensorflow.keras.models import load_model
import pickle
import tensorflow as tf
#graph = tf.get_default_graph()
with open(r'count_vec.pkl','rb') as file:
   cv=pickle.load(file)
cla = load_model('phone.h5')
#cla.compile(optimizer='adam',loss='binary_crossentropy')
app = Flask(__name__)
@app.route('/')
def index():
   return render_template('index.html')
@app.route('/tpredict')
@app.route('/', methods = ['GET','POST'])
def page2():
   if request.method == 'GET':
      img_url = url_for('static',filename = 'style/3.jpg')
      return render_template('index.html',url=img_url)
```

```python
if request.method == 'POST':

    topic = request.form['tweet']
    print("Hey " +topic)
    topic=cv.transform([topic])
    print("\n"+str(topic.shape)+"\n")

    y_pred = cla.predict(topic)
    print("pred is "+str(y_pred))
    if(y_pred > 0.5):
        img_url = url_for('static',filename = 'style/1.jpg')
        topic = "Positive Tweet"
    else:
        img_url = url_for('static',filename = 'style/2.jpg')
        topic = "Negative Tweet"

    return render_template('index.html',ypred = topic)




if __name__ == '__main__':
    app.run(host = 'localhost', debug = False , threaded = False)
```

c. index.html

```html
<html>
<title >Amazon Cellphone Sentimental Review Analysis</title>
<head>
<link rel="stylesheet" type="text/css" href="{{ url_for('static', filename =
'style/index12.css') }}">
<style>

body {
background: #acb6e5;  /* fallback for old browsers */
background: -webkit-linear-gradient(to right, #86fde8, #acb6e5);  /* Chrome 10-25,
Safari 5.1-6 */
background: linear-gradient(to right, #86fde8, #acb6e5); /* W3C, IE 10+/ Edge,
Firefox 16+, Chrome 26+, Opera 12+, Safari 7+ */

}
</style>
</head>
<body  class="body1">
<form  method="POST">
  <div class="container">
    <h1 align = 'center' style="color:black; font-size:30px" >Amazon Cellphone
Sentiment Analysis</h1>

    <p align = 'center'style="font-size:25px">Type your Review here</p>
    <form action='/tpredict' method="post">
        <p align = "center">
        <input type="text"    align = "Center" placeholder="Enter the review"
name="tweet" id="rcorners1" required>
```

```html
</p>
<table align='center' style='padding:0.5px; border-spacing: 10px; font-size:20px'>
<tr>
<td colspan="2" align="center" ><button class="button5" type="submit" name="predict" align='center' value="Predict" onclick="window.location.href = 'D:\ArtificiaI Intelligence with Flask\Amazon phone review Main\New\templates\index.html'" ><font size="4">Predict</font></button></td>
</tr>
<tr>
<td colspan="2" align = "center"><b>{{ypred}}</b></td>
</tr>
</table>
</form>
<!--<img src="/static/happy.png" alt="Positive" class="tab3">-->
<p>
{% if ypred == "Positive Review" %}
<p>Positive</p>
{% else %}
{% if ypred == "Negative Review" %}

{% endif %}
{% endif %}
</p>

</div>

</form>
</html>
```