Contents

Analysis Of Amazon Cell Phone Reviews Using IBM Watson Services

By: Lee Chen Fei

Introduction

According to the Cambridge Dictionary, Review is defined as “to think or talk about something again, in order to make changes to it or to make a decision about it”. For every existing product and project, Reviews are essential in their improvement overtime, as when taken seriously it may help the ones developing those products, projects or provide services to identify the flaws of the services, which are not detected by the people involved. Review can be in many shapes and forms, ranging from just giving feedbacks on the product directly to the developers, to addressing the issue via third party helps such as surveys, video proofs or feedbacks, and discussing the feedback as a community to further discuss the root problem of the product. Overall, the objectives are the same, which is to let them know the current state of the services provided, whether positive or negative, so that they can learn the weakness and adjustments accordingly.

There are many ways the people who provided such services can know whether most reviews from tons, if not millions of reviews to determine whether the reviews are mostly positive or negative. One way is by utilizing Machine Learning which help the Artificial Intelligence (AI) to learn from the reviews and learn whether the reviews are positive or negative. Through AI, one can maintain the task of identifying the density of positive or negative reviews in a short time compared to when a human done the exact task.

From this report, the AI machine learning capabilities will be displayed and used to complete a task in analysing the Amazon Cell Phone Reviews.

Problem Statement

Mobile phones have changed the way customers buy things online by putting all the information at the receiver’s fingertips. Customers reviews and ratings have become an important part of the decision-making process for potential buyers or customers. eCommerce players' review and rating platforms create a transparent system for consumers to make decisions and feel confident about them.

It is, however, difficult to read all the feedback for a certain item, particularly for popular goods with several comments. Each item could contain tens of thousands of reviews which takes a lot of time for a normal human to read through, but this requirement is needed so that the developers or creators could learn the reviews more quickly.

Solution to the Problem Statements

To work on the project, below are some applications used in solving the problem statement:

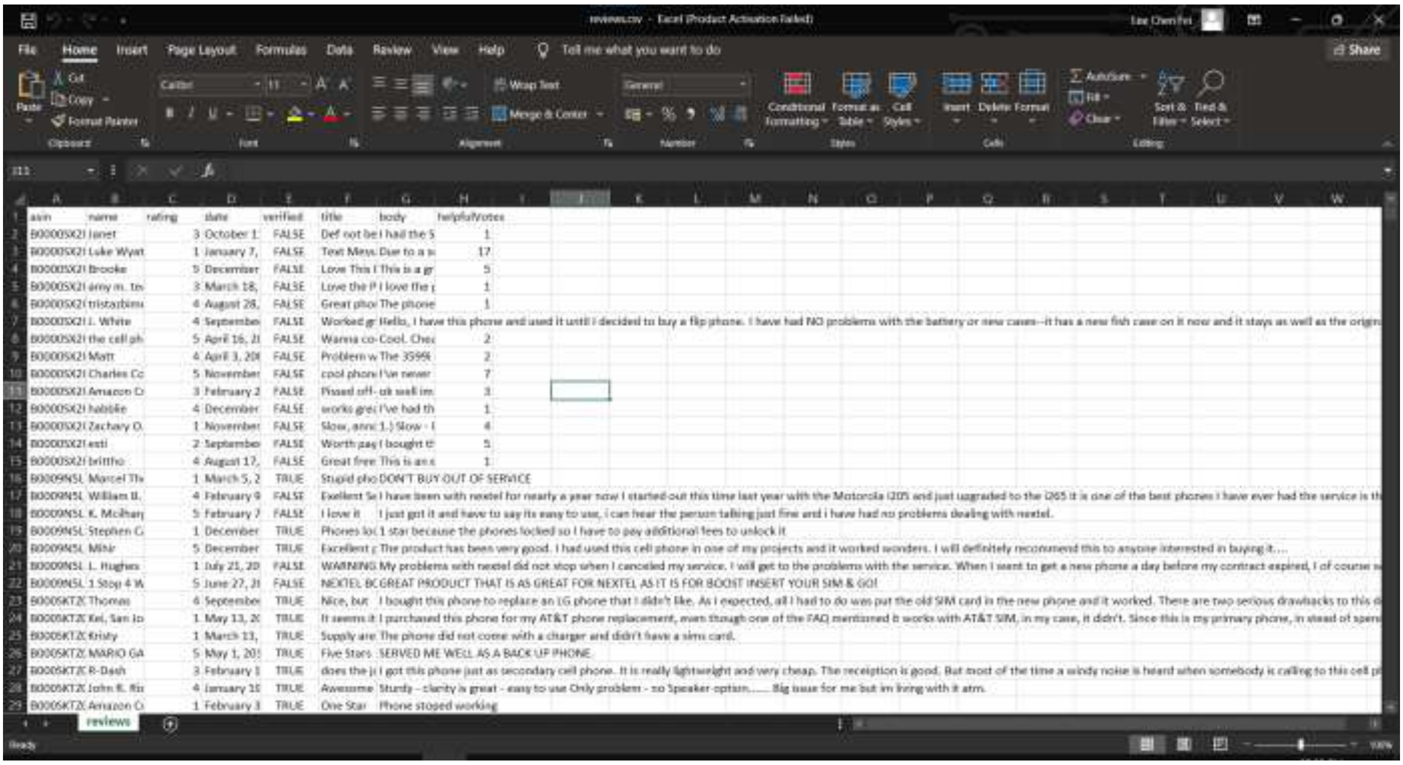
- The Amazon Cell Phone Review Datasets (Main Dataset which is being working on)
- Jupyter Notebook file (Model Creation)
- Python Flask App file (Flask App Creation)
- Html file (App Template Creation)

For the html file, python file and the Jupyter Notebook file, the Visual Studio Code (VSC) software will used for programming and coding purposes.

1. The Amazon Cell Phone Review Datasets

A dataset is a collection of information. It corresponds to one or more database tables in the case of tabular data, where each column of a table represents a specific variable and each row represents a specific record of the data set in question. For each column of the dataset, the data set provides values for each of the variables, such as an object's height and weight.

The screenshot above shows the dataset which is used as a "practise dummy" for the model to train and differentiate positive and negative reviews. From the dataset, there are a total of 8 columns in it, namely Asin, name, rating, date, verified, title, body and helpful notes.



reviews.csv - Excel (Product Activation Failed)
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Help
Tell me what you want to do
Share
reviews
Ready

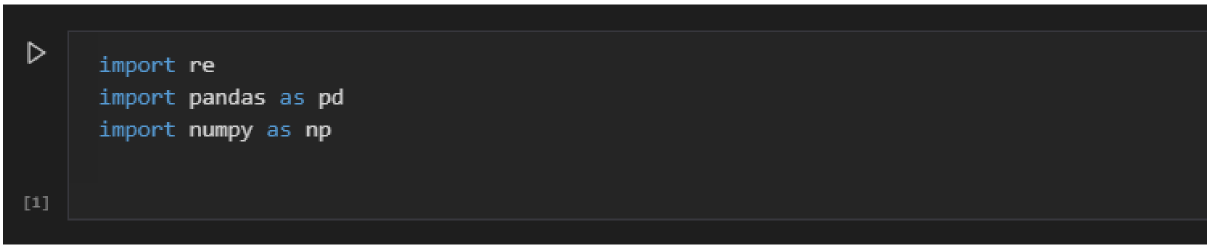
	A	B	C	D	E	F	G	H
1	asin	name	rating	date	verified	title	body	helpfulVotes
2	B0000SX2I	Janet	3	October 1	FALSE	Def not be	I had the S	1
3	B0000SX2I	Luke Wyat	1	January 7,	FALSE	Text Mess	Due to a s	17
4	B0000SX2I	Brooke	5	December	FALSE	Love This I	This is a gr	5
5	B0000SX2I	amy m. te	3	March 18,	FALSE	Love the P	I love the p	1
6	B0000SX2I	tristazbim	4	August 28,	FALSE	Great pho	The phone	1
7	B0000SX2I	J. White	4	Septembe	FALSE	Worked gr	Hello, I have this phone and used it until I decided to buy a flip phone. I have had NO problems with the battery or new cases--it has a new fish case on it now and it stays as well as the origi	
8	B0000SX2I	the cell ph	5	April 16, 2	FALSE	Wanna co-	Cool. Chec	2
9	B0000SX2I	Matt	4	April 3, 20	FALSE	Problem w	The 3599i	2
10	B0000SX2I	Charles Co	5	November	FALSE	cool phone	I've never	7
11	B0000SX2I	Amazon C	3	February 2	FALSE	Pissed off-	ok well im	3
12	B0000SX2I	habibie	4	December	FALSE	works grea	I've had th	1
13	B0000SX2I	Zachary O.	1	November	FALSE	Slow, anno	1.) Slow - I	4
14	B0000SX2I	esti	2	Septembe	FALSE	Worth pay	I bought th	5
15	B0000SX2I	britttho	4	August 17,	FALSE	Great free	This is an e	1
16	B0009N5L	Marcel Th	1	March 5, 2	TRUE	Stupid pho	DON'T BUY OUT OF SERVICE	
17	B0009N5L	William B.	4	February 9	FALSE	Exellent Se	I have been with nextel for nearly a year now I started out this time last year with the Motorola i205 and just upgraded to the i265 it is one of the best phones I have ever had the service is th	
18	B0009N5L	K. Mcilhar	5	February 7	FALSE	I love it	I just got it and have to say its easy to use, i can hear the person talking just fine and i have had no problems dealing with nextel.	
19	B0009N5L	Stephen C	1	December	TRUE	Phones loc	1 star because the phones locked so I have to pay additional fees to unlock it	
20	B0009N5L	Mihir	5	December	TRUE	Excellent p	The product has been very good. I had used this cell phone in one of my projects and it worked wonders. I will definitely recommend this to anyone interested in buying it....	
21	B0009N5L	L. Hughes	1	July 21, 20	FALSE	WARNING	My problems with nextel did not stop when I canceled my service. I will get to the problems with the service. When I went to get a new phone a day before my contract expired, I of course w	
22	B0009N5L	1 Stop 4 W	5	June 27, 2	FALSE	NEXTEL BC	GREAT PRODUCT THAT IS AS GREAT FOR NEXTEL AS IT IS FOR BOOST INSERT YOUR SIM & GO!	
23	B000SKTZ	Thomas	4	Septembe	TRUE	Nice, but	I bought this phone to replace an LG phone that I didn't like. As I expected, all I had to do was put the old SIM card in the new phone and it worked. There are two serious drawbacks to this d	
24	B000SKTZ	Kel, San Jo	1	May 13, 20	TRUE	It seems it	I purchased this phone for my AT&T phone replacement, even though one of the FAQ mentioned it works with AT&T SIM, in my case, it didn't. Since this is my primary phone, in stead of spen	
25	B000SKTZ	Kristy	1	March 13,	TRUE	Supply are	The phone did not come with a charger and didn't have a sims card.	
26	B000SKTZ	MARIO GA	5	May 1, 20	TRUE	Five Stars	SERVED ME WELL AS A BACK UP PHONE	
27	B000SKTZ	R-Dash	3	February 1	TRUE	does the ju	I got this phone just as secondary cell phone. It is really lightweight and very cheap. The receiption is good. But most of the time a windy noise is heard when somebody is calling to this cell p	
28	B000SKTZ	John R. Ri	4	January 1	TRUE	Awesome	Sturdy - clarity is great - easy to use Only problem - no Speaker option...... Big issue for me but im living with it atm.	
29	B000SKTZ	Amazon C	1	February 3	TRUE	One Star	Phone stoped working	

2. Jupyter Notebook file

The Jupyter Notebook file is used for creating the model, training the model and testing the model. The reason why the Jupyter Notebook file is used for such purpose is that it allows for analysis the code step by step instead of executing all at once. This in turn makes it easier to monitor each line of code and document the thought process while developing the analysis process.

The following segment shows the screenshots of each line (or cells) of the Jupyter Notebook used for model training and testing:

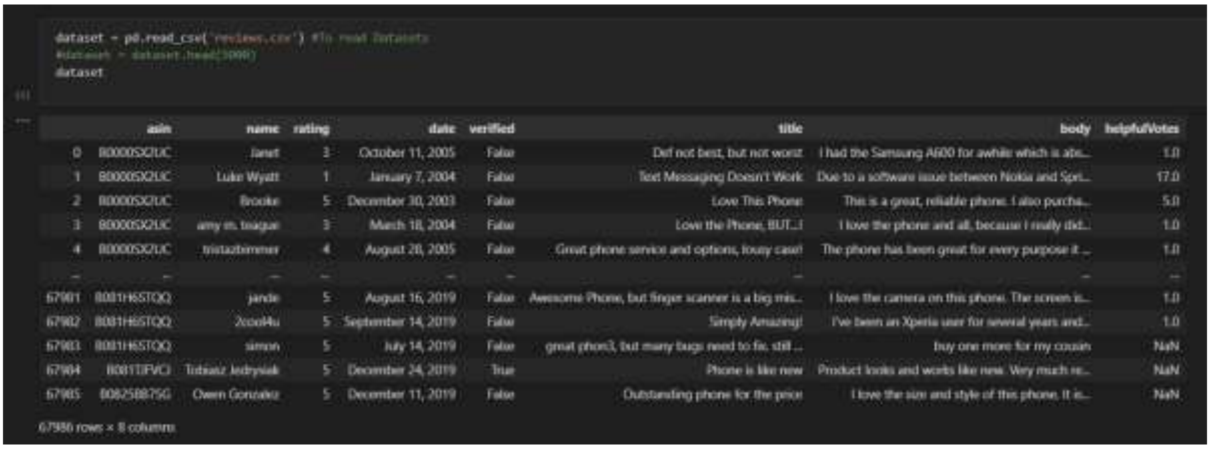
a. Import Libraries



```
import re
import pandas as pd
import numpy as np
```

[1]

b. Reading the Dataset

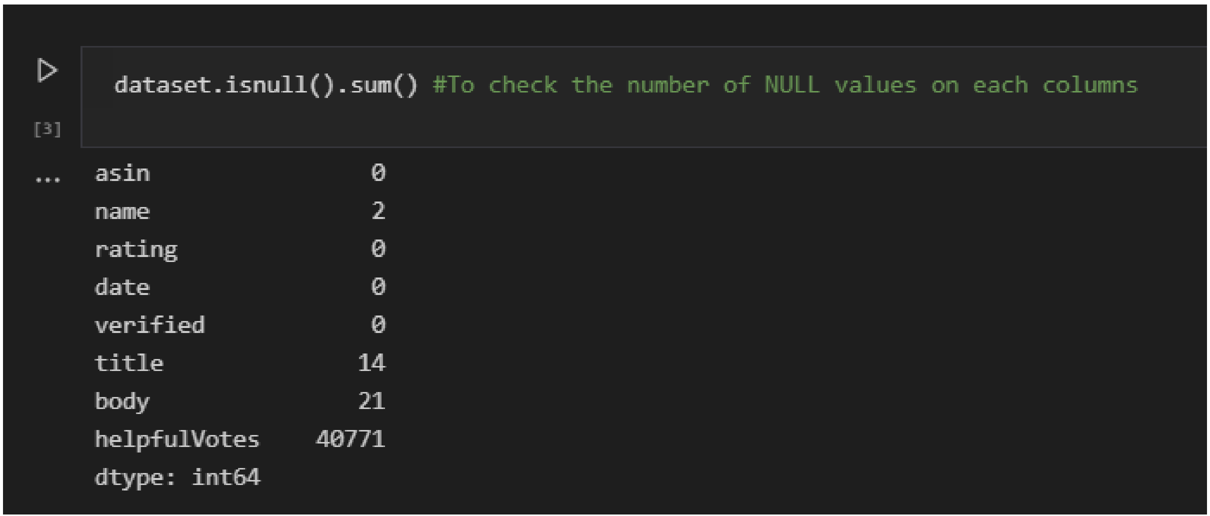


```
dataset = pd.read_csv('reviews.csv') #To read Datasets
#dataset = dataset.head(3000)
dataset
```

	asin	name	rating	date	verified	title	body	helpfulVotes
0	B0000SX2UC	Janet	3	October 11, 2005	False	Def not best, but not worst	I had the Samsung A600 for awhile which is abs...	1.0
1	B0000SX2UC	Luke Wyatt	1	January 7, 2004	False	Text Messaging Doesn't Work	Due to a software issue between Nokia and Spri...	17.0
2	B0000SX2UC	Brooke	5	December 30, 2003	False	Love This Phone	This is a great, reliable phone. I also purcha...	5.0
3	B0000SX2UC	amy m. teague	3	March 18, 2004	False	Love the Phone, BUT...!	I love the phone and all, because I really did...	1.0
4	B0000SX2UC	tristazbimmer	4	August 28, 2005	False	Great phone service and options, lousy case!	The phone has been great for every purpose it ...	1.0
...	...	...	...	...	...	...	...	...
67981	B081H6STQQ	jande	5	August 16, 2019	False	Awesome Phone, but finger scanner is a big mis...	I love the camera on this phone. The screen is...	1.0
67982	B081H6STQQ	2cool4u	5	September 14, 2019	False	Simply Amazing!	I've been an Xperia user for several years and...	1.0
67983	B081H6STQQ	simon	5	July 14, 2019	False	great phon3, but many bugs need to fix. still ...	buy one more for my cousin	NaN
67984	B081TJFVCJ	Tobiasz Jedrysiak	5	December 24, 2019	True	Phone is like new	Product looks and works like new. Very much re...	NaN
67985	B0825BB7SG	Owen Gonzalez	5	December 11, 2019	False	Outstanding phone for the price	I love the size and style of this phone. It is...	NaN

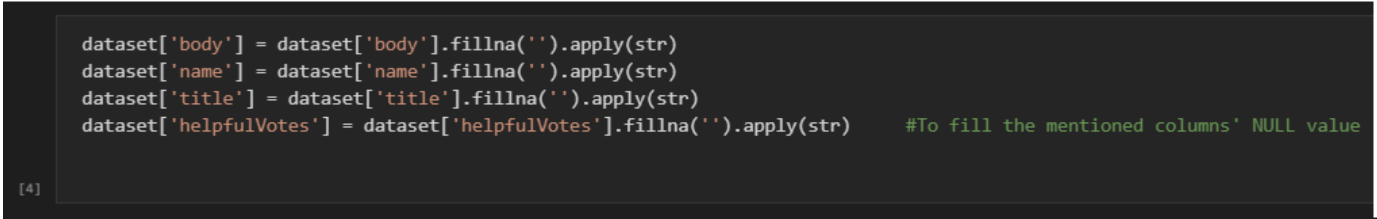
67986 rows × 8 columns

c. Checking for Null values and filling it accordingly

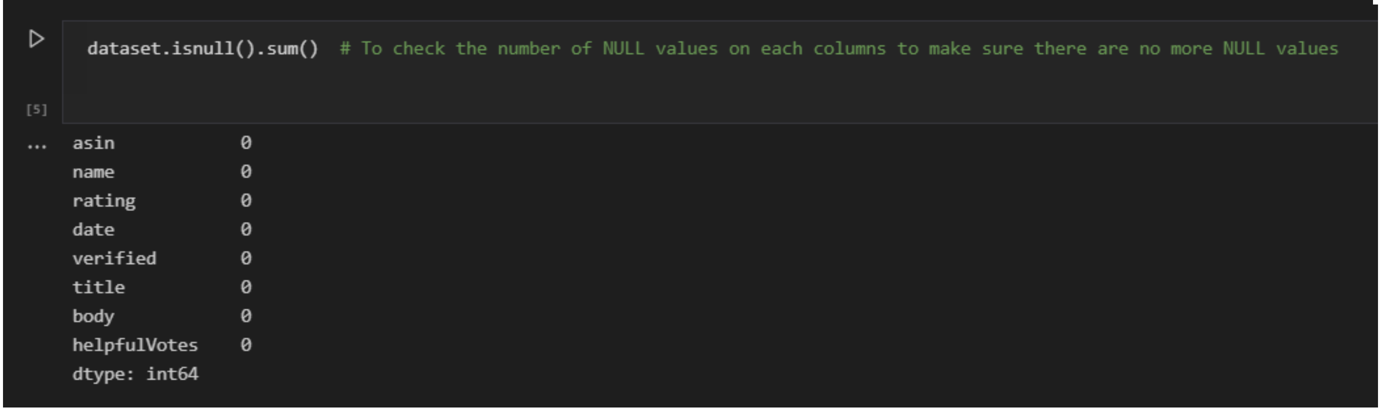


```
dataset.isnull().sum() #To check the number of NULL values on each columns
```

```
asin               0
name               2
rating             0
date               0
verified           0
title             14
body              21
helpfulVotes   40771
dtype: int64
```



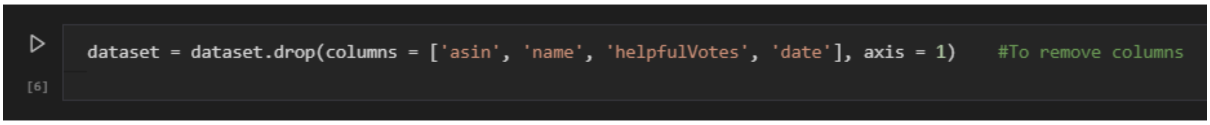
```
dataset['body'] = dataset['body'].fillna('').apply(str)
dataset['name'] = dataset['name'].fillna('').apply(str)
dataset['title'] = dataset['title'].fillna('').apply(str)
dataset['helpfulVotes'] = dataset['helpfulVotes'].fillna('').apply(str)     #To fill the mentioned columns' NULL value
```



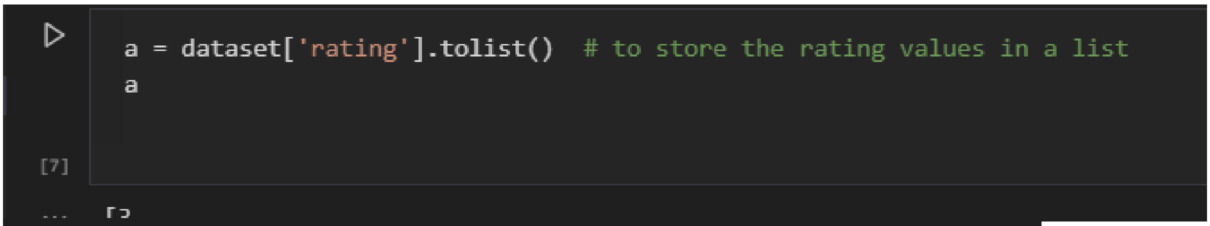
```
dataset.isnull().sum()  # To check the number of NULL values on each columns to make sure there are no more NULL values
```

```
asin            0
name            0
rating          0
date            0
verified        0
title           0
body            0
helpfulVotes    0
dtype: int64
```

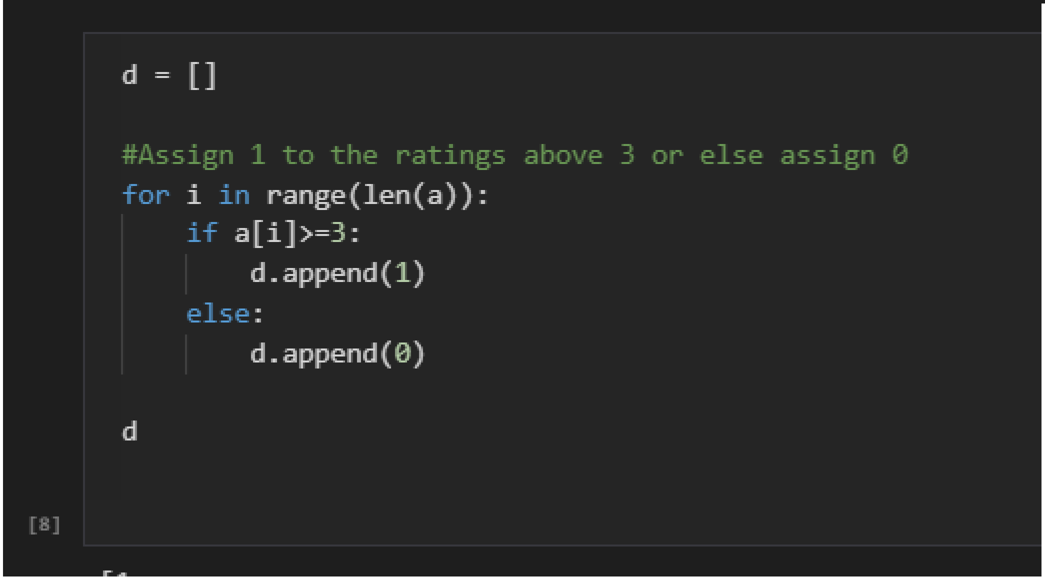
d. Column Manipulations (Dropping Column(s) / Concatenate Column(s))



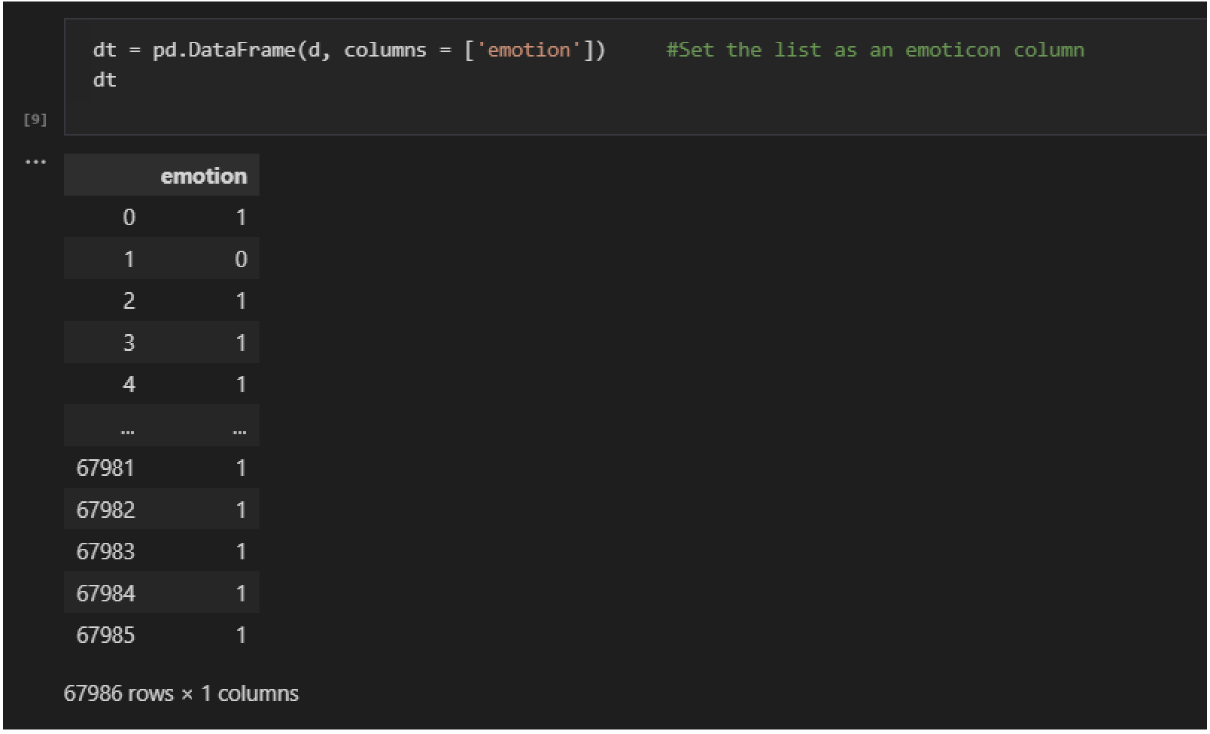
```
[6]
dataset = dataset.drop(columns = ['asin', 'name', 'helpfulVotes', 'date'], axis = 1)    #To remove columns
```



```
[7]
a = dataset['rating'].tolist()  # to store the rating values in a list
a
```

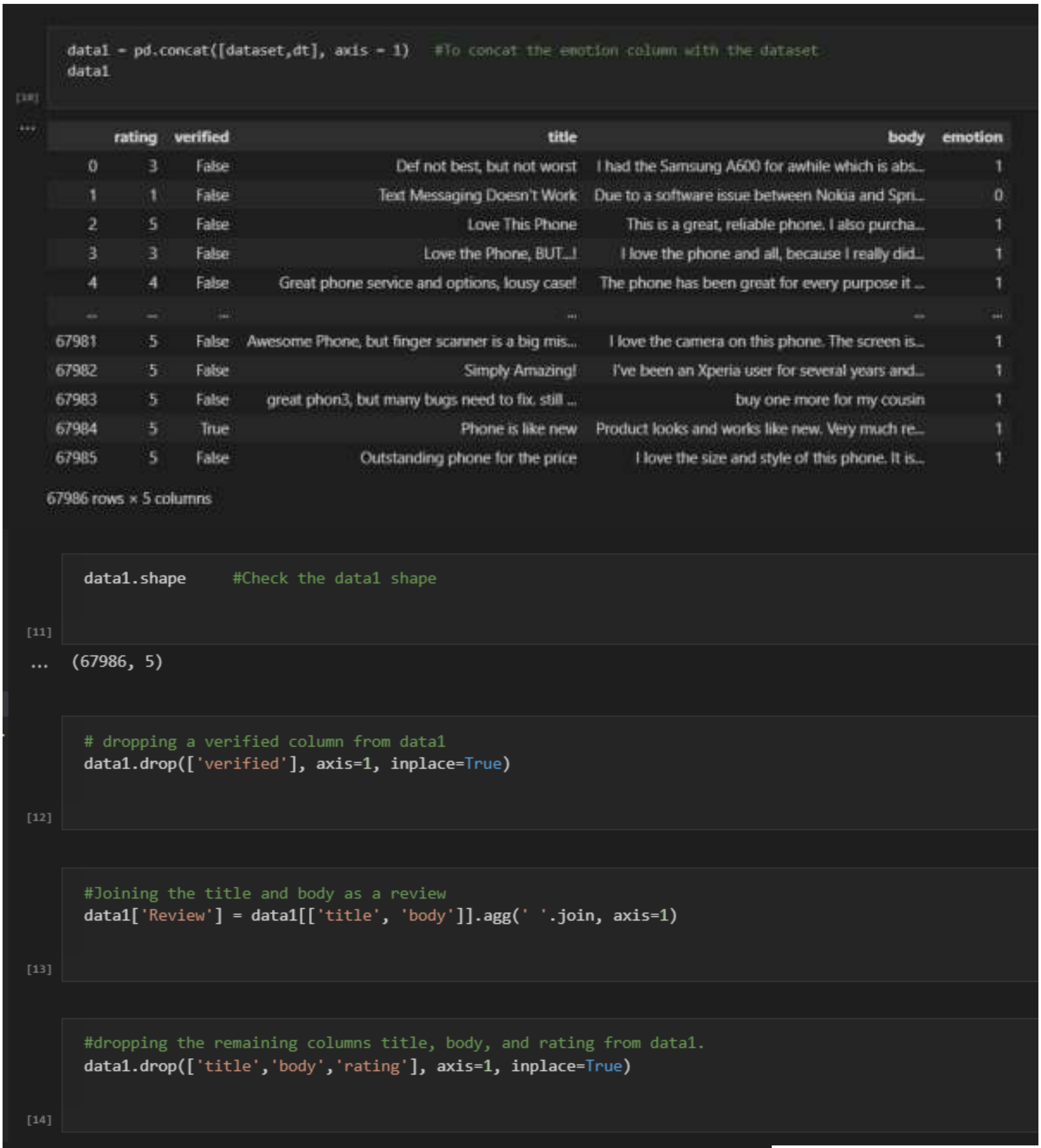


```
[8]
d = []

#Assign 1 to the ratings above 3 or else assign 0
for i in range(len(a)):
    if a[i]>=3:
        d.append(1)
    else:
        d.append(0)

d
```



```
[9]
dt = pd.DataFrame(d, columns = ['emotion'])     #Set the list as an emoticon column
dt
```

```
       emotion
    0        1
    1        0
    2        1
    3        1
    4        1
  ...      ...
67981        1
67982        1
67983        1
67984        1
67985        1

67986 rows × 1 columns
```



[10]

```
data1 = pd.concat([dataset,dt], axis = 1)   #To concat the emotion column with the dataset
data1
```

	rating	verified	title	body	emotion
0	3	False	Def not best, but not worst	I had the Samsung A600 for awhile which is abs...	1
1	1	False	Text Messaging Doesn't Work	Due to a software issue between Nokia and Spri...	0
2	5	False	Love This Phone	This is a great, reliable phone. I also purcha...	1
3	3	False	Love the Phone, BUT...!	I love the phone and all, because I really did...	1
4	4	False	Great phone service and options, lousy case!	The phone has been great for every purpose it ...	1
...	...	...	...	...	...
67981	5	False	Awesome Phone, but finger scanner is a big mis...	I love the camera on this phone. The screen is...	1
67982	5	False	Simply Amazing!	I've been an Xperia user for several years and...	1
67983	5	False	great phon3, but many bugs need to fix. still ...	buy one more for my cousin	1
67984	5	True	Phone is like new	Product looks and works like new. Very much re...	1
67985	5	False	Outstanding phone for the price	I love the size and style of this phone. It is...	1

67986 rows × 5 columns

[11]

```
data1.shape     #Check the data1 shape
```

```
(67986, 5)
```

[12]

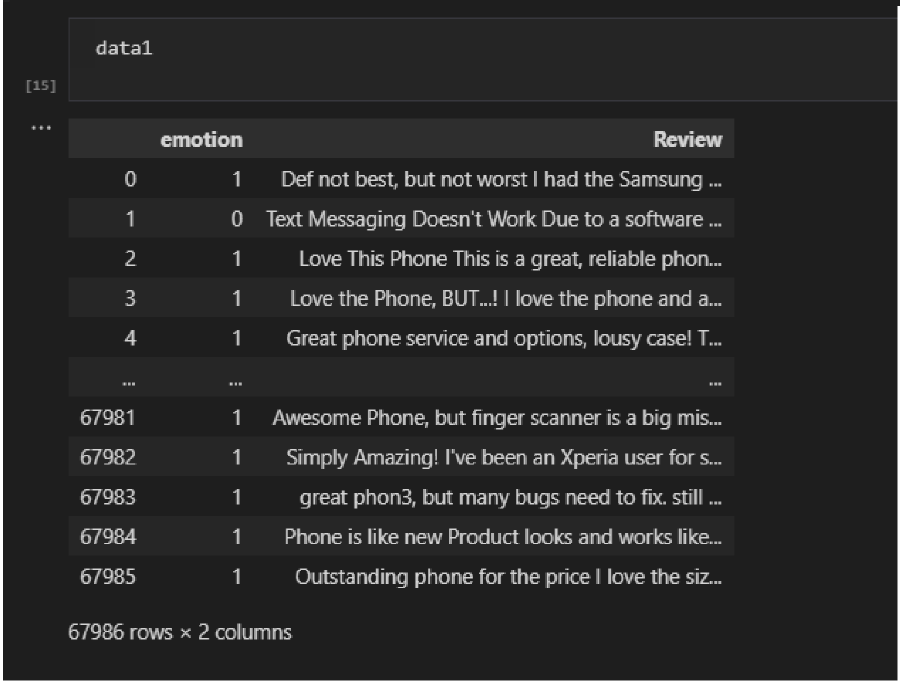
```
# dropping a verified column from data1
data1.drop(['verified'], axis=1, inplace=True)
```

[13]

```
#Joining the title and body as a review
data1['Review'] = data1[['title', 'body']].agg(' '.join, axis=1)
```

[14]

```
#dropping the remaining columns title, body, and rating from data1.
data1.drop(['title','body','rating'], axis=1, inplace=True)
```



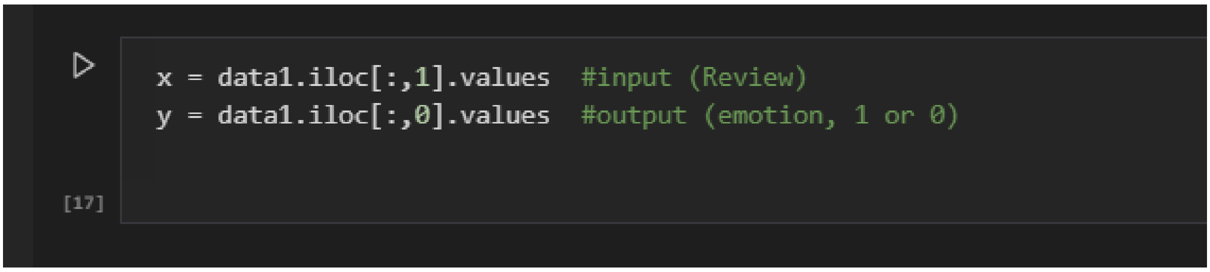
[15]

```
data1
```

	emotion	Review
0	1	Def not best, but not worst I had the Samsung ...
1	0	Text Messaging Doesn't Work Due to a software ...
2	1	Love This Phone This is a great, reliable phon...
3	1	Love the Phone, BUT...! I love the phone and a...
4	1	Great phone service and options, lousy case! T...
...	...	...
67981	1	Awesome Phone, but finger scanner is a big mis...
67982	1	Simply Amazing! I've been an Xperia user for s...
67983	1	great phon3, but many bugs need to fix. still ...
67984	1	Phone is like new Product looks and works like...
67985	1	Outstanding phone for the price I love the siz...

67986 rows × 2 columns

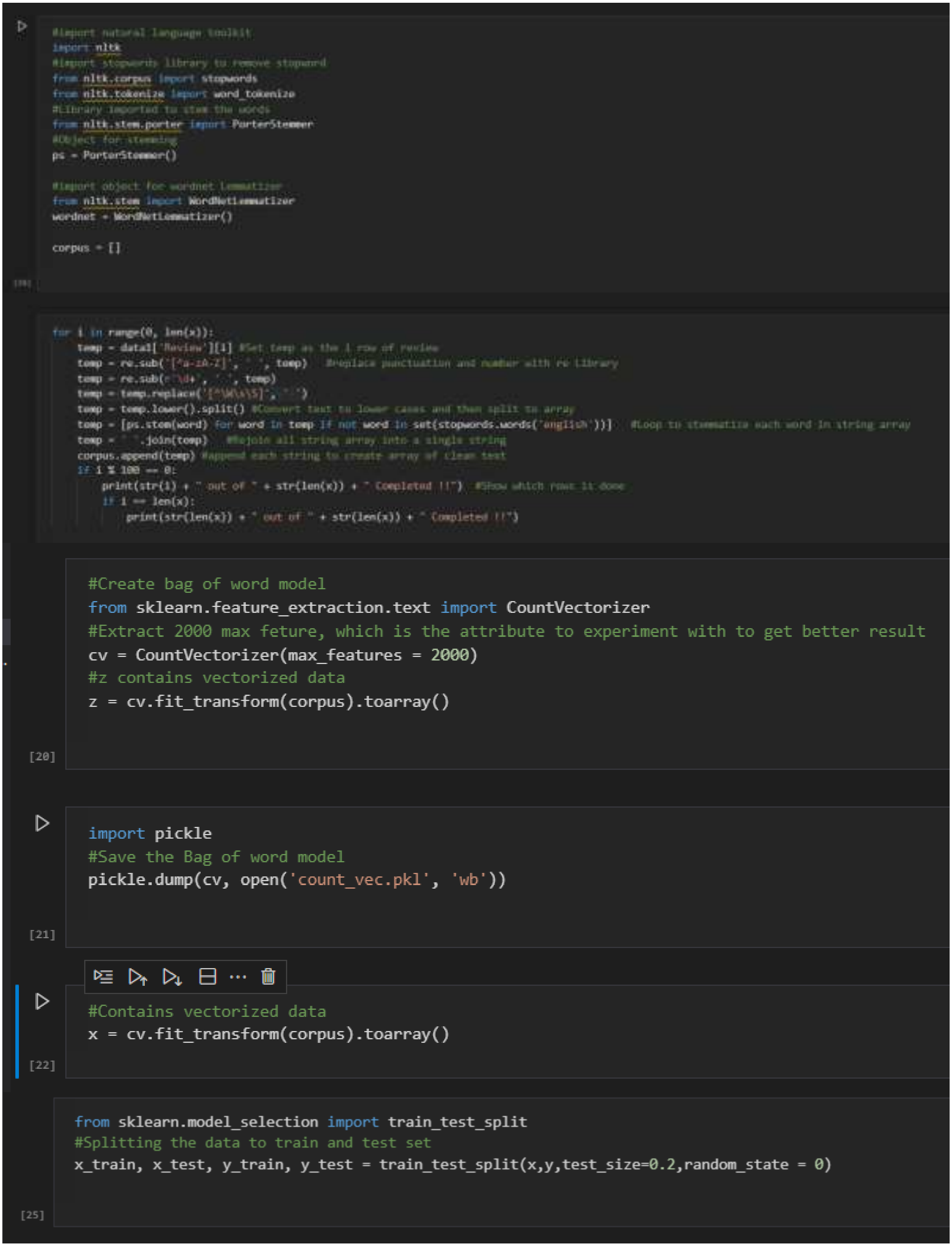
e. Data separation into input (x) and output (y)



```
x = data1.iloc[:,1].values  #input (Review)
y = data1.iloc[:,0].values  #output (emotion, 1 or 0)
```

[17]

f. Text Preprocessing/ Cleaning



```
#import natural language toolkit
import nltk
#import stopwords library to remove stopword
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
#Library imported to stem the words
from nltk.stem.porter import PorterStemmer
#Object for stemming
ps = PorterStemmer()

#import object for wordnet Lemmatizer
from nltk.stem import WordNetLemmatizer
wordnet = WordNetLemmatizer()

corpus = []
```

[19]

```
for i in range(0, len(x)):
    temp = data1['Review'][i] #Set temp as the i row of review
    temp = re.sub('[^a-zA-Z]', ' ', temp)   #replace punctuation and number with re library
    temp = re.sub(r'\d+', ' ', temp)
    temp = temp.replace('[^\W\s\S]', '')
    temp = temp.lower().split() #Convert text to lower cases and then split to array
    temp = [ps.stem(word) for word in temp if not word in set(stopwords.words('english'))]   #Loop to stemmatize each word in string array
    temp = ' '.join(temp)   #Rejoin all string array into a single string
    corpus.append(temp) #append each string to create array of clean text
    if i % 100 == 0:
        print(str(i) + " out of " + str(len(x)) + " Completed !!")  #Show which rows it done
        if i == len(x):
            print(str(len(x)) + " out of " + str(len(x)) + " Completed !!")
```

```
#Create bag of word model
from sklearn.feature_extraction.text import CountVectorizer
#Extract 2000 max feture, which is the attribute to experiment with to get better result
cv = CountVectorizer(max_features = 2000)
#z contains vectorized data
z = cv.fit_transform(corpus).toarray()
```

[20]

```
import pickle
#Save the Bag of word model
pickle.dump(cv, open('count_vec.pkl', 'wb'))
```

[21]

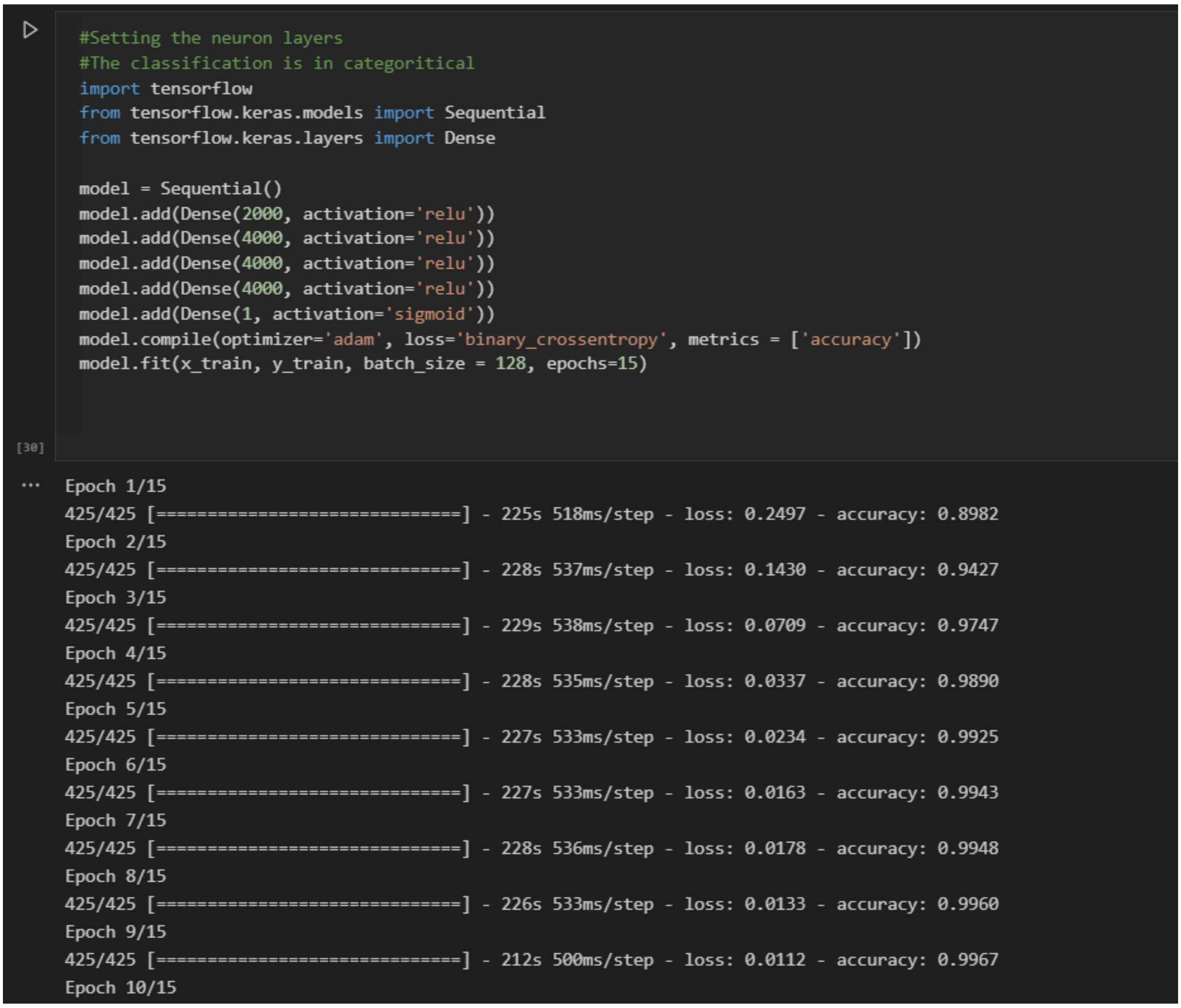
```
#Contains vectorized data
x = cv.fit_transform(corpus).toarray()
```

[22]

```
from sklearn.model_selection import train_test_split
#Splitting the data to train and test set
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.2,random_state = 0)
```

[25]

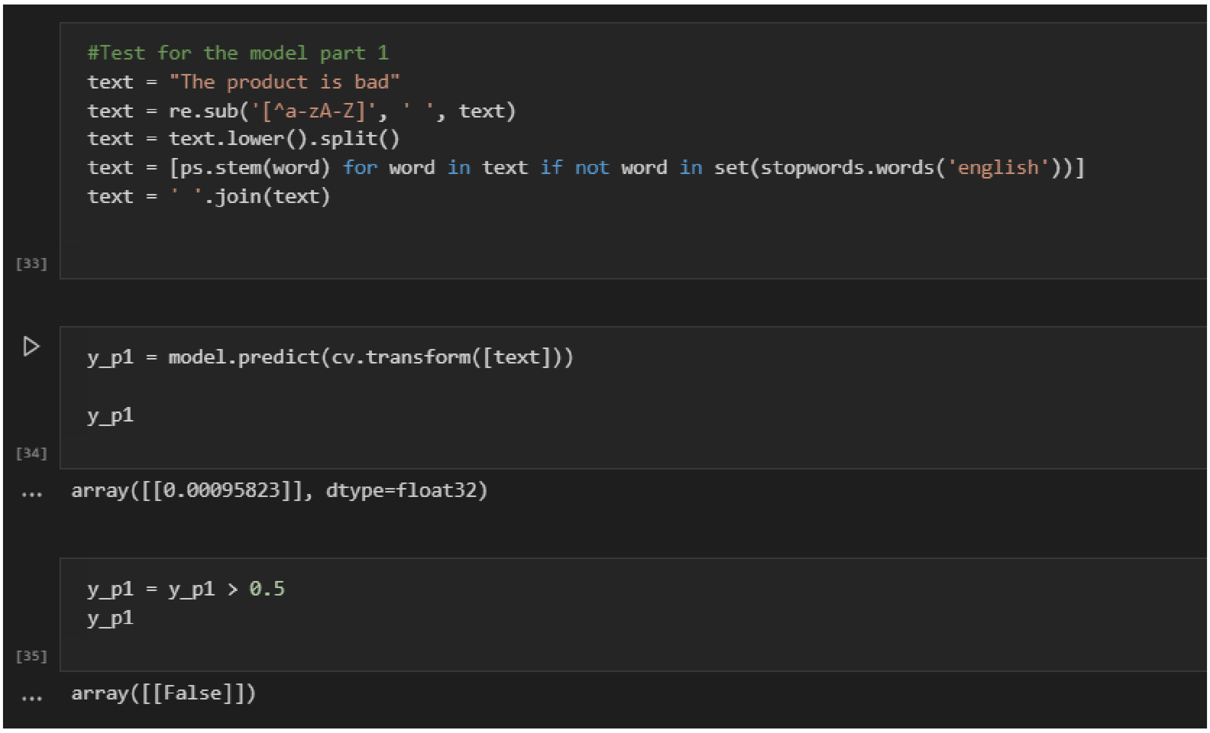
g. Model Building



```
#Setting the neuron layers
#The classification is in categoritical
import tensorflow
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense

model = Sequential()
model.add(Dense(2000, activation='relu'))
model.add(Dense(4000, activation='relu'))
model.add(Dense(4000, activation='relu'))
model.add(Dense(4000, activation='relu'))
model.add(Dense(1, activation='sigmoid'))
model.compile(optimizer='adam', loss='binary_crossentropy', metrics = ['accuracy'])
model.fit(x_train, y_train, batch_size = 128, epochs=15)
```

```
Epoch 1/15
425/425 [==============================] - 225s 518ms/step - loss: 0.2497 - accuracy: 0.8982
Epoch 2/15
425/425 [==============================] - 228s 537ms/step - loss: 0.1430 - accuracy: 0.9427
Epoch 3/15
425/425 [==============================] - 229s 538ms/step - loss: 0.0709 - accuracy: 0.9747
Epoch 4/15
425/425 [==============================] - 228s 535ms/step - loss: 0.0337 - accuracy: 0.9890
Epoch 5/15
425/425 [==============================] - 227s 533ms/step - loss: 0.0234 - accuracy: 0.9925
Epoch 6/15
425/425 [==============================] - 227s 533ms/step - loss: 0.0163 - accuracy: 0.9943
Epoch 7/15
425/425 [==============================] - 228s 536ms/step - loss: 0.0178 - accuracy: 0.9948
Epoch 8/15
425/425 [==============================] - 226s 533ms/step - loss: 0.0133 - accuracy: 0.9960
Epoch 9/15
425/425 [==============================] - 212s 500ms/step - loss: 0.0112 - accuracy: 0.9967
Epoch 10/15
```

h. Model Testing



```
#Test for the model part 1
text = "The product is bad"
text = re.sub('[^a-zA-Z]', ' ', text)
text = text.lower().split()
text = [ps.stem(word) for word in text if not word in set(stopwords.words('english'))]
text = ' '.join(text)
```

```
y_p1 = model.predict(cv.transform([text]))

y_p1
```

```
array([[0.00095823]], dtype=float32)
```

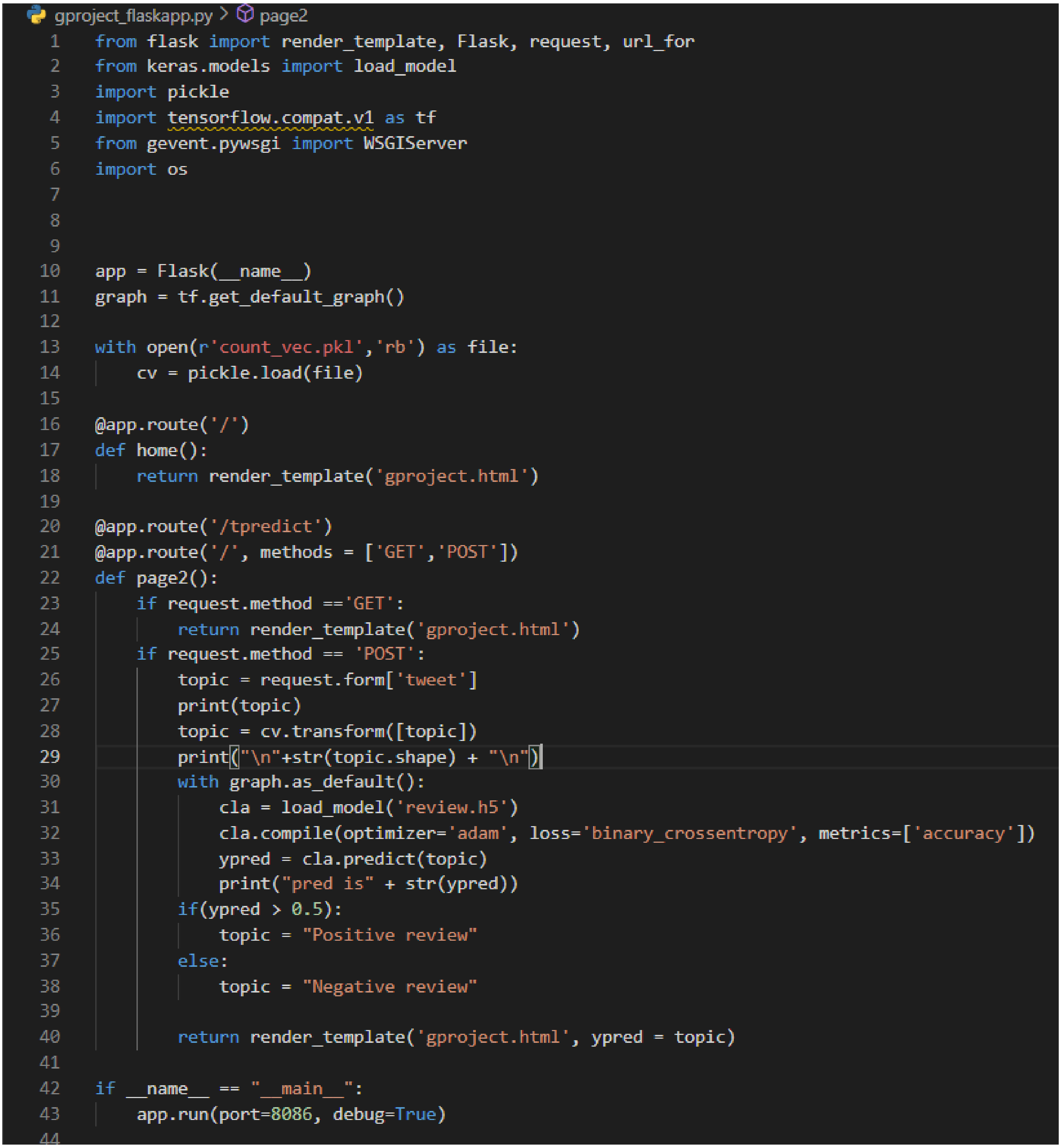
```
y_p1 = y_p1 > 0.5
y_p1
```

```
array([[False]])
```

3. Python Flask App File

By using the python file, the flask app is created which included the model.h5 file and the pkl file saved from the Jupyter Notebook file, as well as the html files which acts as the backbone for the web appearance of the app.

Flask is an open-source web framework. This means flask gives tools, frameworks, and technologies needed to create a web app. This online application might be as simple as a set of web pages, a blog, or a wiki, or as complex as a web-based calendar or a commercial website.

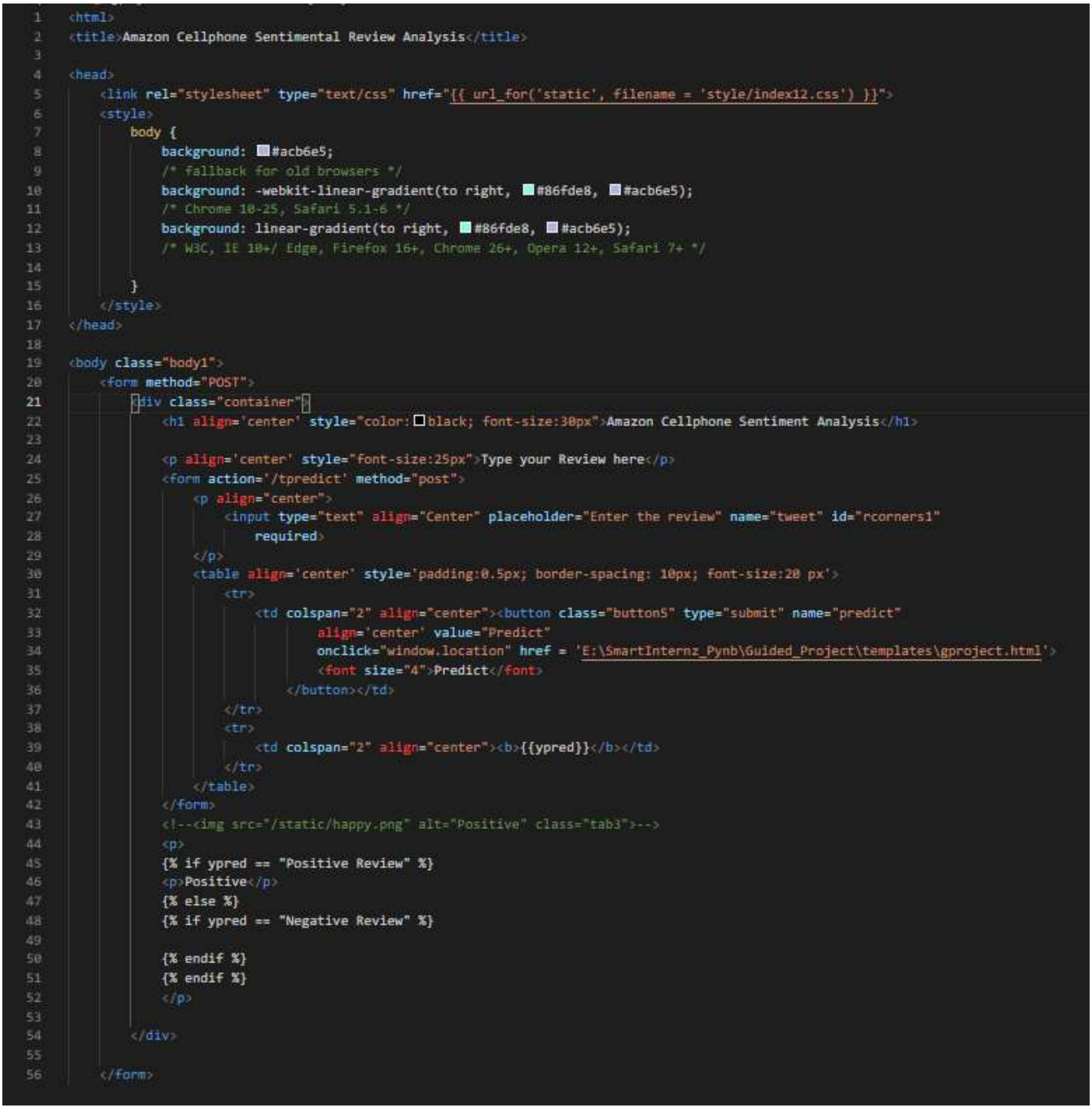


```python
from flask import render_template, Flask, request, url_for
from keras.models import load_model
import pickle
import tensorflow.compat.v1 as tf
from gevent.pywsgi import WSGIServer
import os



app = Flask(__name__)
graph = tf.get_default_graph()

with open(r'count_vec.pkl','rb') as file:
    cv = pickle.load(file)

@app.route('/')
def home():
    return render_template('gproject.html')

@app.route('/tpredict')
@app.route('/', methods = ['GET','POST'])
def page2():
    if request.method =='GET':
        return render_template('gproject.html')
    if request.method == 'POST':
        topic = request.form['tweet']
        print(topic)
        topic = cv.transform([topic])
        print("\n"+str(topic.shape) + "\n")
        with graph.as_default():
            cla = load_model('review.h5')
            cla.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
            ypred = cla.predict(topic)
            print("pred is" + str(ypred))
        if(ypred > 0.5):
            topic = "Positive review"
        else:
            topic = "Negative review"

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

if __name__ == "__main__":
    app.run(port=8086, debug=True)

```

4. Html File

The purpose of the html file is to create a blueprint for the flask web applications. To use the html files alongside the Flask app, a folder called “templates” will be used to store available html files.

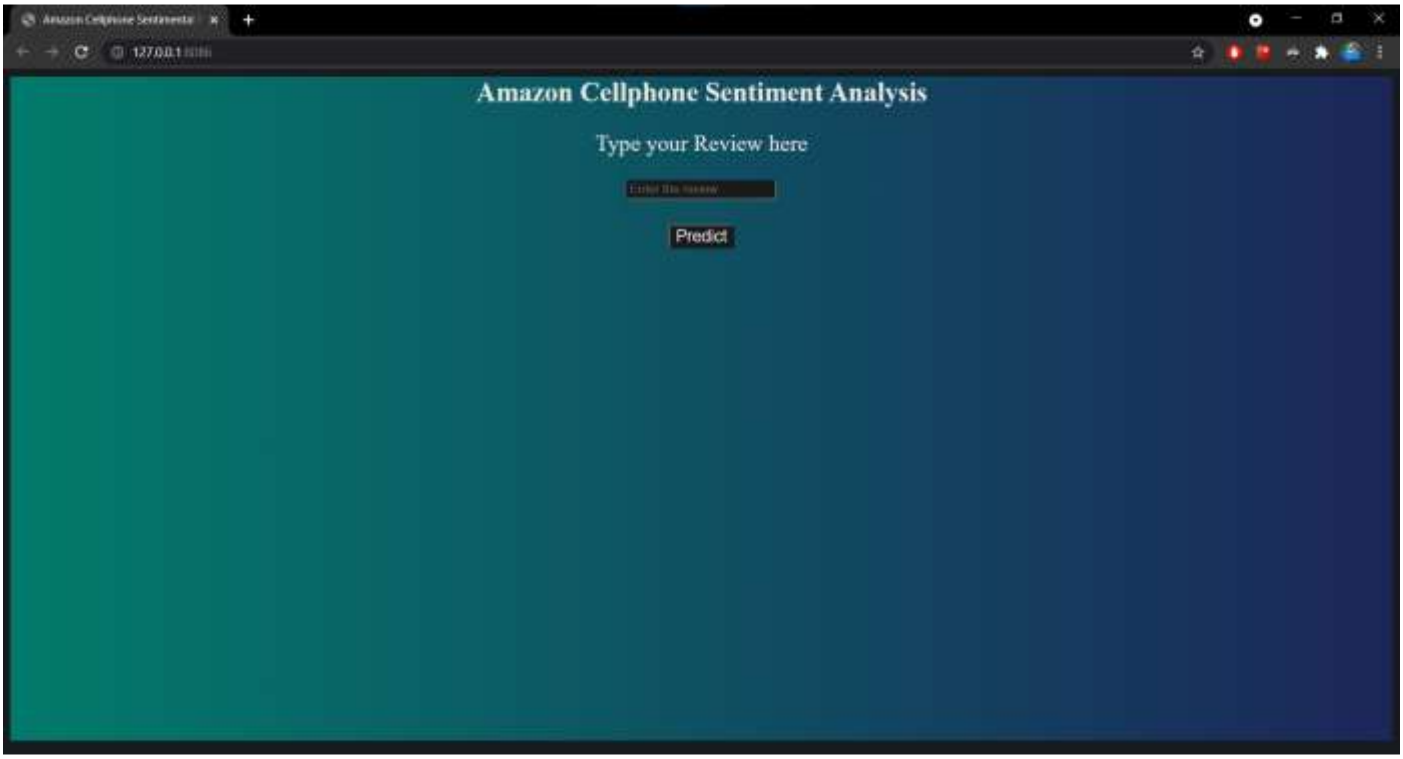
HTML, or HyperText Markup Language, is the standard markup language for texts that are intended to be viewed on a web browser. Technologies such as Cascading Style Sheets (CSS) and programming languages like JavaScript can help. Web browsers accept HTML documents from a web server or locally stored files and convert them to multimedia web pages. HTML initially provided cues for the document's look and described the structure of a web page logically.



```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>
                </p>
                <table align='center' style='padding:0.5px; border-spacing: 10px; font-size:20 px'>
                    <tr>
                        <td colspan="2" align="center"><button class="button5" type="submit" name="predict"
                                align='center' value="Predict"
                                onclick="window.location" href = 'E:\SmartInternz_Pynb\Guided_Project\templates\gproject.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>
```

Output

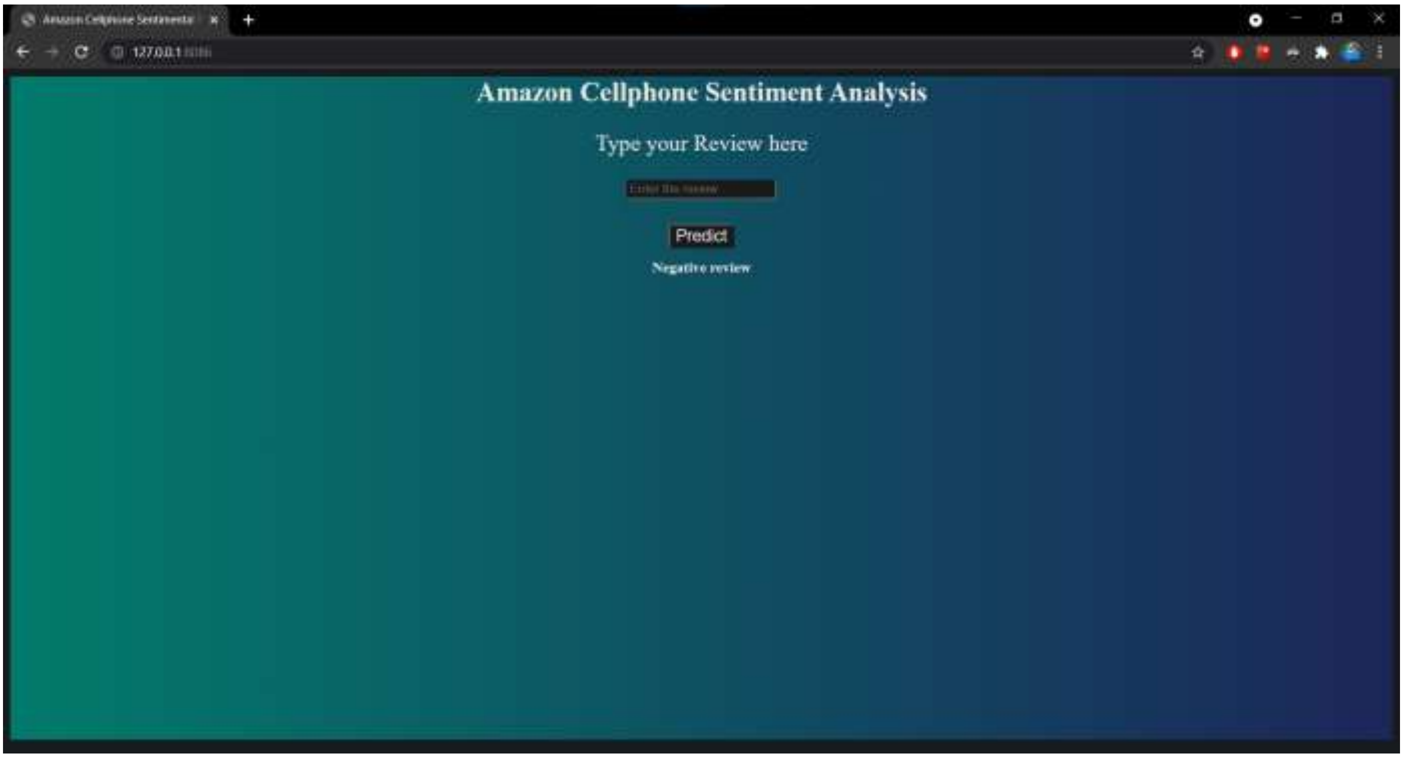
Default Webpage



Amazon Cellphone Sentiment Analysis

Type your Review here

After a negative review is inputted

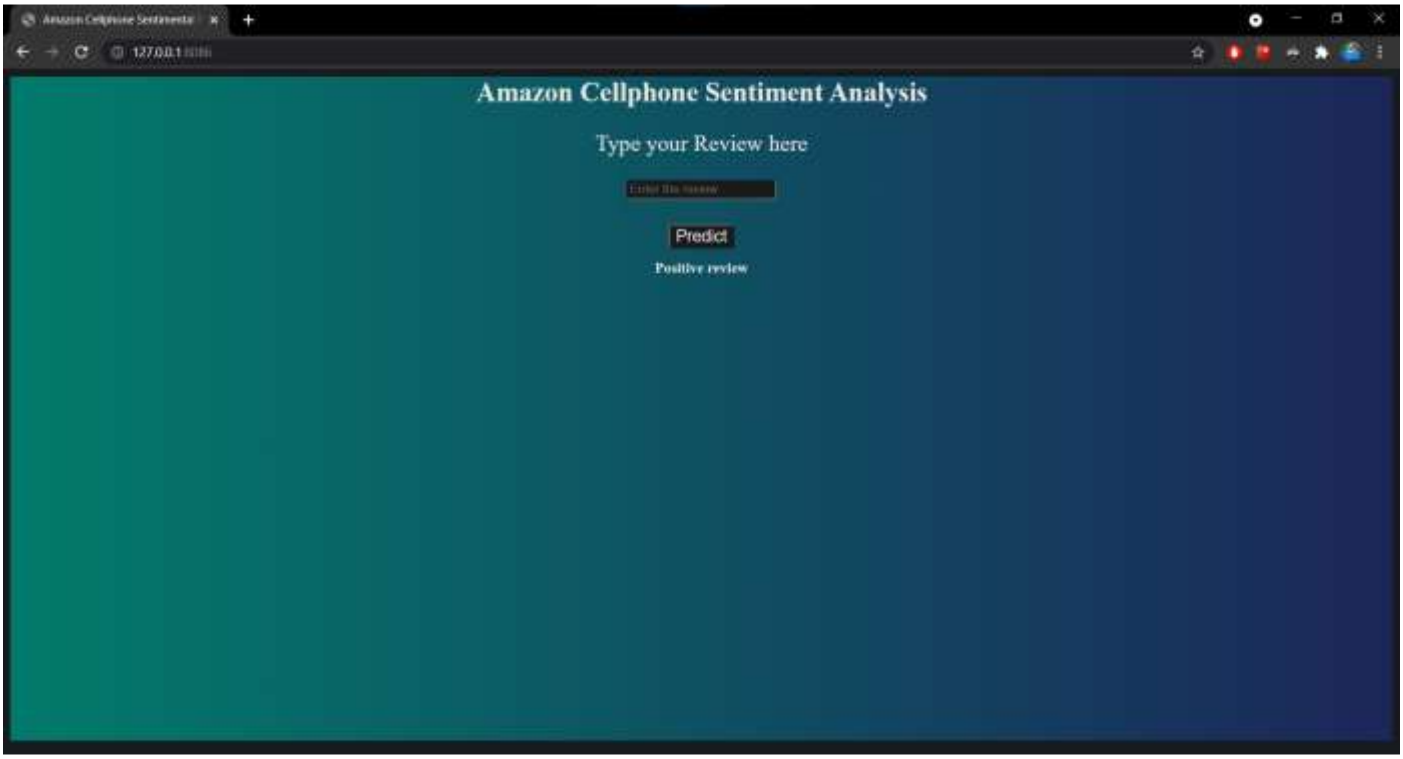


Amazon Cellphone Sentiment Analysis

Type your Review here

Negative review

After a positive review is inputted



Amazon Cellphone Sentiment Analysis

Type your Review here

Positive review

Future Improvements

Although the Flask App created using AI model training had proven to be successful in differentiating reviews from positive and negative, improvement can still be made to increase effectiveness of the app as well as the accuracy of the prediction.

One of the current issues faced is that during the prediction, the page load slowly and thus will produce the prediction of the inputted review slowly. Thus, the model training program can be look at once again and even refined via introducing new methods such as including other libraries such as joblib and Ray, changing the optimization function etc, so that the model can predict much faster while maintaining high prediction accuracy.

All AI systems, big or little, will include machine learning. Connected AI systems will allow machine learning algorithms to "continuously learn" from new data on the internet. Machine Learning will aid AI in deciphering data context and meaning.

Conclusion

In every aspect of the world, model learning is essential including in business, education and so on. According to some of the most prominent AI researchers, it has the potential to enhance network robustness, uncertainty estimating abilities, and machine learning model training costs.

From the project, the goal is to train a model to learn on which review is positive or negative, all while predicting whether the inputted reviews are positive or negative, then integrated it into the Flask app. And according to testing, it shows that for most cases, the app can manage to predict it correctly. Thus, proving that the model training and flask application is indeed successful.