

# **ANALYSIS OF AMAZON CELL PHONE REVIEWS**

Using Natural Language Processing

**Developed by: Rishitha Gajula, Pramod chaithanya Padidala, Sindhu Ette, Rakesh Nallatigala**  
**Smart Bridge-Remote Summer Internship Program**

## **1. INTRODUCTION**

Purchasing a product is an interaction between two entities, consumers and business owners. Consumers often use reviews to make decisions about what products to buy, while businesses, on the other hand, not only want to sell their products but also want to receive feedback in terms of consumer reviews. Consumers reviews about purchased products shared on the internet have great impact. Human nature is generally structured to make decisions based on analysing and getting the benefit of other consumer experience and opinions because others often have a great influence on our beliefs, behaviours, perception of reality, and the choices we make. Hence, we ask others for their feedback whenever we are deciding on doing something. Additionally, this fact applies not only to consumers but also to organizations and institutions. In the last few years, consumer ways of expressing their opinions and feelings have changed according to changes in social networks, virtual communities and other social media communities. Discovering large amounts of data from unstructured data on the web has become an important challenge due to its importance in different areas of life. To allow better information extraction from the plethora of data available sentiment analysis has emerged to be able to predict the polarity (positive, negative, neutral) of consumer opinion. This in turn would help consumers to better analyse the textual data providing useful information. We study in this research sentiment analysis of mobile phone reviews taken from the Amazon website, and how these reviews help consumers to have confidence that they have made the right decision about their purchases. Also, the research in this work aims to help companies understand their consumers' feedback to maintain their products/services or enhance them. In addition, giving them insights about them in providing offers on specific products to increase their profits and customer satisfaction.

### **1.1 Overview**

90 percent of online reviews before they decide to purchase any Mobile phone from any e commerce website. Online mobile applications has revolutionised the way consumers purchase mobile phones online as these apps have all the information regarding any mobile phone at users finger tips. Amazon is one of the best mobile applications which is considered as a treasure trove of all mobile reviews, and their review system is accessible across all the

channels presenting reviews in an easy to use format So there should be a system which analysis thousands of reviews of unlocked mobile phones sold on Amazon.com to find insights with respect to reviews.

## **1.2 Purpose**

Amazon reviews serve a few different purposes

Reviews can:

- Push on the dense buyers into a purchase.

- Convince consumers to buy your product over alternatives..

- Assure consumers of quality.

- Serves as word of mouth recommendations.

Amazon customer reviews about the products are one of the main reasons to attract consumers on Amazon.It basically helps them understand almost every detail of the product.

## **2.LITERATURE SURVEY**

Sentiment analysis involves a combination of natural language processing, computational linguistics and textual analysis in order to detect positive, negative or neutral feelings about the subject of the text . It is used in different areas such as marketing, customer services, and amongst others. Sentiment analysis can be performed on both document-level or sentence-level depending on the unit of information being considered.In this project, sentence-level was considered. Sentiment analysis has several applications in different areas including advertisement where sentiment analysis contributes in selecting specific advertisements to be shown on commercial and social media channels according to particular users opinions on particular products. Sentiment analysis can also be utilized for opinion retrieval, i.e. build search systems to search for specific views on specific topics .Since this work is interested in studying the sentiments of mobile phones reviews on Amazon, the work related to analysing the sentiments of mobile phones or Amazon reviews have been considered in the review. In the following, these researches are reviewed in terms of pre-processing techniques, feature extraction methods, proposed methodologies, and evaluation metrics.

### **2.1 Existing Problem**

Amazon's problem with fake reviews shows no sign of abating,with an investigation by the which?consumer group releasing a flood of fake five star reviews for tech products.Unknown brands of headphones,dash cams,fitness trackers and smart watches are receiving thousands of reviews;as they are unverified,there is no evidence that the reviewer has even bought or used the product.

Ex: ITSHINY, Voguek and Aitalk.

Of the 1.8 million unverified Reviews posted , 99.6% were 5 star.By comparison ,

the no of unverified reviews averaged fewer than 3Lakhs per month,only 75%of which were 5 star

## **2.2 Proposed Solution**

### **Machine Learning(Natural Language Processing)**

NLP refers to AI method of communicating with an intelligent systems using a natural language such as English.Processing of Natural language is required when you want an intelligent system likerobot to perform as per your instructions,when you want to hear decision from a dialogue based clinical expert system.The field of NLP involves making computers to perform useful tasks with the natural language humans use.The input & output of an NLP system can be : 1.Speech 2.Written Text

We will be using the Natural Language Processing to analyse the sentiment (positive or a negative) of the given review.A sample web application is integrated dynamically.And also we have created an UI using the Flask for the Analysis ofAmazon cell phone Reviews, this UI will allow the users to predict the review status very easily and the User interface is user friendly not at least one complication in using the interface, and it can be used just by entering some necessary details into the UI in real time it'll give the predicted value like either given review is positive or negative.

## **3.THEORITICAL ANALYSIS**

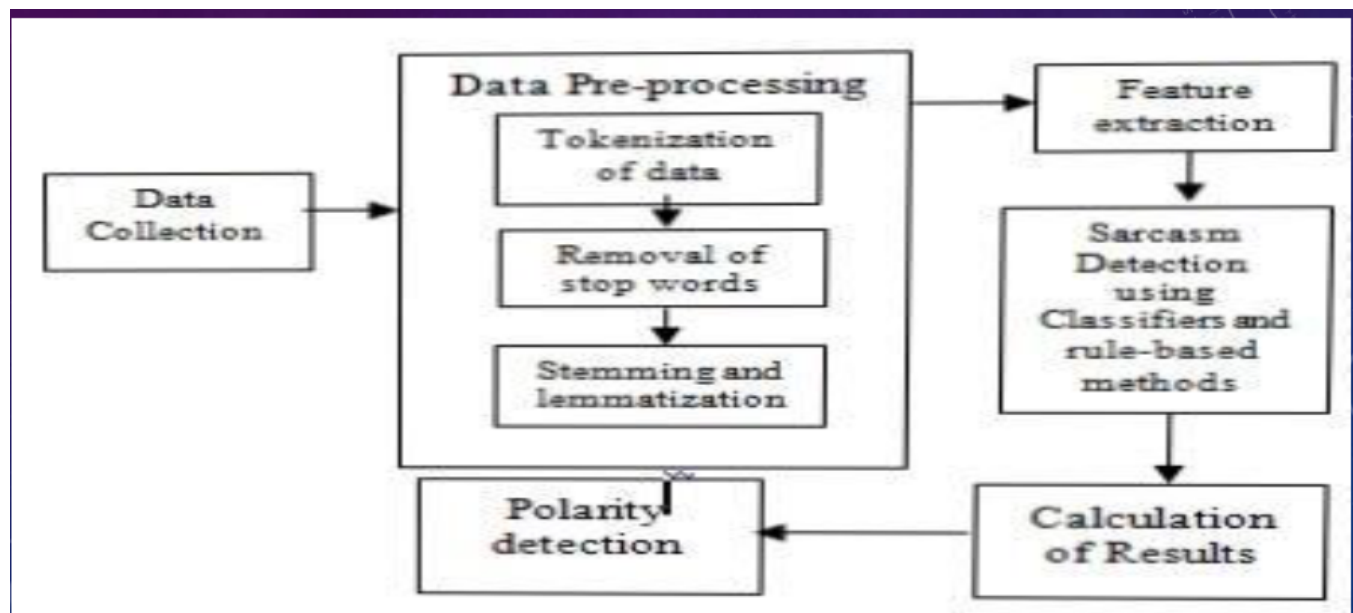
The employed dataset is textual, so it needs to be represented in numerical formats to be fed to the machine learning algorithms to build the desired classifiers. To achieve this,different vectorisation techniques are performed including term frequency which involves counting all the occurrences of all the terms in the document or sentence. A term can be expressed as a single word i.e. unigram, or any arbitrary number of words, namely, n-grams. Term frequency or count vectoriser (BOW) methodsuffer from a major pitfall, as it takes into account all the terms without taking into consideration the fact that some terms are very frequent in the corpus. Those terms do not capture document specific information since they occur in the majority of the documents. Such a drawback can be tackled by defining a maximum threshold for document frequency. However, the tuning of this threshold can be tricky, therefore, term frequency- inverse document frequency (TF-IDF) is introduced. TF-IDF is a weighting scheme that works by giving low weight to the terms that occur frequently in the given corpus. Inverse document frequency (IDF) is the inverse of the number of times a specific term appeared in the entire corpus. It captures how a particular term is document specific, and when multiplied by term frequency (TF) the result should give a measure of how this term is of particular

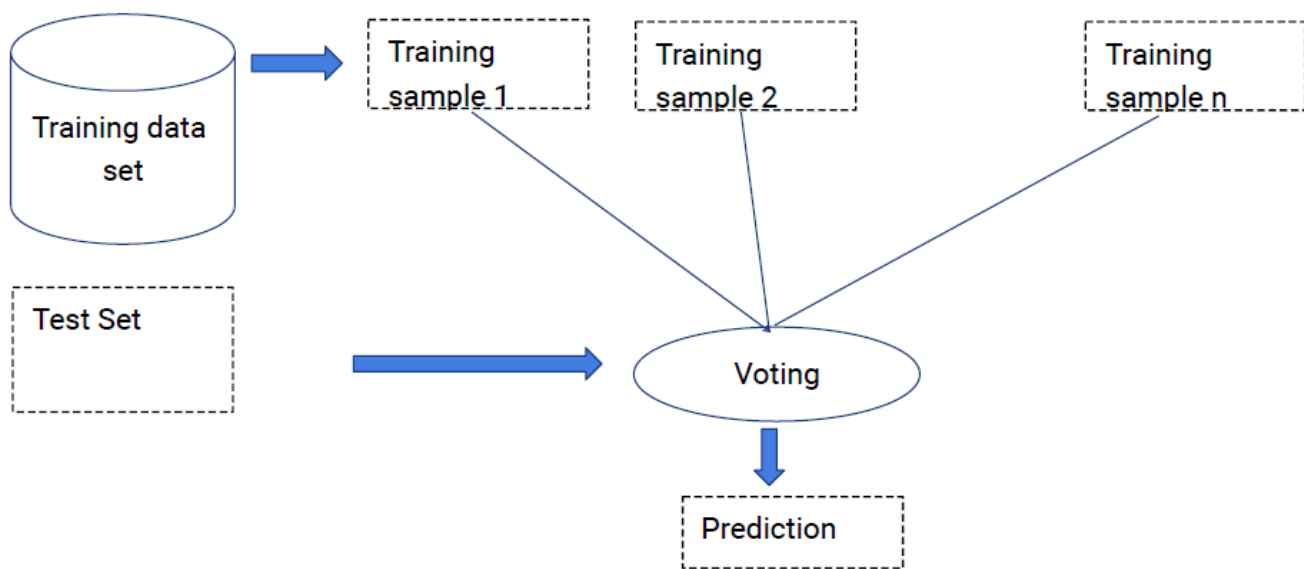
importance to the document at hand. Equation (1) demonstrates the main formula used for computing the TF-IDF for each term in each document.

$$TF - IDF = TF_{w,d} * IDF_w \quad (1)$$

Although TF, and TF-IDF are popular feature representation techniques in various natural language processing tasks, they define the vocabulary over a given corpus as a set of unique words, ignoring the semantic and syntactic similarities between those words. For example, in both TF and TF-IDF extraction techniques, the words pretty and beautiful are represented as two different words although they are nearly synonyms. Therefore, distributed words representations, namely word-embeddings were introduced as an alternative features extraction technique. Word-embeddings are extracted from huge corpora using different algorithms including deep learning algorithms [30]. The main idea behind word embeddings is to convert each word to a mathematical vector. In addition, each word will be represented by a vector, words with similar meaning have similar representations and this word is represented as positive and negative decimal number.

### 3.1 Block Diagram





### 3.2 Software Designing

- Jupyter Notebook Environment
- Spyder Ide
- Machine Learning Algorithms
- Neural Networks
- Python (pandas, numpy, matplotlib, seaborn, sklearn)
- HTML
- ● Flask

We developed this loan status prediction by using the Python language which is a interpreted and high level programming language and usng the Machine Learning algorithms and Neural Networks.for coding we used the Jupyter Notebook environment of the Anaconda distributions and the Spyder, it is an integrated scientific programming in the python language.For creating an user interface for the prediction we used the Flask. It is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions, and a scripting language to create a webpage is HTML by creating the templates to use in th functions of the Flask and HTML.

#### 4.EXPERIMENTAL INVESTIGATION:

In this paper, the dataset we used is derived from [www.kaggle.com/grikomsn/amazon cell phones reviews](https://www.kaggle.com/grikomsn/amazon-cell-phones-reviews). It contains more than 60000 data of users with a lot of Reviews. After that the missing values are filled in by means of mode interpolation, and the duplicate or S attributes are deleted, finally we have retained to 8 attributes. Those attributes were shown below in the screenshot of the data set we used.

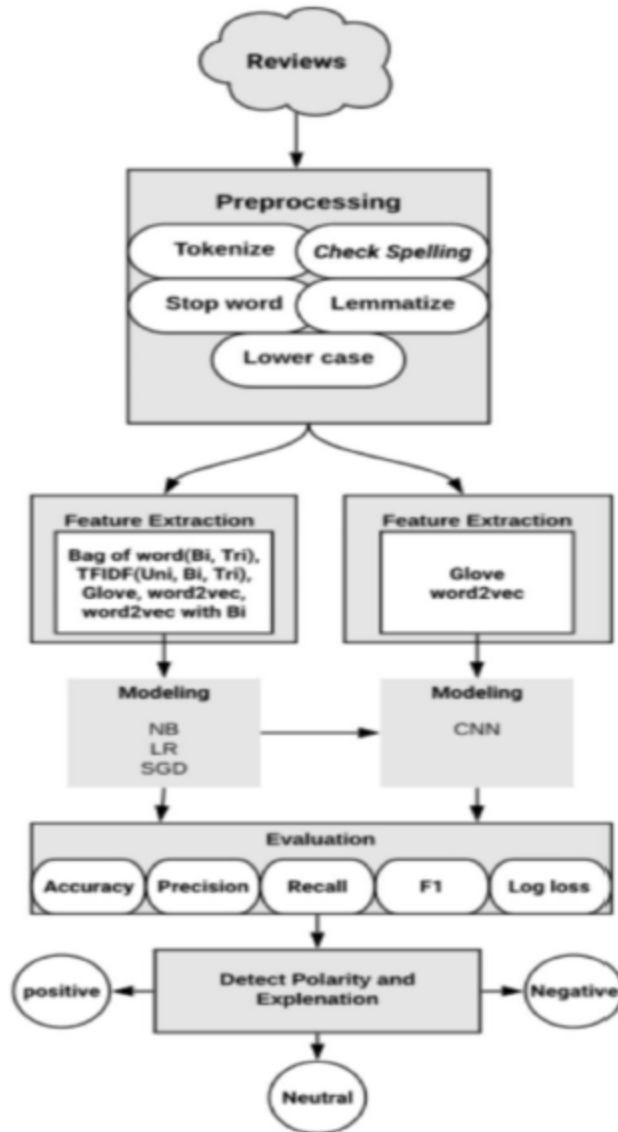
**File Home Insert Page Layout Formulas Data Review View Help LOAD TEST TEAM Tell me what you want to do**

**Clipboard** Paste Cut Copy Format Painter Font Alignment Number Styles Cell Styles Insert Delete Format AutoSum Filter Sort & Find & Select Editing

A1	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
asin	name	rating	date	verified	title	body	helpful	Votes												
1	B0000SX2IJanet	3	11-Oct-05	FALSE	Def not bc I had the S		1													
2	B0000SX2ILuke Wyr	1	7-Jan-04	FALSE	Text Mess Due to a s		17													
3	B0000SX2IBrooke	5	*****	FALSE	Love This This is a gr		5													
4	B0000SX2Iamy m. te	3	*****	FALSE	Love the f i love the		1													
5	B0000SX2Itristabim	4	*****	FALSE	Great pho The phone		1													
6	B0000SX2IJWhite	4	*****	FALSE	Worked g Hello, I have this phone and used it until t decided to buy a flip phone. I have had NO problems with the battery or new cases--it has a new fish case on i															
7	B0000SX2Ithe cell p	5	*****	FALSE	Wanna co Cool. Che		2													
8	B0000SX2IMatt	4	3-Apr-04	FALSE	Problem vThe 3599!		2													
9	B0000SX2ICharles Cr	5	*****	FALSE	cool phon i've never		7													
10	B0000SX2IAmazon C	3	2-Feb-04	FALSE	Pissed off ok well im		3													
11	B0000SX2Ihabblie	4	*****	FALSE	works gre i've had th		1													
12	B0000SX2IZachary O	1	*****	FALSE	Slow, ann 1.) Slow -		4													
13	B0000SX2Iesti	2	8-Sep-04	FALSE	Worth pay i bought tl		5													
14	B0000SX2Ibrittho	4	*****	FALSE	Great free This is an i		1													
15	B0009NSLMarcel Th	1	5-Mar-16	TRUE	Stupid phv DON'T BUY OUT OF SERVICE															
16	B0009NSLWilliam B	4	9-Feb-06	FALSE	Excellent \$! I have been with nextel for nearly a year now I started out this time last year with the Motorola Q205 and just upgraded to the i265 it is one of the best ph															
17	B0009NSLCMcIlhar	5	7-Feb-06	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.															
18	B0009NSLSStephen C	1	*****	TRUE	Phones io 1 star because the phones locked so I have to pay additional fees to unlock it															
19	B0009NSLMihir	5	*****	TRUE	Excellent 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 int															
20	B0009NSLLHughes	1	21-Jul-05	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 be															
21	B0009NSL1Stop d W	5	27-Jun-09	FALSE	NEXTEL BG GREAT PRODUCT THAT IS AS GREAT FOR NEXTEL AS IT IS FOR BOOST INSERT YOUR SIM & GO!															
22	B0005KX2IThomas	4	*****	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. Ther															

**Reviews1**

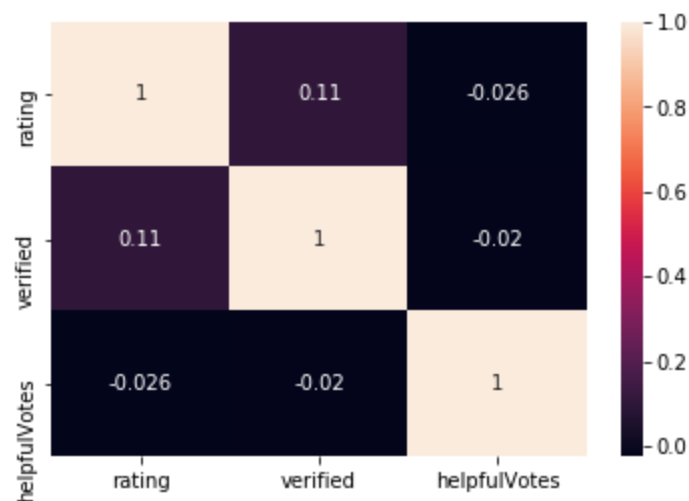
## 5.FLOWCHART



## 6.RESULT

In this paper, the Natural Language processing algorithm is used to predict its performance, and compared with another other machine learning methods namely the decision tree, the logistic regression, Naive Bayes, KNN, Random Forest and the SVM. The

results show that, the performance of NLP have comparable performance than that of logistic regression, random forest, SVM, Naive Bayes and decision tree, but the NLP still performs the best, with an accuracy of 91%, higher than the Naive Bayes with an accuracy of 70%. The loss of the prediction model based on NLP is 0.5.



```
Epoch 1/10
4000/4000 [=====] - 21s 5ms/step - loss: 1.5945 - acc: 0.7350
Epoch 2/10
4000/4000 [=====] - 19s 5ms/step - loss: 0.7718 - acc: 0.8067
Epoch 3/10
4000/4000 [=====] - 20s 5ms/step - loss: 0.4284 - acc: 0.8648
Epoch 4/10
4000/4000 [=====] - 20s 5ms/step - loss: 0.3012 - acc: 0.8900
Epoch 5/10
4000/4000 [=====] - 20s 5ms/step - loss: 0.3130 - acc: 0.8965
Epoch 6/10
4000/4000 [=====] - 20s 5ms/step - loss: 0.2794 - acc: 0.9027
Epoch 7/10
4000/4000 [=====] - 20s 5ms/step - loss: 0.2535 - acc: 0.9120
Epoch 8/10
4000/4000 [=====] - 20s 5ms/step - loss: 0.2296 - acc: 0.9177
Epoch 9/10
4000/4000 [=====] - 21s 5ms/step - loss: 0.2536 - acc: 0.9147
Epoch 10/10
4000/4000 [=====] - 21s 5ms/step - loss: 0.2768 - acc: 0.9143
```



## **7.ADVANTAGES &DISADVANTAGES:**

### **ADVANTAGES:**

- Many potential buyers are looking for products on Amazon based on reviews given by customers.
- Well established Amazon partner program for affiliates of Amazon reviews.
- By using reviews we don't need any own shop software for better understanding of the products.
- We will not have to pay extra charges for renewing of Reviews.
- According to a report, reviews might impact the purchasing decisions even more.Power reviews also found that positive reviews increase the sale by 20%.

### **DISADVANTAGES:**

- There are only possible limited opportunities to directly influence the display of product.
- Several service providers on same product can result a price battle due to reviews given by customers.
- Due to reviews given by a customer whole product may gets bad review.

## **8.APPLICATIONS**

- The development of Amazon reviews explains us a lot of benefits.
- It helps the user to estimate the quality of the product.
- It estimates that it is liable according to the customer.
- Through this product,we can assure that reviews play a key role in estimating the product.

## **9.CONCLUSION**

- Amazon's product review platform shows that most of the reviewers have given

4 star and 3 star ratings  
to unlocked mobile phones.

- The average length of the reviews comes close to 230 characters. We also uncovered that lengthier reviews tend to be more helpful and there is a positive correlation between price & rating.
- Sentiment analysis shows that positive sentiment is prevalent among the reviews and in terms of emotions, 'trust', 'anticipation' and 'joy' have highest scores.It'd be interesting to perform further analysis based on the brand (example: Samsung vs. Apple).
- We can also look 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. There are many more insights to be unveiled from the Amazon reviews.

## **10.FUTURE SCOPE:**

- In this NLP is used further to understand the reviews in a simple language.
- It is used to analyse the reviews in a realistic way.
- In future, reviews are made identical to the product.
- In future reviews are made such that it is made according to beneficiary of a customer &any such fake reviews will be eliminated or discarded.
- A software is used to understand & demolish the fake reviews.

## **11.BIBLIOGRAPHY:**

- Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining In: Proceedings of the Seventh

conference on International Language Resources and Evaluation.. European Languages Resources Association, Valletta, Malta.

- Python Machine learning :Machine learning &Deep learning with Python,Scikit learn,and tensorflow 2,3 rd Edition by Sebastian Raschka&Vahid Mirjalili
- Deep learning for Natural Language Processing:Creating neural networks with python by Sumit pandey,Karan Jain,Palash goyal.
- <https://www.w3schools.com>
- <https://www.python.org>

## APPENDIX

### HTML:

```
<!DOCTYPE html>
<html >
<!--From https://codepen.io/frytyler/pen/EGdtg-->
<head>
  <meta charset="UTF-8">
  <title>ML API</title>
  <link href='https://fonts.googleapis.com/css?family=Pacifico' rel='stylesheet'
type='text/css'>
<link href='https://fonts.googleapis.com/css?family=Arimo' rel='stylesheet'
type='text/css'>
<link href='https://fonts.googleapis.com/css?family=Hind:300' rel='stylesheet'
type='text/css'>
<link
href='https://fonts.googleapis.com/css?family=Open+Sans+Condensed:300'
rel='stylesheet' type='text/css'>
<link rel="stylesheet" href="{{ url_for('static', filename='css/style.css') }}">
```

```
<style>
.login{
top: 20%;
}
</style>
</head>
```

```
<body>
```

```
    <div class="container">
```

```
        <h1>Analysis of Amazon cell phone Reviews</h1>
```

```
    <div class="con2">
```

```
        <!-- Main Input For Receiving Query to our ML -->
```

```
        <form action="{{url_for('predict')}}" method="POST">
```

```
            <input type="text" name="message" placeholder="Enter your
Review" required="required" />
```

```
            <input type="submit" class="btn btn-primary btn-block btn-large"
value="Predict"/>
```

```
            <p><b>{{prediction}}</b></p>
```

```
        </form>
```

```
    </div>
```

```
    <p><b>
```

```
        {% if prediction == "Positive Review" %}
```

```
        <h2 style="color:blue;">Positive</h2>
```

```
        
    {% else %}
    <h2 style="color:red;">Negative</h2>
    

    {% endif %}
    </b>

</p>
</div>
</body>

</form>
</html>

```

### **APP.PY:**

```

import numpy as np
from flask import Flask, request,url_for, render_template
from keras.models import load_model
import pickle
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.externals import joblib
from keras.preprocessing import sequence
from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_sequences
tk=Tokenizer()
import tensorflow as tf
global graph
graph = tf.get_default_graph()
filename = r"E:\Bunny\VEC-rsip\Project\Amazon.h5"
cla=load_model(filename)

```

```
with open(r'E:\Bunny\VEC-rsip\Project\cv_transform.pkl','rb') as file:
    cv=pickle.load(file)
```

```
cla.compile(optimizer='adam',loss='binary_crossentropy')
app = Flask(__name__)
@app.route('/')
def home():
    return render_template('index.html')
@app.route('/predict',methods=['GET','POST'])
def predict():
    if request.method == 'GET':
        img_url = url_for('static',filename = 'css/style/0123.png')
        return render_template('index.html',url=img_url)
    if request.method == 'POST':
        data = request.form['message']
        print("Hey " +data)
        data=cv.transform([data])
        print("\n"+str(data.shape)+"\n")
        with graph.as_default():
            y_pred = cla.predict(data)
            print("pred is "+str(y_pred))
        if(y_pred > 0.5):
            img_url = url_for('static',filename = 'css/style/static2.png')
            data = "Positive Review"
        else:
            img_url = url_for('static',filename = 'css/style/static1.jpg')
            print(img_url)
            data = "Negative Review"

    return render_template('index.html',prediction=data)
if __name__ == '__main__':
    app.run(host = 'localhost', debug = True, threaded = False)
```

