**SENTIMENT ANALYSIS OF HOTEL REVIEW**

**A UG PROJECT PHASE -1 REPORT**

Major Project documentation submitted to

**JAWAHARLAL NEHRU TECNOLOGICAL 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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **PADIDALA SRINIKHIL**  **GADDAM DEEKSHITHA**  **DONTHULA ARUN** |  |  |  | **19UK1A05M5**  **19UK1A05P2**  **20UK5A0506** |

Under the guidance of

**Dr.K. SHARMILA REDDY**

(Associate Professor)



**DEPARTMENT OF COMPUTER SCIENCE OF ENGINEERING**

**VAAGDEVI ENGINEERING COLLEGE**

Affiliated to JNTUH, HYDERABAD

BOLLIKUNTA, WARANGAL (T.S)-506005

**2019-2023**

**I**

**DEPARTMENT OF COMPUTER SCIENCE OF ENGINEERING**

**VAAGDEVI ENGINEERING COLLEGE WARANGAL**

**2019-2023**



# CERTIFICATE OF COMPLETION

**UG PROJECT PHASE-1**

This is to certify that the project report entitled **“SENTIMENT ANALYSIS OF HOTEL REVIEW”** is a bonafide record of work carried out by **PADIDALA.SRINIKHIL(19UK1A05M5),GADDAM.DEEKSHITHA(19UK1A05P2),DONTHULA.ARUN(20UK5A0506)** under the guidance and supervision of **Mr. INDRA PRAKASH CHAUHAN** in the partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology in Computer Science and Engineering of Jawaharlal Nehru Technological University, Hyderabad** during the academic year 2019-23.

**Project Guide Head of the Department**

**Dr. K.SHARMILA REDDY**     **Dr.R.NAVEENKUMAR**

(Associate 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 Technical Seminar 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 Technical Seminar.

We express heartfelt thanks to the UG PROJECT PHASE-1 Coordinator,

**Dr. K.Sharmila Reddy**, Assistant Professor, Department of CSE for her constant support and giving necessary guidance for completion of this mini project .

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

**TEAM MEMBERS**

**P.SRINIKHIL(19UK1A05M5)**

**G.DEEKSHITHA(19UK1A05P2)**

**D.ARUN(20UK5A0506)**

**III**

**INTERNSHIP REPORT APPROVAL FORM**

Date

With immense pleasure, this is to approve that the students of Vaagdevi Engineering College i.e,

P.SRINIHKIL(19UK1A05M5),G.DEEKSHITHA(19UK1A05P2),D.ARUN(20UK5A0506)

Successfully Completed Their Project And Project Report on **“Sentiment Analysis Of Hotel Reviwe”**

Under Your Guidance.

We are highly impressed with the work that they have done and commend them on their quick grasping skills. They have shown good intent to learn and have put the knowledge gained into application in the form of this project. We appreciate the hard work and commitment shown by them.

We, hereby approve that this document is completely checked and accepted by SmartBridge

Technical Team. been an absolute pleasure to educate and mentor these students. We hope

that this document will also serve as a Letter of Recommendation, to whomsover applied.

We wish them success in all future endeavors and a great career ahead.

**IV**

# ABSTRACT

We consider the problem of classifying a hotel review as a positive or negative and thereby analysing the sentiment of a customer. Using Hotel review data from Kaggle, we find that standard Machine Learning techniques can definitely outperform human-produced sentiment analysis baselines. We will explore wide range of probabilistic models including Naive Bayes Classifier, Logistic Regression Classifier to classify a review. To extract the frequent words from the reviews we have used stop words , word net. We conclude by comparing accuracy of different strategic models and discuss about scope for future work.

**Keywords**: Hotel Review , NLTK, Naïve Bayes Classifier , Logistic Regression Classifier

**V**

# CONTENTS

# Pgno:

**1.INTRODUCTION 01**

1.1. Overview

1.2. Purpose

### 2.LITERATURE SURVEY 02

2.1 . Existing Problem

2.2. Proposed Solution

### 3.THEORETICAL ANALYSIS 03

3.1. Block Diagram

3.2. Hardware and Software Design

**4.DESIGN**  **04**

**5.** **EXPERIMENTAL INVESTIGATIONS 05**

**6.RESULT** **06-08**

**7.ADVANTAGES AND DISADVANTAGES** **09**

**8.APPLICATIONS** **10**

**9.CONCLUSION 11**

**10. FUTURE SCOPE 12**

**11. REFERENCES 13**

**12. APPENDIX 14-15**

# 

**VI**

## 1.INTRODUCTION

**1.1OVERVIEW**

Reviews have become a prominent factor that affects people’s bookings. people spend hours and even days sifting through dozens if not hundreds of options. The number of things to consider and the variety of reviews from previous guests is mind-blowing. guest reviews clearly influence people’s booking decisions, which means, one should pay attention to what people are saying about their hotel. Not only should hoteliers strive for good reviews, but also implement it in a way that can help them learn the most about their customers. Reviews can tell the management if they are keeping up with the customers’ expectations, which is crucial for developing marketing strategies based on the personas of the customers. Reviews are important and hotel owners, need to start leveraging it. Sentiment analysis is a classification problem which will be implemented using Naïve Bayes classification and logistic regression algorithm. Flask integration and IBM deployment will also be done.

### 1.2 PURPOSE

Recent years have seen rapid growth in online discussion groups and review sites (e.g.www.tripadvisor.com) where a crucial characteristic of a customer’s review is their sentiment or overall opinion — for example if the review contains words like ‘great’, ‘best’, ‘nice’, ‘good’, ‘awesome’ is probably a positive comment. Whereas if reviews contains words like ‘bad’, ‘poor’, ‘awful’, ‘worse’ is probably a negative review. However, Trip Advisor’s star rating does not express the exact experience of the customer. Most of the ratings are meaningless, large chunk of reviews fall in the range of 3.5 to 4.5 and very few reviews below or above. We seek to turn words and reviews into quantitative measurements. We extend this model with a supervised sentiment component that is capable of classifying a review as positive or negative with accuracy (Section 4). We also determine the polarity of the review that evaluates the review as recommended or not recommended using semantic orientation. A phrase has a positive semantic orientation when it has good associations (e.g., “excellent, awesome”) and a negative semantic orientation when it has bad associations (e.g., “terrific, bad”). Next step is to assign the given review to a class, positive or negative, based on the average semantic orientation of the phrases extracted from the review. If the average is positive, the prediction is that the review posted is positive. Otherwise, the prediction is that the item is negative.

1

**2. LITERATURE SURVEY**

**2.1 EXISTING PROBLEM**

Aiming at the lack of specific domain corpus in text sentiment polarity analysis, the inaccurate classification accuracy of the naïve Bayes algorithm due to the independence assumption and the sparse word vector matrix, a text sentiment analysis method based on the improved naïve Bayes algorithm is proposed. Combining machine learning methods with domain sentiment dictionary weighting methods. The improved word frequency inverse file frequency algorithm is used to extract the feature word vector of hotel review text, and the weight of the feature word vector of the domain dictionary after regression test is introduced to weaken the influence of the independence assumption. The singular value decomposition algorithm realizes the dimensionality reduction of the word vector sparse matrix and eliminates redundancy. The remaining features are used to construct a polynomial model of Naïve Bayes. The results of simulation research show that this method can effectively improve the effect of text sentiment classification

### 2.2 PROPOSED SOLUTION

Machine learning is a subfield of artificial intelligence (AI). The goal of machine learning generally is to understand the structure of data and fit that data into models that can be understood and utilized by people. Although machine learning is a field within computer science, it differs from traditional computational approaches. In traditional computing, algorithms are sets of explicitly programmed instructions used by computers to calculate or problem solve. Machine learning algorithms instead allow for computers to train on data inputs and use statistical analysis in order to output values that fall within a specific range. Because of this, machine learning facilitates computers in building models from sample data in order to automate decision making processes based on data inputs. In machine learning, tasks are generally classified into broad categories. These categories are based on how learning is received or how feedback on the learning is given to the system developed.

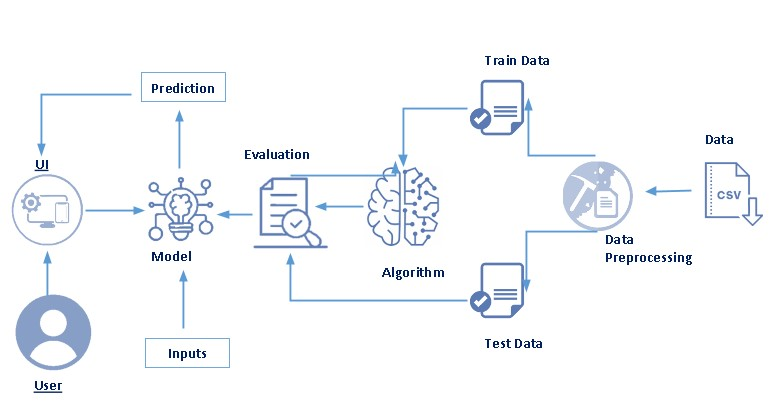
You must have prior knowledge of the following topics to complete this project.

* Supervised learning
* Unsupervised learning
* Regression and classification
* Naïve Bayes Classifier
* Logistic Regression
* Evaluation metrics (Precision, recall, fbeta)

2

**3.THEORITICAL ANALYSIS**

### 3.1 BLOCK DIAGRAM



**3.2 HARDWARE/SOFTWARE DESIGNING**

In this model we used some software and hardware tools follow are there

### Software Requirements

• Operating System : Microsoft Windows 10

* Anaconda3
* Jupyter
* Notebook
* Google collab
* Spyder

## Hardware Requirements

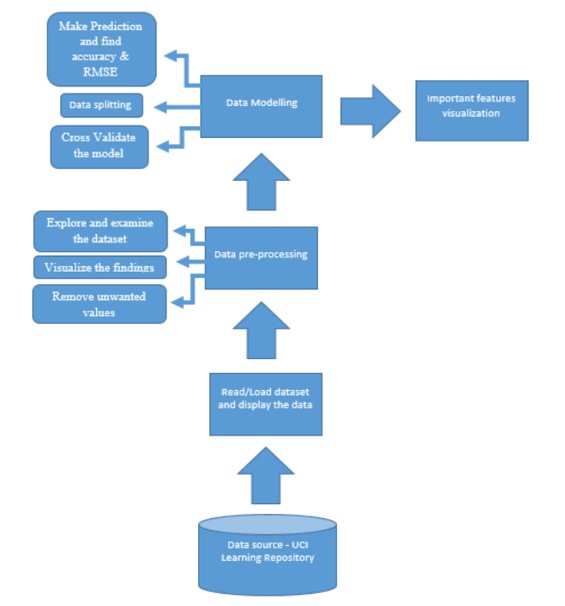
* Main Processor : Intel core i3, Intel core i5

* RAM Size : 4.00 GB

* Processor Speed : 2.60 GHz

3

**4.DESIGNE**



4

**5.EXPERIMENTAL INVESTIGATION**

For the implementation of the project, we have gone through several research papers from the "Google Scholar" website. We have gone through several websites including towardsdatascience.com, tutorials point, geeks for geeks, etc

**RESEARCH PAPER 1:**

Sentiment analysis is the task of identifying opinions expressed in any form of text. With the widespread usage of social media in our daily lives, social media websites became a vital and major source of data about user reviews in various fields. The domain of tourism extended activity online in the most recent decade. In this paper, an approach is introduced that automatically perform sentiment detection using Fuzzy C-means clustering algorithm, and classify hotel reviews provided by customers from one of the leading travel sites. Hotel reviews have been analysed using various techniques like Naïve Bayes, K-Nearest Neighbour, Support Vector Machine, Logistic Regression, and Random Forest. An ensemble learning model was also proposed that combines the five classifiers, and results were compared.

**RESEARCH PAPER 2:**

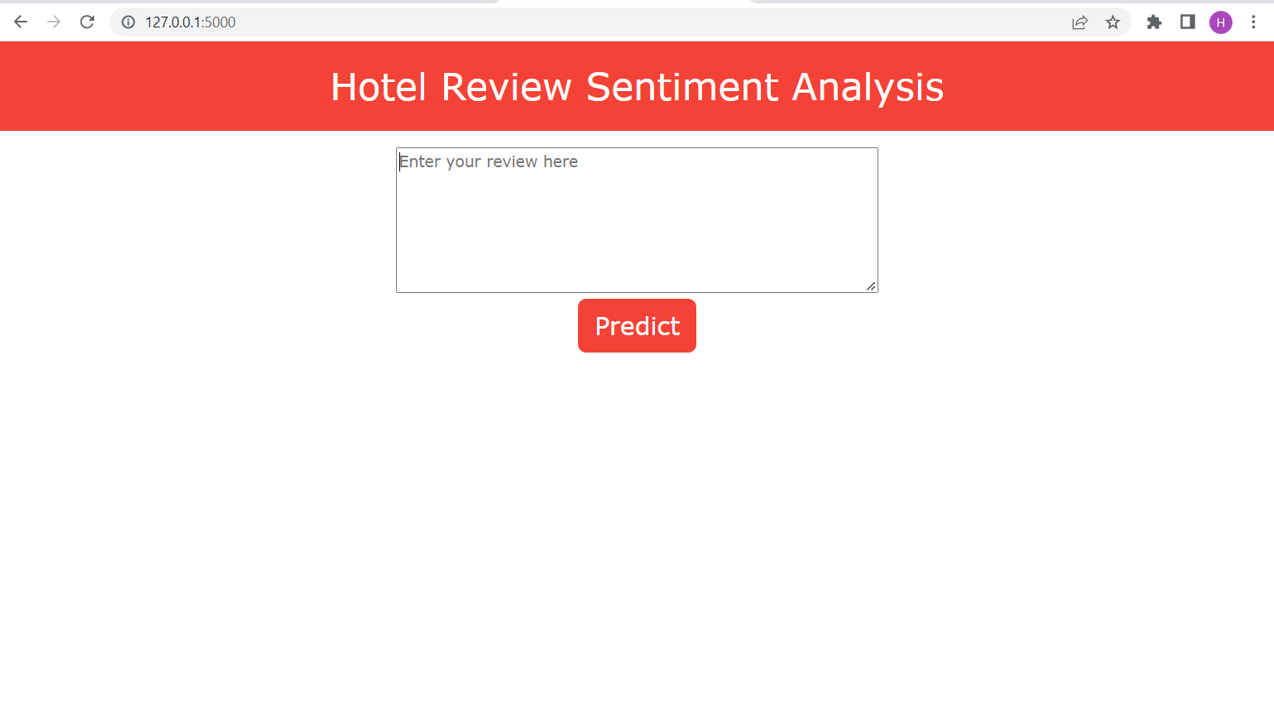
In today’s scenario online reviews on various digital platforms plays a vital role for customers to buy products. Based on the reviews and ratings by the consumer on E-commerce platform like flipkart, amazon etc. products are widely accepted or rejected. Apart from products people also look for the reviews of the services provided from restaurants, hotels, airlines etc. Sentiment analysis helps the developers to easily analyse the reviews and categorize them as positive or negative. In this paper, service of a hotel is analysed by finding out the polarity of the reviews in order to get the subject information. Aspect detection and sentiment classification are the main tasks focused here. For aspect detection latent irichlet allocation (LDA) is used for building the topics. Different machine learning classifiers like naïve bayes classifier, SVM, decision tree and logistic regression are used for classification of reviews. Evaluation is done by computing the accuracy, recall, precision and F score of these algorithms.

5

**6.RESULT**

WEBSITE

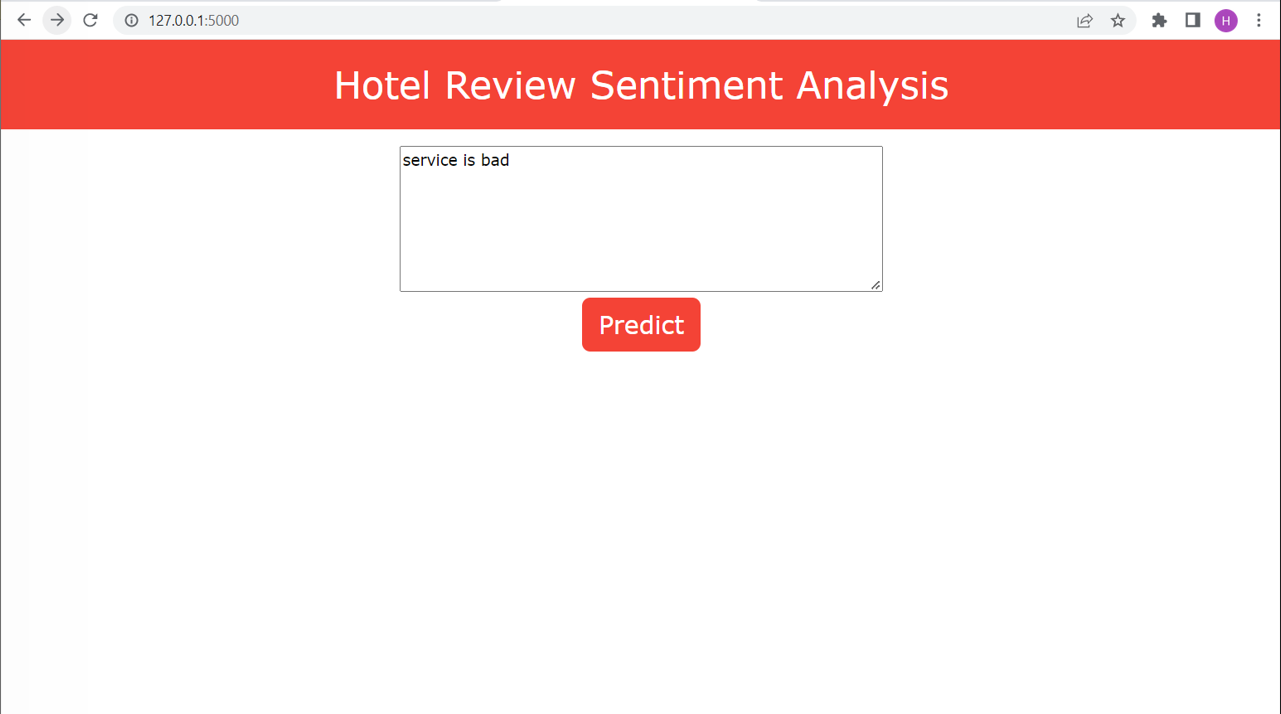
**FIGURE 6.1:**



The Hotel Review Sentiment Analysis Homepage Provides A TextBox With a Button named Predict

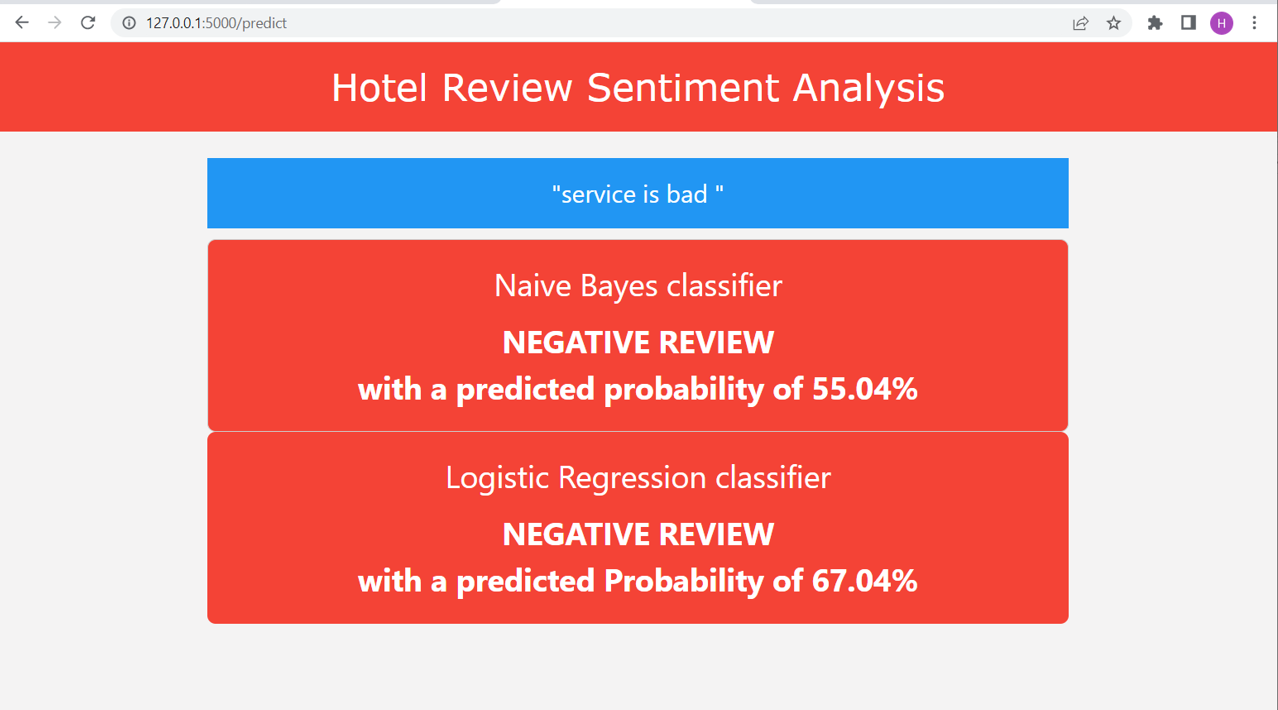
You Enter The review and Click Predict

**FIGURE 6.2**:



6

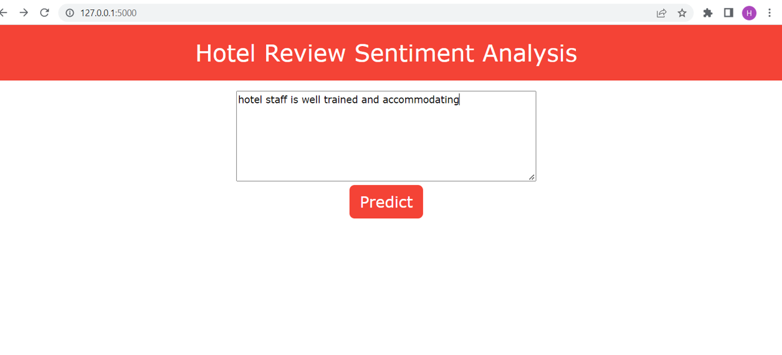
**FIGURE 6.3**:



In this we predicting the reviews of the user is bad or good using Naïve Bayes Classifier and Logistic Regression Classifer it shows the probability percentage according to the review entered in the textbox

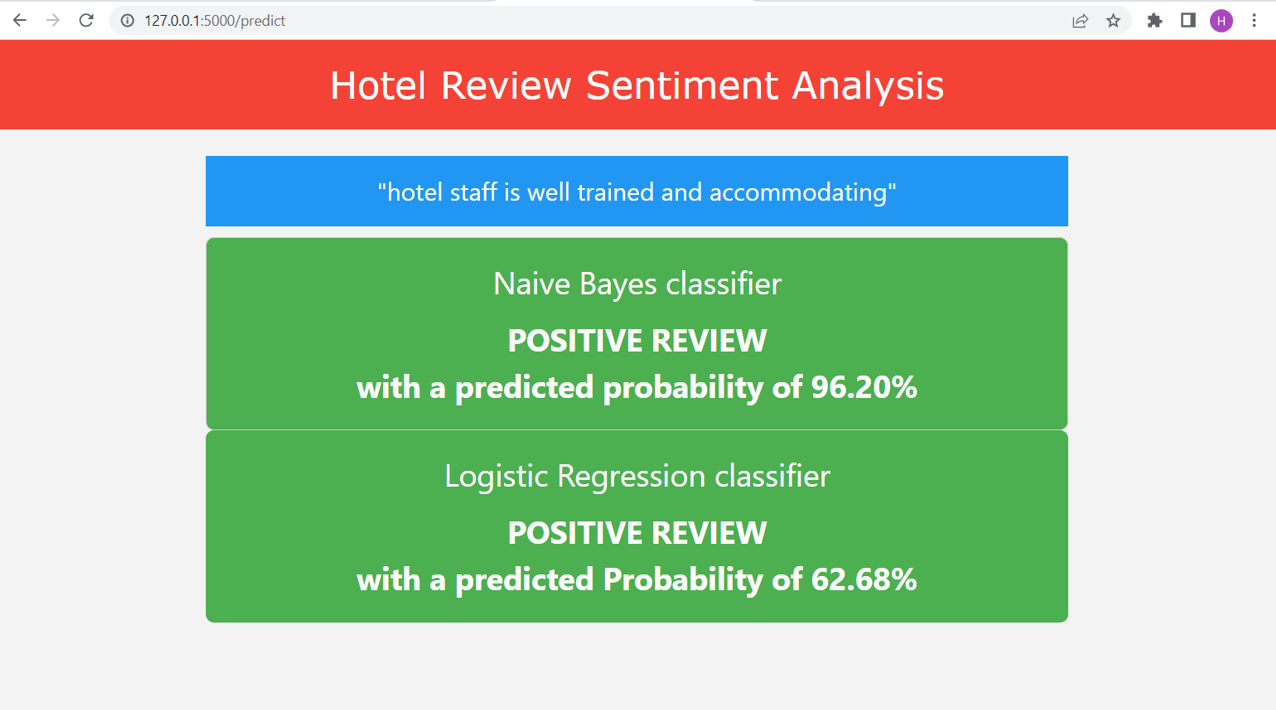
In the above figure 6.3 it is bad

**FIGURE 6.4**:

****

7

**FIGURE 6.5:**



In this we predicting the reviews of the user is bad or good using Naïve Bayes Classifier and Logistic Regression Classifer it shows the probability percentage according to the review entered in the textbox

In the above figure 6.5 it is good

8

### 7.ADVANTAGES&DISADVANTAGES

**ADVANTAGES:**

* They Predict Faster and More Accurately
* The model is trained to predict the bad and negative probability of the Reviews
* Thus with a rapid growth in the deep learning architecture, an objective of reviews of how much good and bad is easy to predict
* This will allow a non-contact, easy to use and low cost that can be performed routinely anywhere in the world
* The Trained model and also used anywhere remotely

**DISADVANTAGES:**

* For Training We need some good dataset from reputed sites like trip advisor
* It is not completely possible to use only raw text as input for making predictions
* We need to extract the relevant features from this raw source of data
* This kind of data can often come as a good complementary source in data science projects in order to extract more learning features and increase the predictive power of the models
* The result of accuracy is only 89%

9

**8.APPLICATIONS**

* This can be used to help the websites to understand.The user reviews with a single click which helps more accurately and improve their websites and hotel amenities according to the user needs
* The goal of opinion mining is to identify the text polarity, which means to classify it as positive, negative, or neutral is easy with this model
* There are some ready-to-use datasets labelled for more than twenty amenities. So, facing time and budget limitations, we took shortcut approach. So applying of the machine learning can evaluate the review probability to improve the respective amenity

10

**9.CONCLUSION**

Know fundamental concepts and techniques used for machine learning.Gain a broad understanding of data.Have knowledge on pre-processing textual data and classification algorithms.This can be used to help the websites to understand The user reviews with a single click which helps more accurately and improve their websites and hotel amenities according to the user needs.This will allow a non-contact, easy to use and low cost that can be performed routinely anywhere in the world

11

### 10.FUTURE SCOPE

In this project we have used the deep learning with NLTK, Naïve Bayes Classifier , Logistic regression classifier but Sentiment analysis and AI could be the answer to mental health treatment, according to **[TDWI](https://tdwi.org/articles/2020/04/07/adv-all-why-emotion-ai-key-to-mental-health-treatment.aspx)** With the ability to read emotions and learn responses, it is believed to be possible.Some think that it might be dangerous to use AI in the mental health field. However, this trend is popping up more as a serious consideration. Consumers are the most important part of a business. With unhappy customers, a company can receive a bad reputation. Sentiment analysis can help with monitoring customer service, and experience. For instance, using AI technology to analyze customer feedback and customer service exchanges, a company can adjust their service to improve customer satisfaction and loyalty.

12

### 11.REFERENCES

* H X Shi and X J Li "A sentiment analysis model for hotel reviews based on supervised learning," in in International Conference on Machine Learning and Cybernetics China 2011
* P Goyal "What is Laplacian smoothing and why do we need it in a Naive Bayes classifier?," Quora 27 September 2017 [Online] Available:https://www.quora.com/What-is-Laplacian-smoothing-and-why-do-we-need-it-in-a-Naive-Bayes-classifier [Accessed 22 April 2018]
* A Goel J Gautam and S Kumar "Real time sentiment analysis of tweets using naïve bayes," 2nd International Conference on Next Generation Computing Technologies (NGCT) pp 257-261 2016
* "6 Tren Wisata Utama Tahun 2016," tripadvisor 14 December 2015 [Online] Available:https://www.tripadvisor.co.id/TripAdvisorInsights/w665 [Accessed 19 February 2018]
* T Ghorpade and L Ragha "Featured Based Sentiment Classification for Hotel Reviews using NLP and Bayesian Classification,” in international conference on communication , information & computing technology (ICCICT) Mumbai india 2012

13

**12.APPENDIX**

**SOURCE CODE OF FLASK:**

import  pickle

filename = 'logistic\_regression\_model.pkl'

pickle.dump(model, open('/content/drive/MyDrive/Hotel review sentiment analysis/Flask/logistic\_regression\_model.pkl', 'wb'))

pickle.dump(cv, open('/content/drive/MyDrive/Hotel review sentiment analysis/Flask/tranform\_logistic.pkl', 'wb'))

pickle.dump(cv, open('/content/drive/MyDrive/Hotel review sentiment analysis/Flask/tranform\_naive.pkl', 'wb'))

filename = 'naive\_bayes\_model.pkl'

pickle.dump(model1, open('/content/drive/MyDrive/Hotel review sentiment analysis/Flask/naive\_bayes\_model.pkl', 'wb'))

!pip install Flask

from flask import Flask,render\_template

import pandas as pd

import pickle

from sklearn.feature\_extraction.text import CountVectorizer

from sklearn.naive\_bayes import MultinomialNB

import pickle

from sklearn.feature\_extraction.text import TfidfVectorizer

filename = 'naive\_bayes\_model.pkl'

naive\_bayes\_model = pickle.load(open('naive\_bayes\_model.pkl','rb'))

logistic\_regression\_model = pickle.load(open('logistic\_regression\_model.pkl','rb'))

cv\_naive = pickle.load(open('tranform\_naive.pkl','rb'))

cv\_logistic = pickle.load(open('tranform\_logistic.pkl','rb'))

app = Flask(\_\_name\_\_)

@app.route('/')

def home():

    return render\_template('home.html')

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

def predict():

    if request.method == 'POST':

        message = request.form['message'].lower()

        data = [message]

        vect\_naive = cv\_naive.transform(data).toarray()

        vect\_logistic = cv\_logistic.transform(data).toarray()

        my\_nb\_prediction = naive\_bayes\_model.predict(vect\_naive)

        nb\_percentage = naive\_bayes\_model.predict\_proba(vect\_naive)

        if my\_nb\_prediction ==1:

            nb\_percentage = nb\_percentage[0][1]

 else:

            nb\_percentage = nb\_percentage[0][0]

        my\_lg\_prediction = logistic\_regression\_model.predict(vect\_naive)

        lg\_percentage = logistic\_regression\_model.predict\_proba(vect\_naive)

        if my\_lg\_prediction ==1:

            lg\_percentage = lg\_percentage[0][1]

        else:

            lg\_percentage = lg\_percentage[0][0]

        return render\_template('result.html',message = message,my\_nb\_prediction = my\_nb\_prediction,nb\_percentage = nb\_percentage,my\_lg\_prediction = my\_lg\_prediction,lg\_percentage = lg\_percentage)

if \_\_name\_\_ == '\_\_main\_\_':

    app.run(debug=False)

15