

KAKATIYA GOVERNMENT COLLEGE HANAMKONDA

STUDENTS' STUDY PROJECT

on

“RESTAURANT REVIEW ANALYSIS USING MACHINE LEARNING “

Under the Supervision of
K.Sravana Kumari



DEPARTMENT OF COMPUTER SCIENCE AND APPLICATIONS

2021-22

CERTIFICATE

This is to certify that the project report entitled “**Restaurant Review Analysis Using Machine Learning** “ submitted to the Department of Computer Science and Applications, Kakatiya Government College, Hanamkonda and it was carried out by the following students under my guidance.

006222214	MOHAMMAD TAYYAB KHAN	BCOM I YEAR
006222233	NEELA AKHILA	BCOM I YEAR
006222185	MADISHETTI ROHITH	BCOM I YEAR
006222224	MULUKUNTLA SAIPAVAN	BCOM I YEAR
006222220	MORLE NITHIN	BCOM I YEAR
006222235	NERELLA REVATHI	BCOM I YEAR
006214606	GADDAM MANASA	B.SC II YEAR
006214615	KODIPELLY HARISH	B.SC II YEAR
006214623	PEGADA SUCHARITHA	B.SC II YEAR
006214627	VEERAMALLA BHARATH KUMAR	B.SC II YEAR

Supervisor

(K.Sravana Kumari)

Incharge

Principal

CERTIFICATE

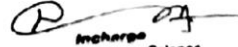
This is to certify that the project report entitled “**Restaurant Review Analysis Using Machine Learning** “ submitted to the Department of Computer Science and Applications, Kakatiya Government College, Hanamkonda and it was carried out by the following students under my guidance.

006222214	MOHAMMAD TAYYAB KHAN	BCOM I YEAR
006222233	NEELA AKHILA	BCOM I YEAR
006222185	MADISHETTI ROHITH	BCOM I YEAR
006222224	MULUKUNTLA SAIPAVAN	BCOM I YEAR
006222220	MORLE NITHIN	BCOM I YEAR
006222235	NERELLA REVATHI	BCOM I YEAR
006214606	GADDAM MANASA	B.SC II YEAR
006214615	KODIPELLY HARISH	B.SC II YEAR
006214623	PEGADA SUCHARITHA	B.SC II YEAR
006214627	VEERAMALLA BHARATH KUMAR	B.SC II YEAR



Supervisor

(K.Sravana Kumari)



Incharge
Dept. of Computer Science
Kakatiya Government College
Hanamkonda, Warangal.

Incharge .



PRINCIPAL
KAKATIYA GOVT. COLLEGE
Hanamkonda.

Principal

ABSTRACT

One of the most effective tools any restaurant has is the ability to track food and also in beverage sales daily. Currently, Recommender systems plays an important role in both academia and industry. These are very helpful form an aging information overload .In this paper ,we applied machine learning techniques for user reviews and analyze valuable information in the reviews. Reviews are useful for making decisions for both customers and owners. We build a Machine learning model with Natural Language Processing techniques that can capture the user's opinions from users' reviews. For experimentation, the python language was used.

Keywords: recommender systems, machine learning, python

INTRODUCTION

1.1 ABOUT PROJECT

The growth of the internet due to social networks such as facebook ,twitter,

LinkedIn, Instagram etc .has led to significant user's interaction and has empowered users to express their opinions about products, services, events, their preferences among others. It has also provided opportunities to the users to share their wish to demand experiences with each other. The faster development of social networks is causing explosive growth of digital content. It has turned online opinions, blogs, tweets, and posts in to a very valuable asset for the corporate to get insight from the data and plan their strategy. Business organizations need to process and study these sentiments to investigate e data and t to gain business insights. Traditional approach to manually extract complex features, identify which feature is relevant, and derivet he patterns from this huge information is very time consuming and require significant human efforts. However, Deep Learning can exhibit excellent performance via Natural Language Processing (NLP) technique to perform sentiment analysis on this massive information. The core idea of Deep Learning techniques is to identify complex features extracted from this vast amount of data with out much external intervention using deep neural networks . These algorithms automatically learn new complex features. Both automatic feature extraction and availability of resources are very important when comparing the traditional machine learning approach and deep learning techniques. Here the goal is to classify the opinions and sentiments expressed by users.

It is a set of techniques/ algorithms used to detect the sentiment(positive ,negative, or neutral)of a given text. It is a very powerful application of natural language processing(NLP) and finds usage in a large number of industries .It refers to the use of NLP, text analysis, computational linguistics, and biometrics to systematically identify ,extract, quantify, and study different states and subjective information. The sentiment analysis sometimes goes beyond the categorization of texts to find opinions and categorizes the more positive or negative, desirable or undesirable. Below figure describe the architecture of sentiment classification on texts. In this, we modify the provided

reviews by applying specific filters, and we use the prepared data set by applying the parameters and implement our proposed model for evaluation.

Another challenge of micro blogging is the incredible breadth of topic that is covered. It is not an exaggeration to say that people tweet about any thing and every thing. Therefore, to be able to build systems to mine sentiment about any given topic, we need a method for quickly identifying data that can be used for training. In this paper, we explore one method for building such data: using hash tags (e.g., #bestfeeling, #epicfail, #news) to identify positive, negative, and neutral reviews to use for training three way sentiment classifiers.

The online medium has become a significant way for people to express their Opinions and with social media, there is an abundance of opinion information available. Using sentiment analysis, the polarity of opinion can be found, such as positive, negative, or neutral by analyzing the text of the opinion. Sentiment analysis has been useful for companies to get their customer's opinions on their products predicting outcomes of elections, and getting opinions from movie reviews. The information gained from sentiment analysis useful for companies making future decisions. Many traditional approaches in sentiment analysis uses the bag of words method. The bag of word technique does not consider language morphology, and it could incorrectly classify two phrases of having the same meaning because it could have the same bag of words. The relationship between the collections of words is considered instead of the relationship between individual words. When determining the overall sentiment, the sentiment of each word is determined and combined using a function. Bag of words also ignores word order, which leads to phrases with negation in them to be incorrectly classified.

1.1 EXISTING SYSTEM WITH DRAWBACKS

Existing approaches to Sentimental Analysis are Knowledge-based techniques. Knowledge-based techniques make use of prebuilt lexicon sources containing polarity of sentiment words SentiWordNet(SWN) for determining the polarity of a tweet .Lexicon based approach suffers from poor recognition of sentiment.

1.2 PROPOSED SYSTEM WITH FEATURES

In the proposed system, sentimental analysis is done using natural language processing, which defines a relation between user posted tweet and opinion and in addition, suggestions of people.

Truly listening to a customer's voice requires deeply understanding what they have expressed in natural language. NLP is a best way to understand natural language used and uncover the sentiment behind it. NLP makes speech analysis easier.

Without NLP and access to the right data, it is difficult to discover and collect insight necessary for driving business decisions. Deep Learning algorithms are used to build a model

ADVANTAGES OF PROPOSED SYSTEM

The advanced techniques like natural language processing is used for the sentiment analysis. It make our project very accurate.

NLP defines a relation between user posted tweet and opinion and in addition suggestions of people.

NLP is a best way to understand natural language used by the people and uncover the sentiment behind it. NLP makes speech analysis easier.

CODE:

Model.py:

```
#Importing Libraries import pandas as pd import numpy as np
From sklearn.metrics import classification_report, confusion_matrix, accuracy_score from
nltk.corpus
import stopwords
from nltk.stem.porter
import PorterStemmer
import nltk
#import nltk nltk.download()
#ImportingDataset
df=pd.read_csv('Restaurantreviews.csv', encoding="ISO-8859-1")
df=df.drop(columns=["Restaurant", "Reviewer", "Metadata", "Time", "Pictures"])

#Transforming&Cleaning
Data y=df["Rating"]
X=df.drop(columns=["Rating"])y=y.replace({'Like':3})
y=y.fillna(y.median())y=pd.to_numeric(y)foriinrange(0,len(y)):
y.iloc[i]=round(y.iloc[i],0)
for i in range(0,len(y)):
if(y[i]>=3):
y[i]="Positive"
else:
y[i]="Negative"

#Applying Stemming with excluding StopWordsps=PorterStemmer()

corpus=[]
for I in range(0,len(X)):
```



```

review=re.sub('[^a-zA-Z]',",",str(X['Review'][i]))review=review.lower()
review=review.split()
review=[ps.stem(word)forwordinreviewifnotwordinstopwords.words('english')]
review=" ".join(review)corpus.append(review)

#Creating Matrix of Count Vectorizer
From sklearn.feature_extraction.text
import CountVectorizer cv=CountVectorizer(max_features=9000)
X=cv.fit_transform(corpus).toarray()

#Train-TestSplit
From sklearn.model_selection
Import train_test_split X_train, X_test, y_train, y_test
=train_test_split(X,y,test_size=0.25,random_state=0)

#Applying MultinomialNB
Froms klearn.naive_bayes
Import Multinomial NBclassifier=MultinomialNB().fit(X_train,y_train)

#Making Predictions
y_pred=classifier.predict(X_test)

#Creating Consusion Matrix
From sklearn.metrics
import confusion_matrix
confusion_m=confusion_matrix(y_test,y_pred)print(confusion_m)

#Getting the Accuracy
fromsklearn.metricsimportaccuracy_scoreaccuracy=accuracy_score(y_test,y_pred)print(a
ccuracy)

#Dumping Models import pickle
pickle.dump(classifier,open('model.pkl','wb'))

```

```
pickle.dump(cv,open('cv-model.pkl','wb'))
```

app.py:

```
import numpy as np
```

```
from flask
```

```
import Flask, request,jsonify,render_template
```

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

```
import pickle
```

```
#Loading the Multinomial Naïve Bayes model and Count Vectorizer object
```

```
filename='model.pkl'
```

```
classifier=pickle.load(open(filename,'rb'))
```

```
cv=pickle.load(open('cv-model.pkl','rb'))
```

```
app=Flask(_____name)
```

```
@app.route('/')defhome():
```

```
Return render_template('form.html')
```

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

```
If request.method=='POST':message=request.form['message']data=[message]
```

```
vect=cv.transform(data).toarray()
```

```
my_prediction=classifier.predict(vect)
```

```
return
```

```
render_template('form.html')
```

```
prediction =my_prediction)
```

6. RESULTS

We use Machine Learning models to evaluate a model. At back end Deep learning algorithms like ANN (Artificial Neural Network) is used to evaluate a model .The following algorithms are used

1. *Naive Bayes*

It is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

2. *Random Forest*

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression in ML.

It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset". Instead of relying on one decision tree, the random forest takes the prediction from each tree based on the majority votes of prediction, and it predicts the final output.

At Backend ANN algorithm is used

1. *ANN (Artificial Neural Network)*

Artificial Neural Networks are computational models and inspired by the human brain. Many of the recent advancements have been made in the field of Artificial Intelligence, including Voice

Recognition, Image Recognition, and Robotics using Artificial Neural Networks. Artificial Neural Networks are the biologically

Inspired simulations performed on the computer to perform certain specific tasks like

- Clustering
- Classification
- Pattern Recognition

In the below figure 6.1 a positive input is given and the model predicts the review as a positive review which is above 3 star rating (figure 6.2)

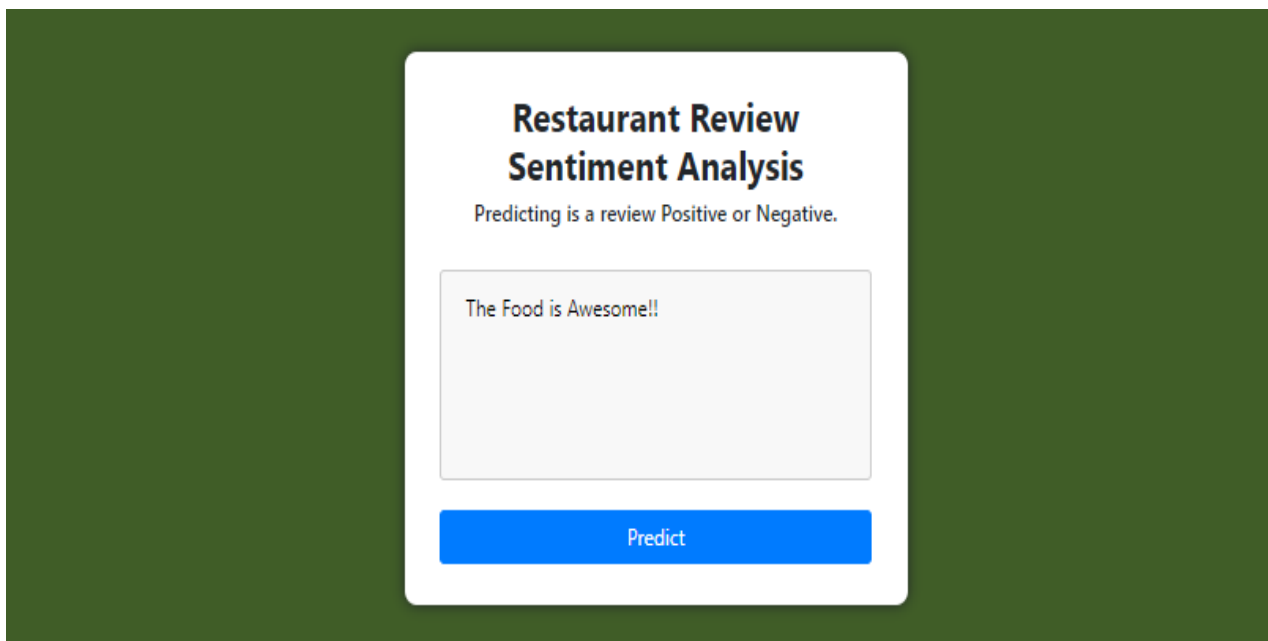


Figure 6.1: Output Screen 1

The below is the output of the given review (Food is very tasty) that the given input is good prediction with 3 and above star rating

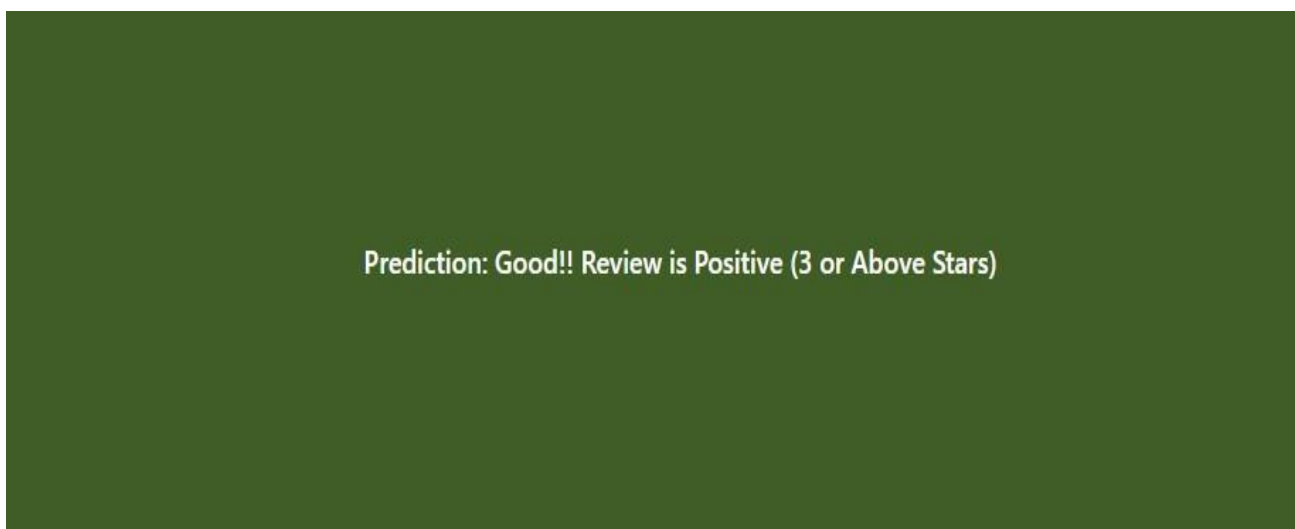


Figure 6.2 Output Screen 2

In the below figure 6.3 the machine learning model using Naïve Bayes algorithm classifies the input given as a negative review which is below 3 star rating (fig 6.4)

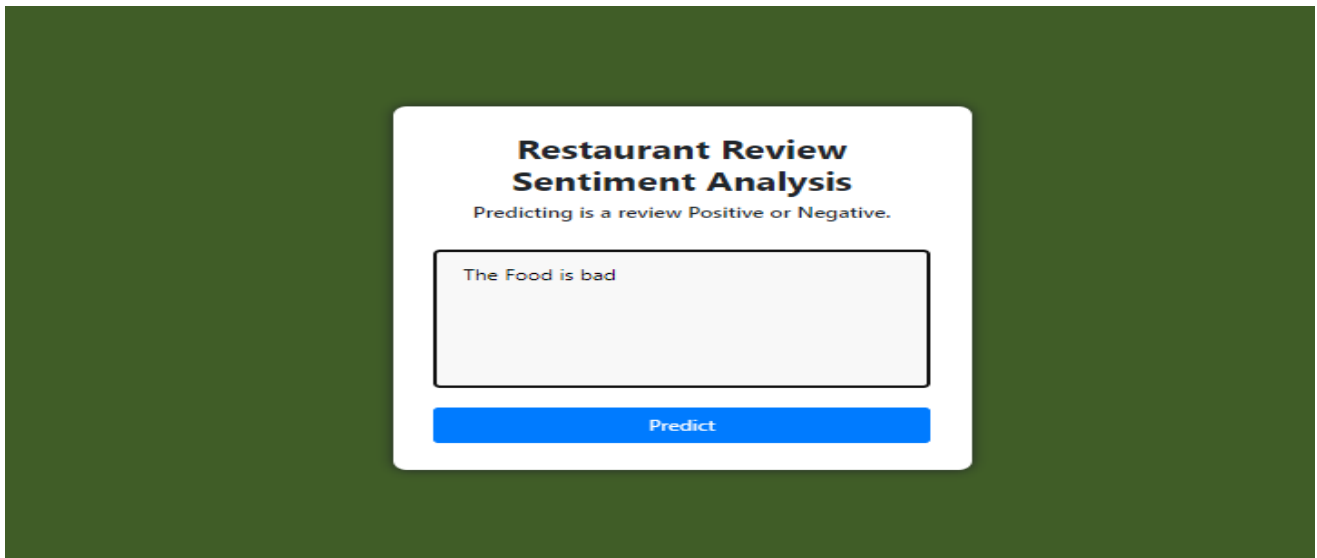


Figure: 6.3 Output Screen 3

The below is the output of the given review (Food is bad) that the given input is bad prediction with less than 3 star rating



Figure: 6.4 Output Screen 4

7. CONCLUSION

In this project, we use a supervised machine learning algorithm – MultinomialNB for the classification of restaurant reviews. We took all of the stop words out of the dataset that was provided to us, then we did stemming on it to optimize it. Next, vectorization was applied to textual data to make it numerical. For the machine learning model- MultinomialNB, we later adopted this vectorization technique. We achieved an accuracy of 89.36% with the MultinomialNB classification model. Reviews are placed into one of two categories: positive or negative. A robust method like MultinomialNB is arguably the most optimistic one due to it being more versatile. We've compared many review samples, and based on this information, we believe the MultinomialNB model is better than alternative algorithms for virtually all assessment metrics. There are many available prediction techniques, however fitting and forecasting the output requires minimal time, which is why it may be used in real-time classification systems.

8. FUTURE SCOPE AND ENHANCEMENT

In this project, internet evaluations have demonstrated the relevance of NLP and text analysis technologies are impacting businesses in a major way. Text summary accuracy and readability may be improved with machine learning model designs that employ Naïve Bayes. Possessing the capacity to succinctly summarize text reviews and uncover the significance of information is bound to drastically impact analytics and the use of data in businesses. Because there are still so many other ways to look into something, it is more exciting to study and grow more. Additionally, research in this field has a lot of potential for further development. However, results from this study may also be applicable to many other aspects of life, including hospitality, transportation, education, and healthcare. In future, one can even collect data which includes videos and images for analyzing and we can use more Deep Learning and Neural network to implement our project in Future