iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52148

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Opinion and Suggestion Mining on Customer Reviews

Shraddha Verma[1], Rutvik Kshirsagar[2], Khushboo Kulkarni[3], Sanyukta Deshmukh[4], Shubhangi Bhagat[5]
[1, 2, 3, 4]*B.E. Student, Tssm's Bhivarabai Sawant College of Engineering and Research, Pune (Maharashtra), India*
[5]*Assistant Professor, Tssm's Bhivarabai Sawant College of Engineering and Research, Pune (Maharashtra), India*

Abstract: ***Opinion and suggestion mining on customer reviews is a system providing both organizations and customers with valuable insights into the quality of products and services. The system operates on two perspectives, the customer, and the organization. For new customers, the system analyses reviews from previous customers to help them determine whether the product or service offered by the organization is worth purchasing. For organizations, the system performs suggestion mining on negative reviews to provide actionable feedback for improvements. Additionally, the system provides organizations with the percentage of customers who have liked or disliked their products or services, offering valuable insights into customer satisfaction. Overall, this project offers a valuable tool for organizations looking to improve their offerings and for customers seeking reliable information to inform their purchasing decisions.***
Keywords: opinion mining, evaluation metrics, sentiments, features, frequency, Online Reviews

I. INTRODUCTION

In today's world for an organization to grow it is very important to satisfy the needs and expectations of the customer. To understand what a customer expects from the product or the service that the organization provides it is very important for the organization to analyze the customer reviews and get working to improve the quality of service or product according to the reviews.

Also, when we talk on the part of the customer, customer gets confused over a wide range of reviews given by the people who've already used the service or the product.

This wide range of reviews may contain some positive as well as negative reviews. In order for a new customer to analyse if he/she should buy the product or service or not they have to go through a bulk of reviews and that is nearly impossible to be analysed by an individual and come to a conclusion.

The main motive of this project is to help the organization to analyse the percentage of positive and negative reviews over their particular product or the service and based on the negative reviews our project gives them improvement tips and suggestions using which they can improve the quality of the service or the product.

II. AIMS AND OBJECTIVE

Our aim is to provide good improvement suggestions for the organizations so that they can work upon the improvement suggestions provided based upon the negative reviews that the customers have given for their product or service and act accordingly in applying those improvement suggestions on their product or service. On the customer perspective our aim is to provide a new customer an easy way to analyze the bulk of reviews and help them make a decision if they can opt to buy service or the product that the organization offers. A ready-to-use model that works as stated above on the organization as well as the customer's perspective is made for industries to analyze different customer reviews according from the variant opinions made for the product or the service that the organization offers.

III. LITERATURE REVIEW

Opinion mining is a computational technique used to extract subjective information from textual data. In recent years, opinion mining has become increasingly important due to the proliferation of social media platforms, which generate a large volume of unstructured data containing opinions, reviews, and feedback. Previously there have been many projects based on opinion mining for customer review which gave you the sentiment of a single review.

©IJRASET: All Rights are Reserved

The following are the review papers on opinion mining for single customer review

1) In 2020, Zhang et al. proposed a novel approach to improve the accuracy of opinion mining in social media posts by incorporating domain-specific knowledge. They used a graph-based approach to represent domain-specific knowledge and achieved significant improvements in sentiment classification accuracy compared to traditional methods.

2) In 2021, Jha et al. presented a study on emotion detection in Indian social media data that uses a deep learning-based approach. They used a multi-layer perceptron network to classify emotions and achieved high accuracy on a Hindi emotion detection dataset.

3) In 2022, Liu et al. proposed a novel approach to aspect-based sentiment analysis that integrates human attention mechanisms and domain-specific knowledge. They used a hierarchical attention mechanism to capture the importance of different aspects and achieved state-of-the-art results on several benchmark datasets.

IV. METHODOLOGY

A. Dataset

In this study, we employed a dataset made up of 38932 reviews which has happy and unhappy sentiments associated to each.

B. Method

Opinion mining and suggestion mining is a machine learning program in which we have used the classification algorithm and trained our machine learning model based on reviews associated to happy and unhappy classes.

	Description	Is_Response
1	Description	Is_Response
2	The room was kind of clean bu	not happy
3	I stayed at the Crown Plaza Ap	not happy
4	I booked this hotel through	not happy
5	Stayed here with husband and	happy
6	My girlfriends and I stayed	not happy
7	We had - rooms. One was very	happy
8	My husband and I have stayed	not happy
9	My wife & I stayed in this	happy
10	My boyfriend and I stayed at	happy
11	Wonderful staff, great locatior	not happy
12	Steps off Times Square, nice ro	happy

Fig 1 – Head rows of dataset used

C. Data Mining

We have collected customer reviews related to the product or brand from various platforms such as Amazon, Flipkart, Twitter, Facebook and Instagram. The data was collected using a web scraping tool and stored in a database for further processing. We have used web mining which is the method that applies data mining procedures for analyzing patterns from the Web. Web usage mining, web content mining and web structure mining are three different types of web mining.

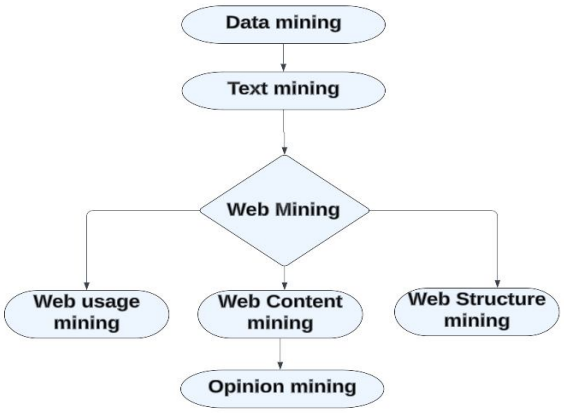


Fig 2 – Data mining and classification

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

D. Input feature

Description from the dataset shown in Fig1 are the reviews that have been extracted during the data collection process, associated to them are the response sentiments which are also the expected output that the project will be predicting.

E. Data Preprocessing

The collected data has been preprocessed to remove noise and irrelevant information such as URLs, hashtags, mentions also emoticons. The text data is further tokenized, lowercased, and lemmatized to reduce the vocabulary size and ensure consistency in the text data. All the rows that contained null values are also dropped to get clean data for model training.

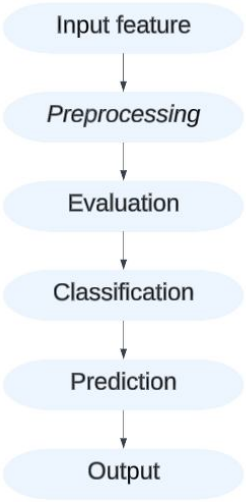


Fig 3- Pipeline of proposed system

F. Model Training

For the working of customer perspective, we have trained our model using the classification algorithm which helped us segregate the positive and negative reviews and on similar note for the working of organization's perspective we have further trained our model using Apriori algorithm which helped us remove unnecessary stop words from negative reviews and get suitable and sensible words from the negative reviews.

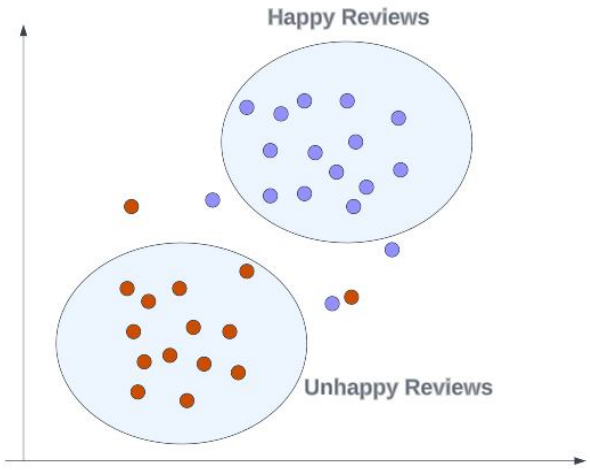


Fig 4 -Model trained to classify reviews into happy unhappy reviews.

©IJRASET: All Rights are Reserved

G. Model Evaluation

We evaluated the trained model using a test dataset that contains customer reviews with sentiment labels as 'happy' and 'not happy'. The evaluation metrics we used included accuracy, precision, recall, and F1-score.

V. WORKING OF MODEL

Our project titled by opinion mining of customer review deals with providing solutions to two perspective the customer perspective and the organizations perspective

Customer reviews play a major role in identifying if the service or the product provided by organization is good to use or not.

A. Model Working Based on Customer Perspective

Our project helps the new customers to know if the product or service is good or not , based upon the bulk of customer reviews that have previously used the product or the service of the organization. It becomes hard for the new customers to analyze if the product or service is good or not if they have to go through each and every review and sit analyzing if the product has positive reviews to the maximum or negative reviews to the maximum.

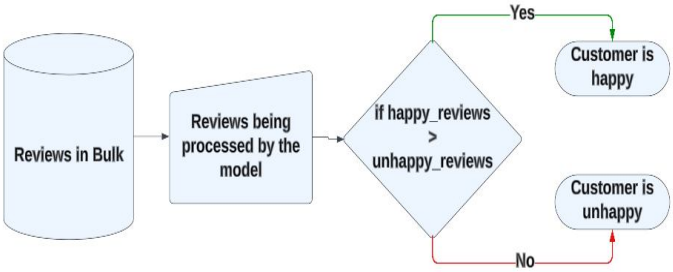


Fig.5- Processing flow of reviews in bulk

Our project deals with solving this problem of the customers by providing a single line of review stating that the product is good or bad based upon the analysis of the whole bulk of reviews uploaded to our project. This is how a project solve the problem on the basis of customer perspective.

B. Model Working Based on Organization's Perspective

Our project not only helps the customers to get an easy understanding about the product of the service provided by organization but also helps the organization to understand where are they getting wrong in providing the service or the product and how can they improve the quality of the service or the product

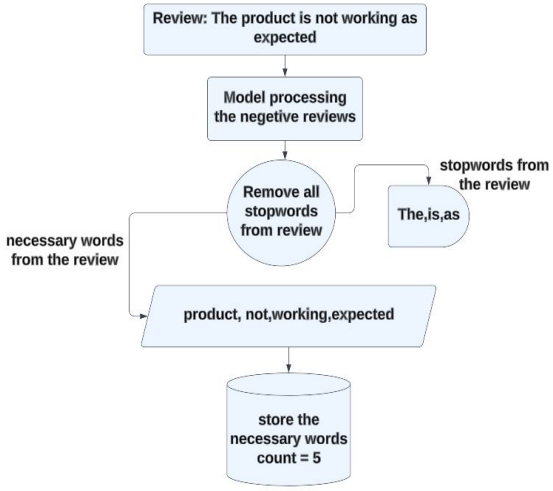


Fig 6- Removal of stop word and getting necessary word count from negative reviews.

Once the bulk of the reviews is analyzed and we get the final outcome if the product or the service provided by the organization is good or not then our project does is analyses all the reviews and gives out the top 5 worst rated reviews.
This part of a project is only provided to the organizations so that they can analyze where are they going wrong in providing their service or product in terms of quality. Once the top 5 worst reviews are displayed on the screen our project also provides a suggestion about how to improve the quality of the service or the product which can lead to customer satisfaction.
Our project also provides the percentage of positive and negative reviews from the bulk of the reviews so that its easy for the company to understand how much percent of people are not liking their product or service.

```
Improvement Tips for Westin Los Angeles Airport:
1. Ensure the front desk staff is adequately trained to provide quick, efficient service.
2. Improve the ticket and pay system for self-parking to make it easier for guests.
3. Provide concierge services to help guests find local attractions and dining options.
4. Have attendants at the front desk to assist guests.
5. Clean and maintain the rooms to a higher standard.
6. Invest in soundproofing materials to reduce jet engine noise from the nearby airport.
7. Ensure hot water is available on demand to all guests.
8. Post staff members on each floor to help guests with any needs or concerns.
9. Offer discounts or free nights to guests who experience issues with the hotel.
10. Proactively contact guests in the event of an alarm malfunction to ensure their safety.
```

Fig 7- Suggestion mining based on negative reviews

This is how a project deals with providing solution on the organization's perspective and working in hand to provide best services and products for their customers.

VI. SOFTWARE REQUIREMENTS

1) Operating System: Windows, Linux or MacOS
2) Programming Language: Python
3) Tool: Anaconda IDE or Jupyter Notebook
4) Database: MYSQL, CSV File

VII. CODE

```
Opinion and Suggestion Mining on Customer Reviews
```

In [2]:

```
import nltk
import pandas as pd
from nltk import FreqDist
import seaborn as sns
nltk.download('stopwords')
import openai
Reviewdata=pd.read_csv('train.csv')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\Lenovo\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

In [3]:

```
Reviewdata.shape
```

Out[3]:

```
(38932, 5)
```

In [4]:

```
Reviewdata.head()
```

Out[4]:

	User_ID	Description	Browser_Used	Device_Used	Is_Response
0	id10326	The room was kind of clean but had a VERY stro...	Edge	Mobile	not happy
1	id10327	I stayed at the Crown Plaza April -- - April -...	Internet Explorer	Mobile	not happy
2	id10328	I booked this hotel through Hotwire at the low...	Mozilla	Tablet	not happy
3	id10329	Stayed here with husband and sons on the way t...	InternetExplorer	Desktop	happy
4	id10330	My girlfriends and I stayed here to celebrate ...	Edge	Tablet	not happy

In [4]:

```
Reviewdata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 38932 entries, 0 to 38931
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   User_ID       38932 non-null  object
 1   Description   38932 non-null  object
 2   Browser_Used  38932 non-null  object
 3   Device_Used   38932 non-null  object
 4   Is_Response   38932 non-null  object
dtypes: object(5)
memory usage: 1.5+ MB
```

In [5]:

```
Reviewdata.describe().transpose()
```

Out[5]:

	count	unique	top	freq
User_ID	38932	38932	id17445	1
Description	38932	38932	From the reviews on this site, my boyfriend an...	1
Browser_Used	38932	11	Firefox	7367
Device_Used	38932	3	Desktop	15026
Is_Response	38932	2	happy	26521

In [6]:

```
count=Reviewdata.isnull().sum().sort_values(ascending=False)
percentage= ((Reviewdata.isnull().sum()/len(Reviewdata)*100)).sort_values(ascending=Fals
missing_data= pd.concat([count,percentage],axis=1,keys=['Count','Percentage'])
print('Count and percentage of missing values for columns : ')
missing_data
```

```
Count and percentage of missing values for columns :
```

Out[6]:

	Count	Percentage
User_ID	0	0.0
Description	0	0.0
Browser_Used	0	0.0
Device_Used	0	0.0
Is_Response	0	0.0

In [13]:

```
from sklearn.model_selection import train_test_split

Independent_var = Reviewdata.cleaned_description_new
Dependent_var=Reviewdata.Is_Response
IV_train, IV_test ,DV_train , DV_test =train_test_split(Independent_var,Dependent_var,te

print('IV_train :',len(IV_train))
print('IV_test :',len(IV_test))
print('IV_train :',len(DV_train))
print('IV_test :',len(DV_test))
```

```
IV_train : 35038
IV_test : 3894
IV_train : 35038
IV_test : 3894
```

In [14]:

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression

tvec = TfidfVectorizer()
clf2 = LogisticRegression(solver = "lbfgs")

from sklearn.pipeline import Pipeline
```

In [15]:

```
model = Pipeline([('vectorizer',tvec),('classifier',clf2)])
model.fit(IV_train , DV_train)

from sklearn.metrics import confusion_matrix

predictions = model.predict(IV_test)

confusion_matrix(predictions, DV_test)
```

```
C:\Users\Lenovo\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:763: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown i
n:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sc
ikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-reg
ression (https://scikit-learn.org/stable/modules/linear_model.html#logisti
```

In [16]:

```
from sklearn.metrics import accuracy_score , precision_score , recall_score

print("Accuracy :", accuracy_score(predictions ,DV_test))
print("Precision : ", precision_score(predictions,DV_test,average="weighted"))
print("Recall :",recall_score(predictions, DV_test , average="weighted"))
```

```
Accuracy : 0.8823831535695943
Precision :  0.8889271415963718
Recall : 0.8823831535695943
```

Trying new reviews

In [17]:

```
example=['its a worst product']
result=model.predict(example)
print(result)
```

```
['not happy']
```

In [75]:

```
#seggregating happy unhappy reviews customer perspective

def happy_unhappy(cols_as_np):
    #All review predictions
    arr=[]
    #negetive reviews added
    bad_review_array_index=[]
    bad_review_array=[]

    for i in range(0,len(cols_as_np)):

        exp=[cols_as_np[i]]
        results=model.predict(exp)
        arr.append(results[0])
        if(results[0]=='not happy'):
            bad_review_array_index.append(i)
            bad_review_array.append(cols_as_np[i])


    h=arr.count('happy')
    nh=arr.count('not happy')
    h_percent=(h*100)/(h+nh)
    nh_percent=(nh*100)/(h+nh)

    if(h>nh):
        print('Customer is happy with your product')
    else:
        print('Customer is not happy with your product')
    print(f"{round(h_percent)}% customers are satisfied with your product")
    print(f"{round(nh_percent)}% customers are not satsified with your product")



    import matplotlib.pyplot as plt
    import numpy as np

    y = np.array([round(h_percent), round(nh_percent)])
    mylabels = ["Happy Customers", "Unhappy Customers"]

    plt.pie(y, labels = mylabels)
    plt.show()
```

In [68]:

```
# this function calls  happy_unhappy fuction after inputting the file.
#input a file function

def input_file_customer():
    import os
    path=input("Enter Path:")
    os.chdir(path)

    user_input=input("Enter File name:")
    a=user_input + ".csv"

    if os.path.exists(a):
        Reviewdata=pd.read_csv(a)
    else:
        print("File does not exists")


    cols_as_np = Reviewdata['Reviews'].to_numpy()
    cols_as_np



    happy_unhappy(cols_as_np)
```

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

In [70]:

```
def sugesstion(arr):

    import os
    import openai

    openai.api_key = "sk-6QJBCWYFbCr4tw416MavT3BlkkFJ6sBiBgOCXVeu9gMV203C"

    response = openai.Completion.create(
      model="text-davinci-003",
      prompt="Give me 10 improvement tips based on following review"+ arr[0]+arr[1]+arr[
      temperature=0.7,
      max_tokens=256,
      top_p=1,
      frequency_penalty=0,
      presence_penalty=0
    )


    sugesstions=response.choices[0].text.strip()
    return sugesstions
```

In [71]:

```
from nltk.corpus import stopwords
#action function
def Action(arr):
    ans = dict()

    for i in range(0,len(arr)):
        string= arr[i]
        temp=string.split()
        stop_words = stopwords.words('english')

        final_word_count=[]

        for j in range(0,len(temp)):
            if(temp[j] in stop_words):
                continue
            else:
                final_word_count.append(temp[j])
        ans[i]=len(final_word_count)

        final_word_count=[]
#key= index of the negetive review  value= count of suitable words

    sorted_suitable_word_count_with_index= sorted(ans.items(),key=lambda x:x[1],reverse=

    return sorted_suitable_word_count_with_index
```

In [72]:

```
#bad reviews seggregation


def unhappy(cols_as_np):

    #negetive reviews added
    bad_review_array=[]

    for i in range(0,len(cols_as_np)):

        exp=[cols_as_np[i]]
        results=model.predict(exp)
        if(results[0]=='not happy'):
            bad_review_array.append(cols_as_np[i])

    sorted_suitable_word_count_with_index=Action(bad_review_array)      # calls action f
    top5_reviews=[]
    for k in range(0,5):
        top5_reviews.append(bad_review_array[sorted_suitable_word_count_with_index[k][0]
    #print(bad_review_array[sorted_suitable_word_count_with_index[k][0]])
    top5=sugesstion(top5_reviews)
    return top5
```

In [73]:

```
#input file organization
def input_file_organization():
    Reviewdata=[]
    import os
    path=input("Enter Path:")
    os.chdir(path)

    user_input=input("Enter File name:")
    a=user_input + ".csv"

    if os.path.exists(a):
        Reviewdata=pd.read_csv(a)
    else:
        print("File does not exists")


    cols_as_np = Reviewdata['Reviews'].to_numpy()
    value_dictionary=unhappy(cols_as_np)
    print(value_dictionary)
```

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

In [74]:

```
#call this function when we click on submit dataset for opinion mining (customer)


input_file_customer()
```

```
Enter Path:C:\Users\Lenovo\Desktop\New folder
Enter File name:reviews_dataset
Customer is happy with your product
85% customers are satisfied with your product
15% customers are not satsified with your product
```



In [66]:

```
#call this function when we click on submit dataset for suggestion mining (organization)


input_file_organization()
```

```
Enter Path:C:\Users\Lenovo\Desktop\New folder
Enter File name:reviews_dataset
1. Consider using a lighter moisturizer specially formulated for those wit
h oily, acne-prone skin.
2. Look for products that are free of fragrances, alcohols, and other irri
tating ingredients.
3. Avoid over-moisturizing, as it can clog pores and contribute to breakou
ts.
4. Cleanse your skin twice a day with a gentle cleanser.
5. Use a light moisturizer in the morning, and a heavier one at night.
6. Exfoliate your skin regularly to remove dead skin cells and unclog pore
s.
7. Use a toner after cleansing to help balance the skin's pH level.
8. Apply a spot treatment to help reduce inflammation and prevent future b
reakouts.
9. Avoid touching your face throughout the day to prevent transferring bac
teria to your skin.
10. Use a non-comedogenic sunscreen to protect your skin from the sun's ha
rmful rays.
```

VIII. CONCLUSION

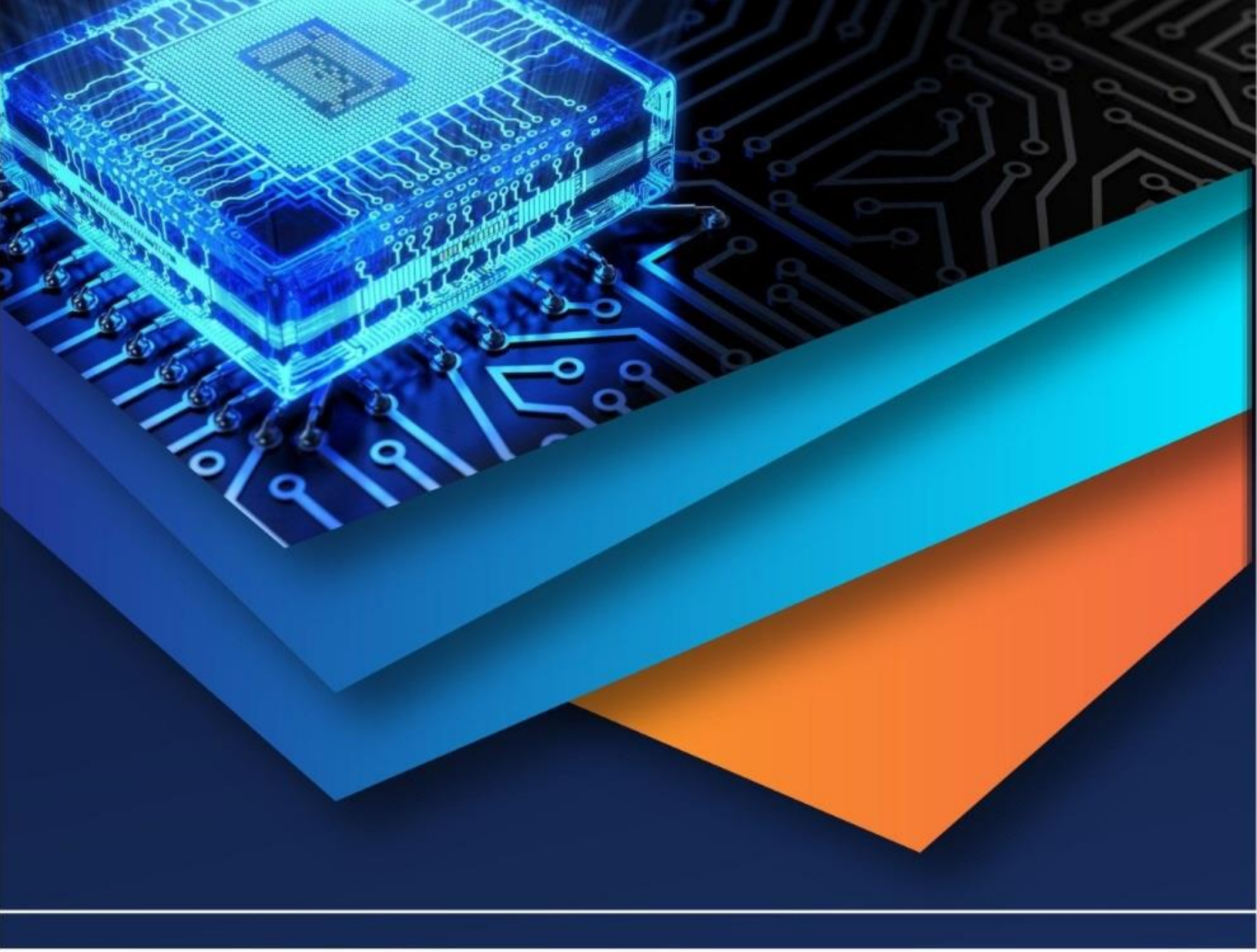
In conclusion, the implementation of our machine learning model for opinion and suggestion mining has proven to be effective in providing valuable insights to both customers and seller organizations. The model accurately analyzes and presents the percentage of positive and negative reviews for products to customers, allowing them to make informed decisions when purchasing. At the same time, organizations can get the percentage of positive and negative reviews for product, additionally model provide benefit from the suggestions provided by customers, which can help them improve their products and services enabling organizations to address specific concerns or issues that may arise. The accuracy of the suggestions provided by the model can help organizations make data-driven decisions, leading to more satisfied customers and increased profits. Overall, this project highlights the potential of machine learning in improving customer satisfaction and enhancing the performance of seller organizations.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] Opinion Mining Using Multi-Dimensional Analysis (IEEE 2023)- SATARUPA BISWAS AND G. POORNALATHA - This review addresses emerging literature in the field of Opinion Mining with a particular emphasis on user-generated textual content
[2] Opinion Mining using Machine Learning Approaches: A Critical Study (IEEE 2020) - Jayabrabu Ramakrishnan, Dinesh Mavaluru, Karthik Srinivasan, Azath Mubarakali
[3] Opinion Mining on Food Services using Topic Modelling+ and Machine Learning Algorithms (IEEE 2020) - R. Akila; S. Revathi; G. Shreedevi - real time McDonald's dataset has been used for this work to determine the positive, negative and neutral reviews.
[4] "Customer opinion mining for product improvement using machine learning" by S. R. Wadhwa et al. (2020) - The authors propose a machine learning-based approach to customer opinion mining that can be used to improve product development and customer satisfaction.
[5] "Sentiment analysis of social media data using natural language processing techniques" by S. Singh et al. (2020) - The authors explore the use of natural language processing techniques for sentiment analysis of social media data.
[6] "Opinion mining using machine learning techniques: A review" by P. Kumar et al. (2020) - This paper provides a comprehensive review of the literature on opinion mining using machine learning techniques.
[7] "Opinion mining of news articles using machine learning techniques" by A. Singh et al. (2021) - This paper investigates the use of machine learning techniques for opinion mining of news articles.
[8] "Opinion mining for political analysis: A case study of Indian election data" by S. Goyal et al. (2021) - This paper presents a case study of opinion mining for political analysis using Indian election data.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)