

**SENTIMENTAL ANALYSIS**

**WITH NLP USING PYTHON AND FLASK**

##### 

##### A MINI PROJECT REPORT

###### ***Submitted by***

**BALASA CHENCHU SAI VYSHNAVI [REG NO:211418104036]**

*in partial fulfillment for the award of the degree*

***of***

**BACHELOR OF ENGINEERING**

**IN**

**COMPUTER SCIENCE AND ENGINEERING**

**PANIMALAR ENGINEERING COLLEGE,CHENNAI-600123**

**ANNA UNIVERSITY: CHENNAI 600 025**

**APRIL 2020**

**BONAFIDE CERTIFICATE**

Certified that this project report **“SENTIMENTAL ANALYSIS WITH NLP USING PYTHON AND FLASK”**

is the bonafide work of “**BALASA CHENCHU SAI VYSHNAVI(211418104036)”** who carried out the project work under my supervision.

**SIGNATURE SIGNATURE**

**Dr.S. MURUGAVALLI, M.E., Ph.D., Dr.L. JabaSheela**

**HEAD OF THE DEPARTMENT SUPERVISOR**

**Professor-CSE Department**

DEPARTMENT OF CSE DEPARTMENT OF CSE,

PANIMALAR ENGINEERING COLLEGE PANIMALAR ENGINEERING COLLEGE,

NAZARATHPETTAI, NAZARATHPETTAI,

POONAMALLEE, POONAMALLEE,

CHENNAI-600 123. CHENNAI-600 123.

Certified that the above candidate(s) was/ were examined in the Anna University Project Viva-Voce Examination held on...........................

**INTERNAL EXAMINER EXTERNAL EXAMINER**

**ACKNOWLEDGEMENT**

We express our deep gratitude to our respected Secretary and Correspondent **Dr.P.CHINNADURAI, M.A., Ph.D.** for his kind words and enthusiastic motivation, which inspired us a lot in completing this project.

We would like to extend our heartfelt and sincere thanks to our Directors **Tmt.C.VIJAYARAJESWARI**, **Dr.C.SAKTHIKUMAR,M.E., Ph.D.** and **Tmt. SARANYASREE SAKTHIKUMAR B.E.,M.B.A.,** for providing us with the necessary facilities for completion of this project.

We also express our gratitude to our Principal **Dr.K.Mani, M.E., Ph.D.** for his timely concern and encouragement provided to us throughout the course.

We thank the HOD of CSE Department, **Dr. S. MURUGAVALLI , M.E.,Ph.D., and Dr. L. JABASHEELA** for the support extended throughout the project.

We would like to thank my all the faculty members of the Department of CSE for their advice and suggestions for the successful completion of the project.

**B.C.SAI VYSHNAVI**

**ABSTRACT**

##### Sentiment analysis or opinion mining is the computational study of people’s opinions, sentiments, attitudes, and emotions expressed in written language. It is one of the most active research areas in natural language processing and text mining in recent years. Its popularity is mainly due to two reasons. First, it has a wide range of applications because opinions are central to almost all human activities and are key influencers of our behaviors. Whenever we need to make a decision, we want to hear others opinions. Second, it presents many challenging research problems, which had never been attempted before the year 2000. Part of the reason for the lack of study before was that there was little opinionated text in digital forms. It is thus no surprise that the inception and the rapid growth of the field coincide with those of the social media on the Web. In fact, the research has also spread outside of computer science to management sciences and social sciences due to its importance to business and society as a whole. In this talk, I will start with the discussion of the mainstream sentiment analysis research and then move on to describe some recent work on modeling comments, discussions, and debates, which represents another kind of analysis of sentiments and opinions.

**TABLE OF CONTENTS**

| **CHAPTER NO.** | **TITLE** | **PAGE NO.** |
| --- | --- | --- |
|  |  |  |
|  | **ABSTRACT** |  |
|  | **LIST OF TABLES** |  |
|  | **LIST OF FIGURES** |  |
| **1.** | **INTRODUCTION** |  |
|  | 1.1 Overview |  |
|  | 1.2 Problem Definition |  |
| **2.** | **LITERATURE SURVEY** |  |
| **3.** | **SYSTEM ANALYSIS** |  |
|  | 3.1 Existing System |  |
|  | 3.2 Proposed system |  |
|  | 3.3 Technology Stack |  |
| **4.** | **SYSTEM DESIGN** |  |
|  | 4.1 ER diagram / Flow Chart |  |
|  | 4.2 Data dictionary |  |
| **5.** | **SYSTEM ARCHITECTURE** |  |
|  | 5.1 Architecture Overview |  |
|  | 5.1 Description of the Modules |  |
| **6.** | **SYSTEM IMPLEMENTATION** |  |
| **7.** | **SYSTEM TESTING** |  |
|  | Test Cases & Reports / Performance Analysis |  |
| **8.** | **CONCLUSION** |  |
|  | Conclusion and Future Enhancements |  |
|  | **APPENDICES** |  |
|  | A.1 Sample Screens |  |
|  | **REFERENCES** |  |
|  |  |  |

**1. INTRODUCTION:**

**1.1 Overview:**

Sentiment analysis is the process of using natural language processing, text analysis, and statistics to analyze customer sentiment. The best businesses understand the sentiment of their customers—what people are saying, how they’re saying it, and what they mean. Customer sentiment can be found in tweets, comments, reviews, or other places where people mention your brand. Sentiment Analysis is the domain of understanding these emotions with software, and it’s a must-understand for developers and business leaders in a modern workplace.

As with many other fields, advances in deep learning have brought sentiment analysis into the foreground of cutting-edge algorithms. Today we use natural language processing, statistics, and text analysis to extract, and identify the sentiment of words into positive, negative, or neutral categories.

One of the most well documented uses of sentiment analysis is to get a full 360 view of how your brand, product, or company is viewed by your customers and stakeholders. Widely available media, like product reviews and social, can reveal key insights about what your business is doing right or wrong. Companies can also use sentiment analysis to measure the impact of a new product, ad campaign, or consumer’s response to recent company news on social media. Private companies like Unamo offer this as a service.

Customer service agents often use sentiment or intent analysis to automatically sort incoming user email into “urgent” or “not urgent” buckets based on the sentiment of the email, proactively identifying frustrated users. The agent then directs their time toward resolving the users with the most urgent needs first. As customer service becomes more and more automated through machine learning, understanding the sentiment and intent of a given case becomes increasingly important.

**1.2 Problem Definition:**

A basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral.

**2. Literature survey:**

**Base paper: Liu B (2010) Sentiment analysis and subjectivity In: Handbook of Natural Language Processing, Second Edition.. Taylor and Francis Group, Boca.**

There are two techniques widely used to detect the sentiments from text. They are Symbolic techniques and Machine Learning techniques [3, 4]. A. Sentiment analysis using Symbolic Techniques A symbolic technique uses the availability of lexical resources. Turney [5] suggested an approach for sentiment analysis called „bag of words‟. In the mentioned approach, individual words are neglected and only collections of words are considered. He gathered word having adjectives or adverb for the polarity of review from a search engine Altavista. A lexical database called WordNet [7] was used by Kamps et al [6] which determines an emotional matter in a word. WordNet carries synonyms and distance metric to find the orientation of adjectives. To overcome obstacles in lexical substitution task, Baroni et al [8] developed a system supported by word space model formalism thereby representing local words. EmotiNet conceptually represented the text that stored the structure of real events in a domain. This was introduced by Balahur et al [9]. B. Sentiment analysis using Machine Learning Techniques Under this technique, there are two sets, namely a training set and a test set.

Generally the dataset which is collected from different sources and whose behavior and output values are known to us falls into the category of training data sets. In contrast with this, the datasets whose values or behavior are unknown to us are called as test data sets. Here different classifiers are trained with training data and then unknown data or we can say a test data is given to this model to get desired results. Machine Learning consists of various different classifiers such as Ensemble classifier, k-means, Artificial Neural Network etc. These are used to classify reviews [9]. Y.Mejova et al [1] in his research work proposed that we can use presence of each character, frequency of occurrences of each character, word which is considered as negation etc. as features for creating feature vector. He also shows that we can effectively use unigram and bigram approaches to make feature vector in Sentiment analysis. Domingos et al [11] suggested that Naive Bayes works well for dependent features for certain problem. Zhen Niu et al [12] found a new model. This model is based on Bayesian algorithm. In this model, some efficient approaches are used for selecting feature, computation of weight and classification. Barbosa et al [13] designed a 2 step analysis method which is an automatic sentiment analysis for classifying tweets. In the first step, tweets are classified into subjective and objective tweets. After that, in a second step, subjective tweets are classified as positive and negative tweets. Celikyilmaz et al [14] developed one method as pronunciation based word clustering. This method normalizes noisy tweets. There are some words which have the same pronunciation but having different meanings. So, for eliminating this conflict, there is method mentioned above. In this mentioned method, words having same pronunciation are clustered and assigned common tokens. Wu et al [15] in his paper recommended model, namely, the influence probability to analysis the sentiment tweets. In this, if @username is found in the tweet, it takes influencing action and helps to influencing probability. By collecting automatic tweets, Pak et al [16] developed a method for sentiment analysis by creating twitter corpus. In his proposed work he shows that, while creating feature vector, we can use emoticons as a feature. He used a Naïve Bayesian classifier to do the sentiment analysis. Some researches made to identify the public opinion about movies, news etc. from twitter tweets. V.M. Kiran et al [17] had taken the information from other publicly available databases like IMDB and Blippr.

**3. System Analysis:**

**3.1 Existing system:**

Existing approaches to sentiment analysis can be grouped into three main categories: knowledge-based techniques, statistical methods, and hybrid approaches. Knowledge-based techniques classify text by affect categories based on the presence of unambiguous affect words such as happy, sad, afraid, and bored.

**Disadvantage of existing system:**

But computer programs have problems recognizing things like sarcasm and irony, negations, jokes, and exaggerations - the sorts of things a person would have little trouble identifying. And failing to recognize these can skew the results.

**3.2 Proposed system:**

In theproposed system, sentiment analysis   is done using Natural Language Processing, which defines a relation between user posted tweets and opinion on the drug, and in addition, suggestions of much better medicines can be provided to the users.

**Advantages of proposed system:**

It will be able to detect changes in the overall opinion towards your brand. Because it provides insight into the way your customers are feeling when they approach you, you can monitor trends and see if overall opinion towards your company drops or rises.

**3.3 Technology Stack:**

* Python
* NLP
* Text blob
* Flask
* Html

**SYSTEM DESIGN:**

**4.1 ER Diagram/ Flow chart:**

Geo-Coordinates

Extracted from DB

Sentiment analyzer

Sentimental analysis

Data base

Stored in DB

Comment

Request for comment

festest for

Comment collector

**Figure 1: Entity relationship diagram**

Negative comment

Negative

Positive

Comment

Feature extractor

Classifier

Training set

Feature extractor

Word features

Positive comment

**Figure 2: Data flow diagram**

**4.2 DATA DICTIONARY:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Entity** | **Attribute** | **Data type** | **Description** |
| User | Text | String | Text typed by user |
| User | Input | String | Comment from user |
| User | Analysized\_comment | int[] | Positve and negative or Neutral comment respectively |
| Text System | Text | String | Comment typed by user |
| Text System | Comment\_detected | int | Positve and negative or Neutral comment respectively |

**SYSTEM ARCHITECTURE:**

**5.1 Architecture Overview:**

**Fig 3** representswe will pass the unstructed data as an input,It depends on area that we are working on.If u want to categorized movie reviews,data then the input data will be a movie reviews. It might be in plain text or excel sheet format or may be data base.

Category of the data

(output)

Decision tree/naïve bayes/SVM/random forest

(classifier)

Stemming,pos tagging,stopwords,lemmatization

(text analytics)

Unstructured data

(input)

**Figure 3: System Architecture Diagram**

**5.2 DESCRIPTION OF THE MODULES:**

1. **Text Module**

The text is written by the user and text the comment is given as input to the system. Then the data is stored in data base.The comment says whether it is positive,negative or neutral.The input is also tokenized and cleaned and is used to predict the if it is positive,negative or neutral comment.

**SYSTEM IMPLEMENTATION:**

from flask import Flask, render\_template,request,url\_for

from flask\_bootstrap import Bootstrap

**# NLP Packages**

from textblob import TextBlob,Word

import random

import time

app = Flask(\_\_name\_\_)

Bootstrap(app)

@app.route('/')

def index():

return render\_template('index.html')

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

def analyse():

start = time.time()

if request.method == 'POST':

rawtext = request.form['rawtext']

#NLP Stuff

blob = TextBlob(rawtext)

received\_text2 = blob

blob\_sentiment,blob\_subjectivity = blob.sentiment.polarity ,blob.sentiment.subjectivity

number\_of\_tokens = len(list(blob.words))

**# Extracting Main Points**

nouns = list()

for word, tag in blob.tags:

if tag == 'NN':

nouns.append(word.lemmatize())

len\_of\_words = len(nouns)

rand\_words = random.sample(nouns,len(nouns))

final\_word = list()

for item in rand\_words:

word = Word(item).pluralize()

final\_word.append(word)

summary = final\_word

end = time.time()

final\_time = end-start

return render\_template('index.html',received\_text = received\_text2,number\_of\_tokens=number\_of\_tokens,blob\_sentiment=blob\_sentiment,blob\_subjectivity= blob\_subjectivity,summary=summary,final\_time=final\_time)

if \_\_name\_\_ == '\_\_main\_\_':

app.run(debug=True)

**index.html**

{% extends "bootstrap/base.html" %}

{% block content %}

<style type="text/css">

body{

font:15px/1.5 Arial, Helvetica,sans-serif;

}

.spinner-1:before{

content: "";

box-sizing: border-box;

position: absolute;

top:50%;

left: 50%;

height: 60px;

width: 60px;

margin-top: -30px;

margin-left: -30px;

border-radius: 50%;

border:6px solid transparent;

border-top-color: blue;

animation: spinner 0.7s linear infinite;

}

.jumbotron text-center{

background-color:blue;

text-color:white;

}

@keyframes spinner {

to {

transform: rotate(360deg);

}

}

li { background-color:#BDBDBD; }

li:nth-child(odd) { background-color:green; }

</style>

<div class="container">

<div class="jumbotron text-center">

<h3><b>Natural Language Processing App by Yaswanth<b></h3>

<p>Evaluating Everyday Language</p>

</div>

</div>

<div class="container">

<form method="POST" action="{{ url\_for('analyse')}}" id="myForm">

<label ><strong>Enter Your Text Below<strong></label><br>

<textarea class="form-control" rows="20" cols="40" name="rawtext"></textarea><br>

<input type="submit" onclick="myAnalyser()" value="Submit" class="btn btn-primary ">

<input type="button" onclick="myFunction()" value="Clear" class="btn btn-outline-dark">

<a href="{{ url\_for('index')}}" type="button" class="btn btn-danger" > Reset</a>

</form>

</div>

<br/>

<hr/>

<div class="main">

<div class="container">

<div class="card">

<div class="card-header">

</div>

<div class="card-body">

<h5 class="card-title"><div class="alert alert-primary" role="alert">

This text has {{number\_of\_tokens}} tokens with {{len\_of\_words}}

</div> </h5>

<div class="card-text">

<h5>Your Text</h5>

<p style="color:#0091EA;font-family:sans-serif;">{{ received\_text }}</p>

<hr/>

<br/>

<p>Time Elapsed: <span style="color:#0091EA;">{{ final\_time }} </span> seconds to analyse</p>

<p>This text is about:</p>

{% for i in summary %}

<ul class="list-group ">

<li class="list-group-item list-group-item-info"><span style="color:black">{{i}}</span>

<a href="http://www.dictionary.com/browse/{{i}}?s=" target="\_blank" type="button" class="btn btn-outline-primary btn-sm" style="float:right;font-size:9px;color:#fff;">View</a>

</li>

</ul>

{% endfor %}

</div>

<div class="card-footer text-muted">

<table class="table table-striped table-dark" >

<thead>

<tr>

<th scope="col">Sentiment</th>

<th scope="col">Polarity</th>

<th scope="col">Subjectivity</th>

</tr>

</thead>

<tbody>

<tr>

<th scope="row">Score:</th>

<td>{{blob\_sentiment}}</td>

<td>{{blob\_subjectivity}}</td>

</tr>

</tbody></table>

</div>

</div>

</div>

{% endblock %}

<!-- Scripts starts here -->

{% block scripts %}

{{ super() }}

<script>

function myFunction() {

document.getElementById("myForm").reset();

}

</script>

<script>

function myAnalyser() {

document.querySelector('.main div').style.display = 'none';

**//Hide the main division**

document.querySelector('.main').classList.add('spinner-1');

**// Server request**

setTimeout(() => {

document.querySelector('.main').classList.remove('spinner-1');

**//Remove the animation**

document.querySelector('.main div').style.display = 'block';

//Show the main division

},5000);//Number of seconds to last

}

</script>

<!-- Prevent it from being overwritten -->

{% endblock %}

**SYSTEM TESTING:**

**Test Cases & Reports / Performance Analysis**

1. **Text Module**

|  |  |
| --- | --- |
| **Customer Feedback Text** | **Sentiment** |
| “This café is great,the staff are really friendly and the coffee is delicious” | Positive |
| “I would not recommend this café to anyone.Their coffee is terrible and is really expensive” | Negative |
| “ This café is ok and the coffee is not bad” | Neutral |

**Figure 4: customer feedback**

**CONCLUSION:**

**Conclusion and Future Enhancements:**

Sentiment analysis or opinion mining is a field of study that analyzes people’s sentiments, attitudes, or emotions towards certain entities. This paper tackles a fundamental problem of sentiment analysis, sentiment polarity categorization. Online product reviews from Amazon.com are selected as data used for this study. A sentiment polarity categorization process (Figure [2](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-015-0015-2#Fig2)) has been proposed along with detailed descriptions of each step. Experiments for both sentence-level categorization and review-level categorization have been performed.

The future of sentiment analysis is going to continue to dig deeper, far past the surface of the number of likes, comments and shares, and aim to reach, and truly understand, the significance of social media interactions and what they tell us about the consumers behind the screens. Sentiment analysis is a uniquely powerful tool for businesses that are looking to measure attitudes, feelings and emotions regarding their brand. To date, the majority of sentiment analysis projects have been conducted almost exclusively by companies and brands through the use of social media data, survey responses and other hubs of user-generated content. By investigating and analyzing customer sentiments, these brands are able to get an inside look at consumer behaviors and, ultimately, better serve their audiences with the products, services and experiences they offer.

**APPENDICES**

**A.1 Sample Screens**

|  |
| --- |
| **Customer Feedback Text** |
| “This café is great,the staff are really friendly and the coffee is delicious” |

**Figure 7: Input for text module**

|  |  |  |
| --- | --- | --- |
| **SNO** | **POLARITY** | **SUBJECTIVITY** |
| 1 | 0.7 | 0.132445647431 |

**Figure 8: output for text module**

**REFERENCES**

1. Kim S-M, Hovy E (2004) Determining the sentiment of opinions In: Proceedings of the 20th international conference on Computational Linguistics, page 1367.. Association for Computational Linguistics, Stroudsburg, PA, USA.
2. Liu B (2010) Sentiment analysis and subjectivity In: Handbook of Natural Language Processing, Second Edition.. Taylor and Francis Group, Boca