

ONLINE PRODUCT REVIEW USING SENTIMENT ANALYSIS

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Engineering degree in Computer Science and Engineering

by

GARAGA BALA SIVA SURYA HARSHA (38110066)
B.DEEPAK REDDY (38110065)



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SCHOOL OF COMPUTING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade "A" by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

MARCH - 2022



SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of GARAGA BALA SIVA SURYA HARSHA(38110066),B.DEEPAK REDDY(38110065) who carried out the project entitled“ ONLINE PRODUCT REVIEW USING SENTIMENT ANALYSIS” under my supervision from October 2021 to May 2022

**InternalGuide
(Dr.S.L.JANY SHABU)**

Head of the Department

Submitted for Viva voce Examination held on_____

Internal Examiner

External Examiner

DECLARATION

I GARAGA BALA SIVA SURYA HARSHA, B.DEEPAK REDDY hereby declare that the Project Report entitled ONLINE PRODUCT REVIEW USING SENTIMENT ANALYSIS done by me under the guidance of Dr./Prof./Mr./Ms. Dr.S.L.JANY SHABU (Internal) is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering / Technology degree In COMPUTER SCIENCE AND TECHNOLOGY.

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E., Ph.D, Dean, School of Computing** **Dr. L. Lakshmanan M.E., Ph.D. ,** and **Dr.S.Vigneshwari M.E., Ph.D. Heads** of the Department of Computer Science and Engineering for providing necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide Dr./Mr./Ms _____ for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of _____ the _____ project.

ABSTRACT

Sentiment analysis or opinion mining is the computational study of people's opinions, sentiments, attitudes and emotions expressed in written language. Create word net dictionary is all positive words are written out separately and all negative words are written out at one place. Extraction of dataset is First dataset of publicly available product reviews were downloaded from the internet and then the passage extraction framework identifies important sections of the text which is most representative of the content of the document. More specifically, involves identifying and extracting those specific product features and the opinions on them.

TABLE OF CONTENTS

CHAPTER No.	TITLE	PAGE No
	ABSTRACT	i
	LIST OF FIGURES	iv
1	INTRODUCTION	5-7
2.	LITERATURE SURVEY	8-10
3.	METHODOLOGY	8
	3.1	10-11

	3.2	11
	3.3	11-12
	3.4	12
	3.5	12-26
	3.6	27
	3.7	28
	3.8	28-31
	3.9	31-35
	3.10	35
4.	RESULTS AND DISCUSSION	40-41
5	CONCLUSION AND FUTUREWORK	41
	REFERENCES	42-44
	APPENDIX	44
	A.SOURCE CODE	44-48
	B. SCREENSHOTS	48-58

LIST OF FIGURES

FIGURE NO.	FIGURE NAME	PAGE NO.
3.7	System Architecture	28
3.10	Data Flow Diagram	35

CHAPTER-1

INTRODUCTION

Social media website is defined as “a website that facilitates meeting people, finding like minds, communicating and sharing content, and building community”; this kind of website allows or encourages various types of activities, such as commercial, social, or a combination of the two. Social media categories include digital library, e-commerce, entertainment, forum, geolocation, social bookmark, social review, social game, and social network. Social network is the subcategory of social media, which is the social structure of people who are joined by common interest. Social media are social channels of communication using web-based technologies, desktop computers, and mobile technologies. These technologies create highly interactive platforms through which individuals, communities, and organizations can share information, discuss, rate, comment, and modify user-generated and online contents. These advancements enable communication among businesses, organizations, communities, and individuals. Social media technologies change the way individuals and large organizations communicate, and they are increasingly being developed.

Wide range of applications in business and public policy uses sentiment analysis. Sentimental analysis is now being used from specific product marketing to antisocial behaviour recognition. Businesses and organizations have always been concerned about how they are perceived by the public. This concern results from a variety of motivations, including marketing and public relations. Before the era of Internet, the only way for an organization to track its reputation in the media was to hire someone for the specific task of reading newspapers and manually compiling lists of positive, negative and neutral references to the organization, it could undertake expensive surveys of uncertain validity. Today, many newspapers are published online. Some of them

publish dedicated online editions, while others publish the pages of their print edition in PDF. In addition to newspapers, there are a wide range of opinionated articles posted online in blogs and other social media. This opens up the possibility of automatically detecting positive or negative mentions of an organization in articles published online, thereby dramatically reducing the effort required to collect this type of information. To this end, Organizations are becoming increasingly interested in acquiring fine sentiment analysis from news articles. Fine-grained sentiment analysis is an extremely challenging problem because of the variety of ways in which opinions can be expressed. News articles present an even greater challenge, as they usually avoid overt indicators of attitudes. However, despite their apparent neutrality, news articles can still bear polarity if they describe events that are objectively positive or negative. Many techniques used for sentiment analysis involve naïve approaches based on spotting certain keywords which reveal the author or speaker's emotions. We use naïve performs fine-grained sentiment analysis to classify sentences as positive, negative or neutral.

Levels of Sentiment Analysis

- **Document level Sentiment Analysis**

In this Sentiment Analysis level whole document has analyzed and classify whether a whole opinion document expresses a positive or negative sentiment. In one document only reviews of one product has been reviewed. And task is to find out the opinion about that product. So this task is broadly known as document-level sentiment classification. In this level, expressed opinion is on single entity. This is not applicable when there is document which contains multiple product reviews.

- **Sentence Level Sentiment Analysis**

In this level, task goes to every sentence and determine whether the sentence expresses the positive, negative or neutral opinion. This level attentively related to Subjectivity Classification, which distinguishes objective sentences and subjective sentences. Objectives sentences express factual information about sentences where Subjective sentences express the subjective information about sentences. Many objective sentence can involve opinions. This task is known as Sentence Level Sentiment Analysis.

- **Aspect level Sentiment Analysis**

Aspect Level sentiment Analysis was earlier called Feature level (feature-based opinion mining and summarization) Sentiment Analysis. Document and Sentence Level Sentiment Analysis do not find out what exactly people like or did not like. It achieves finer-grained analysis. In this level directly looks at the opinion itself instead of looking to documents, paragraphs, sentences, clauses or phrases. This level consider the entity, aspect of that entity, opinion of aspect, opinion holder and time. Because of these parameters this level can find what actually people like means which feature of product mostly likes by customers and also on which time. This task is more interesting and more difficult too.

CHAPTER 2

LITERATURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system

such as what type of operating system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations.

'J. Serrano-Guerrero et al [1] states that Today, Digital Libraries 2.0 depends on client connections through coordinated efforts, for example and wikis. Today, language transmission frameworks with Google Wave are required for the purpose of speaking with specialists inspired by the field of exploration. This framework takes into consideration the formation of a typical space through waves for of joint effort and trade of thoughts between numerous scientists intrigued by a solitary theme. One of the fundamental impacts of this technique is hard to foresee. Z. Zhang et al [2] the review consolidates client focused substance and channel based coordinated effort with fluffy innovation and fosters an incorporated way to deal with item-based administrations.

Z.B. Sun et al [3] in this article he explains, we recommend ways of incorporating data, for example, web-based media to profit from the warning framework. is. The connection between the two clients and the meaning of the mark are utilized to decide the misrepresentation in the network of the client object. Specifically, we utilize a two-bunch calculation to decide the fitting gathering to create various cravings. Investigation dependent on genuine numbers shows that the normal technique accomplishes preferred outcomes over existing strategies. Notwithstanding, this strategy has its limits and shortcomings, which should be tended to. Another obstacle isn't knowing whether two distinct things are addressed by a similar name.

H.L. Zheng et al. The subject of how to work on the realness of the store network has turned into a significant issue in modern examination. Numerous models have been approached to resolve this issue, yet frequently center around a similar idea. In contrast to great work, in this review, we request a method for giving input blended in with the interpersonal organization of shoppers. The proposed technique enjoys numerous upper hands over typical arrangements. To begin with, it gives a solid two-venture process for distinguishing clients involving direct trust between clients in the informal community and defining limits. Second, create the conventional sifting technique in light of the Pearson

coefficient to lessen the worth of the forecast. Third, it fuses Individual Effects into the CF framework to further develop proficiency. Here we use network techniques. There is a method for tackling this issue notwithstanding other data, since it is extremely challenging to comprehend the issue that will be brought by clients who are beginning to freeze.

L.O. Colombo-Mendoza [5] clarified that internet business verification is a generally involved strategy for selling items and fulfilling clients. This article examines the Hybrid way to deal with web-based business counseling. This Hybrid methodology has been acquainted with address the issue of the customary counseling framework. In outrageous cases, the advocate doesn't fit the unexpected however proper support; Customers' thoughts and thoughts ought to be added together effectively use-history

CHAPTER 3

METHODOLOGY

3.1 EXISTING SYSTEM

- E-commerce reviews reveal the customers' attitudes on the products, which are very helpful for customers to know other people's opinions on interested products. Meanwhile, producers are able to learn the public sentiment on their products being sold in E-commerce platforms.
- Generally, E-commerce reviews involve many aspects of products, e.g., appearance, quality, price, logistics, and so on.

- Therefore, sentiment analysis on E-commerce reviews has to cope with those different aspects.
- The problem with public auction is that the participation of the general public is very limited.

3.2 DISADVANTAGES OF EXISTING SYSTEM

- Cyber Bullying
- Higher Risk of Fraud and identity Theft
- Privacy issues

3.3 PROPOSED SYSTEM

- Recommender System (RS): Special type of information filtering system that provides a prediction that assists the user in evaluating items from a large collection that the user is likely to find interesting or useful.
- Status update (micropost): Short message, shared in an online social platform, expressing an activity, state of mind or opinion.

- Folksonomy: Whole set of tags that constitutes an unstructured collaborative knowledge classification scheme in a social tagging system.
- This step involves identifying and extracting those specific product features and the opinions on them.
- The aim of the project is to socialize the auction so that people from far & wide and even across the continent can participate in it.

3.4 ADVANTAGES OF PROPOSED SYSTEM

- It is easily understand the similarities between user and system.
- It is easy to segregate groups with similar traits and assign them into clusters.

3.5 SOFTWARE ENVIRONMENT

Python:

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

- **Python is Interpreted** – Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- **Python is Interactive** – You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

- **Python is Object-Oriented** – Python supports Object-Oriented style or technique of programming that encapsulates code within objects.
- **Python is a Beginner's Language** – Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

History of Python

Python was developed by Guido van Rossum in the late eighties and early nineties at the National Research Institute for Mathematics and Computer Science in the Netherlands.

Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, SmallTalk, and Unix shell and other scripting languages.

Python is copyrighted. Like Perl, Python source code is now available under the GNU General Public License (GPL).

Python is now maintained by a core development team at the institute, although Guido van Rossum still holds a vital role in directing its progress.

Python Features

Python's features include –

- **Easy-to-learn** – Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.

- **Easy-to-read** – Python code is more clearly defined and visible to the eyes.
- **Easy-to-maintain** – Python's source code is fairly easy-to-maintain.
- **A broad standard library** – Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.
- **Interactive Mode** – Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
- **Portable** – Python can run on a wide variety of hardware platforms and has the same interface on all platforms.
- **Extendable** – You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.
- **Databases** – Python provides interfaces to all major commercial databases.
- **GUI Programming** – Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.
- **Scalable** – Python provides a better structure and support for large programs than shell scripting.

Apart from the above-mentioned features, Python has a big list of good features, few are listed below –

- It supports functional and structured programming methods as well as OOP.
- It can be used as a scripting language or can be compiled to byte-code for building large applications.
- It provides very high-level dynamic data types and supports dynamic type checking.
- It supports automatic garbage collection.
- It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

Python is available on a wide variety of platforms including Linux and Mac OS X. Let's understand how to set up our Python environment.

Getting Python

The most up-to-date and current source code, binaries, documentation, news, etc., is available on the official website of Python <https://www.python.org>.

Windows Installation

Here are the steps to install Python on Windows machine.

- Open a Web browser and go to <https://www.python.org/downloads/>.
- Follow the link for the Windows installer python-XYZ.msifile where XYZ is the version you need to install.
- To use this installer python-XYZ.msi, the Windows system must support Microsoft Installer 2.0. Save the installer file to your local machine and then run it to find out if your machine supports MSI.

- Run the downloaded file. This brings up the Python install wizard, which is really easy to use. Just accept the default settings, wait until the install is finished, and you are done.

The Python language has many similarities to Perl, C, and Java. However, there are some definite differences between the languages.

First Python Program

Let us execute programs in different modes of programming.

Interactive Mode Programming

Invoking the interpreter without passing a script file as a parameter brings up the following prompt –

```
$ python

Python2.4.3(#1,Nov112010,13:34:43)

[GCC 4.1.220080704(RedHat4.1.2-48)] on linux2

Type"help","copyright","credits"or"license"for more information.

>>>
```

Type the following text at the Python prompt and press the Enter –

```
>>>print"Hello, Python!"
```

If you are running new version of Python, then you would need to use print statement with parenthesis as in **print ("Hello, Python!");**. However in Python version 2.4.3, this produces the following result –

```
Hello, Python!
```

Script Mode Programming

Invoking the interpreter with a script parameter begins execution of the script and continues until the script is finished. When the script is finished, the interpreter is no longer active.

Let us write a simple Python program in a script. Python files have extension **.py**. Type the following source code in a test.py file –

```
print"Hello, Python!"
```

We assume that you have Python interpreter set in PATH variable. Now, try to run this program as follows –

```
$ python test.py
```

This produces the following result –

```
Hello, Python!
```

Flask Framework:

Flask is a web application framework written in Python. Armin Ronacher, who leads an international group of Python enthusiasts named Pocco, develops it. Flask is based on Werkzeug WSGI toolkit and Jinja2 template engine. Both are Pocco projects.

Http protocol is the foundation of data communication in world wide web. Different methods of data retrieval from specified URL are defined in this protocol.

The following table summarizes different http methods –

Sr.No	Methods & Description
1	GET Sends data in unencrypted form to the server. Most common method.
2	HEAD Same as GET, but without response body
3	POST Used to send HTML form data to server. Data received by POST method is not cached by server.
4	PUT Replaces all current representations of the target resource with the uploaded content.
5	DELETE Removes all current representations of the target resource given by a URL

By default, the Flask route responds to the **GET** requests. However, this preference can be altered by providing methods argument to **route()** decorator.

In order to demonstrate the use of **POST** method in URL routing, first let us create an HTML form and use the **POST** method to send form data to a URL.

Save the following script as login.html

```
<html>

<body>

<formaction="http://localhost:5000/login"method="post">

<p>Enter Name:</p>

<p><inputtype="text"name="nm"/></p>

<p><inputtype="submit"value="submit"/></p>

</form>

</body>

</html>
```

Now enter the following script in Python shell.

```
from flask import Flask, redirect, url_for, request

app=Flask(__name__)

@app.route('/success/<name>')

def success(name):

return'welcome %s'% name

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

def login():

if request.method=='POST':
```

```
user=request.form['nm']

return redirect(url_for('success',name= user))

else:

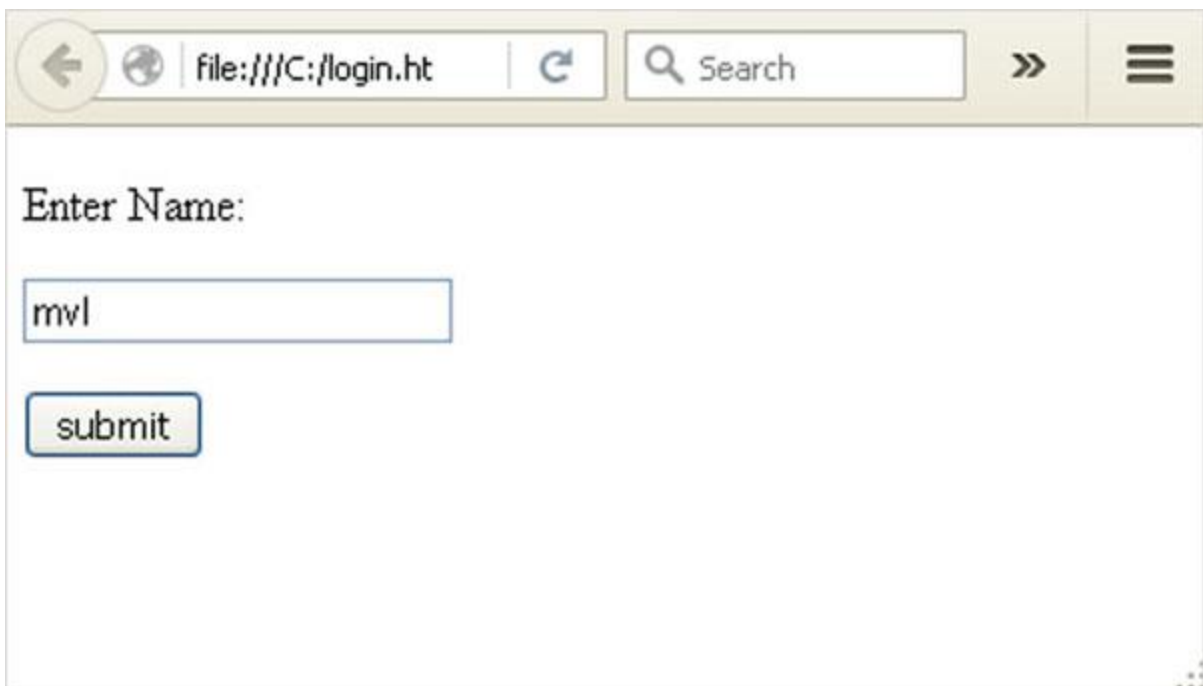
user=request.args.get('nm')

return redirect(url_for('success',name= user))

if __name__ == '__main__':

app.run(debug =True)
```

After the development server starts running, open **login.html** in the browser, enter name in the text field and click **Submit**.

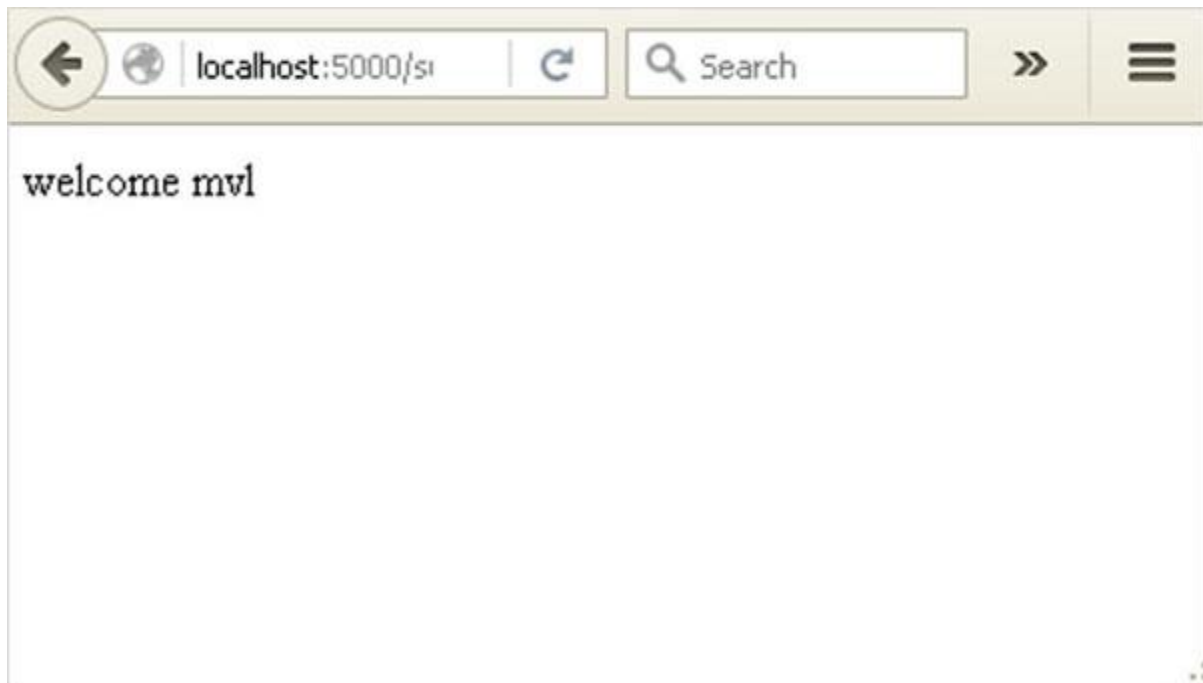
A screenshot of a web browser window. The address bar shows the file path 'file:///C:/login.ht'. The page content includes a label 'Enter Name:', a text input field with the value 'mvl', and a button labeled 'submit'.

Form data is POSTed to the URL in action clause of form tag.

http://localhost/login is mapped to the **login()** function. Since the server has received data by **POST** method, value of 'nm' parameter obtained from the form data is obtained by –

```
user = request.form['nm']
```

It is passed to **‘/success’** URL as variable part. The browser displays a **welcome** message in the window.



Change the method parameter to **‘GET’** in **login.html** and open it again in the browser. The data received on server is by the **GET** method. The value of **‘nm’** parameter is now obtained by –

```
User = request.args.get('nm')
```

Here, **args** is dictionary object containing a list of pairs of form parameter and its corresponding value. The value corresponding to **‘nm’** parameter is passed on to **‘/success’** URL as before.

What is Python?

Python is a popular programming language. It was created in 1991 by Guido van Rossum.

It is used for:

- web development (server-side),
- software development,
- mathematics,
- system scripting.

What can Python do?

- Python can be used on a server to create web applications.
- Python can be used alongside software to create workflows.
- Python can connect to database systems. It can also read and modify files.
- Python can be used to handle big data and perform complex mathematics.
- Python can be used for rapid prototyping, or for production-ready software development.

Why Python?

- Python works on different platforms (Windows, Mac, Linux, Raspberry Pi, etc).
- Python has a simple syntax similar to the English language.
- Python has syntax that allows developers to write programs with fewer lines than some other programming languages.
- Python runs on an interpreter system, meaning that code can be executed as soon as it is written. This means that prototyping can be very quick.

- Python can be treated in a procedural way, an object-orientated way or a functional way.

Good to know

- The most recent major version of Python is Python 3, which we shall be using in this tutorial. However, Python 2, although not being updated with anything other than security updates, is still quite popular.
- In this tutorial Python will be written in a text editor. It is possible to write Python in an Integrated Development Environment, such as Thonny, Pycharm, Netbeans or Eclipse which are particularly useful when managing larger collections of Python files.

Python Syntax compared to other programming languages

- Python was designed to for readability, and has some similarities to the English language with influence from mathematics.
- Python uses new lines to complete a command, as opposed to other programming languages which often use semicolons or parentheses.
- Python relies on indentation, using whitespace, to define scope; such as the scope of loops, functions and classes. Other programming languages often use curly-brackets for this purpose.

Python Install

Many PCs and Macs will have python already installed.

To check if you have python installed on a Windows PC, search in the start bar for Python or run the following on the Command Line (cmd.exe):

```
C:\Users\Your Name>python --version
```

To check if you have python installed on a Linux or Mac, then on linux open the command line or on Mac open the Terminal and type:

```
python --version
```

If you find that you do not have python installed on your computer, then you can download it for free from the following website: <https://www.python.org/>

Python Quickstart

Python is an interpreted programming language, this means that as a developer you write Python (.py) files in a text editor and then put those files into the python interpreter to be executed.

The way to run a python file is like this on the command line:

```
C:\Users\Your Name>python helloworld.py
```

Where "helloworld.py" is the name of your python file.

Let's write our first Python file, called helloworld.py, which can be done in any text editor.

```
helloworld.py
```

```
print("Hello, World!")
```

Simple as that. Save your file. Open your command line, navigate to the directory where you saved your file, and run:

```
C:\Users\Your Name>python helloworld.py
```

The output should read:

```
Hello, World!
```

Congratulations, you have written and executed your first Python program.

The Python Command Line

To test a short amount of code in python sometimes it is quickest and easiest not to write the code in a file. This is made possible because Python can be run as a command line itself.

Type the following on the Windows, Mac or Linux command line:

```
C:\Users\Your Name>python
```

From there you can write any python, including our hello world example from earlier in the tutorial:

```
C:\Users\Your Name>python
Python 3.6.4 (v3.6.4:d48eceb, Dec 19 2017, 06:04:45) [MSC v.1900 32
bit (Intel)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> print("Hello, World!")
```

Which will write "Hello, World!" in the command line:

```
C:\Users\Your Name>python
Python 3.6.4 (v3.6.4:d48eceb, Dec 19 2017, 06:04:45) [MSC v.1900 32
bit (Intel)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> print("Hello, World!")
Hello, World!
```

Whenever you are done in the python command line, you can simply type the following to quit the python command line interface:

```
exit()
```

Execute Python Syntax

As we learned in the previous page, Python syntax can be executed by writing directly in the Command Line:

```
>>>                                print("Hello,                                World!")
Hello, World!
```

Or by creating a python file on the server, using the .py file extension, and running it in the Command Line:

```
C:\Users\Your Name>python myfile.py
```

Python Indentations

Where in other programming languages the indentation in code is for readability only, in Python the indentation is very important.

Python uses indentation to indicate a block of code.

Example

```
if 5 > 2:
```

```
    print("Five is greater than two!")
```

Python will give you an error if you skip the indentation:

Example

```
if 5 > 2:
```

```
print("Five is greater than two!")
```

Comments

Python has commenting capability for the purpose of in-code documentation.

Comments start with a #, and Python will render the rest of the line as a comment:

Example

Comments in Python:

```
#This is a comment.
```

```
print("Hello, World!")
```

Docstrings

Python also has extended documentation capability, called docstrings.

Docstrings can be one line, or multiline.

Python uses triple quotes at the beginning and end of the docstring:

Example

Docstrings are also comments:

```
"""This is a
multiline docstring."""
print("Hello, World!")
```

3.6 SYSTEM REQUIREMENTS

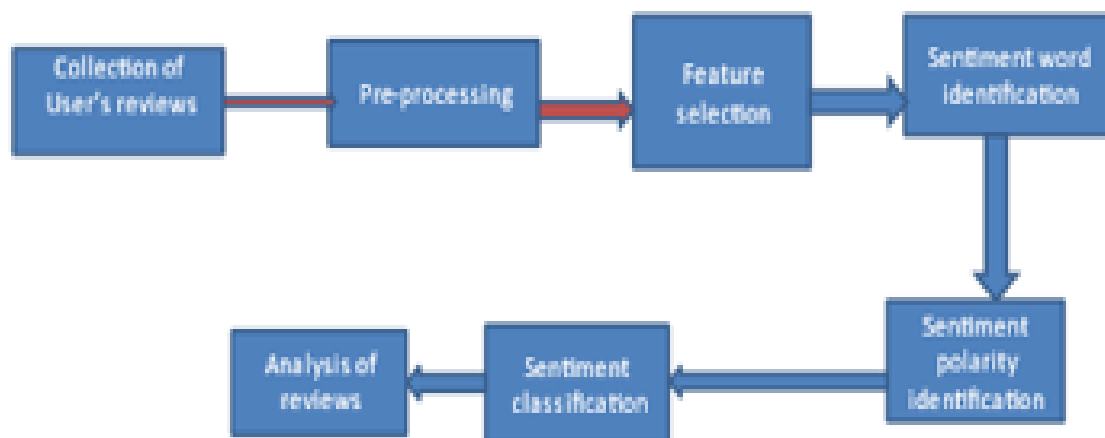
HARDWARE REQUIREMENTS

- System : i5 Processor
- Hard Disk : 500 GB.
- Input Devices : Keyboard, Mouse
- Ram : 8GB.

SOFTWARE REQUIREMENTS

- Operating system : Windows
- Coding Language : Python
- IDE : Flask web app

3.7 ARCHITECTURE DIAGRAM



3.8 MODULES

- Collection of User's Reviews
- Pre-Processing
- Feature Selection

- Sentiment Word Identification
- Sentiment Polarity Identification
- Sentiment Classification
- Analysis of Reviews

Modules Description

Collection of User's Reviews

Reviews are necessary for doing the Sentiment Analysis Task. For the Collection of reviews there are different techniques which are used in this survey. The reviews can be a structured, semi-structured and unstructured type. Sentiment Analysis research, there are open source framework where researcher can get their data for the research purpose. R is one of the programming language and software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. By installing required packages and authentication process of social website, to crawl the reviews from that site is easy task. Once we have our text data with us then we can use that data for Pre-processing purpose.

Pre-Processing

In pre-processing following are some tasks:

- Removing URLs, Special characters, Numbers, Punctuations etc.
- Removing Stop words
- Removal of Retweets (in case of twitter dataset)
- Stemming

- Tokenization

Feature Selection

Feature selection from pre-processed text is the difficult task in sentiment analysis. The main goal of the feature selection is to decrease the dimensionality of the feature space and thus computational cost. Feature selection will reduce the over fitting of the learning scheme to the training data. In different machine learning algorithms were analysed on a news review dataset with different feature selection techniques features are usually unigrams, bigrams and grams. POS tagging is used in feature selection techniques.

Sentiment Word Identification

Sentiment word identification is a fundamental work in numerous applications of sentiment analysis and opinion mining, such as review mining, opinion holder finding, and review classification. Sentiment words can be classified into positive, negative and neutral words.

Sentiment Polarity Identification

The basic task in SA is classifying the polarity of a given text at the document, sentence, or feature. The polarity is in three category i.e. Positive, Negative and Neutral. Polarity identification is done by using different lexicons which help to calculate sentiment score, sentiment strength etc.

Sentiment Classification

Sentiment classification of news review dataset and product review dataset is done using supervised machine learning approaches like naïve Bayes, SVM, Maximum Entropy etc. Accuracy is depends on which dataset is used for which classification methods. In the case of Supervised machine learning approaches Training dataset is used to train the classification model which then help to classify the test data.

Analysis of Reviews

Finally Analysis of result is important to make decision to individual and industry. In case of news reviews if more result is positive then user can decide to go that news event. Analysis is used in business intelligence.

3.9 ALGORITHMS:

Guiltlessness is an arranging cycle dependent on Base's hypothesis, which expects the autonomy of theories. More or less, Naive Bayes records it as one of the classes that doesn't has anything to do with different elements. The Vector Machine (SVM) support is an AI calculation that can be utilized in stages and cycles. Notwithstanding, it is generally expected utilized in arranging issues. In the SVM calculation, we configuration each item as an article in the n-layered space (n is the quantity of resources it has), and the worth of each item has a comparing esteem. We promptly look for and sort the hyper plane that separates the two classes well overall.

1. Naïve bayes Algorithm:

$P(X)$ due to an earlier case. The technique relies on split Bayes associated with the conclusion of the first step on the assumption of free predictors. In the presence of the fixed function of defined limits I am in the presence of a simple categorizer Bayes too much foreign matter, and the other part of the bed. Even if it is the fruit of the well of the well - to shine and the properties of each other's special occasions, a companion of the opposites of one, or to confer the degree of his evil, whence it is said, `` which is good. Words for a Naive Bayes is a simple example, so that significantly terribly useful, and huge sets for the sake of knowledge. Simplicity is still attached to a more subtle kind of nice Bayes, the developer thought.

The diagram shows the Naive Bayes formula: $P(c|x) = \frac{P(x|c)P(c)}{P(x)}$. Arrows point from labels to the terms in the formula: 'Likelihood' points to $P(x|c)$, 'Class Prior Probability' points to $P(c)$, 'Posterior Probability' points to $P(c|x)$, and 'Predictor Prior Probability' points to $P(x)$.

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Above,

- $P(c|x)$ offered the prophets the last mechanical chance (c, objective) (x, characteristic).
- $P(c)$ is the main chance to watch out.
- $P(x|c)$ is the capacity to anticipate the stage.

2. SVM Algorithm:

SVM upholds vector machines. For an informational index comprising of choices designed on an introduced name, the A SVM records models that anticipate another example order. Relegate other level/data displayed in classification 1. Assuming there are just two classifications, it tends to be shown as a paired SVM list. Here are a few kinds of SVM:

- SVM line
- Lines without SVM lines

SVM Linear Classifier:

As far as enlistment, we will more often than not accept that the mentor gives a model at home. These information focuses are planned to overcome any issues. Hyperplane forecast is straightforwardly partitioned into two phases. The main thing to do when planning a hyperplane is to diminish the separation from the hyperplane to the closest information in two stages. The hyper-plane outline is displayed as the greatest hyper-plane.

SVM Non-Linear Classifier:

Our data bundles are broadly appropriated all over the planet. Getting this data from totally various classes of hyperplants ought not be viewed as a decent choice. That is the reason Vapnik recommended making a nonlinear classifier utilizing a hyper-plane stunt. In the nonlinear SVM list, information focuses are relied upon to surpass the breaking point.

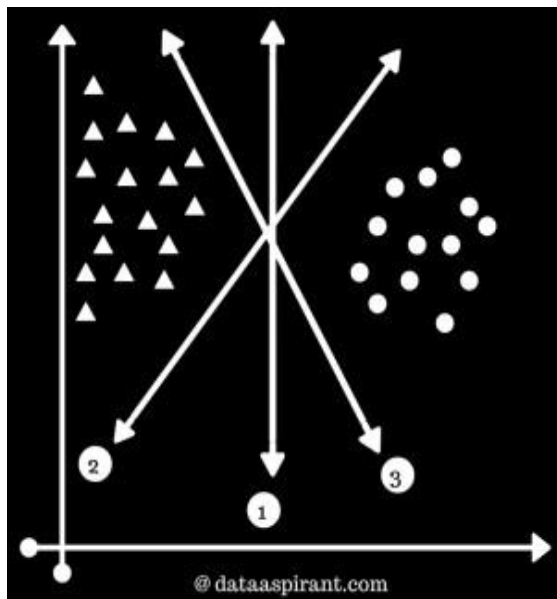
Examples of SVM boundaries:

In this section, we will figure out how to pick the best hyperplan to execute. We will show you Category 2 data. The classes are displayed in triangles and circles.

Case 1:

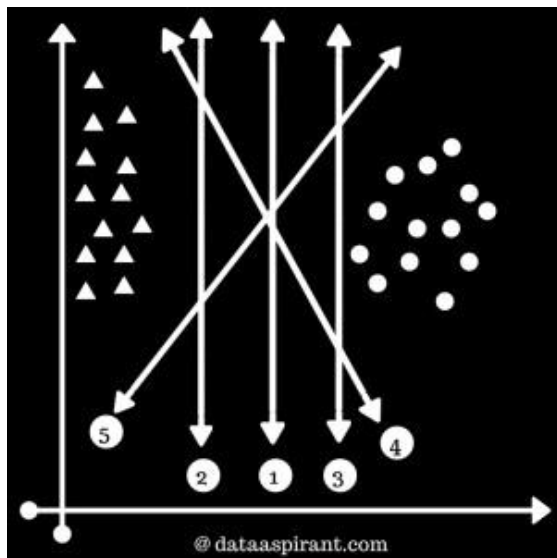
- Take a gander at the issue in Figure 2 and the data in the two unique classes. Presently we need to observe a decent hyper plane that can isolate the two classifications.
- For this situation, see Figure 1. on the option to see as the proper hyper plane In SVM, we attempt to build the distance between the hyper-plane and the closest information. This is known as an edge.
- Arrangement 1 is restricted, so it is more than the distance between the left and right sides of the example. So, our most elevated hyperplan edge will be "first".

- hyperplan edge will be "first".

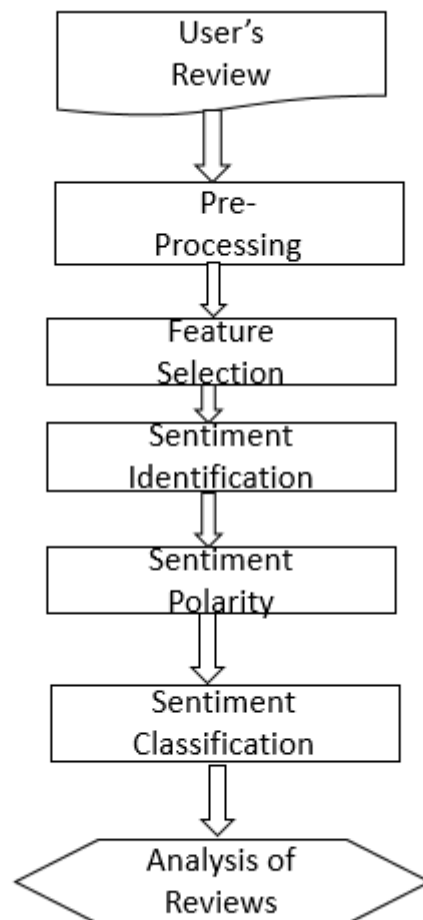


Case 2:

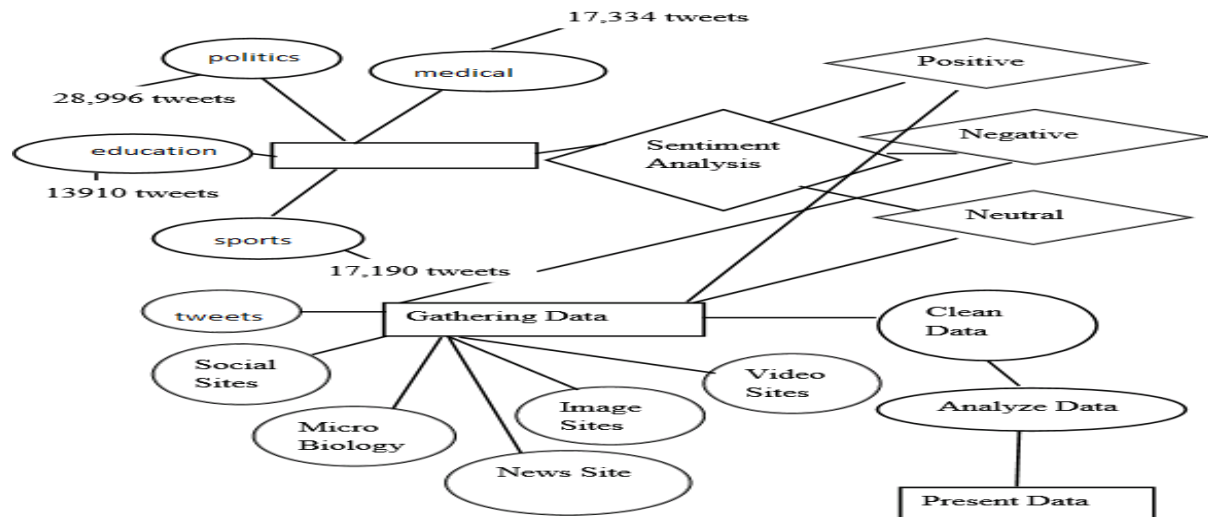
- In Figure 2, we think about two distinct classes of media. Presently we need to find a decent hyperplane that can separate between the two classes.
- Information for every class is circulated to the left or right. We will probably pick a hyperplan that can separate between classes for most extreme contrasts.
- For this situation, the choice limits are ordered, yet the limits of choice 1 demonstrate the most extreme distinction between $\bigtriangleup$ and $\bigcirc$.



3.10 DATA FLOW DIAGRAM

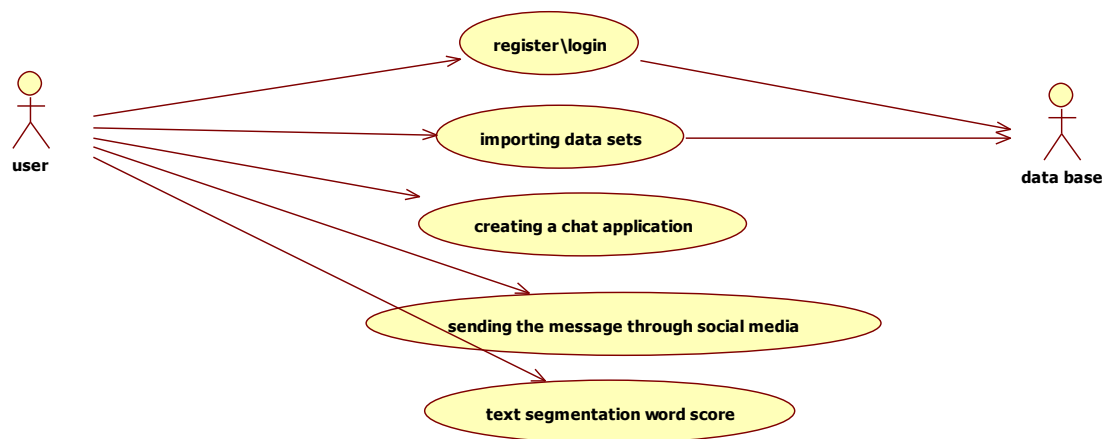


ER DIAGRAMS

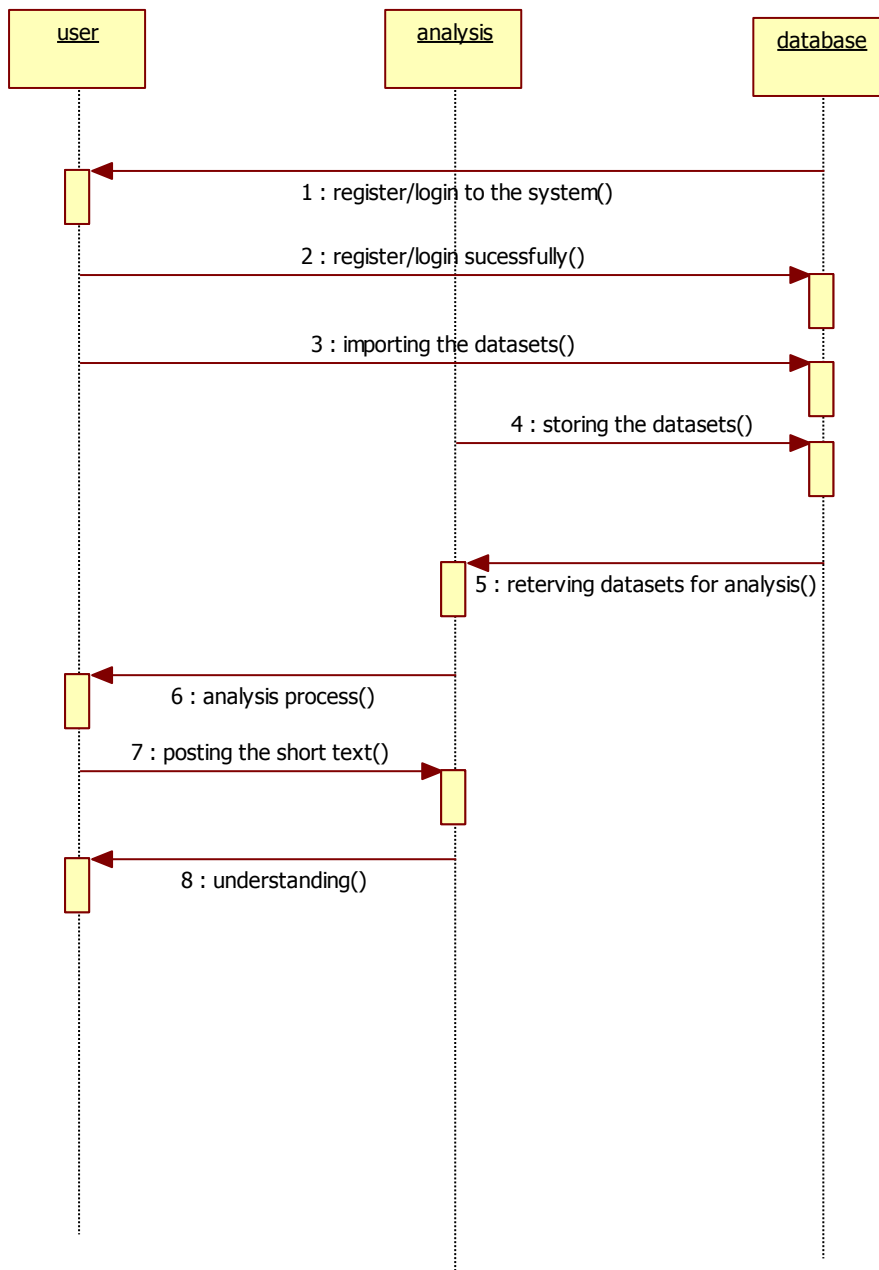


UML DIAGRAM

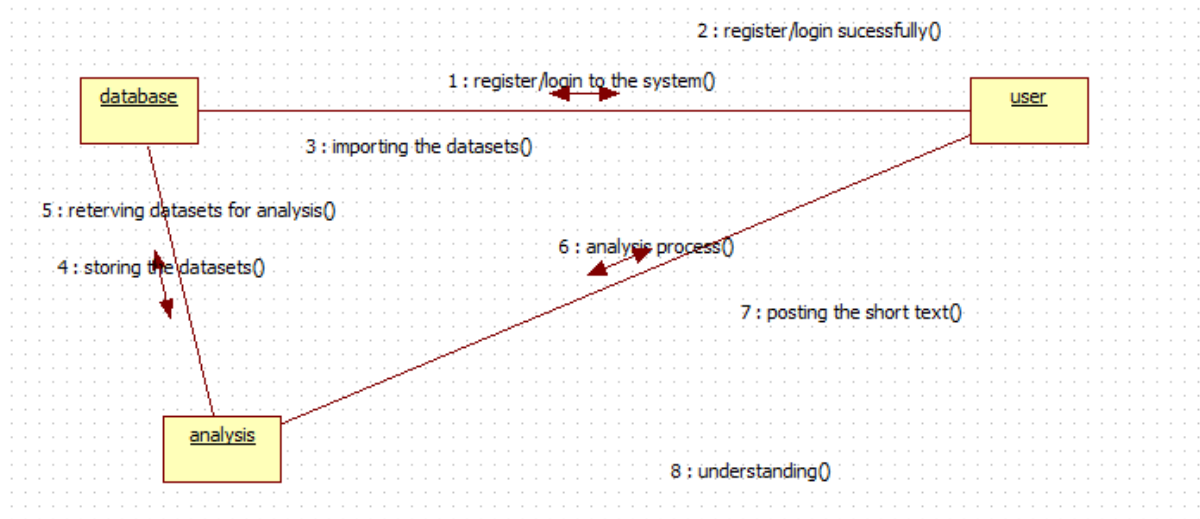
USE CASE DIAGRAM



SEQUENCE DIAGRAM

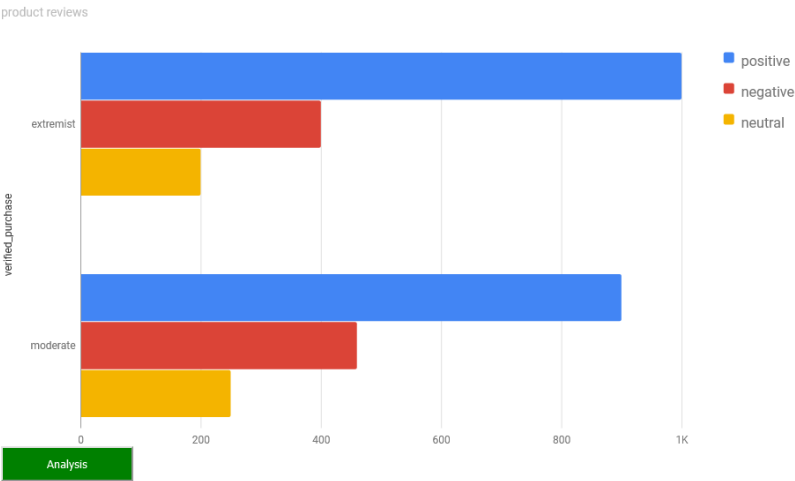
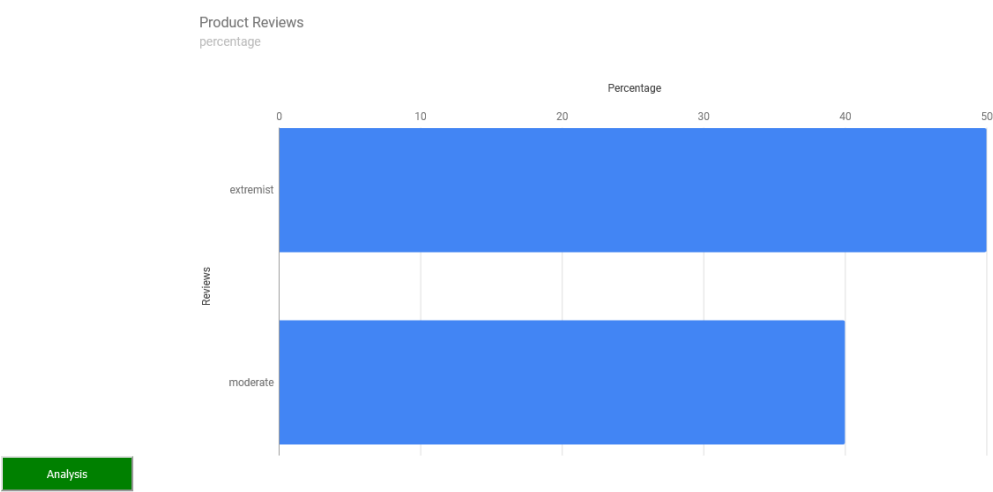


COLLABRATION DIAGRAM

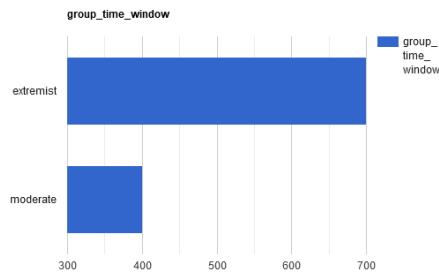


CHAPTER – IV

RESULT AND DISCUSSION



product reviews



FUTURE

CONCLUSION AND FUTURE WORK

CONCLUSION

The primary reason for this new proposed strategy is essential for the word, which was tried utilizing six notable data control frameworks at Stanford. Obviously joining text, action words, and action words has been great for consolidating various grammatical features. In this undertaking, we utilized the Twitter API utilizing open-source devices R. The Twitter account we have gathered is additionally an instrument for pre-handling. R utilizes apparatuses to open message handling instruments and download data sent by means of web-based media destinations like twitter and Facebook. The film survey was likewise set up in the R examination and analysis materials. There are different controlled and uncontrolled strategies, various word references, word references, and corpus-based techniques that are extremely useful in enthusiastic examination. Different snippets of data can be utilized, for example, film surveys, audit items, and Epinions news. This technique permits you to ascertain the quantity of positive, negative and one-sided hashtags and to foresee general assessment on what occurred. The above examination, directed on different #Hashtags of feeling investigation, shows that industry individuals can get an overall thought of what occurred.

FUTURE WORK

The subsequent stage in examining the item is to recognize the item in abundance, which can assist clients with choosing to purchase the item online at the internet business website. The responsiveness level examination gives definite data about the item or survey of the film, so the producer should know about what the client knows in the film, perhaps with a film-based investigation. Likewise with any film audit, the inn proprietor should realize what individuals like in the inn and what different clients need

REFERENCES

- [1] J. Serrano-Guerrero et al., "A google wave-based fuzzy recommender system to disseminate information in University Digital Libraries 2.0," *Information Sciences*, Vol. 181, no.9, pp. 1503- 1516, May. 2011.
- [2] Z. Zhang et al., "A hybrid fuzzy-based personalized recommender system for telecom products/services," *Information Sciences*, Vol. 235, pp. 117-129, Jun. 2013
- [3] Z.B. Sun et al., "Recommender systems based on social networks," *Journal of Systems and Software*, Vol. 99, pp. 109-119, Jan. 2015
- [4] X.L. Zheng et al., "A Hybrid Trust-Based Recommender System for Online Communities of Practice," *IEEE Transactions on Learning Technologies*, Vol. 8 no. 4, pp. 345-356. Apr. 2015.
- [5] L.O. Colombo-Mendoza et al., "RecomMetz: A context-aware knowledge-based mobile recommender system for movie showtimes," *Expert Systems with Applications*, Vol. 42, no. 3, pp. 1202- 1222, Feb. 2015.
- [6] Ji fang, Bi Chen, "Incorporating Lexicon Knowledge into SVM Learning to Improve Sentiment Classification", *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP), IJCNLP 2019*
- [7] Mika V. Mantyla, Daniel Graziotin and Miikka Kuutila, "The Evolution of Sentiment Analysis-A Review of Research Topics",

- [8] Muhammad Taimoor Khan, Mehr Durrani², Armughan Ali, Irum Inayat, Shehzad Khalid and Kamran Habib Khan “Sentiment analysis and the complex natural language” Khan et al. Complex Adapt Syst Model (2016)
- [9] Anvar Shathik,, Krishna Prasad Karani “A Literature Review on Application of Sentiment Analysis Using Machine Learning Techniques” Aug 2020
- [10] G.M.L. Sarne, “A novel hybrid approach improving effectiveness of recommender systems,” Journal of Intelligent Information Systems, Vol. 44, no. 3, pp. 397-414, Jun. 2015.
- [11] A. Tejeda-Lorente et al., “REFORE: A recommender system for researchers based on bibliometrics,” Applied Soft Computing, Vol. 30, pp. 778-791, May 2015.
- [11] Jessie Richie Naval de Los Santos “Depression among Public School Teachers in the Philippines: A News Sentiment Analysis” Jan 2021
- [12] B. Amini et al., “A reference ontology for profiling scholar's background knowledge in recommender systems,” Expert Systems with Applications, Vol. 42, no. 2, pp. 913-928, Feb. 2015.
- [13] Understanding Sentiment Analysis: What It Is and Why It's Used
Understanding Sentiment Analysis: What It Is and Why It's Used,
Available at: <https://www.brandwatch.com/blog/understanding-sentiment-analysis/>

[14] Sentiment Analysis Explained, Available at:
<https://www.lexalytics.com/technology/sentiment-analysis>

[15] Hafiz Muhammad Ahmed, Mazhar Javed Awan, Faisal Shehzad
“Sentiment Analysis of Online Food Reviews using Big Data Analytics”

Apr 2021

APPENDIX

A.SAMPLE CODE:

```
from flask import Flask,render_template,url_for,request
import pandas as pd
import pickle
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.externals import joblib
```

```
app = Flask(__name__)
```

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

```
@app.route('/index')
```

```
def index():
```

```
    return render_template('index.html')
```

```
@app.route('/login')
```

```
def login():
```

```
    return render_template('login.html')
```

```

@app.route('/abstract')
def abstract():
    return render_template('abstract.html')

@app.route('/future')
def future():
    return render_template('future.html')

@app.route('/chart')
def chart():
    return render_template('chart.html')

@app.route('/upload')
def upload():
    return render_template('upload.html')

@app.route('/preview',methods=["POST"])
def preview():
    if request.method == 'POST':
        dataset = request.files['datasetfile']
        df = pd.read_csv(dataset,encoding = 'unicode_escape')
        df.set_index('Id', inplace=True)
        return render_template("preview.html",df_view = df)

@app.route('/home')
def home():
    return render_template('home.html')

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

```

```

def predict():
    df= pd.read_csv("D:/amazon.csv", encoding="latin-1")

    # Features and Labels
    df['LABEL']
    df['REVIEW_TEXT']

    X = df['REVIEW_TEXT']
    y = df['LABEL']

    # Extract Feature With CountVectorizer
    cv = CountVectorizer()
    X = cv.fit_transform(X) # Fit the Data
    from sklearn.model_selection import train_test_split
    X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.33, random_state=42)
    #Naive Bayes Classifier
    from sklearn.naive_bayes import MultinomialNB

    clf = MultinomialNB()
    clf.fit(X_train,y_train)
    clf.score(X_test,y_test)
    #Alternative Usage of Saved Model
    # joblib.dump(clf, 'NB_spam_model.pkl')
    # NB_spam_model = open('NB_spam_model.pkl','rb')
    # clf = joblib.load(NB_spam_model)

```

```

if request.method == 'POST':
    message = request.form['message']
    data = [message]
    vect = cv.transform(data).toarray()
    prediction = clf.predict(vect)
    if prediction[0] == "__label1__":
        label="extremist"
    elif prediction[0] == "__label2__":
        label="moderate"

    return render_template('home.html',prediction = label)

@app.route('/chart1')
def chart1():
    return render_template('chart1.html')

@app.route('/uploads')
def uploads():
    return render_template('uploads.html')

@app.route('/previews',methods=["POST"])
def previews():
    if request.method == 'POST':
        dataset = request.files['datasetfile']
        df = pd.read_csv(dataset,encoding = 'unicode_escape')
        df.set_index('Id', inplace=True)
        return render_template("previews.html",df_view = df)

@app.route('/chart2')
def chart2():
    return render_template('chart2.html')

```

```
if __name__ == '__main__':  
    app.run(debug=True)
```

B.SCREENSHOTS:





GROUPS IN ONLINE PRODUCT REVIEWS

UPLOAD DATASET

★

Browse...

No file selected.

Upload

GROUPS IN ONLINE PRODUCT REVIEWS

PREVIEW



	LABEL	REVIEW_TEXT
Id		
1	__label1__	When least you think so, this product will save the day. Just keep it around just in case you need it for something.

GROUPS IN ONLINE PRODUCT REVIEWS

12	__label1__	This kettle is a gift for my daughter. She said that it is easy to use. She likes the design of it, so nice and neat.
13	__label1__	I love this tumbler! Keeps drinks warm or cold for a good period of time, and its very sturdy. I get complements on how sleek it looks all the time, especially because it doesnt leave fingerprints. Highly recommend. Only suggestion when washing the lid, be sure to shake it and dry well to prevent dripping when the lid is on the tumbler.
14	__label1__	Only giving this 3 stars because it is so cheap. Has a very mild sedative effect, other than that it's pretty useless for stress/anxiety/panic attacks. Save your money ad get some ZenRx. Night and day difference. [[ASIN:B00HBGBRY4 ZenRx - Dietary supplement reduces symptoms of anxiety, stress, depression, and panic attacks. Boost your mood, increase relaxation, and beat your anxiety with Kava kava, 5-htp, Theanine, Gaba, and more. Money Back Guarantee!]]

GROUPS IN ONLINE PRODUCT REVIEWS

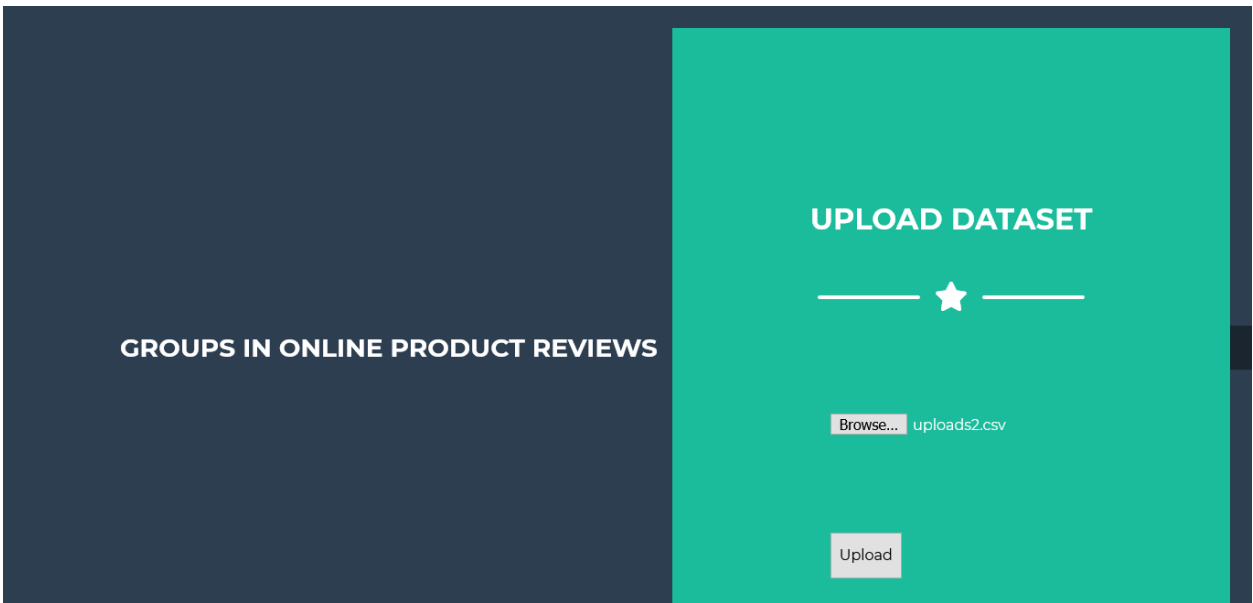
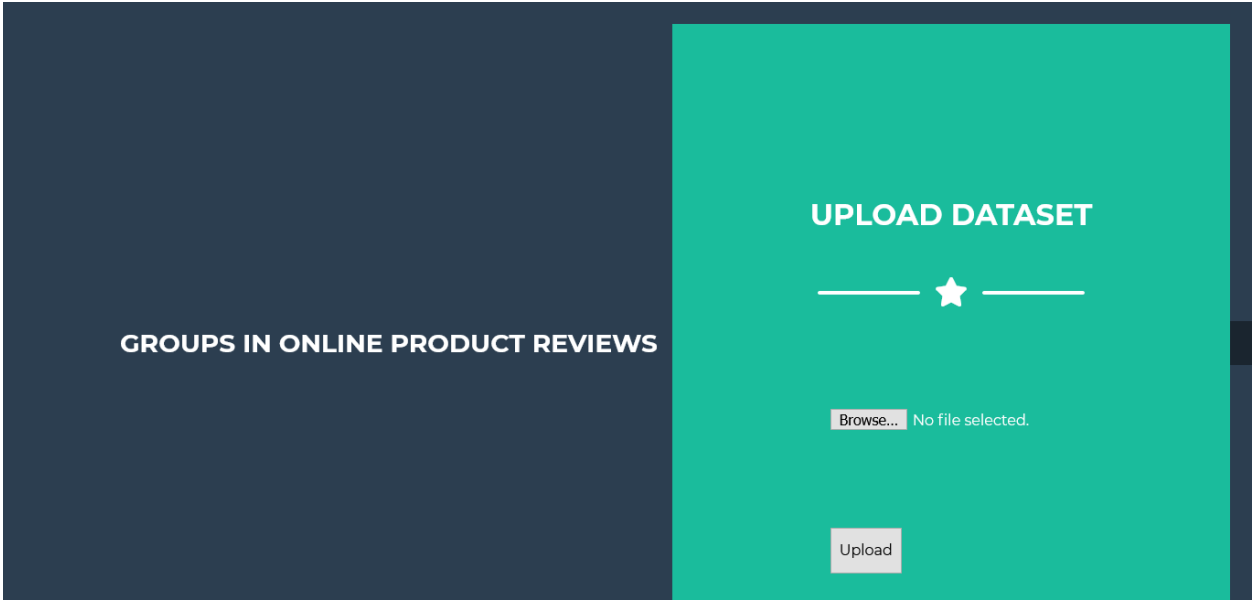
2132	__label2__	I was a bit concerned as this is a 2 pack of cables for about the same price as one cable. Yes. 2 cables. Yes. work just as well as the other cables. These are great and would buy them again.
------	------------	---

Click to Train | Test









GROUPS IN ONLINE PRODUCT REVIEWS

PREVIEW



	avg_upvotes	avg_sentiment	group_time_window	review_count	rating_deviation	early_time_window	verified_purchase	avg_rating	label
Id									
1	0.095238	0.198602	0.000000	63	0.806248	0.984127	0.904762	4.365079	0

GROUPS IN ONLINE PRODUCT REVIEWS

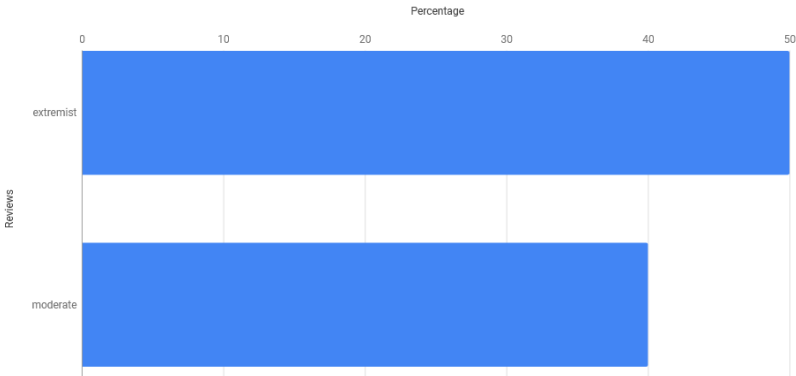
28	0.272727	0.184857	0.000000	11	0.462810	0.181818	1.000000	4.636364	1
29	0.171429	0.252348	0.000000	35	0.107755	0.264706	0.942857	4.942857	0
30	0.241379	0.292993	0.000000	29	1.060642	0.280000	1.000000	3.896552	0
31	0.200000	0.265659	0.000000	15	0.488889	0.384615	0.933333	4.666667	1
32	1.666667	0.241187	0.000000	15	0.773333	1.000000	0.866667	3.466667	1
33	1.444444	0.116904	0.000000	9	0.345679	0.000000	1.000000	4.777778	1

GROUPS IN ONLINE PRODUCT REVIEWS

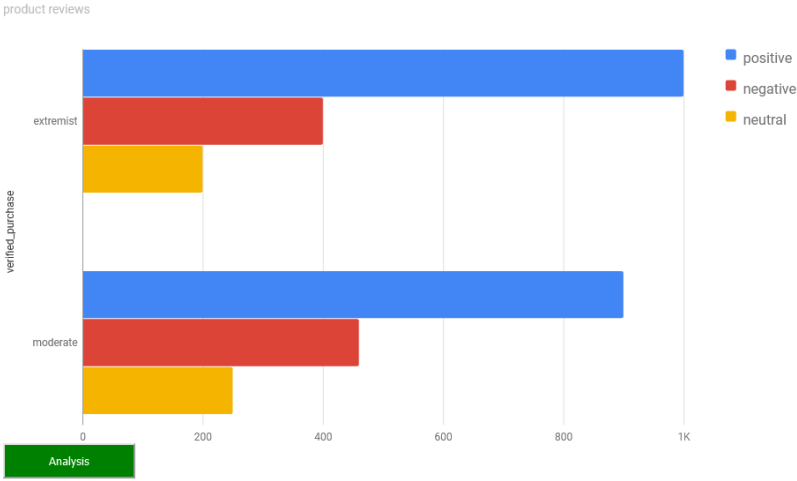
922	0.733333	0.181721	0.000000	13	0.248889	1.000000	1.000000	4.888887	1
923	0.166667	0.024429	0.000000	6	1.666667	0.000000	1.000000	3.500000	1

Click to Train | Test

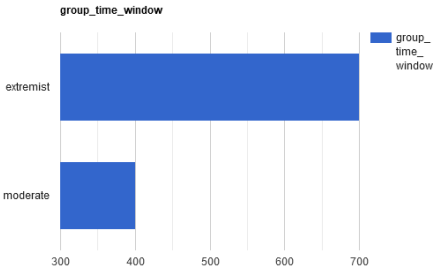
Product Reviews
percentage



Analysis



product reviews



FUTURE

PLAGIARISM REPORT

ONLINE PRODUCT REVIEW USING SENTIMENT ANALYSIS.docx

ORIGINALITY REPORT

1 %

SIMILARITY INDEX

1 %

INTERNET SOURCES

0 %

PUBLICATIONS

0 %

STUDENT PAPERS

PRIMARY SOURCES

1

www.i-scholar.in

Internet Source

<1 %

2

Pathuri Revanth, P. Venkata Krishna, A. C. Santha Sheela. "Chapter 46 Novel Embedding Based Ranking Approach for Effective Product Marketing on E-Commerce Websites", Springer Science and Business Media LLC, 2021

Publication

<1 %

3

serisc.org

Internet Source

<1 %