SENTIMENT ANALYSIS USING SOCIAL MEDIA POSTS

A PROJECT REPORT

Submitted by

MANOJ KUMAR M (410716104051)

MARIA ALEX AMALRAJ T (410716104052)

MOHAN PRASAD P (410716104057)

in partial fulfillment for the award of the degree

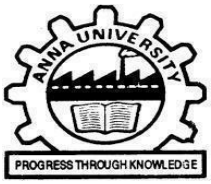
of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

DHANALAKSHMI COLLEGE OF ENGINEERING



ANNA UNIVERSITY :: CHENNAI 600 025

SEPTEMBER 2020

ANNA UNIVERSITY : CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report **"SENTIMENT ANALYSIS USING SOCIAL MEDIA POSTS"** is the bonafide work of **"MANOJ KUMAR.M (410716104051), MARIA ALEX AMALRAJ.T (410716104052), MOHAN PRASAD.P (410716104057)"** who carried out the project work under my supervision.

SIGNATURE

Dr. S Sivasubramanian, M.Tech., Ph.D.,

HEAD OF THE DEPARTMENT,

PROFESSOR

Computer Science and Engineering

Dhanalakshmi College of Engineering,

Dr. VPR Nagar,

Manimangalam,

Chennai – 601 301.

SIGNATURE

Mrs.P.Geetha, M.E., Ph.D.,

SUPERVISOR,

ASSISTANT PROFESSOR

Computer Science and Engineering,

Dhanalakshmi College of Engineering,

Dr. VPR Nagar, Manimangalam,

Chennai – 601 301.

Submitted for the project viva-voce examination held at Dhanalakshmi College of Engineering on _____

INTERNAL EXAMINER

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

We would like to thank the almighty God, beloved Parents and Friends for being a guide and a well-wisher to us throughout the project, along with their constant and invaluable support.

We are fortunate to express our heartfelt thanks to our honourable **Founder and Chairman Dr.V.P.Ramamurthi, Ph.D.,** Dhanalakshmi College of Engineering, for his guiding us and permitting us to do our project by our own.

We express our sincere gratitude and wish to thank our beloved **Principal, Dr V. Krishnakumar, M.Tech., Ph.D, MISTE**. for his support and guidance.

We extend our gratitude and heartfelt thanks to **Head of the Department Dr. Sivasubramanian, M.Tech., Ph.D.** for guiding us in all aspects of our project in each stage and providing us with valuable suggestions.

We owe our deep sense of gratitude to our **Project Coordinator Mr.Y. Sharmasth vali M.E., (Ph.D.),** who rendered valuable guidance and support always.

We would like to thank at most to our **Project Supervisor, Mrs. P.Geetha, M.E., Ph.D.** Assistant Professor who rendered valuable guidance and support throughout the project.

Finally, we take this opportunity to thank all **the Faculty members of the Department of Computer Science and Engineering** for their unwavering support and cooperation which made us keep our zeal and spirits high to complete this project work successfully**.**

TABLE OF CONTENTS

ABSTRACT

The use of Social Network Sites (SNS) like facebook, twitter, instagram is increasing nowadays especially by the younger generations. By using SNS users can express their interests, feelings,emotions and share daily routine. In many researches it is proven that using user-generated content (UGC) in a correct manner may help to determine people's mental health levels. Mining the social media post could help to predict the mental health and depression of the person.

Depression is a serious medical illness in our current generation, which reflects in work, study, eat, sleep and having fun. However, from the user profile in SNS, to collect all the information that relates to person's mood, and negativism. Social media helps to know about individual's thinking, mood, depression & socialization. So by analysing the social media post & by using some algorithms on it, able to determine the depression levels and emotions of that particular individual. Using the traditional techniques, the psychiatrist cannot get the complete and accurate information from the depressed patient.

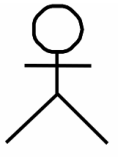
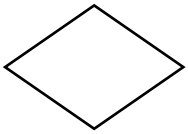
In this research, our aim is to investigate how SNS user's posts can help classify users according to mental health levels and depression.

LIST OF ABBREVIATION

S.NO	ABBREVIATION	EXPANSION
1	SA	Sentiment Analysis
2	SQL	Structured Query Language
3	DFD	Data Flow Diagram
4	SNS	Social Network Sites
5	UGC	User Generated Content
6	OCR	Optical Character Recogination
7	NLP	Natural Language Processing
8	JVM	Java Virtual Machine
9	JRE	Java Runtime Environment
10	UML	Unified Modeling Language
11	TF	Term Frequency
12	OMG	Object Management Group
13	API	Application Programming Interface
14	JCL	Java Class Library

LIST OF FIGURES

LIST OF SYMBOLS

S.NO	SYMBOL	SYMBOL NAME
1.		DATA FLOW
2.		ACTOR
3.		USE CASE
4.		DECISION
5.		PROCESS
6.		CLASS

CHAPTER 1

INTRODUCTION

Web application is a part of this project. In this project the social media posts (Facebook) are considered. Social Network Sites (SNS) is a source of data and screening tool to classify the users according to user generated contents (UGC). By using machine learning algorithms such as Naive Bayes depression level of user is classified into different levels and provides doctor's location near to user location. As depression is very serious problem which is increasing day by day, many people are suffering from this problem. In India, out of total population 7.5% of it facing this problem. It seems to be major issue and that is the reason it motivate us to worked on it .Earlier diagnosis of depressed patient were done on basis of questionnaires and its behaviour reported by his relatives or friends. But the result was not so qualitative and accurate. In contrast with that, social media is powerful tool for predicting depression levels of an individual.

Nowadays, many people are using social media platforms such as Facebook, twitter. They share their thoughts, feelings, emotions, feelings of guilt, worthlessness, helplessness and egoistic nature of individual etc. Whatever they post is related to their daily activities & happenings. Social media helps to know about individual's thinking, mood, activities & socialization. So by analysing the social media data & applying some algorithms on it, we can able to determine the depression levels of that particular individual. So it will help to diagnose that person before he /she gets more affected to it. This motivates us to do the project, so that this will help not only to psychiatrist but also used by patients who want to do self-diagnosis.

CHAPTER 2

LITERATURE SURVEY

[1] Title: A framework for depression dataset to build automatic diagnoses in clinically depressed Saudi patient by Lubana yusuf (2016).

Depression is a public health problem that has high effects on a person's functional and social relationships. Depression is a growing problem in the society. It causes pain and suffering not only to patients, but also to those who care about them. Depression disorder is hard to diagnose, because its symptoms could be confused with other disorders and has different cross-cultural symptoms. This paper proposes a framework that would best solve the problem of automatic depression detection in depressed Saudi patients. This paper particularly focuses on designing the collection of Saudi depression dataset using multiple modalities.

[2] Title: Detection of Clinical Depression in Adolescents' Speech During Family Interactions by Nammana C madague (2015).

The properties of acoustic speech have previously been investigated as possible cues for depression in adults. However, these studies were restricted to small populations of patients and the speech recordings were made during patients' clinical interviews or fixed-text reading sessions. Symptoms of depression often first appear during adolescence at a time when the voice is changing, in both males and females, suggesting that specific studies of these phenomena in adolescent populations are warranted. This study investigated acoustic correlates of depression in a large sample of 139 adolescents (68 clinically depressed and 71 controls).

[3] Title: Analysing Psychosocial Difficulties in Depression: A Content Comparison between Systematic Literature Review and Patient Perspective by Kaloyan Kamenov (2014).

Despite all the knowledge on depression, it is still unclear whether current literature covers all the psychosocial difficulties (PSDs) important for depressed patients. The aim of the present study was to identify the gaps in the recent literature concerning PSDs and their related variables. Psychosocial difficulties were defined according to the World Health Organization International Classification of Functioning, Disability and Health (ICF). A comparative approach between a systematic literature review, a focus group, and individual interviews with depressed patients was used. Literature reported the main psychosocial difficulties almost fully, but not in the same degree of importance as patients' reports.

[4] Title: An improved model for depression detection in micro blog social network by Xingu wang (2013).

Social networks contain a tremendous amount of node and linkage data, providing unprecedented opportunities for a wide variety of fields. As the world's fourth largest disease, depression has become one of the most significant research subjects. Previously, a depression classifier has been proposed to classify the users in online social networks to be depressed or not, however, the classifier takes only node features into account and neglects the influence of linkages. This paper proposes an improved model to calculate the probability of a user being depressed, which is based on both node and linkage features. The linkage features are measured in two aspects: tie strength and interaction content analysis. Moreover, the propagation rule of depression is considered for improving the prediction accuracy.

[5] Title: Toward the development of cost effective e-depression effective system by TaunD Pham (2012).

Diagnosis and prevention of depressive disorders at any scale have been attracting considerable attention of the public healthcare in Japan because depression is one of the most rapidly pervasive mental disorders in the country. A major issue that hinders the feasibility of depression screening for its prevention is the availability of some simple and cost-effective methods for depression detection and monitoring. Here in this paper, we present the development of a computerized tool for depression detection. The tool utilizes the theory of chaos and systems complexity to extract robust dynamically statistical features of physiological signals provided by the low-cost technology of photoplethysmography.

[6] Title: Depression and pain Comorbidity by Matthew J. Bair (2012).

Because depression and painful symptoms commonly occur together, we conducted a literature review to determine the prevalence of both conditions and the effects of comorbidity on diagnosis, clinical outcomes, and treatment. The prevalences of pain in depressed cohorts and depression in pain cohorts are higher than when these conditions are individually examined. The presence of pain negatively affects the recognition and treatment of depression. When pain is moderate to severe, impairs function, and/or is refractory to treatment, it is associated with more depressive symptoms and worse depression outcomes (eg, lower quality of life, decreased work function, and increased health care utilization).

CHAPTER 3

SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

- In the existing system the medical science relies on asking the patients questions about their situations, which does not diagnose the depression in a precise way.

- Electronic Health Record (EHR) systems are not optimally designed to handle integrating behavioural health and primary care.

- Using the traditional techniques, the psychiatrist cannot get the complete and accurate information from the depressed patient.

3.1.1 DISADVANTAGES OF EXISTING SYSTEM

- Its takes time to find the person was in depression.
- The patient has to attend more than one session during a period of two weeks.
- It may cause human error.

3.2 PROPOSED SYSTEM

- In this proposed system we are collect the data sets (user generated content) in Facebook.
- The SNS based system can overcome the problems regarding self-reporting from user's social activities.
- Here we are using the Naïve Bayes Classifier for train the dataset.

3.2.1 ADVANTAGES PROPOSED SYSTEM

- Psychiatrist, parents, and friends, could track the user depression by the proposed tool, which will save the time before the depressed user could get into major depression phase.
- UGC in a correct way might help to maintain people's mental health or diagnose at an early stage.

CHAPTER 4

REQUIREMENT SPECIFICATION

4.1 INTRODUCTION

The requirement specification is a technical specification of requirements for the software products. It is the first step in the requirements analysis process it lists the requirements of a particular software system including functional, performance and security requirements. The requirements also provide usage scenarios from a user, an operational and an administrative perspective. The purpose of software requirements specification is to provide a detailed overview of the software project, its parameters and goals. This describes the project target audience and its user interface, hardware and software requirements.

4.2 HARDWARE AND SOFTWARE SPECIFICATION

4.2.1 HARDWARE REQUIREMENTS

System	: Pentium IV 2.4 GHz
Hard Disk	: 40 GB
Floppy Drive	: 1.44 Mb
Monitor	: 15 VGA Colour
Mouse	: Logitech
Ram	: 512 Mb

4.2.2 SOFTWARE REQUIREMENTS

Operating system	: Windows 10
IDE	: Eclipse
Coding Language	: Java
Backend	: SQLyog

4.3 TECHNOLOGIES USED

4.3.1 JAVA TECHNOLOGY

Java technology is both a programming language and a platform. Java is a general-purpose computer programming language that is concurrent, class-based, object-oriented, and specifically designed to have as few implementation dependencies as possible. It is envisioned to let application developers "write once, run anywhere" (WORA), meaning that code that runs on one platform does not need to be recompiled to run on another. Java applications are typically compiled to byte code that can run on any Java virtual machine (JVM) regardless of computer architecture. Java is, as of 2015, one of the most widespread programming languages in use, predominantly for client-server web applications, with a reported 9 million developers. Java was originally developed by James Gosling at Sun Microsystems (which has since merged into Oracle Corporation) and released in 1995 as a core component of Sun Microsystems' Java platform. The language originates much of its syntax from C and C++, but it has fewer low-level facilities than either of them. The original and reference implementation Java compilers, virtual machines, and class libraries were originally released by Sun under

proprietary licences. As of May 2007, in compliance with the specifications of the Java Community Process, Sun relicensed most of its Java technologies under the GNU General Public License. Others have also developed substitute enactments of these Sun technologies, such as the GNU Compiler for Java (byte code compiler), GNU Class path (standard libraries), and Iced Tea-Web (browser plugin for applets).

JAVA PLATFORM

One strategy goal of Java is portability, which means that programs written for the Java platform must run likewise on any amalgamation of hardware and operating system with adequate runtime provision. This is attained by compiling the Java language code to an in-between illustration called Java byte code, instead of directly to architecture-specific machine code. Java byte code instructions are analogous to machine code, but they are envisioned to be achieved by a virtual machine (VM) written precisely for the host hardware. End users commonly use a Java Runtime Environment (JRE) installed on their own machine for standalone Java applications, or in a web browser for Java applets. Standardized libraries offer a generic way to access host-specific structures such as graphics, threading, and networking.

A foremost benefit of using byte code is porting. However, the upstairs of understanding means that interpreted programs nearly always run more slowly than programs compiled to native executable would. Just-in-Time (JIT) compilers were presented from an primary stage that compiles byte codes to machine code during runtime. Java is platform independent. But as Java virtual machine must convert Java byte code into machine language which depends on the operating system being used, it is platform dependent.

JAVA VIRTUAL MACHINE

The heart of the Java platform is the idea of a "virtual machine" that executes Java byte code programs. This byte code is the same no matter what hardware or operating system the program is running under. There is a JIT (Just in Time) compiler within the Java Virtual Machine, or JVM. The JIT compiler interprets the Java byte code into native processor instructions at run-time and caches the native code in memory during execution. The use of byte code as an middle language permits Java programs to run on any platform that has a virtual machine obtainable. The use of a JIT compiler means that Java applications, after a short postponement during piling and once they have "warmed up" by being all or mostly JIT-compiled, tend to run about as fast as native programs. Since JRE version 1.2, Sun's JVM implementation has comprised a just-in-time compiler instead of an interpreter. Although Java programs are cross-platform or platform independent, the code of the Java Virtual Machines (JVM) that implements these programs is not. Every supported operating platform has its own JVM.

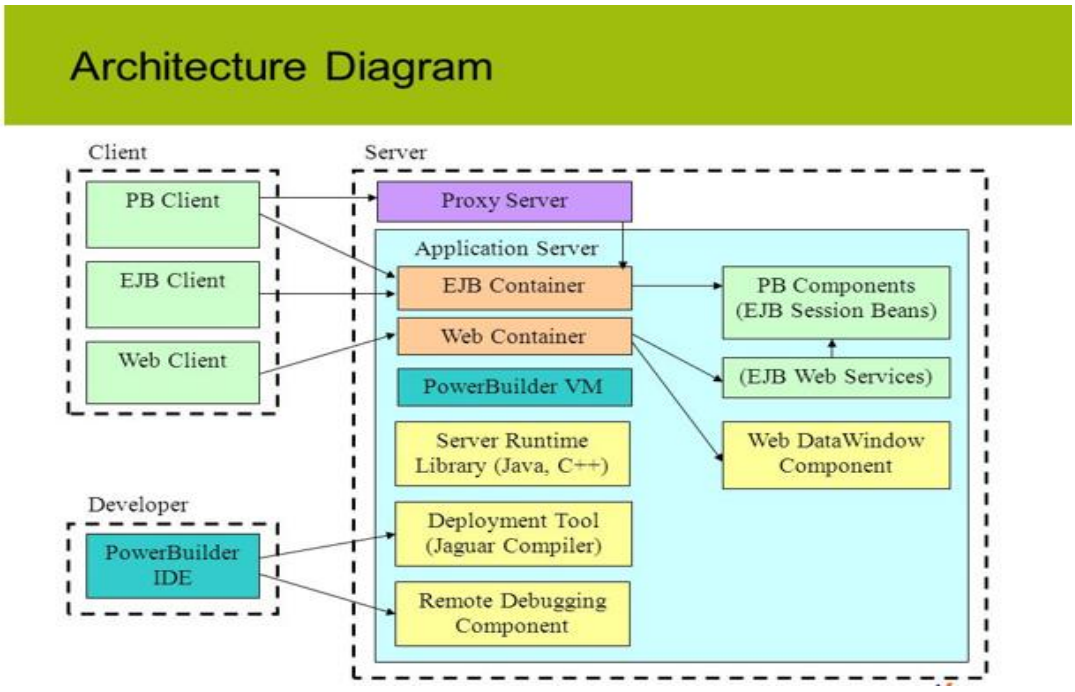


Figure 4.3.1 Java Architecture

The Oracle implementation is packaged into two dissimilar distributions: The Java Runtime Environment (JRE) which comprises the parts of the Java SE platform required to run Java programs and is proposed for end users, and the Java Development Kit (JDK), which is intended for software developers and includes development tools such as the Java compiler, Java doc, Jar, and a debugger. Open JDK is another notable Java SE implementation that is licensed under the GNU GPL.

The implementation started when Sun began releasing the Java source code under the GPL. As of Java SE 7, Open JDK is the official Java reference implementation. The goal of Java is to make all implementations of Java compatible. Historically, Sun's trademark license for usage of the Java brand insists that all implementations be "compatible". This resulted in a legal dispute with Microsoft after Sun claimed that the Microsoft implementation did not support RMI or JNI and had added platform-specific features of their own.

4.3.2 DATA MINING

There is a huge quantity of data available in the Information Trade. This data is of no use until it is changed into valuable information. It is compulsory to analyze this huge quantity of data and extract valuable information from it.

Mining of information is not the only process we need to complete; data mining also involves other processes such as Data Cleaning, Data Integration, Data Transformation, Data Mining, Pattern Evaluation and Data Presentation. Once all these procedures are completed, we would be able to use this info in many applications such as Fraud Detection, Market Analysis, Construction Control, Science Exploration, etc.

WHAT IS DATA MINING?

Data Mining is defined as extracting info from vast sets of data. In other words, we can say that data mining is the way of mining knowledge from data. The information or knowledge extracted so can be used for any of the following applications −

- Market Examination
- Fake Detection
- Customer Retention
- Fabrication Control
- Science Exploration

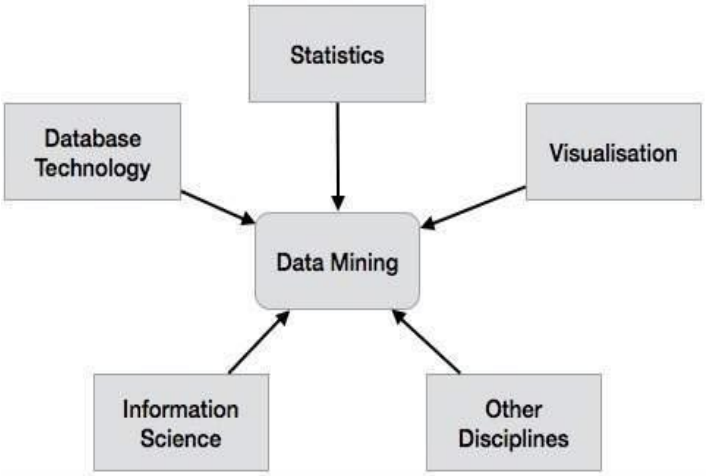


Figure 4.3.2 Data Mining

DATA MINING APPLICATIONS

Data mining is highly useful in the following domains −

- Market Analysis and Management
- Corporate Analysis &Risk Management
- Fraud Detection

CHAPTER 5

SYSTEM DESIGN

5.1 SYSTEM ARCHITECTURE

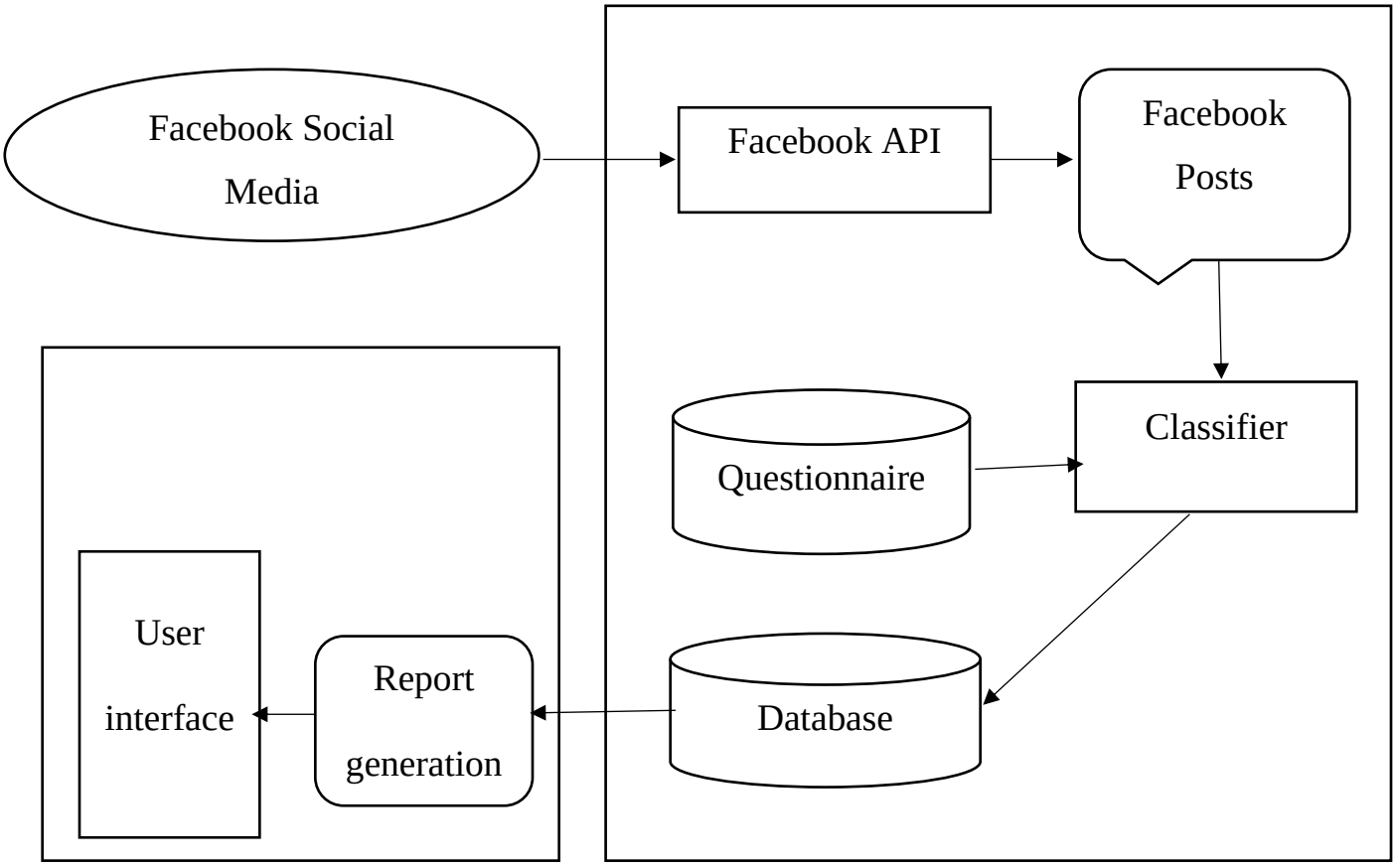


Figure 5.1 System Architecture

System architecture is used to design and develop an web application, which provide an easy and convenient way to get information about depression levels of user by using machine learning algorithms and according to the location of user, the information about doctor is provided. The extraction class will performed the extraction of textual data from facebook through facebook graph API.Preprocessing class is used to preprocess the extracted data. Data must be clear,right and it is

preprocessed for taking care of missing or repetitive attributes. The data ought to be complete and reliable data to deliver the best result from the data mining methodology. Preprocessing of data takes place by using techniques such as tokenization, lower case conversion, word stemming and words removal.Term frequency (tf) has been computed to measure term occurrence. In proposed system user is on Facebook, according to his Facebook post system can find out user in stressed or not as well as different quaternaries which is provided by the system. If user's are not on Facebook they can attempt only quaternaries which is provided by the system according to that we can find out user's in stressed or not.

5.2 UML DIAGRAMS

UML is a standard language for specifying, visualizing, and documenting of software systems and formed by Object Management Group (OMG) in 1997.There are three imperative type of UML modeling are Structural model, Behavioral model, and Architecture model. To model a system the most essential aspect is to capture the dynamic behavior which has some internal or external factors for creating the collaboration. These internal or external agents are known as actors. It consists of actors, use cases and their associations. In this fig we represent the Use Case diagram for our development.

5.2.1 SEQUENCE DIAGRAM

A Sequence diagram is a kind of interaction diagram that shows how processes operate with one another and in what order. A sequence diagram depicts the categorization of actions that occur in a system. The supplication of methods in each object, and the order in which the invocation occurs is captured in a sequence diagram. This makes the sequence diagram a very beneficial tool to easily represent the dynamic behavior of a system.

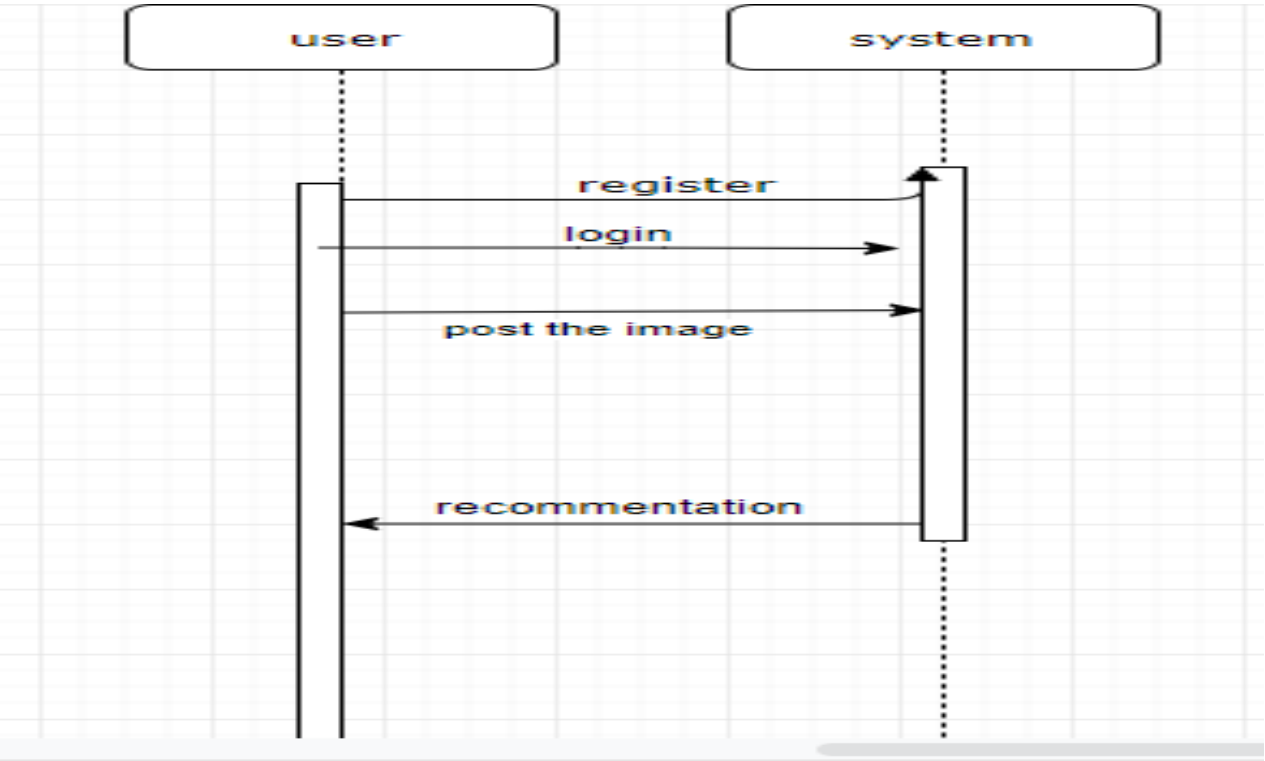


Figure 5.2.1 Sequence Diagram

5.2.2 USE CASE DIAGRAM

Use case diagrams overview the usage requirements for a system. They describe "the meat" of the actual necessities.

A Use case Diagram is used to present a graphical overview of the functionality provided by a system in terms of actors, their goals and any dependencies between those use cases. Use case diagram consists of two parts:

Use case: A use case describes a sequence of actions that provided something of measurable value to an actor and is drawn as a horizontal ellipse.

Actor: An actor is a person, organization or external system that plays a role in one or more interaction with the system.

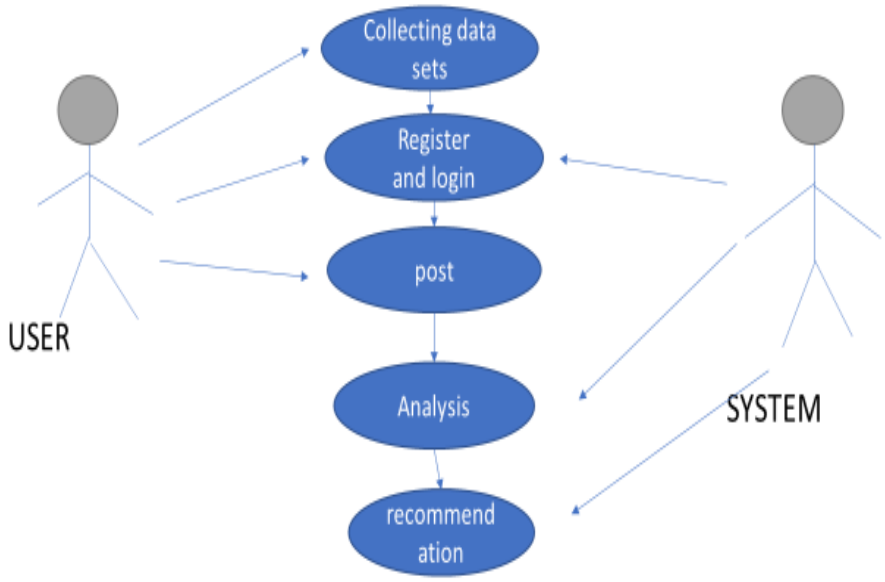


Figure 5.2.2 Use Case Daigram

5.2.3 CLASS DIAGRAM

A class is a collection of things that have alike attributes and common behaviors. Class diagrams are used to label the structure of the system. Objects are instances of classes that are shaped, altered, and destroyed during the execution of the system. An object has state that comprises the values of its attributes and it links with other objects.

A Class diagram in the Unified Modelling Language is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among objects.

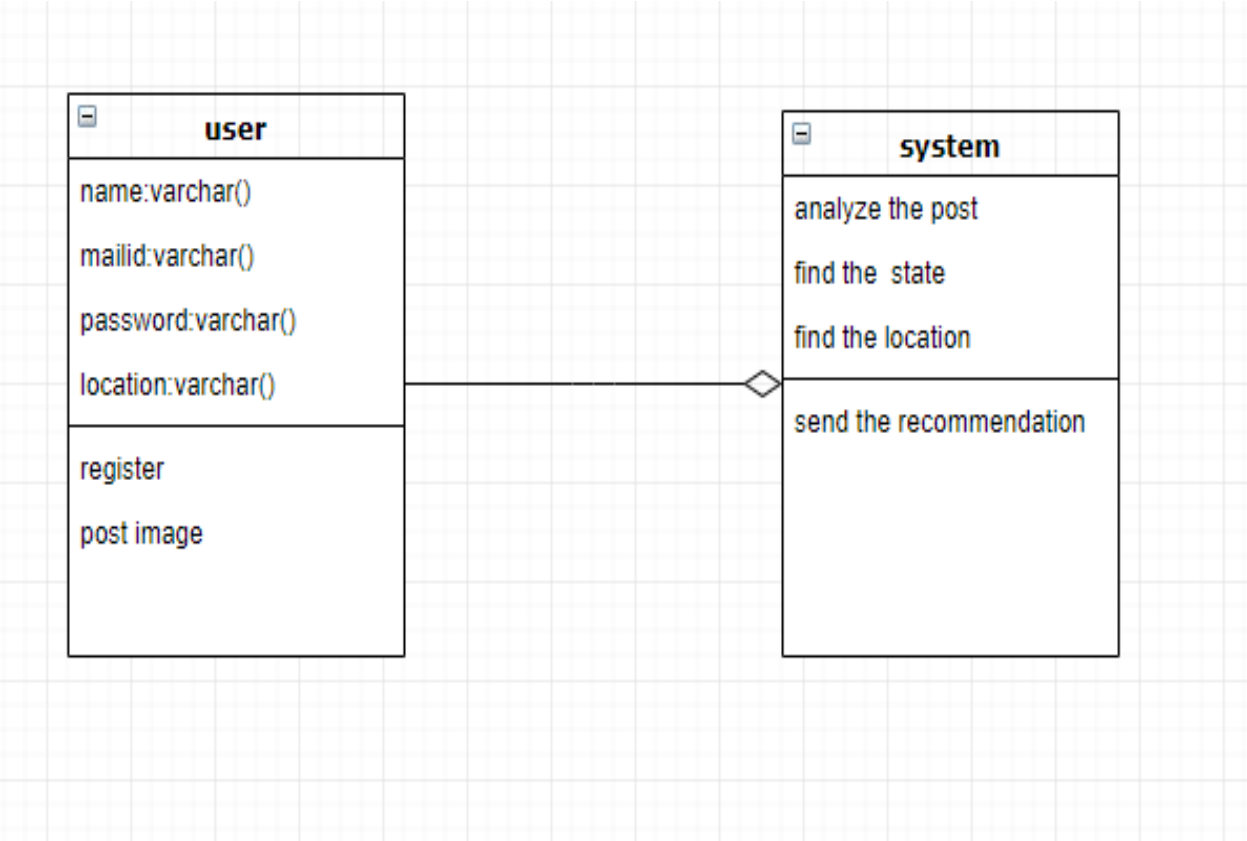


Figure 5.2.3 Class Diagram

5.2.4 DATA FLOW DIAGRAM

A data flow diagram (DFD) is a graphical demonstration of the "flow" of data through an information system, modelling its process characteristics. A DFD is often used as a initial step to create an overview of the system, which can later be explained. It is a preliminary step used to create an overview of the system which can later be elaborated DFDs can also be used for visualization of data processing.

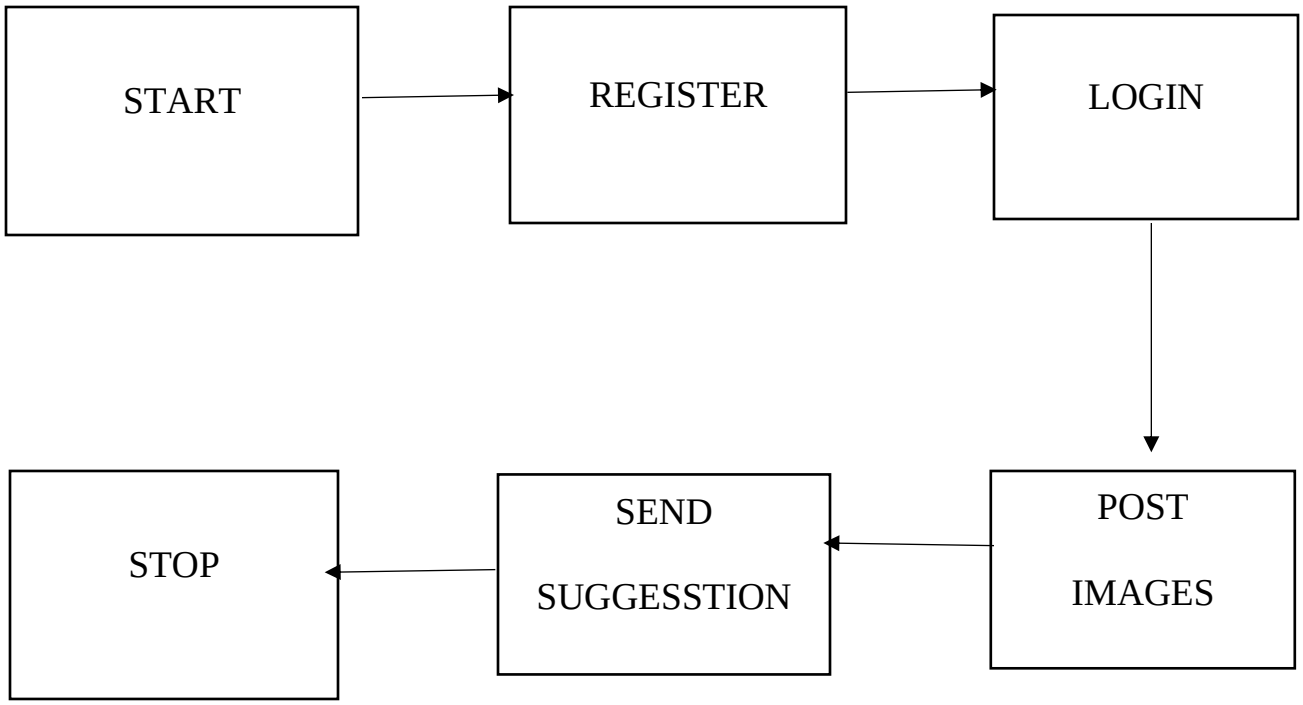


Figure 5.2.4 Data Flow Diagram

CHAPTER 6

SYSTEM IMPLEMENTATION

6.1 ALGORITHM USED

6.1.1 OPTICAL CHARACTER RECOGINITION

Optical character recognition or optical character reader (OCR) is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo (for example the text on signs and billboards in a landscape photo) or from subtitle text superimposed on an image (for example from a television broadcast).

Widely used as a form of data entry from printed paper data records – whether passport documents, invoices, bank statements, computerized receipts, business cards, mail, printouts of static-data, or any suitable documentation – it is a common method of digitizing printed texts so that they can be electronically edited, searched, stored more compactly, displayed on-line, and used in machine processes such as cognitive computing, machine translation, (extracted) text-to-speech, key data and text mining. OCR is a field of research in pattern recognition, artificial intelligence and computer vision.

HOW OCR WORKS ?

The first step of OCR is using a scanner to process the physical form of a document. Once all pages are copied, OCR software converts the document into a two-color, or

black and white, version. The scanned-in image or bitmap is analyzed for light and dark areas, where the dark areas are identified as characters that need to be recognized and light areas are identified as background. The dark areas are then processed further to find alphabetic letters or numeric digits. OCR programs can vary in their techniques, but typically involve targeting one character, word or block of text at a time. Characters are then identified using one of two algorithms:

1. Pattern recognition- OCR programs are fed examples of text in various fonts and formats which are then used to compare, and recognize, characters in the scanned document.
2. Feature detection- OCR programs apply rules regarding the features of a specific letter or number to recognize characters in the scanned document. Features could include the number of angled lines, crossed lines or curves in a character for comparison. For example, the capital letter "A" may be stored as two diagonal lines that meet with a horizontal line across the middle.

6.1.2 NATURAL LANGUAGAE PROCESSING

Natural language processing (NLP) is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.

Natural Language Processing, or NLP for short, is broadly defined as the automatic manipulation of natural language, like speech and text, by software. The study of natural language processing has been around for more than 50 years and grew out of the field of linguistics with the rise of computers.

USES OF NLP

NLP is used to analyze text, allowing machines to understand how human's speak. This human-computer interaction enables real-world applications like automatic text summarization, sentiment analyse, topic extraction, named entity recognition, parts-of-speech tagging, relationship extraction, stemming, and more. NLP is commonly used for text mining, machine translation, and automated question answering.

NLP is characterized as a difficult problem in computer science. Human language is rarely precise, or plainly spoken. To understand human language is to understand not only the words, but the concepts and how they're linked together to create meaning. Despite language being one of the easiest things for the human mind to learn, the ambiguity of language is what makes natural language processing a difficult problem for computers to master.

STEPS IN NLP

There are general five steps −

- **Lexical Analysis** − It involves identifying and analyzing the structure of words. Lexicon of a language means the collection of words and phrases in a language. Lexical analysis is dividing the whole chunk of txt into paragraphs, sentences, and words.
- **Syntactic Analysis (Parsing)** − It involves analysis of words in the sentence for grammar and arranging words in a manner that shows the relationship among the words. The sentence such as "The school goes to boy" is rejected by English syntactic analyzer.

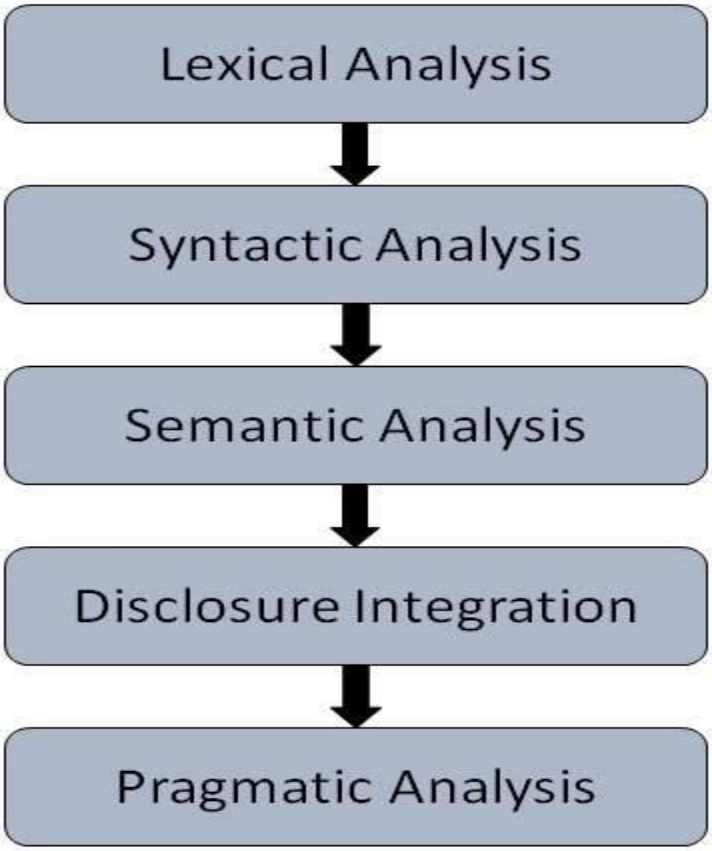


Figure 6.1.2 NLP

- **Semantic Analysis** − It draws the exact meaning or the dictionary meaning from the text. The text is checked for meaningfulness. It is done by mapping syntactic structures and objects in the task domain. The semantic analyzer disregards sentence such as “hot ice-cream”.
- **Discourse Integration** − The meaning of any sentence depends upon the meaning of the sentence just before it. In addition, it also brings about the meaning of immediately succeeding sentence.

- **Pragmatic Analysis** − During this, what was said is re-interpreted on what it actually meant. It involves deriving those aspects of language which require real world knowledge.

6.2 MODULES

- Collection of Real time datasets
- Register and Login
- Classification of Image into OCR
- Sentiment analysis
- Gives suggestions

MODULE DESCRIPTION

6.2.1 COLLECTION OF REAL TIME DATASETS

In this section users can login the organism .If username and password is right login successful otherwise you enter the accurate username and password .If the user do not have any username and password then the user can Registered with the system. In this system user can register the particulars like name ,email id, username , password ,mobile no and then go the login page enter the particulars then advance the process.

6.2.2 REGISTER AND LOGIN

A data set is a collection of data. Most commonly a data set corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question. Here we collect the positive and negative type of datasets. Using this dataset we are analysing with users data and produce the output.

6.2.3 CLASSIFICATION OF IMAGE INTO OCR

OCR (optical character recognition) is the recognition of printed or written text characters by a computer. This involves photo scanning of the text character-by-character, analysis of the scanned-in image, and then translation of the character image into character code. Tesseract is an Optical Character Recognition that's has to be used here for getting text out from the images**.** After getting the string separately the strings are tokenized and compare with the dataset.

6.2.4 SENTIMENT ANALYSIS

Sentiment analysis is one of the most common applications in natural language processing. With Sentiment analysis, we can decide what emotion a text is written. Here we are using the standford-corenlp for finding the users sentimental. The following outline is provided as an overview of and topical guide to natural language processing:

Natural language processing –

computer activity in which computers are entailed to analyze, understand, alter, or generate natural language. This includes the automation of any or all linguistic forms, activities, or methods of communication, such as conversation, correspondence, reading, written,composition, dictation, publishing, translation, lip reading, and so on. Natural language processing is also the name of the branch of computer science, artificial intelligence, and linguistics concerned with enabling computers to engage in communication using natural language(s) in all forms, including but not limited to speech, print, writing, and signing.

6.2.4 GIVES SUGGESTIONS

In that module if the user posted a three continuous negative post means it will gives the suggestion's. In that suggestions it will be shows the near by hospitals to the user.

CHAPTER 7
TESTING

7.1 INTRODUCTION

The purpose of testing is to discover faults. Testing is the process of trying to determine every possible fault or weakness in a work invention. It provides a way to check the functionality of workings, sub-assemblies, meetings and/or a ended product it is the process of training software with the intent of ensuring that the Software system meets its requirements and user opportunities and does not fail in an deplorable manner. There are various types of test. Each test type addresses a specific testing condition.

7.1.1 TESTING OBJECTIVES

- All field entries must work properly.
- Pages must be started from the recognized link.
- The access screen, messages and replies must not be delayed.
- Features to be verified
- Verify that the accesses are of the correct setup
- No replacement passes should be allowed.
- All links must take the user to the right page.

7.2 TYPES OF TESTS

7.2.1 UNIT TESTING

Unit testing includes the design of test belongings that validate that the internal program logic is operative properly, and that program input produces valid outputs.

All choice branches and internal code flow should be authorized. It is the testing of separate software units of the request .it is done after the close of an individual unit before integration. This is a structural testing, that relies on data of its structure and is invasive. Unit tests achieve basic tests at factor level and test a specific commercial process, application, and/or system formation. Unit tests ensure that each single path of business process completes accurately to the documented provisions and contains clearly defined inputs and probable results.

7.2.2 FUNCTIONAL TESTING

Functional tests provide systematic protests that functions tested are accessible as stated by the business and technical necessities, system certification and user guides. Functional difficult is centered on the subsequent items:

Valid Input is used to identified classes of valid input must be accepted.

Invalid Input is used to identified classes of illegal input must be disallowed.

Functions is used to identified purposes must be exercised.

Output is used to classify modules of request outputs.

Systems/Procedures is used to interfacing systems or events must be appealed. Organization and grounding of functional tests is focused on supplies, key functions, or special test belongings. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and succeeding processes must be well-thought-out for testing. Before functional testing is complete, supplementary tests are identified and the operative value of recent tests is resulted.

7.2.3 INTEGRATION TESTING

Integration tests are calculated to test integrated software components to regulate if they actually run as one program. Testing is occasion driven and is more concerned

with the basic result of screens or fields. Integration tests validate that although the workings were individually approval, as shown by positively unit testing, the grouping of components is correct and dependable. Integration testing is specifically aimed at revealing the problems that arise from the grouping of components.

7.2.4 WHITE BOX TESTING

White Box Testing is a challenging in which the software tester has information of the inner workings, construction and language of the software, or at least its drive. It is used to test areas that cannot be stretched from a black box level.

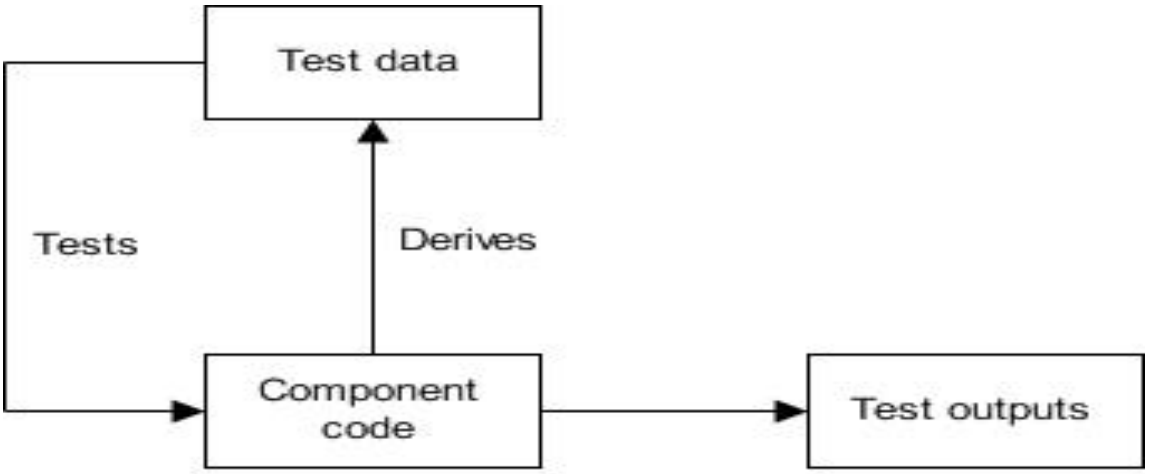


Figure 7.2.4 White Box Testing

7.2.5 BlACK BOX TESTING

Black Box Testing is testing the software short of any knowledge of the inner mechanisms, building or language of the module actuality tested. Black box tests, as most other kinds of tests, must be printed from a final source document, such as requirement or necessities file, such as specification or requirement file. It is a testing in which the software below test is treated, as a black box .you cannot “see” into it. The test provides inputs and respondto outputs without seeing how the software works.

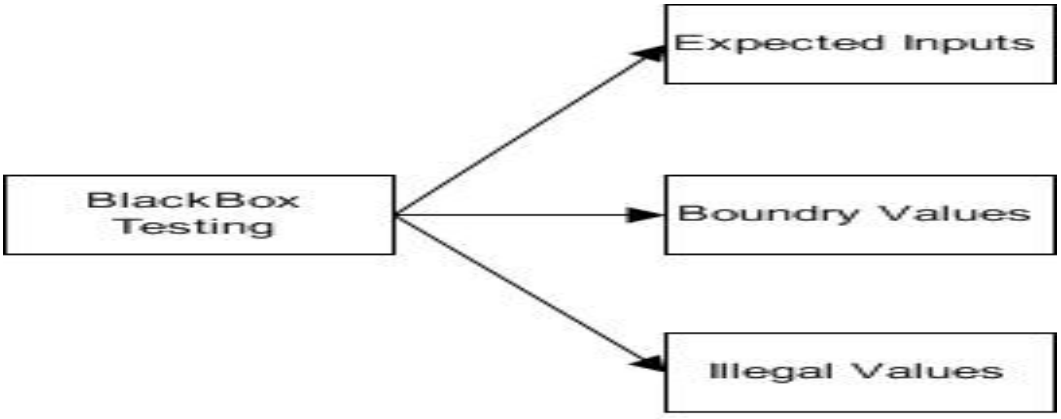


Figure 7.2.5 Black box Testing

7.2.6 SYSTEM TESTING

System testing confirms that the entire combined software system meets supplies. It tests a configuration to ensure known and predictable outcomes. An example of system testing is the arrangement oriented system mixing test. System testing is based on procedure similes and flows, emphasizing pre-driven process links and addition points.

7.2.8 ALPHA TESTING

In software expansion, alpha test will be a test amongst the teams to confirm that your creation works. Originally, the term alpha test destined the first stage of testing in a software development process. The first phase covers component testing, module testing, and scheme testing. It also allows us to test the produce on the last common denominator tackles to make sure transfer times are suitable and preloads work.

7.2.8 BETA TESTING

In software advance, a beta test is the second opinion of software challenging in which a selection of the planned viewers tries the product out. Beta testing can be restrained "pre-release testing." Beta test changes of software are now spread to curriculum establishments and teachers to give the database a "real-world" test.

CHAPTER 8

CONCLUSION AND FUTURE WORKS

The experimental results show that the use of behavioural information on Facebook, both in forms of messages and activities, could predict depression.However, the sample of this research is relatively small because Facebook has limited their permission to collect personal information and the process of gaining approval has become more complicated. Thus, the results getting from this study might not cover all relevant factors. Moreover, as the language-related features had to be translated from Thai to English for analyzing the process, there might be some errors due to this process because some important sentiment polar words might have been eliminated during the translation process.

For the future research, we intend to collect more data to get more relevant and valid features. Manual annotating all complex attributes using crowdsourcing and deeper dimensions should also be analysed in order to be able to create a better depression detection algorithm.

REFERENCES

[1] Abel F, Herder E, Houben G J, Henze N, and Krause D, "Cross-system user modeling and personalization on the social web," User Modeling and User-Adapted Interaction, vol. 23, pp. 169-209, 2013.

[2] Abel F, Bischoff K , Fankhauser P, Iofciu T "Identifying users across social tagging systems," Proc. of the 5th International AAAI Conference on Weblogs and Social Media, pp. 522-525, 2011.

[3] Acquisti A, Gross R and Stutzman F, "Privacy in the age of augmented reality," Proc. National Academy of Sciences, 2011.

[4] Almishari M and Tsudik G, "Exploring linkability of user reviews," Computer Security–ESORICS 2012 (ESORICS'12), pp. 307- 324, 2012.

[5] Anderson A, Corney M, De Vel O, and Mohay G, "Mining email content for author identification forensics," ACM Sigmod Record, vol. 30, no. 4, pp. 55-64, 2001.

[6] Bartunov S, Korshunov A, Lee H, Park S and Ryu W, "Joint link-attribute user identity resolution in online social networks," The 6th SNA-KDD Workshop '12, 2012.

[7] Castelluccia C, Kaafar M A, and Manils P, "How unique and traceable are usernames?," Privacy Enhancing Technologies (PETS'11), pp. 1-17, 2011.

[8] Chbeir R, Dipanda A and Raad E, "User profile matching in social networks," Proc. Of the 13th International Conference on Network-Based Information Systems (NBiS'10), pp.297-304, 2010.

[9] Chen H, Huang Z, Li J and Zheng R, "A framework for authorship identification of online messages: writing-style features and classification techniques," J. of the American Society for Information Science and Technology, vol. 57, no. 3, pp. 378-393,2006.

[10] Cortis K, Handschuh S, Rivera I and Scerri S, "An ontology based technique for online profile resolution," Social Informatics, Berlin: Springer, pp. 284-298, 2013.

[11] Erdős P and Rényi A, "On random graphs I," Publ. Math. Debrecen, vol. 6, pp. 290-297, 2010.

[12] Friedland G, Goga O, Lei H, Parthasarathi S H K, Sommer R, and Teixeira R, "Exploiting innocuous activity for correlating users across sites," Proc. 22nd international conference on World WideWeb (WWW'13),pp. 447-458, 2013.

[13] Goga O, Lei H, Perito D, Sommer R , and Teixeira R , "Largescale Correlation of Accounts across Social Networks," Technical report, 2013.

[14] Hong D, Shen V Y, and Vosecky J, "User identification across multiple social networks," Proc. Of the 1st International Conference on Networked Digital Technologies, pp.360-365, 2009.

[15] Jain P and Kumaraguru P, "Finding Nemo: searching and resolving identities of users across online social networks," arXiv preprint arXiv:1212.6147, 2012.

[16] Jain P, Joshi A, and Kumaraguru P, "@ i seek 'fb. me': identifying users across multiple online social networks," Proc. of the 22nd International Conference on World Wide Web Companion, pp. 1259-1268, 2013.

[17] Kong K, Yu P S and Zhang J, "inferring anchor links across multiple heterogeneous social networks," Proc. of the 22nd ACM International Conf. on Information and Knowledge Management (CIKM'13), pp. 179-188, 2013.

[18] Korula N and Lattanzi S, "An efficient reconciliation algorithm for social networks," arXiv preprint arXiv:1307.1690, 2013.

[19] Liu H and Zafarani R "Connecting corresponding identities across communities," Proc. of the 3rd International ICWSM Conference, pp. 354-357, 2009.

[20] Liu J, Hon H W, Song X, Song Y I, and Zhang F ,"What's in a name an unsupervised approach to link users across communities," Proc. of the 6th ACM international conference on Web search and data mining(WDM'13), pp. 495-504, 2013.

[21] Motoyama M and Varghese G, "I seek you: searching and matching individuals in social networks," Proc. of the 11th international workshop on Web Information and Data Management (WIDM'09), pp. 67-75, 2009.

[22] Narayanan A and Shmatikov V, "De-anonymizing social networks," Proc. Of the 30th IEEE Symposium on Security and Privacy (SSP'09), pp. 173-187, 2009.

[23] Wikipedia, "Twitter, " http://en.wikipedia.org/wiki/Twitter.2014.

[24] Xinhuanet, "Sina Microblog Achieves over 500 Million Users," http://news.xinhuanet.com/tech/2012-02/29/c_122769084.htm. 2014.

[25] Zafarani R and Liu H, "Connecting users across social media sites: a behavioral-modeling approach, " Proc. of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'13), pp.41-49, 2013.

APPENDIX 1

SOURCE CODE

```
package tesseracttest;
import java.awt.Color;
import java.awt.FileDialog;
import java.awt.Frame;
import java.awt.image.BufferedImage;
import java.io.File;
import java.io.IOException;
import java.sql.Connection;
import java.sql.DriverManager;
import java.sql.ResultSet;
import java.sql.Statement;
import java.util.StringTokenizer;
import java.util.logging.Level;
import java.util.logging.Logger;
import javax.imageio.ImageIO;
import mini.readfile;
import net.sourceforge.tess4j.Tesseract;
import net.sourceforge.tess4j.TesseractException;
import tesseracttest.ColorUtils;
public class Main1 extends javax.swing.JFrame {
private Frame mainFrame;
final FileDialog fileDialog = new FileDialog(mainFrame,"Select file");
public static int col=0,sent=0,flag=0;
Connection con=null;
Statement st=null,st1=null;
```

```
ResultSet rs1=null;
public Main1() {
initComponents();
}
@SuppressWarnings("unchecked")
// <editor-fold defaultstate="collapsed" desc="Generated Code">
private void initComponents() {
jPanel1 = new javax.swing.JPanel();
label1 = new java.awt.Label();
textArea1 = new java.awt.TextArea();
label2 = new java.awt.Label();
button1 = new java.awt.Button();
label3 = new java.awt.Label();
jPanel2 = new javax.swing.JPanel();
label4 = new java.awt.Label();
textArea2 = new java.awt.TextArea();
textArea3 = new java.awt.TextArea();
label5 = new java.awt.Label();
label6 = new java.awt.Label();
button2 = new java.awt.Button();
setDefaultCloseOperation(javax.swing.WindowConstants.EXIT_ON_CLOSE);
label1.setText("");
label1.addComponentListener(new java.awt.event.ComponentAdapter() {
public void componentMoved(java.awt.event.ComponentEvent evt) {
label1ComponentMoved(evt);
}
label2.setText("Select Image to post");
button1.setLabel("Open File");
button1.addActionListener(new java.awt.event.ActionListener() {
```

```
public void actionPerformed(java.awt.event.ActionEvent evt) {
button1ActionPerformed(evt);
}
});
label3.setText("File Name");
label4.setText("Suggestion");
label6.setText("Suggestion");
javax.swing.GroupLayout jPanel2Layout = new
javax.swing.GroupLayout(jPanel2);
jPanel2.setLayout(jPanel2Layout);
jPanel2Layout.setHorizontalGroup(
jPanel2Layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEA
DING)
.addGroup(jPanel2Layout.createSequentialGroup()
.addGroup(jPanel2Layout.createParallelGroup(javax.swing.GroupLayout.Align
ment.LEADING)
.addGroup(jPanel2Layout.createSequentialGroup()
.addGroup(jPanel2Layout.createParallelGroup(javax.swing.GroupLayout.Align
ment.LEADING)
.addGroup(jPanel2Layout.createSequentialGroup()
.addContainerGap()
.addComponent(label4, javax.swing.GroupLayout.PREFERRED_SIZE,
javax.swing.GroupLayout.DEFAULT_SIZE,
javax.swing.GroupLayout.PREFERRED_SIZE))
.addGroup(jPanel2Layout.createSequentialGroup()
.addGap(20, 20, 20)
.addComponent(textArea2,
javax.swing.GroupLayout.PREFERRED_SIZE,737,javax.swing.GroupLayout.P
REFERRED_SIZE))).addGap(0, 9, Short.MAX_VALUE))
.addComponent(textArea3, javax.swing.GroupLayout.Alignment.TRAILING,
javax.swing.GroupLayout.DEFAULT_SIZE,
javax.swing.GroupLayout.DEFAULT_SIZE, Short.MAX_VALUE))
```

```
.addContainerGap())
);
jPanel2Layout.setVerticalGroup(
jPanel2Layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
.addGroup(jPanel2Layout.createSequentialGroup()
.addComponent(textArea2, javax.swing.GroupLayout.PREFERRED_SIZE, 144, javax.swing.GroupLayout.PREFERRED_SIZE)
.addPreferredGap(javax.swing.LayoutStyle.ComponentPlacement.RELATED)
.addComponent(label4, javax.swing.GroupLayout.PREFERRED_SIZE, javax.swing.GroupLayout.DEFAULT_SIZE, javax.swing.GroupLayout.PREFERRED_SIZE)
.addPreferredGap(javax.swing.LayoutStyle.ComponentPlacement.RELATED)
.addComponent(textArea3, javax.swing.GroupLayout.PREFERRED_SIZE, 128, javax.swing.GroupLayout.PREFERRED_SIZE)
.addContainerGap(javax.swing.GroupLayout.DEFAULT_SIZE, Short.MAX_VALUE))
);
label5.setText("label5");
button2.setLabel("Post");
button2.addActionListener(new java.awt.event.ActionListener() {
public void actionPerformed(java.awt.event.ActionEvent evt) {
button2ActionPerformed(evt);
}
});
javax.swing.GroupLayout jPanel1Layout = new javax.swing.GroupLayout(jPanel1);
jPanel1.setLayout(jPanel1Layout);
jPanel1Layout.setHorizontalGroup(
jPanel1Layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
```

.addGroup(jPanel1Layout.createSequentialGroup()

.addGroup(jPanel1Layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)

.addGroup(jPanel1Layout.createSequentialGroup()

.addGap(26, 26, 26)

.addGroup(jPanel1Layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)

.addComponent(label3, javax.swing.GroupLayout.DEFAULT_SIZE, javax.swing.GroupLayout.DEFAULT_SIZE, Short.MAX_VALUE)

.addGroup(jPanel1Layout.createSequentialGroup()

.addComponent(label2, javax.swing.GroupLayout.PREFERRED_SIZE, javax.swing.GroupLayout.DEFAULT_SIZE, javax.swing.GroupLayout.PREFERRED_SIZE)

.addGap(57, 57, 57)

.addComponent(button1, javax.swing.GroupLayout.PREFERRED_SIZE, javax.swing.GroupLayout.DEFAULT_SIZE, javax.swing.GroupLayout.PREFERRED_SIZE)

.addGap(87, 87, 87)

.addComponent(button2, javax.swing.GroupLayout.PREFERRED_SIZE, javax.swing.GroupLayout.DEFAULT_SIZE, javax.swing.GroupLayout.PREFERRED_SIZE)

.addPreferredGap(javax.swing.LayoutStyle.ComponentPlacement.RELATED, javax.swing.GroupLayout.DEFAULT_SIZE, Short.MAX_VALUE)

.addComponent(label5, javax.swing.GroupLayout.PREFERRED_SIZE, javax.swing.GroupLayout.DEFAULT_SIZE, javax.swing.GroupLayout.PREFERRED_SIZE)

.addGap(190, 190, 190))))

.addGroup(jPanel1Layout.createSequentialGroup()

.addGap(35, 35, 35)

.addComponent(label1, javax.swing.GroupLayout.PREFERRED_SIZE, javax.swing.GroupLayout.DEFAULT_SIZE, javax.swing.GroupLayout.PREFERRED_SIZE)

String hex = C.getColorNameFromRgb(red, green, blue);

```
System.out.println("image Color name");
System.out.println(".....");
System.out.println("Color name " +  hex.toString());
String col1=hex.toString();
if(col1.contains("Red")||col1.contains("red"))
{
col=1;
}
else if(col1.contains("black")||col1.contains("Black"))
{
col=2;
}
else
{
col=0;
}
File imgfile = new File(imgFilePath);
Tesseract instance = Tesseract.getInstance();
String result = instance.doOCR(imgfile);
String rs=result.substring(1);
//rs=rs.substring(beginIndex, endIndex)
rs=rs.substring(0, rs.length() - 4);
rs=rs.trim();
System.out.println(".....");
System.out.println("String Extraction from Image");
System.out.println(".....");
System.out.println(rs);
textArea2.setText(rs);
//System.out.println();
```

```
System.out.println(".....");
System.out.println("sentiment Extraction from Image Text");
System.out.println(".....");
s=rs;
readfile rf = new readfile();
//tokenizing text
StringTokenizer st = new StringTokenizer(s," =,;!?");
while(st.hasMoreTokens())
{
f=0;
key=st.nextToken();
key=key.toLowerCase();
System.out.println(key);
rf.openFileP();
while(rf.z.hasNext())
{
if(key.equals(rf.readFileP()))
{
psentiment++;
f=1;
break;
}}
rf.closeFileP();
if(f!=1)
{
rf.openFileN();
while(rf.w.hasNext())
{
if(key.equals(rf.readFileN()))
```

```
{
nsentiment++;
break;
}}
rf.closeFileN();
}}
if (psentiment>nsentiment)
{
System.out.println("The overall sentiment of the text is POSITIVE");
sent=0;
}
else if (psentiment==0 && nsentiment==0)
{
System.out.println("There is NO EMOTION");
sent=1;
}
else if (psentiment==nsentiment)
{
System.out.println("The overall sentiment of text is MIXED, as same positive and negative notions");
sent=2;
}
else
{
System.out.println("The overall sentiment of text is NEGATIVE");
sent=3;
}
double x=(double)sent+(double)col;
System.out.println(String.valueOf(x));
```

```
double y=x/5.0;
System.out.println(String.valueOf(y));
System.out.println(".....END.....");
if(y<0.6)
{
label5.setText("No");
}
else
{
label5.setText("Yes");
flag=flag+1;
if(flag==2)
{
textArea3.setText(" YOU ARE SOME WHAT DEPRESSED SO HANG OUT WITH FRIENDS  ");
}
else if(flag==3)
{
flag=0;
try
{
utile u=new utile();
Class.forName("com.mysql.jdbc.Driver");
con = DriverManager.getConnection("jdbc:mysql://localhost:3306/stess","root","password");
st1 = con.createStatement();
rs1 = st1.executeQuery("select * from add1 where loc="'+u.getloc()+'"");
if(rs1.next()) {
```

```
textArea3.setText(" YOU ARE CONTINOUSLY POSTING NEGATIVE
STATUS TRY TO GET SOME CONSULTANT FROM
\r\n"+rs1.getString(2)+"\r\n");
//sn.setAttribute("dpm",department);
}}
catch(Exception ex)
{
}}}
} catch (TesseractException e) {
System.err.println(e.getMessage());
} catch (IOException ex) {
Logger.getLogger(Main.class.getName()).log(Level.SEVERE, null, ex);
}}
else
{
int psentiment=0 ,nsentiment=0,f;
String s = new String();
String key= new String();
String imgFilePath = label3.getText().toString();
File file= new File(imgFilePath);
BufferedImage image;
try {
image = ImageIO.read(file);
// Getting pixel color by position x=100 and y=40
int clr=  image.getRGB(100,40);
int  red   = (clr & 0x00ff0000) >> 16;
int  green = (clr & 0x0000ff00) >> 8;
int  blue  =  clr & 0x000000ff;
ColorUtils C=new ColorUtils();
```

```
String hex = C.getColorNameFromRgb(red, green, blue);
System.out.println("image Color name");
System.out.println(".....");
System.out.println("Color name " +  hex.toString());
String col1=hex.toString();
if(col1.contains("Red")||col1.contains("red"))
{
col=1;
}
else if(col1.contains("black")||col1.contains("Black"))
{
col=2;
}
else
{
col=0;
}
File imgfile = new File(imgFilePath);
Tesseract instance = Tesseract.getInstance(); //
//System.out.println(imgfile.canRead());
String result = instance.doOCR(imgfile);
String rs=result.substring(1);
rs=rs.substring(0, rs.length() - 4);
rs=rs.trim();
System.out.println(".....");
System.out.println("String Extraction from Image");
System.out.println(".....");
System.out.println(rs);
//System.out.println();
```

```
System.out.println(".....");
System.out.println("sentiment Extraction from Image Text");
System.out.println(".....");
s=rs;
readfile rf = new readfile();
//tokenizing text
StringTokenizer st = new StringTokenizer(s," =,;!?");
while(st.hasMoreTokens())
{
f=0;
key=st.nextToken();
key=key.toLowerCase();
System.out.println(key);
rf.openFileP();
while(rf.z.hasNext())
{
if(key.equals(rf.readFileP()))
{
psentiment++;
f=1;
break;
}}
rf.closeFileP();
if(f!=1)
{
rf.openFileN();
while(rf.w.hasNext())
{
if(key.equals(rf.readFileN()))
```

```
{
nsentiment++;
break;
}}
rf.closeFileN();
}}
if (psentiment>nsentiment)
{
System.out.println("The overall sentiment of the text is POSITIVE");
sent=0;
}
else if (psentiment==0 && nsentiment==0)
{
System.out.println("There is NO EMOTION");
sent=1;
}
else if (psentiment==nsentiment)
{
System.out.println("The overall sentiment of text is MIXED, as same positive and negative notions");
sent=2;
}
else
{
System.out.println("The overall sentiment of text is NEGATIVE");
sent=3;
}
double x=(double)sent+(double)col;
System.out.println(String.valueOf(x));
```

```
double y=x/5.0;
System.out.println(String.valueOf(y));
System.out.println(".....END.....");
if(y<0.6)
{
label5.setText("No");
}
else
{
label5.setText("Yes");
flag=flag+1;
if(flag==4)
{
flag=0;
try
{
utile u=new utile();
Class.forName("com.mysql.jdbc.Driver");
con =
DriverManager.getConnection("jdbc:mysql://localhost:3306/stess","root","pass
ward");
st1 = con.createStatement();
rs1 = st1.executeQuery("select * from add1 where loc="'+u.getloc()+"'");
if(rs1.next()) {
textArea3.setText(rs1.getString(2)+"\r\n");
//sn.setAttribute("dpm",department);
}}
catch(Exception ex)
{
```

```
}
}}
} catch (TesseractException e) {
System.err.println(e.getMessage());
} catch (IOException ex) {
Logger.getLogger(Main.class.getName()).log(Level.SEVERE, null, ex);
}}}
public static Color[][] loadPixelsFromImage(File file) throws IOException {
BufferedImage image = ImageIO.read(file);
Color[][] colors = new Color[image.getWidth()][image.getHeight()];
for (int x = 0; x < image.getWidth(); x++) {
for (int y = 0; y < image.getHeight(); y++) {
colors[x][y] = new Color(image.getRGB(x, y));
}}
return colors;
}
public static void main(String args[]) {
try {
for (javax.swing.UIManager.LookAndFeelInfo info :
javax.swing.UIManager.getInstalledLookAndFeels()) {
if ("Nimbus".equals(info.getName())) {
javax.swing.UIManager.setLookAndFeel(info.getClassName());
break;
}}
} catch (ClassNotFoundException ex) {
java.util.logging.Logger.getLogger(Main.class.getName()).log(java.util.logging.
Level.SEVERE, null, ex);
} catch (InstantiationException ex) {
java.util.logging.Logger.getLogger(Main.class.getName()).log(java.util.logging.
Level.SEVERE, null, ex);
```

```
} catch (IllegalAccessException ex) {
java.util.logging.Logger.getLogger(Main.class.getName()).log(java.util.logging.
Level.SEVERE, null, ex);
} catch (javax.swing.UnsupportedLookAndFeelException ex) {
java.util.logging.Logger.getLogger(Main.class.getName()).log(java.util.logging.
Level.SEVERE, null, ex);
}
java.awt.EventQueue.invokeLater(new Runnable() {
public void run() {
new Main1().setVisible(true);
}}};
private java.awt.Button button1;
private java.awt.Button button2;
private javax.swing.JPanel jPanel1;
private javax.swing.JPanel jPanel2;
private java.awt.Label label1;
private java.awt.Label label2;
private java.awt.Label label3;
private java.awt.Label label4;
private java.awt.Label label5;
private java.awt.Label label6;
private java.awt.TextArea textArea1;
private java.awt.TextArea textArea2;
private java.awt.TextArea textArea3;
}
```

APPENDIX 2

SCREENSHOTS

STEP 1

The user collects the real time datasets from the social media sites. Theses datas will be used futher to analyse the depression level of the user.

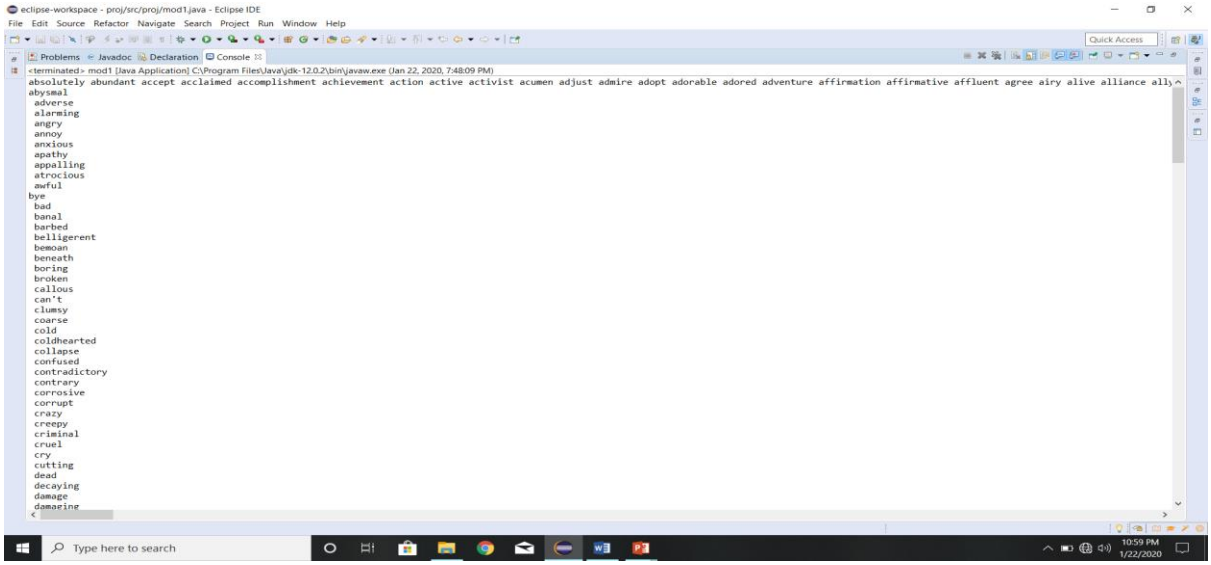


Figure A2.1 Collecting Data sets

STEP 2

Here the user Register their details in the database to login into the process.

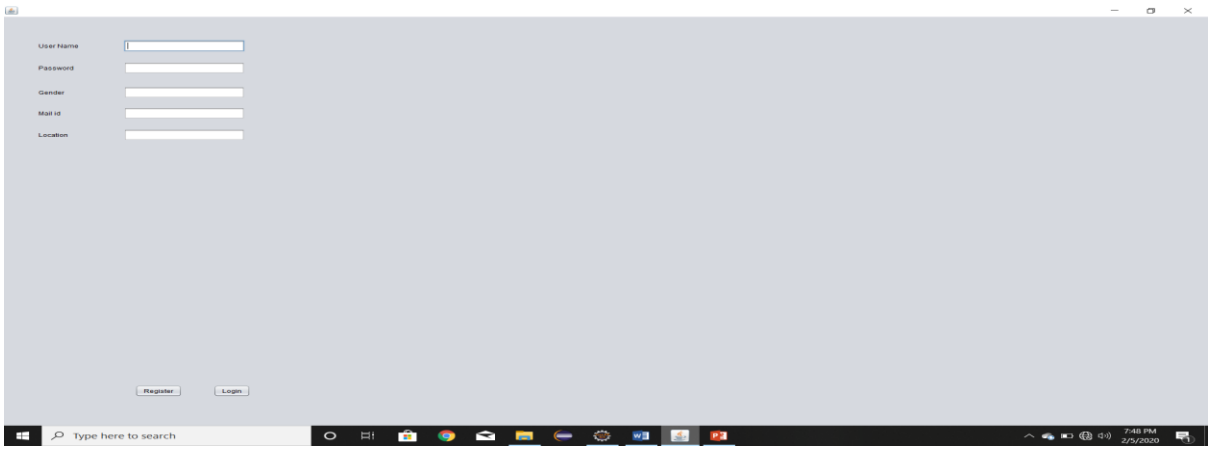


Figure A2.2 Register page

STEP 3

In this phase the user login by giving the user id and password.

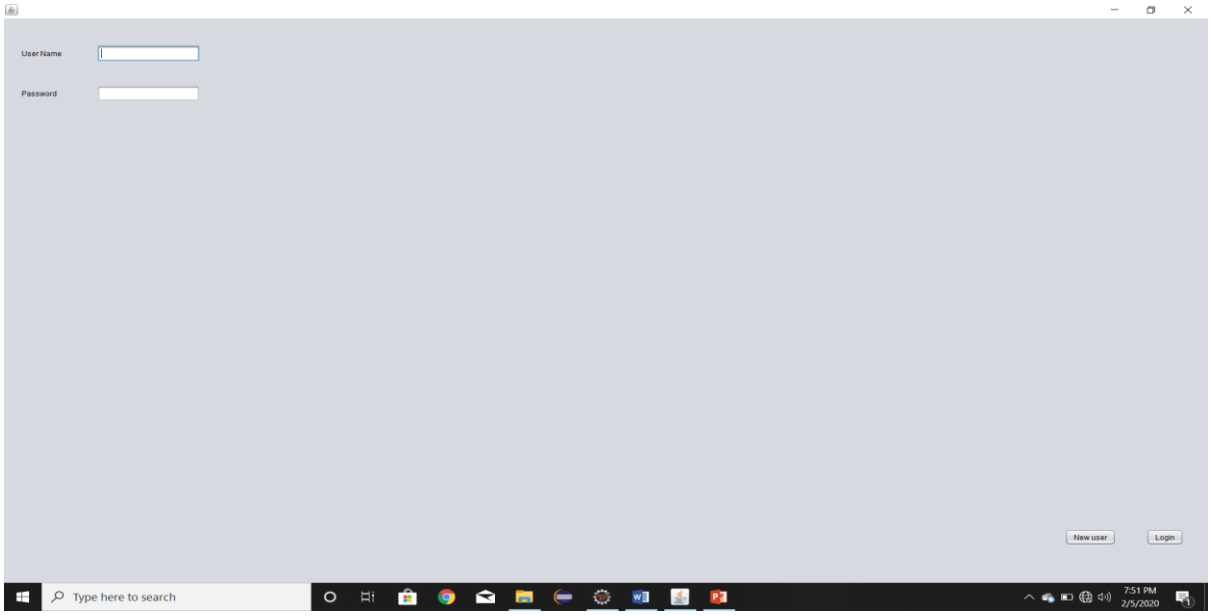


Figure A2.3 Login Page

STEP 4

Then the post that is posted by the user is classified using OCR alogorithm,in these phase the words are extracted from the image post.

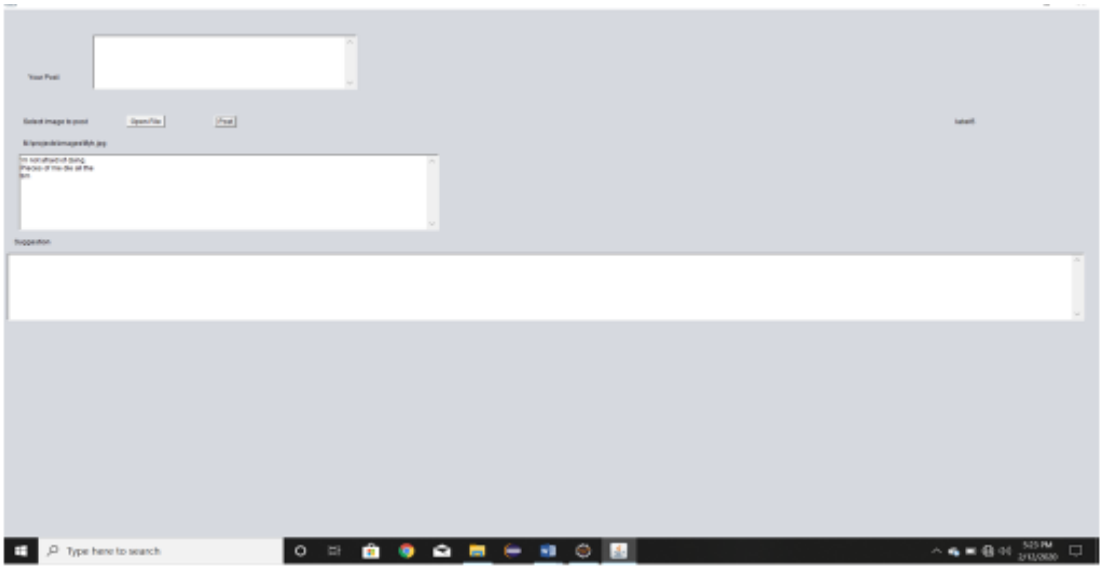


Figure A2.4 Classification of Image into OCR

STEP 5

In this phase the words that are extracted from the image is analysed to predict the deprression level of the user.

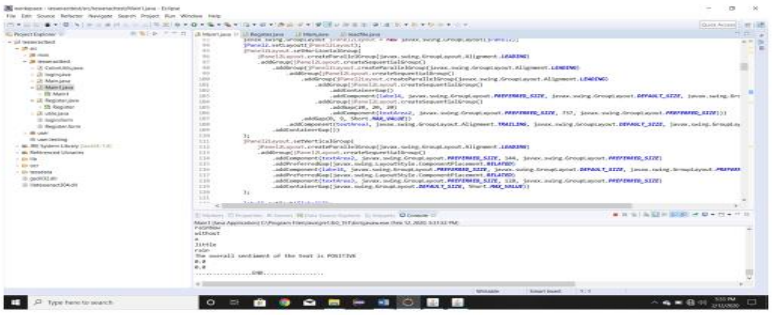


Figure A2.5 Sentiment Analysis

STEP 6

Here the suggestions are given to the user based on the predicted depression level.

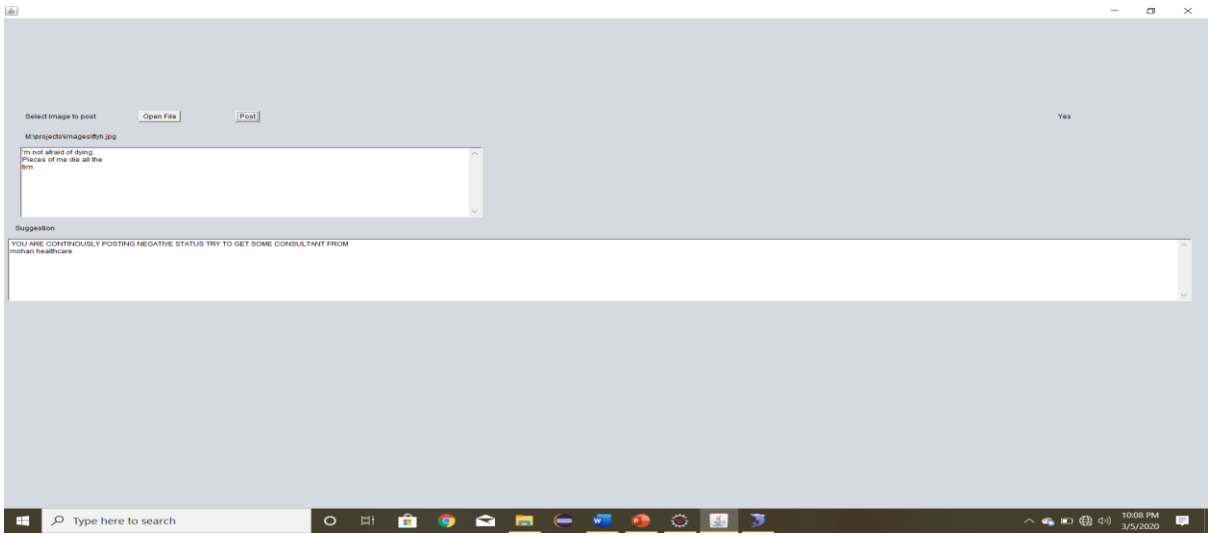


Figure A2.6 Giving Suggestions

Sentiment Analysis using the Social Mediaposts

Maria Alex Amalraj T [1], Mohan prasad P [1], Manoj kumar M [1], Geetha P [2], Sharmasth Vali Y [2]

[1]Student, Computer Science and Engineering, Dhanalakshmi College of Engineering, Chennai-601 301

[2] Assistant Professor, Computer Science and Engineering, Dhanalakshmi College of Engineering,Chennai-601 301

alexamalraj1999@gmail.com, psm.manojkumar1998@gmail.com, pmohanprasad99@gmail.com, geetha_mails@yahoo.com

Abstract - The use of Social Network Sites (SNS) like facebook, twitter, instagram is increasing nowadays especially by the younger generations. By using SNS users can express their interests, feelings,emotions and share daily routine. In many researches it is proven that using user-generated content (UGC) in a correct manner may help to determine people's mental health levels. Mining the social media post could help to predict the mental health and depression of the person. Depression is a serious medical illness in our current generation, which reflects in work, study, eat, sleep and having fun. However, from the user profile in SNS, to collect all the information that relates to person's mood, and negativism. Social media helps to know about individual's thinking, mood, depression & socialization. So by analysing the social media post & by using some algorithms on it, able to determine the depression levels and emotions of that particular individual. Using the traditional techniques, the psychiatrist cannot get the complete and accurate information from the depressed patient. In this research, our aim is to investigate how SNS user's posts can help classify users according to mental health levels and depression.

Key words: Social Network Sites; user-generated content; data mining; optical character recognition; natural language processing.

I INTRODUCTION

Web application is a part of this paper. In this paper the social media posts (Facebook) are considered. Social Network Sites (SNS) is a source of data and information to classify the users according to their user generated contents (UGC). By using machine learning algorithms such as Naive Bayes, depression level of user is classified into various levels and provides doctor's location near to users location. As depression is serious problem which is keep on moving higher day by day. In India, out of total population 8.5% of it facing this problem. It seems to be major issue and that is the reason that motivated us to worked on this problem. Earlier diagnosis of depressed patient were done on basis of noting the behaviour of a person and reported by his relatives or friends. But the result was not so effective and accurate. In contrast to that, SNS is powerful tool for predicting depression levels of an individual. Nowadays, many people are using social media platforms such as Facebook, twitter. They share their thoughts, feelings, emotions, feelings of guilt, worthlessness, helplessness and egoistic nature of individual etc. Whatever they post is related to their behaviour and emotions. So it will help to diagnose that person before he/she gets worse. This motivates to do the paper, so that this will help not only to psychiatrist but also used by general patients who want to do self-diagnosis.

DATA MINING - OVERVIEW

There is a huge quantity of data available in the Information Trade. This data is of no use until it is changed into valuable information. It is compulsory to analyze this huge quantity of data and extract valuable information from it. Mining of information is not the only process need to complete; Data mining also involves some other processes such as Data Collection, Data Integration, Data Transformation, Pattern Evaluation and Data Presentation. Once all these procedures are completed, we would be able to use this info in many real time applications. Data Mining is defined as extracting info from vast sets of data. In other words, we can say that data mining is the way of mining knowledge from data. The information or knowledge extracted so can be used for any of the following applications like market examination, fake detection, customer retention,fabrication control, science exploration.

DATA MINING - TASKS

Data mining deals with the kind of configurations that can be mined. On the basis of the kind of data to be mined, there are two groups of tasks involved in Data Mining −

- Descriptive
- Arrangement and Prediction

DESCRIPTIVE FUCTION

The descriptive function deals with the properties of data in the database. Here are some of the descriptive functions −

- Data Description

- Mining of Repeated Terms
- Mining of Links
- Mining of Correlations
- Mining of Clusters.

DATA MINING PRIMITIVES

- Specify a data mining job in the form of a data mining query.
- This query is input to the system.
- A data mining query is defined in terms of data mining data predicting.
- Set of tasks and relevant data to be mined.
- Kind of information to be mined.
- Background information to be used in finding course.
- Interestingness procedures and thresholds for form evaluation.
- Demonstration for imagining the exposed pattern.

DATA MINING SYSTEM CLASSIFICATION

A data mining system can be classified based on the following criteria −

- Database Technique
- Statistics
- Machine Learning
- Information
- Conception
- Other Disciplines

Apart from these the diagram is mentioned in figure 1, a data mining system can also be categorized based on the kind of
(a) databanks mined,
(c) information mined,
(c) methods utilized, and
(d) tenders adapted.

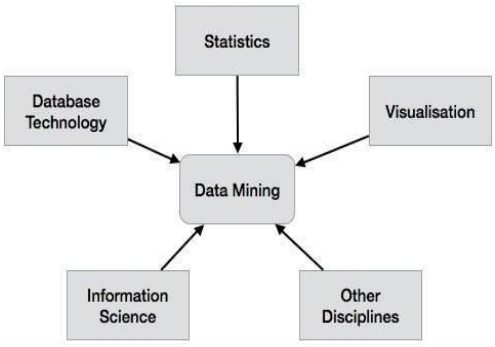


Figure 1 Data mining classification

II LITERATURE SURVEY

In the existing system the medical science relies on asking the patients questions about their situations, which does not diagnose the depression in a precise way. Electronic Health Record (EHR) systems are not optimally designed to handle integrating behavioural health and primary care. Using the traditional techniques, the psychiatrist cannot get the complete and accurate information from the depressed patient.

[1 et al,] Depression is a public health problem that has high effects on a person's functional and social relationships. Depression is a growing problem in the society. It causes pain and suffering not only to patients, but also to those who care about them. Depression disorder is hard to diagnose, because its symptoms could be confused with other disorders and has different cross-cultural symptoms. This paper proposes a framework that would best solve the problem of automatic depression detection in depressed Saudi patients. This paper particularly focuses on designing the collection of Saudi depression dataset using multiple modalities.

[2 et al,]The properties of acoustic speech have previously been investigated as possible cues for depression in adults. However, these studies were restricted to small populations of patients and the speech recordings were made during patients' clinical interviews or fixed-text reading sessions. Symptoms of depression often first appear during adolescence at a time when the voice is changing, in both males and females, suggesting that specific studies of these phenomena in adolescent populations are warranted. This study investigated acoustic correlates of depression in a large sample of 139 adolescents (68 clinically depressed and 71 controls).

[3 et al.,] Social networks contain a tremendous amount of node and linkage data, providing unprecedented opportunities for a wide variety of fields. As the world's fourth largest disease, depression has become one of the most significant research subjects. Previously, a depression classifier has been proposed to classify the users in online social networks to be depressed or not, however, the classifier takes only node features into account and neglects the influence of linkages. This paper proposes an improved model to calculate the probability of a user being depressed, which is based on both node and linkage features. The linkage features are measured in two aspects: tie strength and interaction content analysis. Moreover, the propagation rule of depression is considered for improving the prediction accuracy.

[4 et al,]Diagnosis and prevention of depressive disorders at any scale have been attracting considerable attention of the public healthcare in Japan because depression is one of the most rapidly pervasive mental disorders in the country. A major issue that hinders the feasibility of depression screening for its prevention is the availability of some simple and cost-effective methods for depression detection and monitoring. Here in this paper, we present the development of a computerized tool for depression detection. The tool utilizes the theory of chaos and systems complexity to extract robust dynamically statistical

features of physiological signals provided by the low-cost technology of photoplethysmography.

III PROPOSED PLAN

In this proposed system we are collect the data sets (user generated content) in Facebook. The SNS based system can overcome the problems regarding self-reporting from user's social activities by analysing the user's posts and predicting the emotion status of a person.The emotion is mapped and given values from 3 to -3.Where 3 is the highest point of happiness,and the value decreases to 2 and 1 as the level of happiness decreses.Similarly -3 being the most depressed while 0 is neutral. If the value reaches the most negative for a long time that person is considered to be depressed. Here we are using the Naïve Bayes Classifier for train the dataset.

IV SYSTEM ARCHITECTURE

This design focuses on the data structure software architecture, procedural details, implementations etc., and interface among modules. The design procedure also decode the requirements into presentation of software that can be accessed for excellence before coding begins. Computer software design change continuously as novel methods; improved analysis and border understanding evolved. Software proposal is at relatively primary stage in its revolution. Therefore, software design methodology lacks the depth, flexibility and quantitative nature that are usually associated with more conventional engineering disciplines. However methods for software designs do exit, criteria for design qualities are existing and design notation can be applied. System architecture is used to design and develop an web application, which provide an easy and convenient way to predict the depression levels of user by using machine learning & data mining algorithms and according to the users location, the information about doctor to consult is provided.

The extraction class will performed the extraction of textual data from facebook through facebook graph API.Preprocessing class is used to preprocess the extracted data. Data must be clear, true and it is preprocessed for taking care of missing or repetitive attributes and words. The data ought to be true and reliable data to predict and deliver the best result from the data mining methodology. Preprocessing of data takes place by using techniques such as tokenization, lower case conversion, word stemming and words removal. Term frequency (tf) has been computed to measure term occurrence and calculate the result. In proposed system according to user Facebook post the system can find out user is stressed or not as well as different quaternaries which is provided by the system. If user's are not on Facebook they can attempt only quaternaries which is provided by the system according to that we can find out user's is stressed or not.

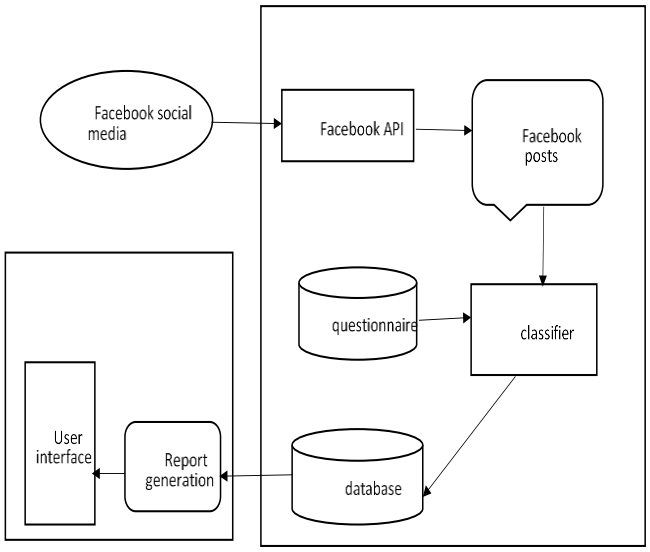


Figure 4.1 System Architecture

V SYSTEM IMPLEMENTATION

This paper shows the implementation of

- User registration
- Collecting data set
- OCR (Optical Character Recognition)
- Sentiment analysis
- Gives the suggestions

DESCRIPTION

User registration

In this section, users can login the organism. If username and password is right login successful otherwise you enter the accurate username and password . If the user do not have any username and password then the user can registered with the system. In this system user can register the particulars like name ,email id, username, password ,mobile no and then go the login page enter the particulars then advance the process.

Collecting data set

A data set is a collection of data. The dataset and information is stored to the contents of a single database table, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in this process. Here we collect the positive and negative type of datasets. Using this dataset we are analysing with users data and produce the output.

OCR

OCR (optical character recognition) is the recognition of printed text or word characters from the image by a computer. This phase involves techniques to image scanning of the text character-by-character, analysis of the scanned-in image, and then translation of the character image into character token. Tesseract is an Optical Character Recognition that's has to be used here for getting text out from the images**.** After getting the string separately the strings are tokenized and compare with the dataset.

Sentiment analysis

Sentiment analysis is one of the most important phase in this paper. By using Sentiment analysis, to can decide in what emotion a text is written. These basic concepts when used in combination, become a very important tool for analyzing millions of brand conversations with human level accuracy. Here to are using the standford-corenlp for finding the users sentimental. The following outline is provided as an overview of a natural language processing:

Natural language processing

Computer activity in which computers are entailed to analyze, understand, predict, and generate natural language.

Gives the suggestions

In that module if the user posted a three continuous negative post means it will gives the suggestion's. In that suggestions it will be shows the near by hospitals to the user.

VI ALGORITHMS USED

OPTICAL CHARACTER RECOGNITION

Optical character recognition (OCR) is an algorithm used for electronic conversion of images of typed, handwritten or printed text into machine understandable form. Optical character recognition, is a method of converting image into text. When a image is scanned, it is intially stored as a bit-mapped file in TIF format. When the image is displayed on the screen, human can read it. But to the computer, it is just a series of zeros and ones. The computer can not recognize any "words" on the image. This is what OCR does. OCR looks at each and every line of the image and attempts to determine whether the black and white dots represent a character or number. OCR technique was actually created originally to assist visually disabled individuals gain access to printed information. OCR algorithm has been updated and improved and is now used to recognize the character from the image in the computer files. OCR can be a very powerful tool for a law firm. The key is to produce a text version of the scanned image. Once a text file has been created, it then becomes possible to launch a text search and locate any page with a given word or set of words.

NATURAL LANGUAGE PROCESSING

Natural language processing helps computers understand the humans in their own language and technique to other language-related tasks. NLP makes it possible for computers to read, hear, interpret the human language, to measure sentiment and determine which parts are important. Today's computers can analyze more language-based data than humans, without any error and in a consistent, efficient way. Considering the large amount of data that is generated every day, from medical records to social media, automation will be critical and difficult to fully analyze text data efficiently.

Natural language processing provides many different techniques for interacting with human, ranging from statistical and data mining methods to rules-based and mathematical approaches. To need a very broad array of approaches because there are many data to be processed like text- and voice-based data varies widely, as do the practical applications.

VII RESULTS

In this paper the IDE used is Eclipse, Swing as front end and the database SQL Yog is used in the backend to store the data sets of users.

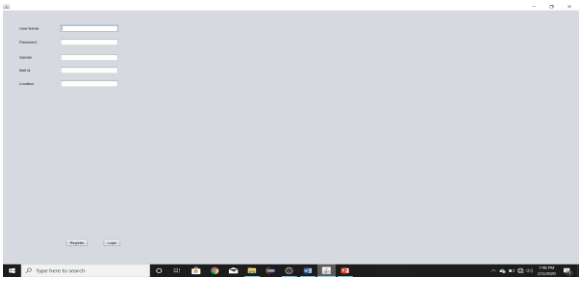


Figure 2 login page

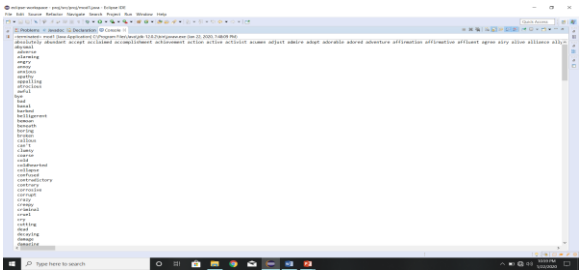


Figure 3 Collecting Data set

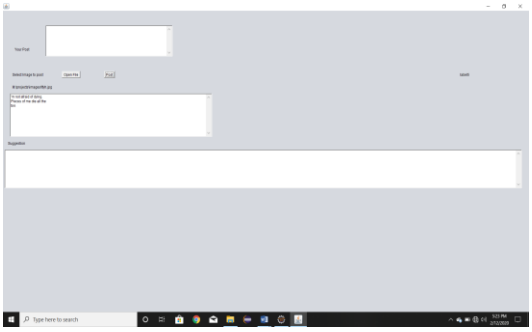


Figure 4 Image posting

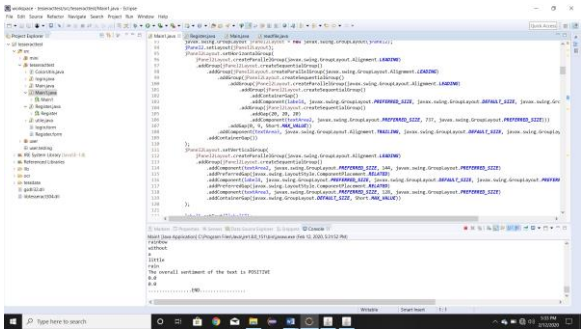


Figure 5 Sentiment Analysis

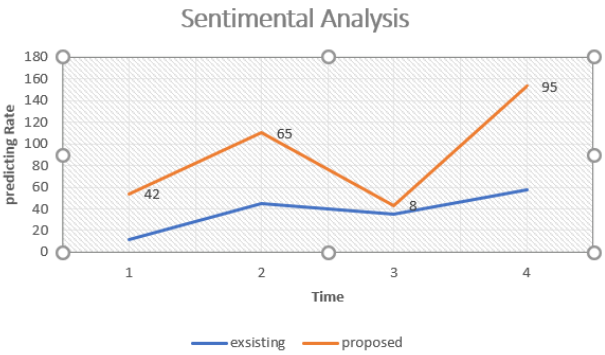


Figure 6 Graph

VIII CONCLUSION

The experiment results show that the use of behavioural information on Facebook, both in forms of messages and activities, could predict depression. Hotover, the sample of this research is relatively small because Facebook has limited their permission to collect personal information and the process gaining approval has become more complicated. Thus, the results getting from this study might not cover all relevant factors. Moreover, as the language-related features had to be translated from Thai to English for analyzing the process, there might be some errors due to this process because some important sentiment polar words might have been eliminated during the translation process. For the future research, to intend to collect more data to get more relevant and valid features so that the result will be more accurate. Manual annotating all complex attributes using crowd sourcing and deeper dimensions should also be analysed in order to be able to create a better depression detection algorithm.

IX REFERENCES

[1]Wikipedia,"Twitter,"http://en.wikipedia.org/wiki/Twitter.2014.

[2] Xinhuanet, "Sina Microblog Achieves over 500 Million User",2014.

[3] D. Perito, C. Castelluccia, M.A. Kaafar, and P. Manils, "How unique and traceable are usernames?," Privacy Enhancing Technologies (PETS'11), pp. 1-17, 2011.

[4] J. Liu, F. Zhang, X. Song, Y.I. Song, C.Y. Lin, and H.W. Hon,"What's in a name?: an unsupervised approach to link users across communities," Proc. of the 6th ACM international conference on Tob search and data mining(WDM'13), pp. 495-504, 2013.

[5] R. Zafarani and H. Liu, "Connecting corresponding identities across communities," Proc. of the 3rd International ICWSM Conference, pp. 354-357, 2009.

[6] R. Zafarani and H. Liu, "Connecting users across social media sites: a behavioral-modeling approach, " Proc. of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'13), pp.41-49, 2013.

[7] A. Acquisti, R. Gross and F. Stutzman, "Privacy in the age of augmented reality," Proc. National Academy of Sciences, 2011.

[8] T. Iofciu, P. Fankhauser, F. Abel, and K. Bischoff, "Identifying users across social tagging systems," Proc. of the 5th International AAAI Conference on Weblogs and Social Media, pp. 522-525, 2011.

[9] M. Motoyama and G. Varghese, "I seek you: searching and matching individuals in social networks," Proc. of the 11th international workshop on Web Information and Data Management (WIDM'09), pp. 67-75, 2009.

[10] O. Goga, D. Perito, H. Lei, R. Teixeira, and R. Sommer, "Largescale Correlation of Accounts across Social Networks," Technical report, 2013.

[11] K. Cortis, S. Scerri, I. Rivera, and S. Handschuh, "An ontology based technique for online profile resolution," Social Informatics, Berlin: Springer, pp. 284-298, 2013.

[12] F. Abel, E. Herder, G.J. Houben, N. Henze, and D. Krause, "Cross-system user modeling and personalization on the social web," User Modeling and User-Adapted Interaction, vol. 23, pp. 169-209, 2013.

[13] O. De Vel, A. Anderson, M. Corney, and G. Mohay, "Mining email content for author identification forensics," ACM Sigmod Record, vol. 30, no. 4, pp. 55-64, 2001.

[14] E. Raad, R. Chbeir, and A. Dipanda, "User profile matching in social networks," Proc. Of the 13th International Conference on Network-Based Information Systems (NBiS'10), pp.297-304, 2010.

[15] J. Vosecky, D. Hong, and V.Y. Shen, "User identification across multiple social networks," Proc. Of the 1st International Conference on Networked Digital Technologies, pp.360-365, 2009.

[16] P. Jain, P. Kumaraguru, and A. Joshi, "@ i seek 'fb. me': identifying users across multiple online social networks," Proc. of the 22nd International Conference on World Wide Web Companion, pp. 1259-1268, 2013.

[17] P. Jain and P. Kumaraguru, "Finding Nemo: searching and resolving identities of users across online social networks," arXiv preprint arXiv:1212.6147, 2012.

[18] R. Zheng, J. Li, H. Chen, and Z. Huang, "A framework for authorship identification of online messages: writing-style features and classification techniques," J. of the American Society for Information Science and Technology, vol. 57, no. 3, pp. 378-393,2006.

[19] M. Almishari and G. Tsudik, "Exploring linkability of user reviews," Computer Security–ESORICS 2012 (ESORICS'12), pp. 307- 324, 2012.

[20] X. Kong, J. Zhang, and P.S. Yu, "inferring anchor links across multiple heterogeneous social networks," Proc. of the 22nd ACM International Conf. on Information and Knowledge Management (CIKM'13), pp. 179-188, 2013.

[21] O. Goga, H. Lei, S.H.K. Parthasarathi, G. Friedland, R. Sommer, and R. Teixeira, "Exploiting innocuous activity for correlating users across sites," Proc. 22nd international conference on World WideWeb (WWW'13),pp. 447-458, 2013.

[22] A. Narayanan and V. Shmatikov, "De-anonymizing social networks," Proc. Of the 30th IEEE Symposium on Security and Privacy (SSP'09), pp. 173-187, 2009.

[23] S. Bartunov, A. Korshunov, S. Park, W. Ryu, and H. Lee, "Joint link-attribute user identity resolution in online social networks," The 6th SNA-KDD Workshop '12, 2012.

[24] N. Korula and S. Lattanzi, "An efficient reconciliation algorithm for social networks," arXiv preprint arXiv:1307.1690, 2013.

[25] P. Erdős and A. Rényi, "On random graphs I," Publ. Math. Debrecen, vol. 6, pp. 290-297, 2010.

VELAMMAL
INSTITUTE OF TECHNOLOGY

Approved by AICTE - New Delhi, Affiliated to Anna University - Chennai
All Programs Accredited by NBA and NAAC
"Velammal Knowledge Park", Chennai - Kolkatta Highway, Ponneri - 601204

NBA NATIONAL BOARD OF ACCREDITATION

NAAC NATIONAL ASSESSMENT AND ACCREDITATION COUNCIL

International Conference on Research Advancements & Challenges in Engineering Sciences (ICRACE'20)

06th - 07th March, 2020

CERTIFICATE

This is to certify that ~~Dr./Mr.~~/Ms. MARIA ALEX AMALRAJ. T

of DHANALAKSHMI COLLEGE OF ENGINEERING has

presented a paper titled SENTIMENT ANALYSIS USING THE SOCIAL MEDIAPOSTS

in the International Conference on Research Advancements & Challenges in Engineering Sciences (ICRACE'20) held during 06th - 07th March 2020, organized by Velammal Institute of Technology, Chennai.

M.S. P S

Coordinator

Principal

VELAMMAL INSTITUTE OF TECHNOLOGY

Approved by AICTE - New Delhi, Affiliated to Anna University - Chennai

All Programs Accredited by NBA and NAAC

"Velammal Knowledge Park", Chennai - Kolkatta Highway, Ponneri - 601204

VELAMMAL 33 years 1986 - 2019

NBA NATIONAL BOARD OF ACCREDITATION

NAAC NATIONAL ASSESSMENT AND ACCREDITATION COUNCIL

International Conference on Research Advancements & Challenges in Engineering Sciences (ICRACE'20)

06th - 07th March, 2020

CERTIFICATE

This is to certify that ~~Dr.~~/Mr./~~Ms.~~ MOHAN PRASAD. P

of DHANALAKSHMI COLLEGE OF ENGINEERING has

presented a paper titled SENTIMENT ANALYSIS USING THE SOCIAL MEDIAPOSTS

in the International Conference on Research Advancements & Challenges in Engineering Sciences (ICRACE'20) held during 06th - 07th March 2020, organized by Velammal Institute of Technology, Chennai.

M.S.R

Coordinator

Principal

VELAMMAL
INSTITUTE OF TECHNOLOGY

Approved by AICTE - New Delhi, Affiliated to Anna University - Chennai

All Programs Accredited by NBA and NAAC

"Velammal Knowledge Park", Chennai - Kolkatta Highway, Ponneri - 601204

VELAMMAL 33 years 1985 - 2018

NBA NATIONAL BOARD OF ACCREDITATION

NAAC NATIONAL ASSESSMENT AND ACCREDITATION COUNCIL

International Conference on Research Advancements & Challenges in Engineering Sciences (ICRACE'20)

06th - 07th March, 2020

CERTIFICATE

This is to certify that ~~Dr./Mr./Ms.~~ MANOJ KUMAR. M

of DHANALAKSHMI COLLEGE OF ENGINEERING has

presented a paper titled SENTIMENT ANALYSIS USING THE SOCIAL MEDIAPOSTS

in the International Conference on Research Advancements & Challenges in Engineering Sciences (ICRACE'20) held during 06th - 07th March 2020, organized by Velammal Institute of Technology, Chennai.

M.S

Coordinator

Principal