

## ABSTRACT

Design is a multi- step that focuses on data structure software architecture, procedural details, procedure etc. and interface among modules. The design procedure also decodes 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 exist, criteria for design qualities are existing and design notation can be applied. System architecture is used to design and develop a 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 perform 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 users 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.

## LIST OF FIGURES

FIGURE NO.	NAME OF THE FIGURE	PAGE NO.
	ARCHITECTURE DIAGRAM	6
	BLOCK DIAGRAM OF OCR	8
	IMPLEMENTATION OF PROJECT	13
	USER REGISTRATION	14
	STORED USER LOGIN DATA IN DATABASE	15
	USER LOGIN	15
	APPLICATION DIALOG BOX	16
	IMAGE SELECTION	16
	DISPLAYING SENTIMENT RESULT	17

# TABLE OF CONTENTS

<b>CHAPTER NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
	<b>ABSTRACT</b>	<b>V</b>
	<b>LIST OF FIGURES</b>	<b>VI</b>
<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
	1.1 OUTLINE OF THE PROJECT	1
<b>2</b>	<b>LITERATURE SURVEY</b>	<b>3</b>
<b>3</b>	<b>METHODOLOGY</b>	<b>5</b>
	3.1 SYSTEM REQUIREMENTS	5
	3.2 SYSTEM ARCHITECTURE	6
	3.3 SENTIEXTRACT	8
	3.4 TESSERACT OCR	8
	3.5 EXTRACTION OF TEXT FROM PICTURE FILES	8
	3.6 UML DIAGRAMS	9
<b>4</b>	<b>RESULTS</b>	<b>14</b>
	4.1 RESULTS	14
	4.2 DISCUSSIONS	14
<b>5</b>	<b>CONCLUSION AND FUTURE WORK</b>	<b>18</b>
	5.1 CONCLUSION	18
	5.2 FUTURE WORK	18
	<b>REFERENCES</b>	<b>19</b>
	<b>APPENDIX</b>	<b>21</b>
	A. SOURCE CODE	21
	B. PAPER WORK	

# CHAPTER 1

## INTRODUCTION

### OUTLINE OF THE PROJECT

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

### *DATA MINING APPLICATIONS*

- Market Analysis and Management
- Corporate Analysis & Risk Management

- Fraud Detection

Apart from these, data mining can also be used in the zones of fabrication control, buyer retention, science exploration, sports, astrology, and Internet Web Surf-Aid.

### ***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 Function***

The descriptive function deals with the general properties of data in the database. Here is the list of descriptive functions –

- Class/Concept Description
- Mining of Repeated Patterns
- Mining of Links
- Mining of Correlations
- Mining of Clusters

#### ***Data Mining Task Primitives***

- We can specify a data mining job in the form of a data mining query.
- This query is input to the system.

## **CHAPTER 2**

### **LITERATURE SURVEY**

#### **A depression dataset to build automatic diagnoses in clinically depressed Saudi patient**

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.

#### **An improved model for depression detection in micro blog social network**

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,

#### **Toward the development of cost effective e-depression effective system**

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.

### **Detection of Clinical Depression in Adolescents' Speech During Family Interactions**

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).

## **CHAPTER 3**

### **METHODOLOGY**

#### **SYSTEM REQUIREMENTS**

The software requirements specification is produced at the culmination of the analysis task. The function and performance allocated to software as part of system engineering are refined by establishing a complete information description as functional representation of system behavior, an indication of performance requirements and design constraints, appropriate validation criteria.

#### ***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

#### ***SOFTWARE REQUIREMENTS***

Operating system : Windows 10

IDE : Eclipse

Coding Language : Java

Backend : SQLyog

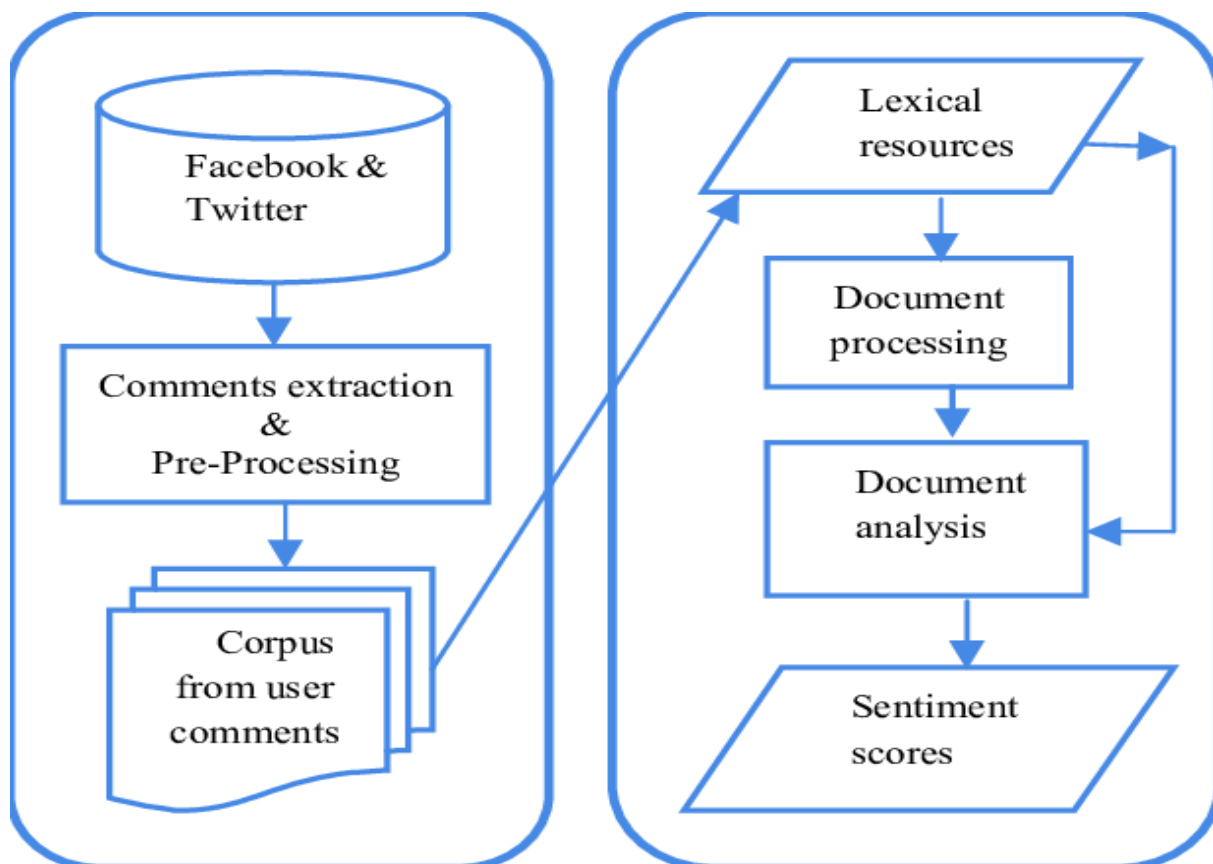


# SYSTEM ARCHITECTURE

## INTRODUCTION

Design is a multi- step that focuses on data structure software architecture, procedural details, procedure 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.

## ARCHITECTURE DIAGRAM:



**Fig 3.1 ARCHITECTURE DIAGRAM**

System architecture is used to design and develop an web application, which provide an easy and convenient way to get information about depression levelsof 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.

### ***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.

## SENTIEXTRACT

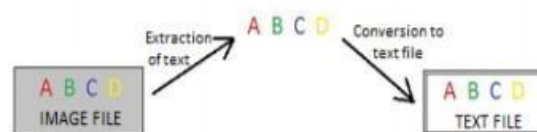
Our system boasts of two highly efficient algorithms namely, tesseract OCR algorithm for image to text conversion and naïve bayes classifier for sentiment extraction of the converted text. A single or multiple image file can be given as input to our system for calculation of their sentiment polarity.

## TESSERACT OCR

Tesseract OCR is an optical character reading engine developed by HP laboratories in 1985 and open sourced in 2005. Since 2006 it is developed by Google. Tesseract has Unicode (UTF-8) support and can recognize more than 100 languages “out of the box” and thus can be used for building different language scanning software also. Latest Tesseract version is Tesseract 4. It adds a new neural net (LSTM) based OCR engine which is focused on line recognition but also still supports the legacy Tesseract OCR engine which works by recognizing character patterns.

## EXTRACTION OF TEXT FROM PICTURE FILES

The image file is firstly converted to a text file using the tesseract ocr algorithm, in our system we are using the pyocr wrapper for this purpose. This wrapper invokes the tesseract algorithm and gives us the required text file, which is then used for further analysis. Image files of any size can be given as input to the system, but care should be taken as blurred or skewed images may not fetch you optimal results.



**Fig 3.2 Block Diagram Of OCR**

Number of words a file may contain has no constraints, but the words should be of English language as it is the only language currently supported by our system