A PROJECT REPORT

on

"ALLEVIATE"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER SCIENCE AND ENGINEERING

BY

SIDHARTH PUROHIT	1705078
SHUBHAM KUMAR MAURYA	1705074
NIWANSHU MAHESWARI	1705051
ABHISH KUMAR ANAND	1705005
SURAJ KUMAR MISHRA	1805951

UNDER THE GUIDANCE OF
PROF. AJAY ANAND

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAR, ODISHA - 751024

May 2020

A PROJECT REPORT

on

"ALLEVIATE"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN COMPUTER SCIENCE AND ENGINEERING

BY

SIDHARTH PUROHIT	1705078
SHUBHAM KUMAR MAURYA	1705074
NIWANSHU MAHESHWARI	1705051
ABHISH KUMAR ANAND	1705005
SURAJ KUMAR MISHRA	1805951

UNDER THE GUIDANCE OF

PROF. AJAY ANAND

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAE, ODISHA -751024

May 2020

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, ODISHA 751024

CERTIFICATE

This is certifying that the project entitled

"ALLEVIATE"

submitted by

SIDHARTH PUROHIT	1705078
SHUBHAM KUMAR MAURYA	1705074
NIWANSHU MAHESHWARI	1705051
ABHISH KUMAR ANAND	1705005
SURAJ KUMAR MISHRA	1805951

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2019-2020, under our guidance.

Date: 30 / 5 / 2020

(Prof. Ajay Anand)
Project Guide

Acknowledgements

We are profoundly grateful to Prof. AJAY ANAND for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion. .....

SIDHARTH PUROHIT
SHUBHAM KUMAR MAURYA
NIWANSHU MAHESHWARI
ABHISH KUMAR ANAND
SURAJ KUMAR MISHRA

ABSTRACT

Alleviate aims to provide some help to people dealing with unhealthy mental health (such as anxiety, depression) by giving them a platform to share their thoughts and get an analysis based on it, i.e., whatever they have shared is either tilting towards positive or negative.
The idea is to give the user insights of their own thoughts not a cure thus to *alleviate* the pain.

Alleviate also aims to spread awareness for a healthy mental health and to be a digital companion to those who need.

The main feature of this project is *"The Dark Diary"* which aims to provide users a platform to write, record and analyse their thoughts, this is done by adding posts to the diary such as adding journal entries (but digitally and that entry then can be analysed to get the sentiment in it).

Keywords: Mental health, alleviate, awareness, digital diary, human sentiment analysis

Contents

List of Figures

Chapter 1

Introduction

1.1 OVERVIEW

According to *World Health Organization* "Mental health refers to cognitive, behavioural, and emotional well-being. It is all about how people think, feel, and behave". People sometimes use the term mental health to signify the absence of a mental disorder. It can also affect daily living, relationships, and physical health. Mental composure is a distinguishable feature of a human being. Various behaviour changes like-*long lasting sadness or irritability*, *high and low mood swings*, *social withdrawal*, *dramatic changes in sleeping and eating habits* are considered as the primary warning signs of Mental Illness. The *WHO* stress that mental health is "more than just the absence of mental disorders or disabilities."

Peak mental health is about not only avoiding active conditions but also looking after ongoing wellness and happiness. They also emphasize that preserving and restoring mental health is crucial on an individual basis, as well as throughout different communities and societies of the world. In the United States, the National Alliance on Mental Illness estimate that for every 5 adults, one adult experience mental health problem each year. In the year 2017, an estimated 11.2 million people in the U.S (about 4.5% of adults) had a severe psychological condition, as per the reports of *National Institute of Mental Health (NIMH)*.The situations like lockdown, unemployment, economic crisis etc. are pertaining to worsen the current situation of mental wellbeing.

The project is designed to tackle the experiences, counter attacks and responses to various situations/ stimulus that a human being confronts in his everyday life.
A result of changes like *Global Pandemics* and situations like *Economic crisis* overhead this year, the mental status and wellbeing is essential to be maintained for anyone, whether it's a school going teenager or an elderly recovering with some prolonged chronic ailment. One of the crucial aspects to study the mental health is by the means of *human confessions*, how a person thinks is what he writes and chalks down, thereby, acts accordingly. We have tried to capture all these natural human instincts to understand the psychology behind it. Without invading any individual privacy, we have extracted and web scrapped our textual data and have made a Machine learning based model to classify the human emotions broadly now into two categories of being sad or happy.

The project also aims to provide a user interface on the internet to give our end user a liberty to provide their confessions and find meaningful insights from the same, related to their hate content and a work to give specific recommendations is currently being done. The data collection, storage, manipulation, analysis and labeling it into a supervised learning binary classification is being done manually by our team. Facing stress, dilemma and anxiety is common for anyone in their lives, so the project has taken data from the real world to come up with plausible solutions without losing the human touch.

Conclusively, the project uses the Machine learning tools of *Natural Language processing* techniques to analyze the human sentiments and has come up with a solution to identify and cope up with the adverse effects of mental health via training a model on the textual confessions data extracted from the popular social media confessions pages on the internet. An effort is being made to create a web hosted application of the project. The data has been cleaned, converted to its base form and is analyzed with the state-of-the-art Natural language processing tools found in the python NLTK library. The model training has been done on Random forest classifiers, Gaussian Naive Bayes classification model, Multiple Layer Perceptron model and the accuracy evaluation on all these has been shown in the notebook as a part of our research.

1.2 INTRODUCTION TO MACHINE LEARNING

What is Machine Learning?

Two definitions of Machine Learning are offered. Arthur Samuel described it as: "the field of study that gives computers the ability to learn without being explicitly programmed." This is an older, informal definition.

Tom Mitchell provides a more modern definition: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

Example: playing checkers.

E = the experience of playing process many confession statements

T = the task of pprocessing statements

P = the probability that the program will give the right output.

In general, any machine learning problem can be assigned to one of two broad classifications:

Supervised learning and Unsupervised learning.

Supervised Learning

In supervised learning, we are given a data set and already know what our correct output should look like, having the idea that there is a relationship between the input and the output.

Supervised learning problems are categorized into "regression" and "classification" problems. In a regression problem, we are trying to predict results within a continuous output, meaning that we are trying to map input variables to some continuous function. In a classification problem, we are instead trying to predict results in a discrete output. In other words, we are trying to map input variables into discrete categories.

The problem we're going to discuss is also a supervised learning problem where we give our model a labels to learn . It's basically a binary classification problem .

Unsupervised Learning

In unsupervised learning,the model is trained on unlabeled data I.e data which is not given any labels. Unsupervised learning focuses on generating patterns in the data and remembering them or learning from them. One of the most common approach in unsupervised learning is "clustering" or grouping of data through means of some kind of analysis. New Data is then assigned to one of those newly found groups.

Reinforcement Learning

In this type of Learning , every good prediction done by the model includes a reward and every wrong prediction includes a penalty.In this way a feedback mechanism is incorporated with the learning process.

1.3 NATURAL LANGUAGE PROCESSING (NLP)

WHAT IS NATURAL LANGUAGE PROCESSING?

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. Natural language refers to the way we, humans, communicate with each other, namely, speech and text. We are surrounded by text. Think about how much text you see each day:

- Confessions
- Symbols
- Literature texts
- Signs
- Menus
- Email
- SMS
- Web Pages
- *and so much more…*

Now think about speech. We may speak to each other, as a species, more than we write. It may even be easier to learn to speak than to write, voice and text are how we communicate with each other.

Given the importance of this type of data, we must have methods to understand and reason about natural language, just like we do for other types of data.

What is Linguistics?

Linguistics is the scientific study of language, including its grammar, semantics, and phonetics. Classical linguistics involved devising and evaluating rules of language. Great progress was made on formal methods for syntax and semantics, but for the most part, the interesting problems in natural language understanding resist clean mathematical formalism. Broadly, a linguist is anyone who studies language, but perhaps more colloquially, a self-defining linguist may be more focused on being out in the field.

Mathematics is the tool of science. Mathematicians working on natural language may refer to their study as mathematical linguistics, focusing exclusively on the use of discrete mathematical formalism and theory for natural language (e.g. formal languages and automata theory).

Computational Linguistics

Computational Linguistics is the modern study of linguistics using the tools of computer science. Yesterday's linguistics may be today's computational linguist as the use of computational tools and thinking has overtaken most fields of study. It is the study of computer systems for understanding and generating natural language.

One natural function for computational linguistics would be the testing of grammars proposed by theoretical linguists, or human sentiment analysis, Twitter's data being very famously explored with the same. We have used this application of Computer Science in the field of Linguistics to generate a machine which predicts the human emotions among the two binary classes, that is happy or sad.

Large data and fast computers mean that new and different things can be discovered from large datasets of text by writing and running software.

In the 1990s, statistical methods and statistical machine learning began to and eventually replaced the classical top-down rule-based approaches to language, primarily because of their better results, speed, and robustness. The statistical approach to studying natural language now dominates the field; it may define the field.

Data-Drive methods and recommendation engines based on natural language processing have now become so popular that they must be considered mainstream approaches to computational linguistics. A strong contributing factor to this development is undoubtedly the increase amount of available electronically stored data to which these methods can be applied; another factor might be a certain disenchantment with approaches relying exclusively on hand-crafted rules, due to their observed brittleness.

Chapter 2

Literature Survey

Natural language processing (NLP) is a major area of artificial intelligence research, which in its turn serves as a field of application and interaction of a number of other traditional AI areas. Until recently, the focus in AI applications in NLP was on knowledge representation, logical reasoning, and constraint satisfaction - first applied to semantics and later to the grammar. In the last decade, a dramatic shift in the NLP research has led to the prevalence of very large-scale applications of statistical methods, such as machine learning and data mining. Naturally, this also opened the way to the learning and optimization methods that constitute the core of modern AI, most notably sentiment analysis and neural networks. In this paper/project we give an overview of the current trends in NLP in health care Analytics and mental health, discussing the possible applications of traditional AI techniques and their combination in this fascinating area.

Transfer learning ,i.e. transferring the knowledge built upon training complex Machine Learning and Deep Learning based models to a web or mobile based application, or sometimes known as domain adaptation, plays an important role in various natural language processing (NLP) applications, especially when we do not have large enough data for the task of interest (called the *target* task *T*). In such scenarios, we would like to transfer or adapt knowledge from other domains (called the *source* domains/tasks *S*) so as to mitigate the problem of Over fitting and to improve model performance in *T*. For traditional feature-rich or kernel-based models, researchers have developed a variety of elegant methods for domain adaptation; examples include EasyAdapt (Daume III, 2007; Daum ´ e III etal., 2010), instance weighting (Jiang and Zhai, 2007; Foster et al., 2010), and structural correspondence learning (Blitzer et al., 2006; Prettenhofer and Stein, 2010).

Recently, deep neural networks are emerging as the prevailing technical solution to almost every field in NLP. Although capable of learning highly nonlinear features, deep neural networks are very prone to over-fitting, compared with traditional methods. Transfer learning therefore becomes even more important. Thus keeping the applications in consideration, we have, in this project have trained our model using Multiple Layer Perceptron, a one layer dense, neural network. A comparative analysis of traditional algorithms versus the MLP classification task has been done.

Abstract Systems that extract structured information from natural language (NL) passages have been highly successful in specialized domains. The time is opportune for developing analogous applications for mental health applications ,as due to scenarios like lockdown and quarantine, a despair is natural for the time being, keeping this in consideration our team has worked upon the sentiment analysis of real life confessions web scrapped from popular sites on the internet and framed a web based application for the same. We present a system, *Alleviate, a machine that understands you!*

Chapter 3

Software Requirements Specification

3.1 Introduction

Good mental health is related to mental and psychological wellbeing. In this era of multiple simultaneous technological changes, people tend to neglect their mental health which leads to problems like social anxiety, obsessive compulsion disorder and various personality disorders.

Emotional and mental health is important because it's a vital part of your life and impacts your thoughts, behaviours and emotions. Being healthy emotionally can promote productivity and effectiveness in activities like work, school or sports.

3.2 Purpose

Our goal is to provide a platform in which people can communicate their mental worries, get help with the issues troubling their minds and rise or "Alleviate" their mental state to strive for a better world. We also like to share your burdens so that you can feel at ease.

Sharing your emotions helps release any anxiety you may be having. It can also help improve communication between people. There is always help and support out there when you need it, and by looking at our self-help recommendations above you could find some tips that maybe useful to yourself.

3.3 Functional Requirements

3.3.1 Dark Diary

3.3.1.1) **Add new entry**- User is given an option to add new confession to their personal diary.

3.3.1.2) **Edit/Update entry** - User can update previous entry or remove them as per their choice.

3.3.1.3) **Mood/Polarity Output** - The emotions reflected by the entries are analysed to estimate and display the results as necessary. Also, the posts made by user is added to the data store.

3.3.2 Dashboard

3.3.2.1) **View Posts -** User is presented with recent history of posts/entries sorted by their creation date which he/she has requested for analysis.

3.4 Non-Functional Requirements

3.4.1) **Privacy** - The data entered by the user will remain private and will not be disclosed in any manner that is against the user's best interests without prior permission.

3.4.2) **Scalability** - The website should be capable of handling multiple users without affecting its performance.

Chapter 4

Requirement Analysis

We aim to create a software for analysing the sentiments of humans and providing and classifying them as on the basis of two Primary emotions - Happy or Sad.

First of all, gathering or Eliciting requirements was done by collecting data from various websites (with due attention to privacy rules). Also, we have collected feedbacks and reviews from various Sources like social media and personal reviews to understand our exact requirements and model them according to several situations.

Manual processing is done on the collected data. Manual processing requires a human element in the analysis, specifically to help interpret language complexities such as context, ambiguity, sarcasm and irony.
Customer experience is also taken into account because it is necessary for customer to feel that his voice has been heard.

We identified the stakeholders of the software. The stakeholders for our software are listed as follows –

1. User - Customer
2. User - Administrator

The various high-level requirements for different stakeholders are listed as –

1. User -Customer
    - Main purpose of User is to keep a personal diary of various events and also view them along with their sentiments whenever necessary.
    - The User is required to log in into our server to accomplish the overmentioned tasks.
2. User - Administrator
    - Main purpose of the administrator is to manage various users and provide them with a seamless interface to interact with the system.
    - In case of any exceptional behaviour of the system, the administrator is responsible for bringing the system to a consistent state to resume operations.

Chapter 5

System Design

5.1 Data Flow Diagram (DFD)

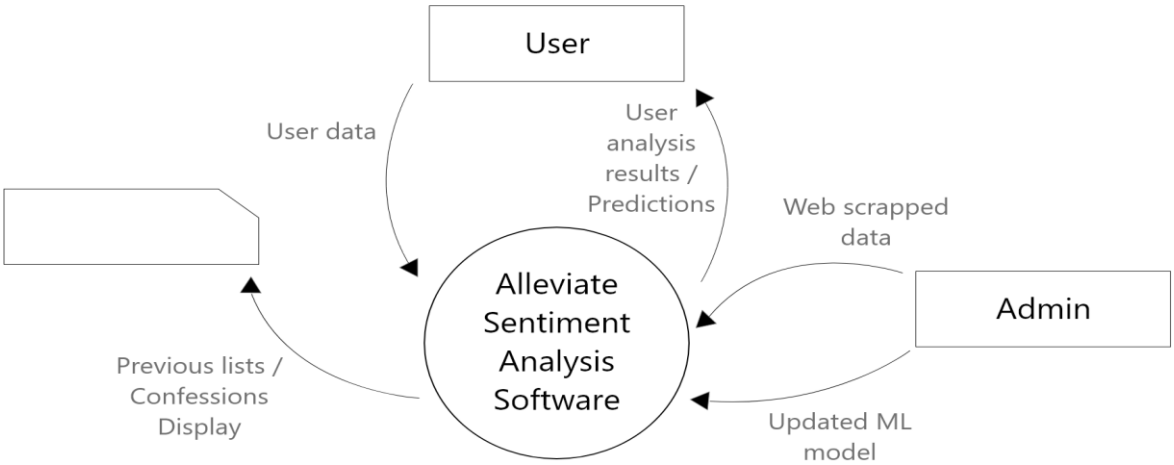


Fig 1 : Level – 0 (DFD)

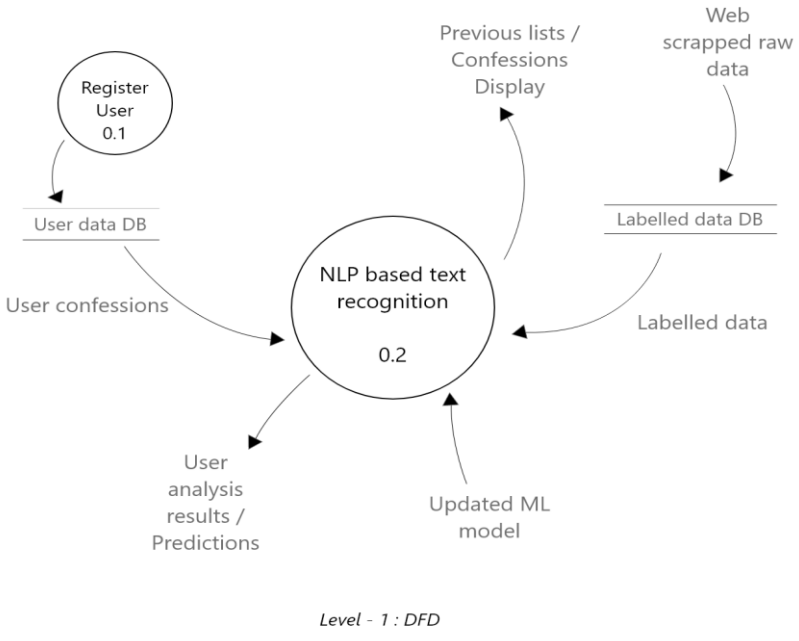


Fig 2 : Level – 1 (DFD)

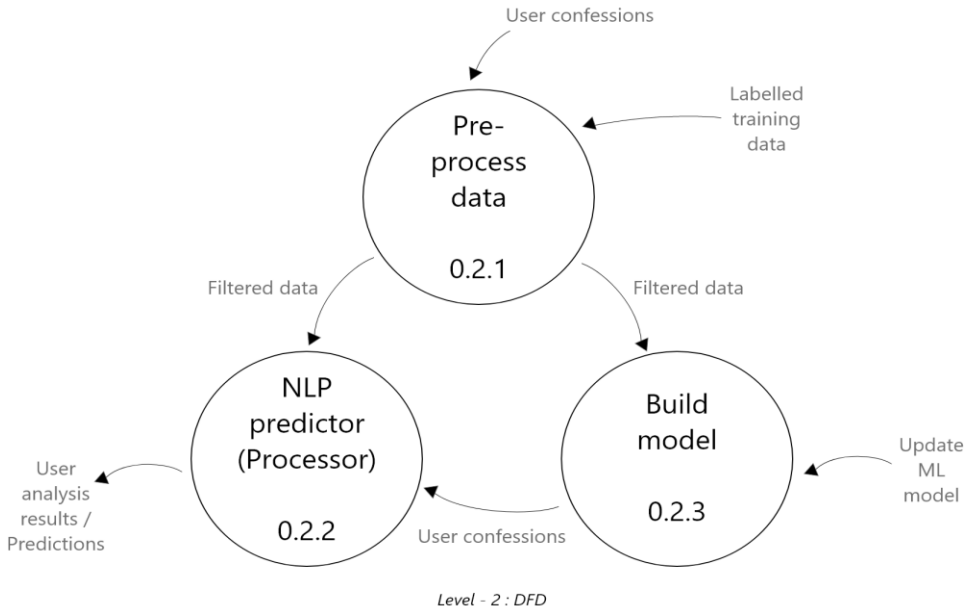
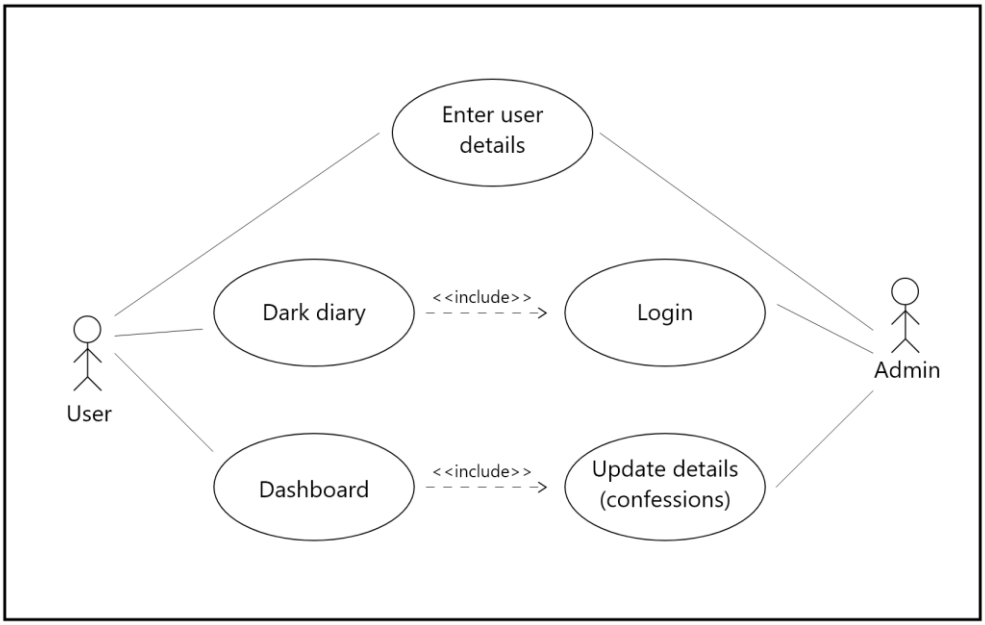


Fig 3 : Level – 2 (DFD)

5.2 Use-Case Diagram



Use case diagram

Fig 4 : Use Case Diagram

5.3 Class Diagram

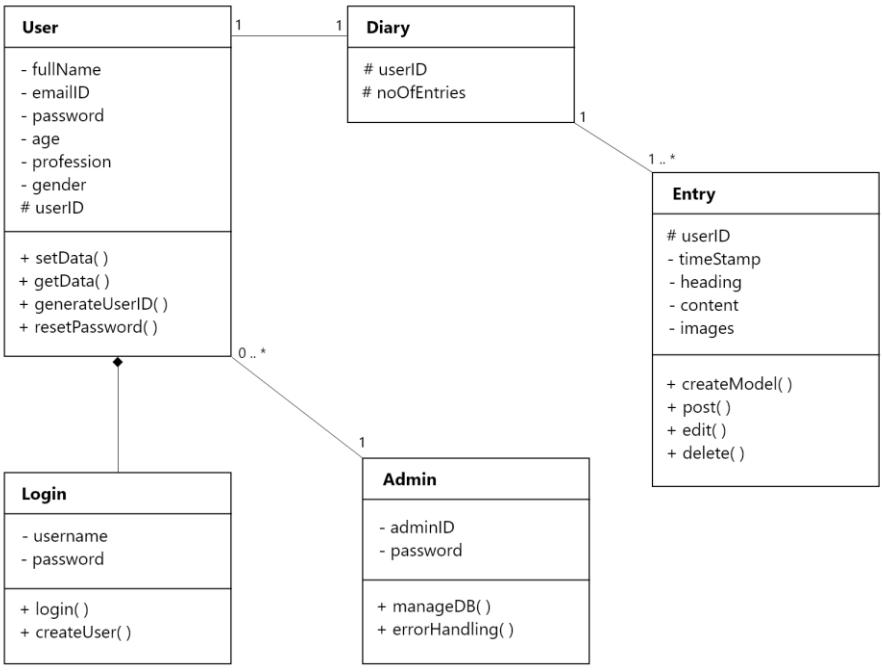


Fig 5 : Class Diagram

Chapter 6

System Testing

PURPOSE

This test plan describes the testing approach and overall framework that will drive the testing of the Alleviate – An ML based Web-app. The document introduces:

Test Strategy: rules the test will be based on, including the givens of the project (e.g.: start / end dates, objectives, assumptions); description of the process to set up a valid test (e.g.: entry / exit criteria, creation of test cases, specific tasks to perform, scheduling, data strategy).

- Execution Strategy: describes how the test will be performed and process to identify and report defects, and to fix and implement fixes.
- Test Management: process to handle the logistics of the test and all the events that come up during execution (e.g.: communications, escalation procedures, risk and mitigation, team roster)

PROJECT OVERVIEW

ALLEVIATE is a product that will help and bring people together who are either suffering from bad mental health and/or are willing to contribute to help spread the awareness for a better mental health & well-being and also wanting to help those who are suffering or going through tough times.
The functionality of this module spans through the entire system, making information available anywhere, anytime. All information is subject to company's defined security policy, where he/she can only view the information he/she is authorized to. An ESS-User can only edit certain fields in the ESS Module, maintaining the security and confidentiality of employee information

TEST ITEMS

Testing should be done on both front end and back end of the application on the Windows/Linux environments.

Test Principles

- Testing will be focused on meeting the business objectives, cost efficiency, and quality.
- There will be common, consistent procedures for all teams supporting testing activities.
- Testing processes will be well defined, yet flexible, with the ability to change as needed.
- Testing activities will build upon previous stages to avoid redundancy or duplication of effort.
- Testing environment and data will emulate a production environment as much as possible.
- Testing will be a repeatable, quantifiable, and measurable activity.
- Testing will be divided into distinct phases, each with clearly defined objectives and goals.
- There will be entrance and exit criteria.

Data Approach

- In functional testing, **Alleviate** will contain pre-loaded test data and which is used for testing activities.

Features to be Tested

The features which are to be tested are Login / Signup Page, Dashboard, Dark Diary(predictions).

Features not to be tested :-

The features which is in the software but not need to be tested as of now is :- a contact form and some database part.

APPROACH

We follow prototyping Methodology in this project And using functional testing to the functions of the software.

PASS/FAIL CRITERIA :-

All the major functionality of the application should work as intended and if the test cases contains positive words the test case will pass the result as happy or if it contains negative the test case will pass the result as sad.

SUSPENSION CRITERIA :-

If the processing takes more time it will lead to suspension or if input is in the form of images or anything other than text it will get automatically suspended

TEST DELIVERABLES :-

The testing will we be done on the test cases that contains in a csv file

TESTING TASKS :-

In this section, we specify the list of testing tasks we need to complete in the current project.
Example: Test environment should be ready prior to test execution phase. Test summary report needs to be prepared.

TEST ENVIRONMENT :-

Alleviate servers will be hosted at X company's site.

A windows environment with Internet Explorer 8, 9 and 10, and with Firefox 27.0, as well as Google Chrome 32.0 and later should be available to each tester.

APPROVALS :-

Who should sign off and approve the testing project.

Chapter 7

Project Planning

Project	Alleviate

Business Need / Problem

Alleviate is a Web based application that will help and bring people together who are either suffering from bad mental health and/or are willing to contribute to help spread the awareness for a better mental health & well-being and also wanting to help those who are suffering or going through tough times. They can use its features for writing their views/opinions/understanding on its Dark diary feature.

Objectives

We aim to create a software for analysing the sentiments of humans and providing and classifying them as on the basis of two Primary emotions - Happy or Sad.

Plan Scope

The idea was to create a Web application because nowadays most of the peoples are using web application. *Alleviate* aims to a companion to soothe all your worries so that you can relax and free yourselves from burdens and help you lead a peaceful life.

Action Steps

STEP 1:-

The very first and foremost important step was in the project is collecting data on which we had to train our model. So, First of all gathering or Eliciting requirements was done by collecting data from various websites (with due attention to privacy rules). Also we have collected feedbacks and reviews from various Sources like social media and personal reviews to understand our exact requirements and model them according to several situations.

STEP 2 :-

Data cleaning and Manual processing is done on the collected data.Manual processing requires a human element in the analysis, specifically to help interpret language complexities such as context, ambiguity, sarcasm and irony.

Customer experience is also taken into account because it is necessary for customer to feel that his voice has been heard.

STEP 3 :-

Trained an ML model and finds its accuracy on the dataset using which we are going to predict the emotion of the user (Happy/Sad) from their entered sentence.

STEP 4 :-

Developed a Login / Signup Page from where the user can get into the website.

STEP 5 :-

Developed the user dashboard and Home page and a text editor to take the user input.

STEP 6 :-

Deployed the Ml Algorithm using Flask and Python.

Technical Project Components

Requirement analysis -
We aim to create a software for analysing the sentiments of humans and providing and classifying them as on the basis of two Primary emotions - Happy or Sad.

First of all gathering or Eliciting requirements was done by collecting data from various websites (with due attention to privacy rules). Also we have collected feedbacks and reviews from various Sources like social media and personal reviews to understand our exact requiremnts and model them according to several situations.
Manual processing is done on the collected data. Manual processing requires a human element in the analysis, specifically to help interpret language complexities such as context, ambiguity, sarcasm and irony.
Customer experience is also taken into account because it is necessary for customer to feel that his voice has been heard.

We identified the stakeholders of the software.The stakeholders for our software are listed as follows -
3. User - Customer
4. User - Administrator

The various high level requirements for different stakeholders are listed as --
3. User -Customer

* Main purpose of User is To keep a personal diary of various events and also view them along with their sentiments whenever necessary.
* The User is required to log in into our server to accomplish the a forth mentioned tasks.
4. User - Administrator

* Main purpose of the administrator is to manage various users and provide them with a seamless interface to interact with the system.
* In case of any exceptional behaviour of the system , The administrator is responsible for bringing the system to a consistent state to resume operations.

-: Used Flask And Python For Backend deployment.
:- Used HTML , CSS , JavaScript , Bootstrap for Developing Frontend.

Signatures

Name/Title	Signature	Date
Abhish Anand (17050)	Abhish anand	29th May 2020
Niwanshu maheswari (1705051)	Niwanshu Maheswari	29th May 2020
Sidharth Purohit (1705077)	Sidharth purohit	29th May 2020
Shubham Kumar Maurya (1705074)	Shubham maurya	29th May 2020
Suraj Kumar Mishra (1805951)	Suraj Mishra	29th May 2020

Chapter 8
Implementation

NLP TOOLKIT TECHNIQUES USED

NLTK (NATURAL LANGUAGE TOOL KIT)

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum.

Thanks to a hands-on guide introducing programming fundamentals alongside topics in computational linguistics, plus comprehensive API documentation, NLTK is suitable for linguists, engineers, students, educators, researchers, and industry users alike. NLTK is available for Windows, Mac OS X, and Linux. Best of all, NLTK is a free, open source, community-driven project.

NLTK has been called "a wonderful tool for teaching, and working in, computational linguistics using Python," and "an amazing library to play with natural language."

Natural Language Processing with Python provides a practical introduction to programming for language processing. Written by the creators of NLTK, it guides the reader through the fundamentals of writing Python programs, working with corpora, categorizing text, analyzing linguistic structure, and more.

Removing stop words with NLTK in Python

The process of converting data to something a computer can understand is referred to as **pre-processing.** One of the major forms of pre-processing is to filter out useless data. In natural language processing, useless words (data), are referred to as stop words.

What are Stop words?

A stop word is a commonly used word (such as "the", "a", "an", "in") that a search engine has been programmed to ignore, both when indexing entries for searching and when retrieving them as the result of a search query. We would not want these words to take up space in our database, or taking up valuable processing time. For this, we can remove them easily, by storing a list of words that you consider to stop words. NLTK(Natural Language Toolkit) in python has a list of stop words stored in 16 different languages. You can find them in the nltk_data directory.

To check the list of stopwords you can type the following commands in the python shell.

```
import nltk

from nltk.corpus import stopwords

print(stopwords.words('english'))
```

Stemming words with NLTK

The idea of stemming is a sort of normalizing method. Many variations of words carry the same meaning, other than when tense is involved.

The reason why we stem is to shorten the look up, and normalize sentences.

Consider:

```
I was taking a ride in the car.
I was riding in the car.
```

This sentence means the same thing. in the car is the same. I was is the same. the “ing” denotes a clear past-tense in both cases, so is it truly necessary to differentiate between ride and riding, in the case of just trying to figure out the meaning of what this past-tense activity was?

This is just one minor example, but imagine every word in the English language, every possible tense and affix you can put on a word. Having individual dictionary entries per version would be highly redundant and inefficient, especially since, once we convert to numbers, the "value" is going to be identical.

One of the most popular stemming algorithms is the Porter stemmer, which has been around since 1979, which is being used in this project as well.

To use the PorterStemmer you can type the following commands in the python shell.

```
from nltk.stem import PorterStemmerfrom nltk.tokenize import sent_tokenize, word_tokenize

ps = PorterStemmer()
```

Making bag of words with NLTK

Whenever we apply any algorithm in NLP, it works on numbers. We cannot directly feed our text into that algorithm. Hence, Bag of Words model is used to preprocess the text by converting it into a *bag of words*, which keeps a count of the total occurrences of most frequently used words.

This model can be visualized using a table, which contains the count of words corresponding to the word itself.

Step #1 : We will first pre-process the data, in order to:

Convert text to lower case.

Remove all non-word characters.

Remove all punctuation.

Step #2 : Obtaining most frequent words in our text.

We will apply the following steps to generate our model.

We declare a dictionary to hold our bag of words.

Next we tokenize each sentence to words.

Now for each word in sentence, we check if the word exists in our dictionary.

If it does, then we increment its count by 1. If it doesn't, we add it to our dictionary and set its count as 1.

Step #3 : Building the Bag of Words model

In this step we construct a vector, which would tell us whether a word in each sentence is a frequent word or not. If a word in a sentence is a frequent word, we set it as 1, else we set it as 0.

The task of vectorizing the words present in the text can also be performed using the CountVectorizer function in NLTK toolkit.

The CountVectorizer provides a simple way to both tokenize a collection of text documents and build a vocabulary of known words, but also to encode new documents using that vocabulary.

You can use it as follows:

1. Create an instance of the *CountVectorizer* class.
2. Call the *fit()* function in order to learn a vocabulary from one or more documents.
3. Call the *transform()* function on one or more documents as needed to encode each as a vector.

An encoded vector is returned with a length of the entire vocabulary and an integer count for the number of times each word appeared in the document.

Because these vectors will contain a lot of zeros, we call them sparse. Python provides an efficient way of handling sparse vectors in the SciPy. Sparse package.

The vectors returned from a call to transform() will be sparse vectors, and you can transform them back to NumPy arrays to look and better understand what is going on by calling the toarray() function.

Data Preparation

This is the first step of our project to collect the raw data. We scraped the dataset from various confessions sites on google . The tool we used to scrap the dataset is **Selenium.**

Selenium is a free open source tool used mainly for automation testing.

- Selenium Integrated Development Environment (IDE)
- Selenium Remote Control (RC)
- WebDriver
- Selenium Grid

Here we used a selenium Web Driver to open a automated google chrome which is controlled by our code . There are some code snippets attached too for reference.
Here we open the automated browser and open the desired url .

```
In [ ]: from selenium import webdriver

browser = webdriver.Chrome('D:\programs\python\Programs\Day8\selenium\chromedriver.exe')
url = 'http://www.confessions.net/'
browser.get(url)
```

These below line are used to scrap the data using xpath .

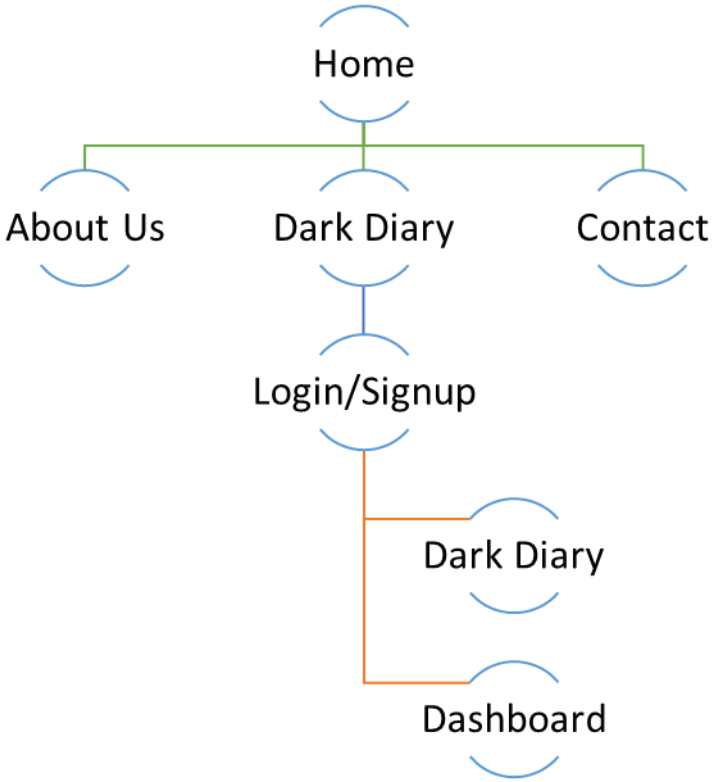
```
In [ ]: data = [] #Empty list
for j in range(1,10): # From Page 1-9
    browser.find_element_by_xpath('/html/body/div[2]/div/div/div[4]/p/a['+str(j)+']').click()
    sleep(5)
    for i in range(1,19):
        data.append(browser.find_element_by_xpath('/html/body/div[2]/div/div/div[3]/div[1]/div['+str(i)+']').text)

np.save('data.npy',data )
```

User Experience of Alleviate

Target Audience

- Youth suffering from poor mental health issues (anxiety, depression etc).
- People trying to spread awareness regarding better mental health.
- People who need help and/or trying to help others.
- Professionals (Psychologists, Psychiatrists etc).

Site Map



Focus Points

- Reliable and efficient sentiment analysis
- Ease to use interface
- Self-intuitive and minimalistic design

User Interface of Alleviate

Colour Scheme

These colours were chosen in such a way to closely represent the emotions (mainly to reflect the range of emotions Alleviate can process) and provide an easier to look at UI and to provide some contrast to the idea.

The Logo

The logo was designed in such a way that captures two main mottos of Alleviate.

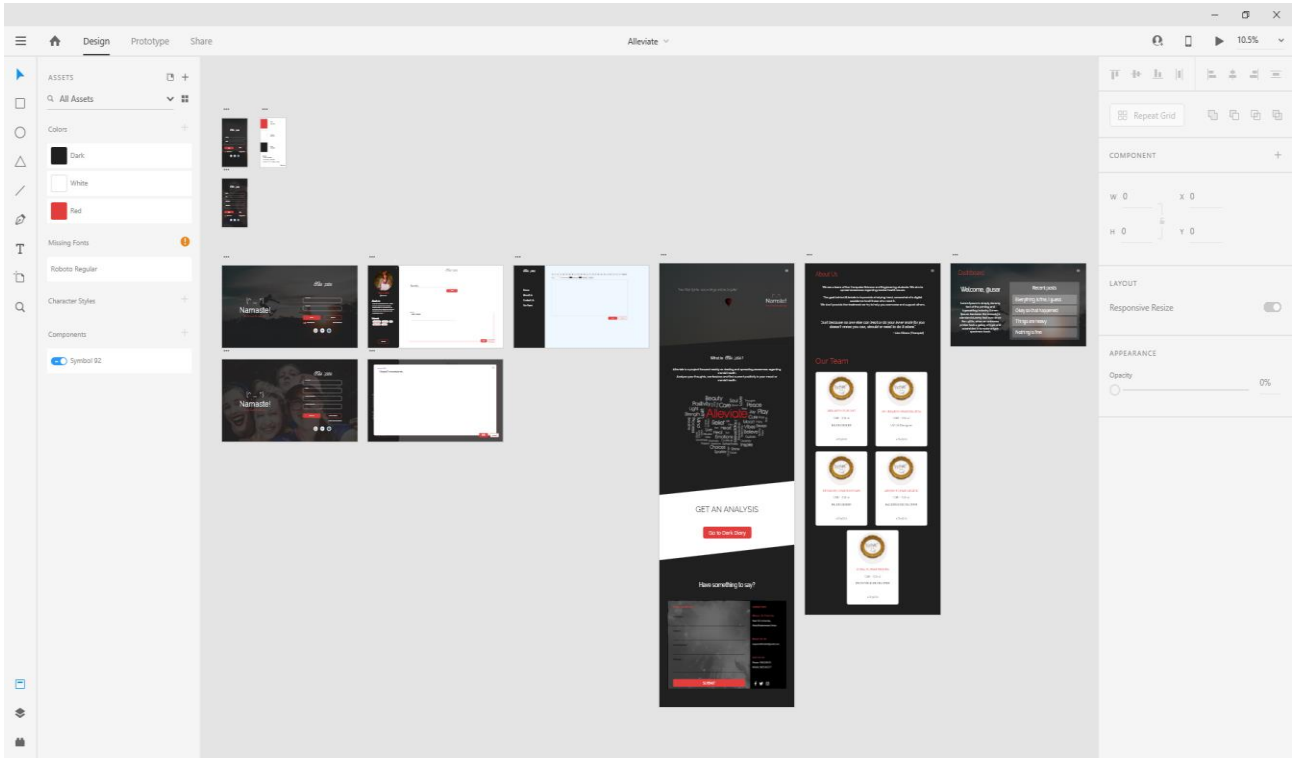
1. **Care** – *Heart in place of 'v'*
2. **Continuity** – Semicolon (;) in place of 'i'

   Semicolon became a symbol for mental health awareness. It symbolizes "where an author could have ended a sentence, but didn't. That author is you and the sentence – your life" (Lakey).

Fonts

Srisakdi – *For logo*
Raleway – *For contents of the website*

Screenshot of the project's UI designs (Prepared in Adobe XD)



Frontend Development of Alleviate

Technologies

The Core Web Technologies we used are :-

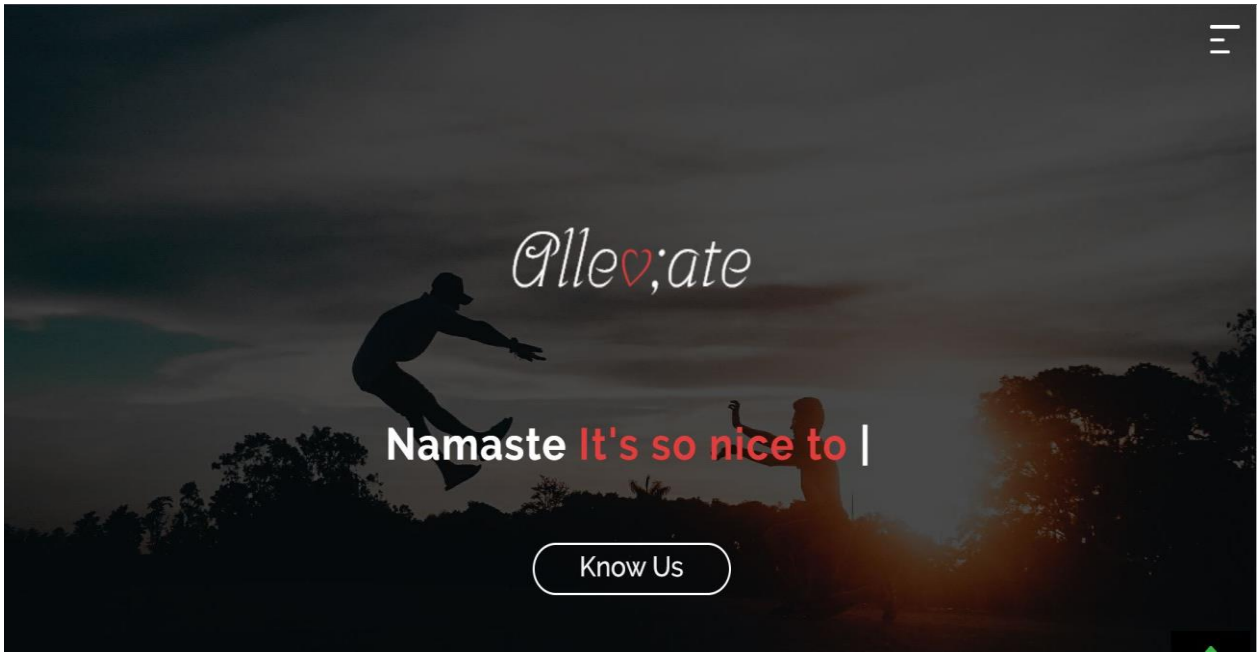
- HTML
- CSS
- JAVASCRIPT

Frameworks and Library

The Frameworks and Libraries used are :-

- Bootstrap
- jQuery
- scroll reveal (JS library)

Screenshot of the Home Page



This floating Animation effect in the Home Page is implemented using the **Scroll reveal** javascript library**.**
ScrollReveal is a JavaScript library for easily animating elements as they enter/leave the viewport. It was designed to be robust and flexible.

```
<script src="https://unpkg.com/scrollreveal"></script>
```

This script we implemented in our html code within the head tag. This will create the global variable ScrollReveal.
We further used this variable to specify the design

```
<script>
  window.sr = ScrollReveal({
    distance: '50px',
    duration: 5000,
    easing:'ease',
    mobile: true,
    reset: true,
    viewFactor: 0.4,
  });
</script>
```

You can see a Typing effect in the Text on the Home page in the above screenshot. It was given using the JavaScript.

```
<script src="js/pre.js"></script>
<script src="js/typed.js"></script>
<script src="js/scroll.js"></script>
```

The following JS script files are included . The **scroll.js** is used for giving the scrolling effects and The **typed.js** is used for giving the typing effect which shown in the starting on the Home page. It is called using the script provided below :-

```
<script>
        var typed = new Typed(".type", {
          strings: [" User !!",
           " It's so nice to see You !!",

           ],
          typeSpeed:60,
          backSpeed:60,
          loop:true
        });
  </script>
```

The **pre.js** is included for the preloader which you can see in the starting of the website.

```
var preloader = document.getElementById("loading");

function loader(){
    preloader.style.display = 'none';
};
```

The loader is further called in the body tag of the html . The designing of the loader is set in the **pre.css** file .

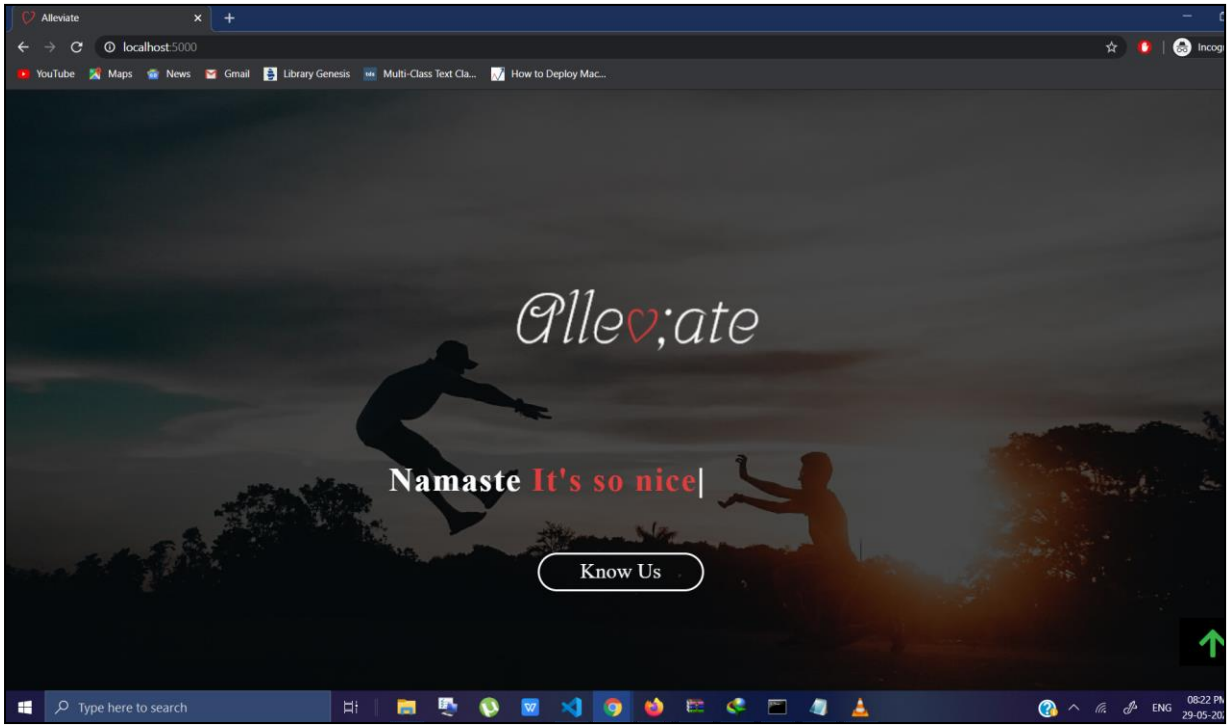
```
<body onload="loader()">
    <!-- Preloader -->
    <div class="container-fluid" id="loading">

    </div>
    <!--/preloader-->
```

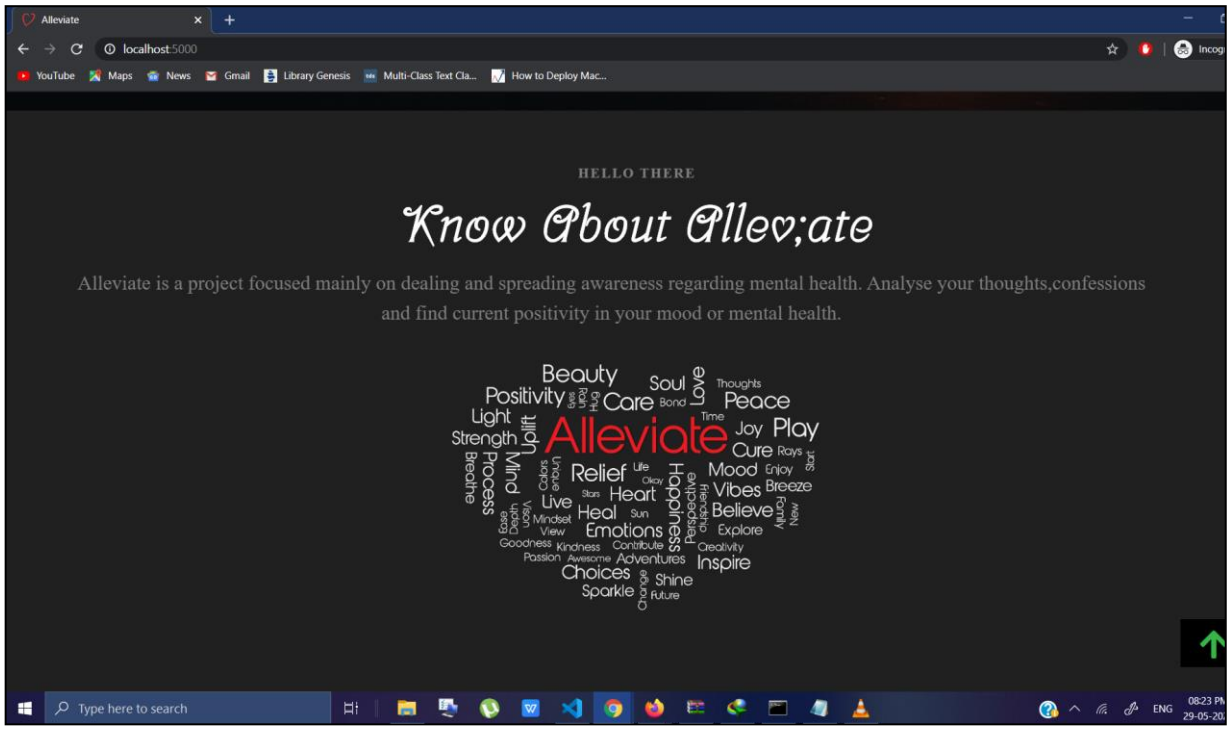
For The rest of the screenshots of the website you can Refer **Chapter :- 9**

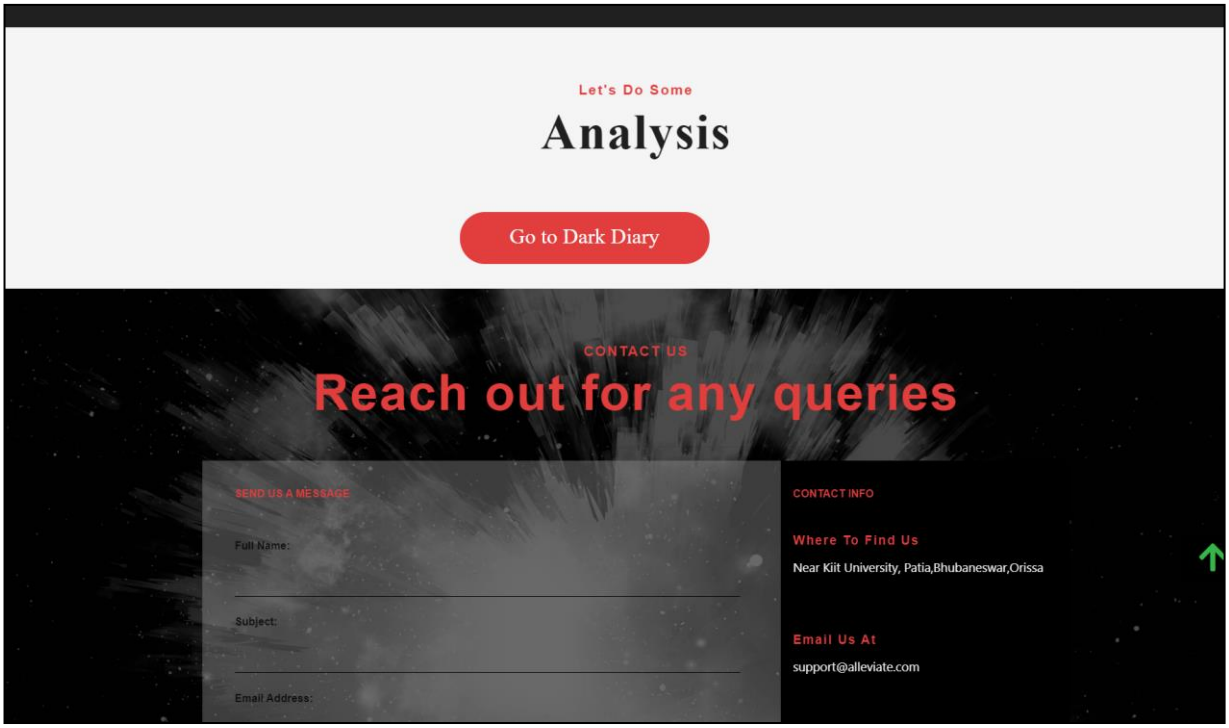
Chapter 9

Screen shots of Project

9.1 HOME PAGE







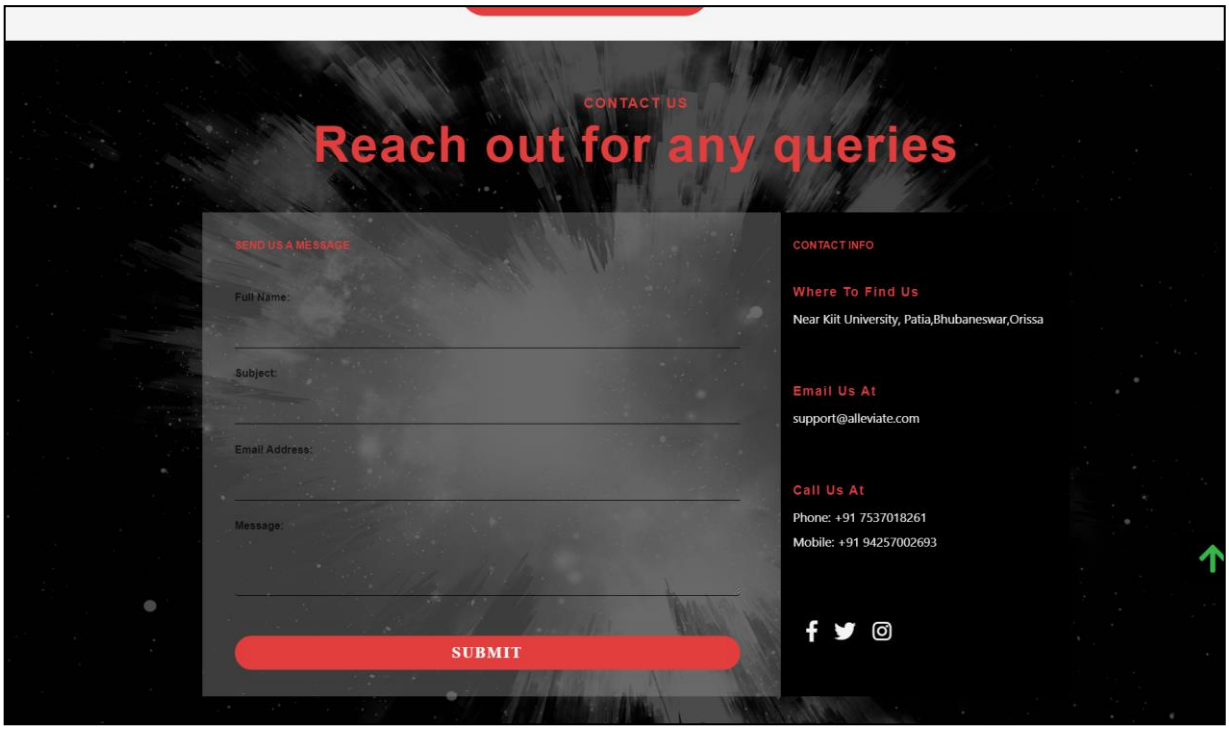


Fig 6 : Home Page Screenshots

9.2 DASHBOARD

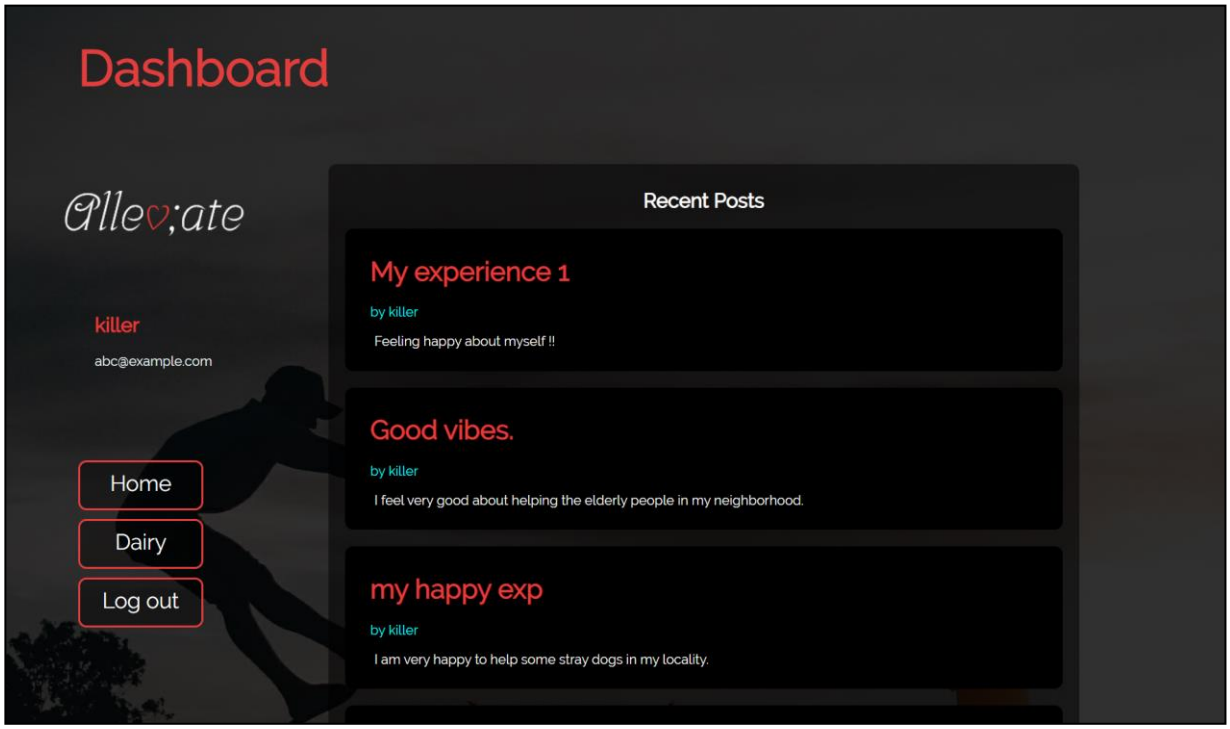


Fig 7 : Dashboard Screenshot

9.3 LOGIN

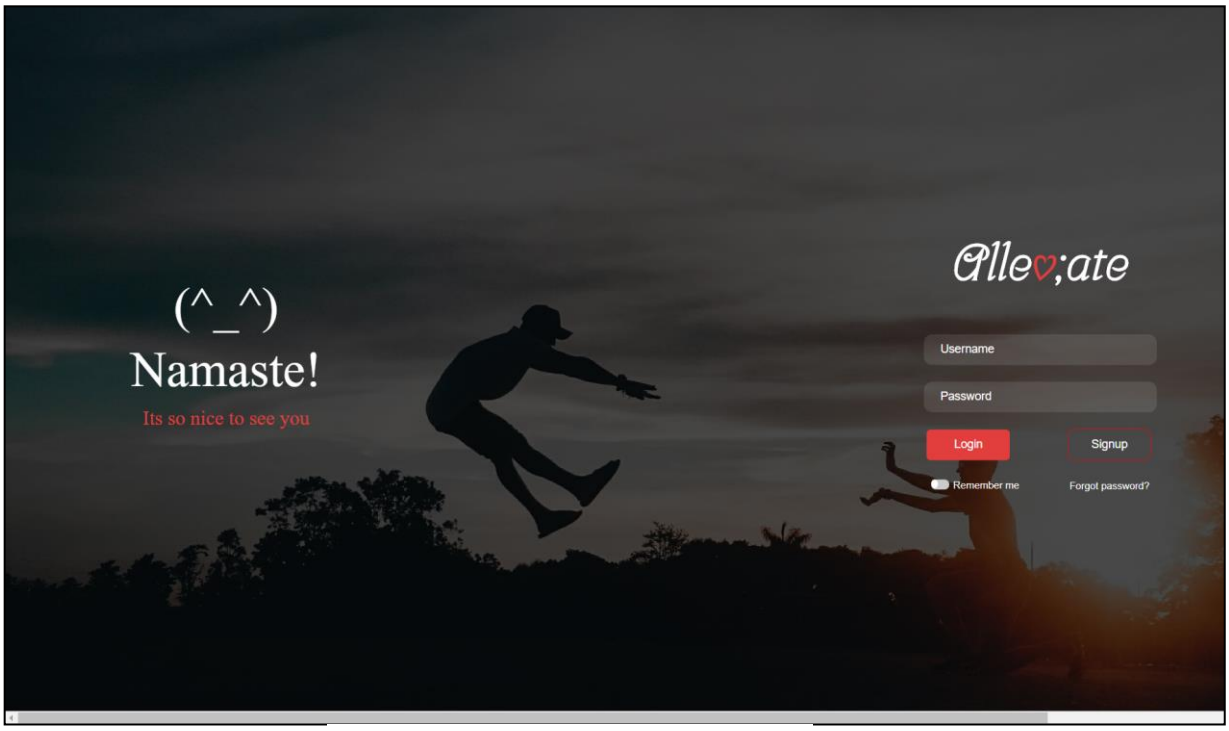


Fig 8 : Login Screenshot

9.4 SIGNUP

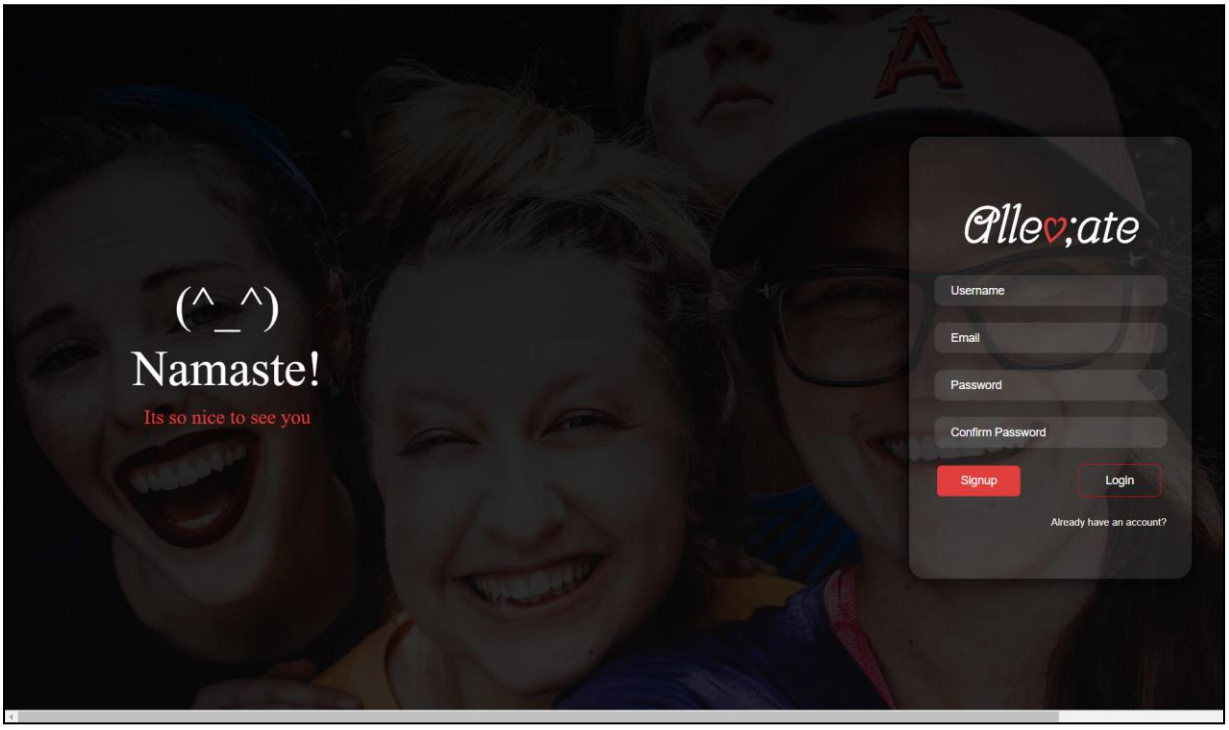


Fig 9: Signup page sceenshot

9.5 DARK DIARY

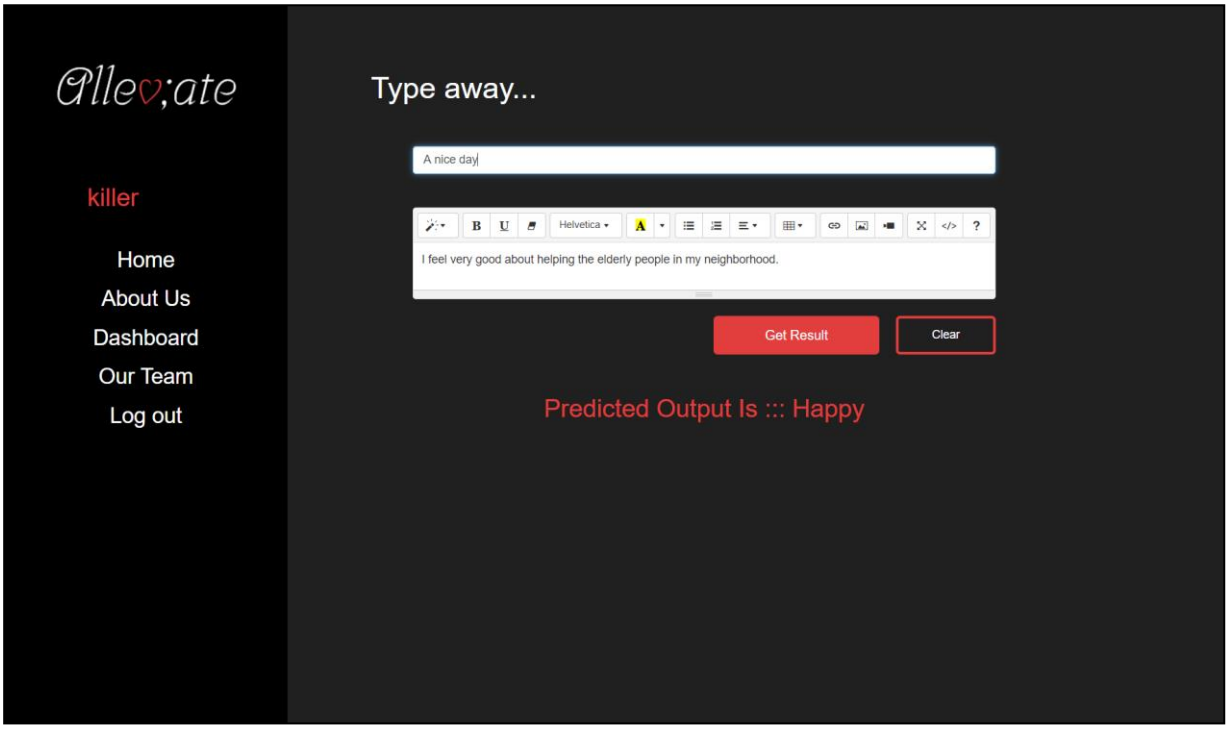


Fig 10: Dark diary screenshot

Chapter 10

Conclusion and Future Scope

10.1 Conclusion

The idea behind Alleviate is to spread awareness regarding good mental health and help those who are need by providing them with the insights of their own thoughts.
This was made possible by using and combing several different technologies to create one whole web based application backed by Machine Learning and authenticated user account based system *Alleviate*.

10.2 Future Scope

1. Better prediction accuracy
2. Complex sentence and sentiment analysis
3. More secure digital journaling system
4. More secure and expandable databases

References

1. https://www.oosay.net/
2. https://www.quora.com/Whats-one-anonymous-confession-you-want-to-make-on-Quora
3. http://www.confessions.net/
4. https://www.secret-confessions.com/
5. http://www.confessions4u.com/