PUCIT
Punjab University College of Information Technology

Social Media Web Application for prevention of mental disorders in the posts using Natural Language Processing

Version 2.0

TABLE OF CONTENTS

FINAL PROJECT PROPOSAL GUIDE 3

1. INTRODUCTION 3
1.1 PROJECT TITLE 3
1.2 PROJECT OVERVIEW STATEMENT 3
1.4 PROJECT GOALS & OBJECTIVES 6
1.5 HIGH-LEVEL SYSTEM COMPONENTS 6
1.6 LIST OF OPTIONAL FUNCTIONAL UNITS 8
1.7 EXCLUSIONS 8
1.8 APPLICATION ARCHITECTURE 8
1.9 GANTT CHART 9
1.10 HARDWARE AND SOFTWARE SPECIFICATION 9
1.11 TOOLS AND TECHNOLOGIES USED WITH REASONING 10

Final Project Proposal Guide

1. Introduction

Mental health is a serious issue in the modern-day world. According to the World Health Organization's 2021 report, more than 20% of the world's children and the adult population suffers from an episode of a mental disorder in their life. The problem grows bigger with the fact that as much as 35–50% of those affected go undiagnosed and receive no treatment for their illness. In line with WHO's Mental Health Action Plan, the natural language processing community helps the gathering of information and evidence on mental conditions, focusing on text analysis of authors affected by mental illnesses.

All the major social media platforms (Instagram, Twitter, Facebook, etc) do not work on this to stop the mental disorder in text or images, or audio/videos. Users are not able to prevent themself from this type of the content.

Here in this project, we want to build a system/application where a user can view the posts/text/images according to his/her interest, but these posts/text/images are prevented from mental disorders like depression, tension, anxiety, etc. Here we use different techniques and tools of Machine Learning, Natural Language Processing, and Deep learning to analyze the text/images of social media posts to find the mental health disorders. After filtering out the text/images, we will recommend/display those posts to the user which will not be harmful to her/his mental health.

Eventually, Our system is a research and development-based project, in which we build a web-based application where users can sign-up/log in with his/her credentials and use the application effectively.

1.1 Project Title

The project title is "**Social Media Web Application for prevention of mental disorders in the posts using Natural Language Processing**".

1.2 Project Overview Statement

Every social media platform provides free hands to its user to view any type of content whether the content is good or bad for his/her mental health. In addition to this, these social media platforms do not work to prevent the material which is harmful to a particular user.

Here we aim to build a system or application where a user can view the text/images which are almost free from mental health issues and he/she will also be able to post or upload image/text of her/his own choice. In this project, firstly we scrape the text/images from social media platforms, after that, we use different machine learning, deep learning, and Natural Language processing techniques tools and techniques to analyze the text/images to infer the nature of the text/images. After these operations, we build a model to perform prediction on new data. After building a model, we will evaluate our model and its performance so that it gives us better results on new data. After successful model building, we deploy we system/project in the form of the web application. There are the following steps we need to perform to create this system.

- Data Acquisition
- Text/Image Preprocessing
- Data Visualization
- Feature Engineering
- Model Building
- Model Evaluation
- NLP Pipeline
- Application Development
- Model Deployment

Project Overview Statement Template

Project Title:
Social Media Web Application for prevention of mental disorders in the posts using Natural Language Processing

Group Leader: Ehtisham Sadiq

Project Members:

Name	Registration #	Email Address	Signature
Zainab Saeed	BSEF19M516	bsef19m516@pucit.edu.pk	
Ehtisham Sadiq	BSEF19M521	bsef19m521@pucit.edu.pk	
Tayyaba Nasir	BSEF19M525	bsef19m525@pucit.edu.pk	

Ayesha Rashid	BSEF19M531	bsef19m531@pucit.edu.pk	

Project Goal:
Build a web application that will be used by the users to view images/textual posts of different social media platforms to prevent/stop mental health issues. Starting from data scraping to model building and evaluation, the system will perform each operation in dynamic or runtime.

Objectives:

Sr.#	
1	Scrape data from different social media platforms using APIs and Selenium
2	Text/Image Pre-Processing on scraped data using POS Tagging, Lemmatization etc
3	Data Visualization is used to represent textual data in the form of plots, charts, graphs using seaborn, scikit-learn and wordcloud.
4	Feature Extraction from Textual/images data using word embedding techniques.
5	Model Building using ML , NLP or DP algorithms.
6	After Model Building, Model Evaluation comes to evaluate performance of model using different evaluation metrics.
7	Concept of Pipeline is used to connect all the components into single flow. Scitkit-learn provides the functionality to build the pipeline.
8	In Development phase, we will create a web application using HTML, CSS and JavaScript for front-end and django for backend.
9	Model deployment , here we integrate our trained model with our web application using django and deploy on server for users.

Project Success criteria:
The system or application detects information related to mental disorders and performs all operations accurately on newly coming data like text/image Pre-Processing, Feature engineering, etc. Web Applications will display the images/texts according to the choice of the user. The most important success criterion for a project is the supervision of an experienced supervisor.

Assumptions, Risks and Obstacles:
Dirty or uncleaned or imbalanced data destroy the performance as well as project workflow. Train one or more algorithms to find the best model. Maintains system integrity and efficiency.

Organization Address (if any):
Faculty of Computing & Information Technology, Lahore

Type of project: ❑Research ❑Development

Target End users: Any users or organization

Development Technology: ❑Object Oriented ❑Structured

Platform: ❑Web based ❑Distributed
❑Desktop based ❑Setup Configurations
❑Other_____

Suggested Project Supervisor: Dr. Arif Butt

Approved By:

Date:

1.4 Project Goals & Objectives

Following are the goals of the project:

- Build a web application which will be used by the users to view images/textual posts of different social media platforms to prevent/stop mental health issues.
- Starting from data scraping to model building and evaluation , the system will perform each operation in dynamic or runtime.
- System must run in online-learning mode.

Following are the objectives of the project:

- Scraping of data from different social media websites like Twitter, Facebook, Instagram etc. using APIs and Selenium.
- Text/Image Pre-Processing on raw data using POS Tagging, Lemmatization etc to convert into raw data into useful information.
- Feature Engineering to extract useful or required features or attributes from pre-processed data using word embedding techniques.
- Model Building to train or test the featured data using ML, NLP or DP algorithms.
- Model Evaluation to evaluate the performance or accuracy of the model using model metrics like confusion matric or RMSV etc.
- Model deployment, in this phase, we put our trained model into production using Web Base technology like Django etc.

1.5 High-level system components

Following are the high level system components

1. **Data Acquisition:**
   - Selenium to automate web browser interaction with python.
   - tweepy is a library to access the Twitter API with Python.
   - instaloader a tool to download images from Instagram.
   - facebook-scraper is a tool to extract publicly available data from Facebook.
2. **Text Preprocessing:**
   - Lower Casing converts text into lower case.

- ➢ Tokenization is breaking down of a paragraph into smaller units like sentence or words.
- ➢ Punctuation Mark Removal to remove punctuation from text.
- ➢ Stop Word Removal which are the most frequently occur in any language.
- ➢ Stemming is reduction of a word into its root or stem word.
- ➢ Lemmatization is a text normalization technique.

3. Data Visualization:

- ➢ Text statistics visualizations contains very insightful techniques. They include: word frequency analysis, sentence length analysis, average word length analysis,
- ➢ Seaborn, Plotly, Bokeh, Matplotlib, Scipy libraries can be used for data visualization
- ➢ Wordcloud is a great way to represent text data. The size and color of each word that appears in the wordcloud indicate it's frequency or importance.

4. Feature Engineering:

- ➢ Parsing
- ➢ PoS Tagging
- ➢ Name Entity Recognition (NER)
- ➢ Bag of Words (BoW)
- ➢ Term Frequency-Inverse Document Frequency (TF-IDF)
- ➢ word2Vec

5. Model Building:

ML Approach:

- ➢ Supervised ML Algorithms
- ➢ Unsupervised ML Algorithms

DP Approach:

- ➢ CNN
- ➢ RNN
- ➢ LSTM
- ➢ GRU

6. Model Evaluation:

- ➢ Accuracy
- ➢ Precision
- ➢ Recall

- F1 Score
- Area Under the Curve (AUC)
- Mean Reciprocal Rank (MRR)
- Mean Average Precision (MAP)
- Root Mean Squared Error (RMSE)

7. NLP Pipeline:

- The set of ordered stages one should go through from a labeled data-set to creating a classifier that can be applied to new samples.
- Scikit-learn

8. Application Development:

- Django
- HTML & CSS
- Java-script

9. Model Deployment:

- Django is a back-end server side web framework.
- Docker building, deploying, and managing containerized applications.
- Server Like AWS, Azure etc

1.6 List of optional functional units

Following are the optional functional units:

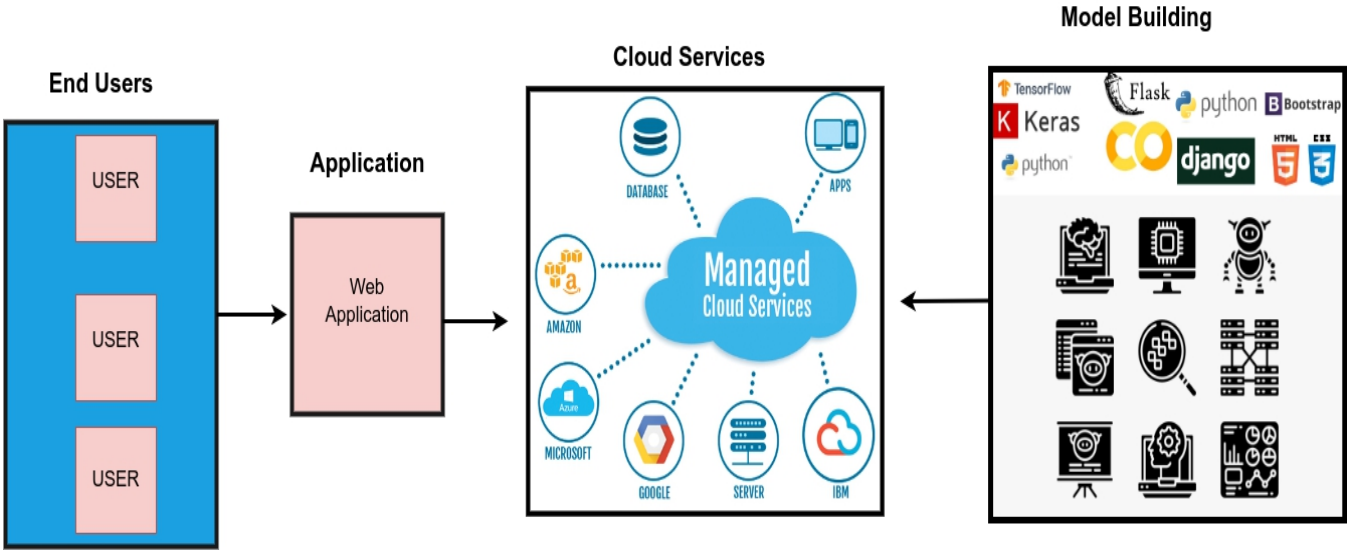
User can post her/his own text/image on his/her profile.

User can recommend some text/images on the basis of likes/dislikes.

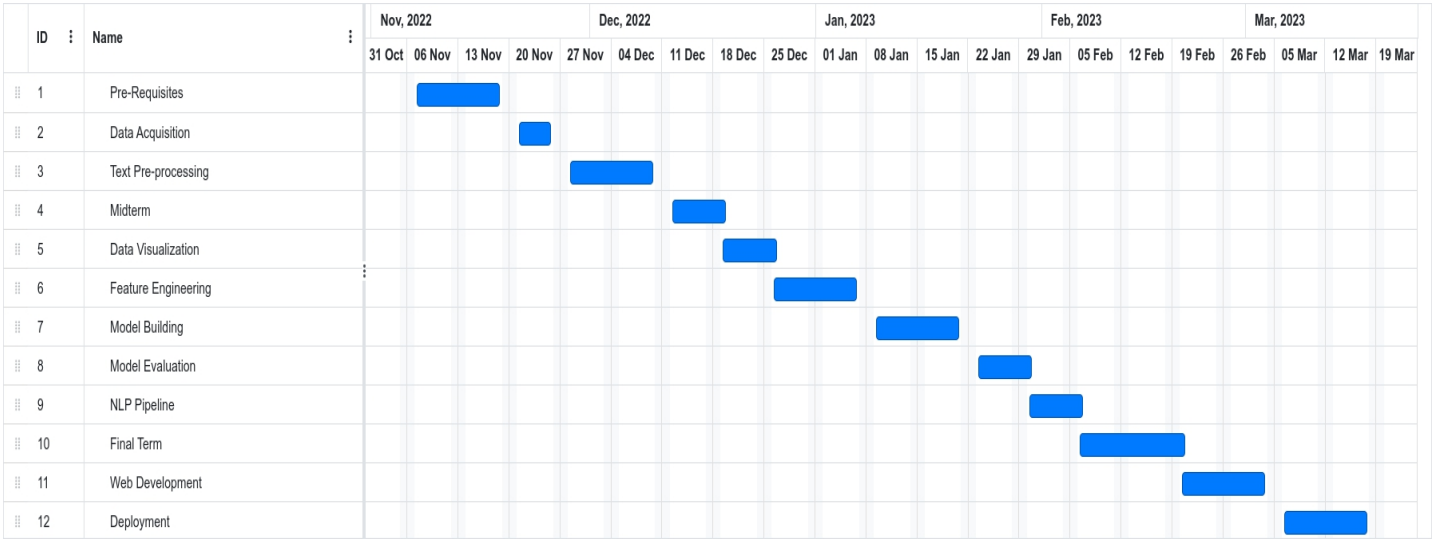
1.7 Exclusions

Audio and Video data will not be treated in this project due to time and cost constraints. You-tube and Reddit platforms also will not be treated in this project.

1.8 Application Architecture



1.9 Gantt chart



1.10 Hardware and Software Specification

Web Application
Memory : 4GB
Graphics Card: AMD Radeon R5 M230
CPU: Intel Core i3-2340UE
Window: 10 or 11

1.11 Tools and technologies used with reasoning

The development process.:

- Iterative development process will be used for this project. Functionalities for different components/modules will be released in an iterative manner.

Host (or development) platform(s).

- Any OS supporting Python and JavaScript, HTML, CSS development.

Target platform(s).

- All OS supporting internet browsers

The programming language(s) to be used.

- **UI:** User interface will be written in JavaScript, HTML, CSS. Java-script provides a much better systematic way of building dynamic client side rendered applications.
- **Server:** Python will be used for the development of server. Python provides relatively simple way of creating back-end servers.
- **Model Building**: Python and its libraries like Tensor-flow , Scikit-learn etc.

Existing tools.

- **Visual Studio Code:** For development of UI. VS code is a flexible code editor for HTML, CSS, JavaScript and other programming and markup languages.
- **Jupyter Notebook:** It is a easiest tool to build ML and DL model.

The size of the development effort.

- A team of 3 to 4 persons. One person for designing and developing the user interface. One person for managing the models. Two or three persons for developing the server.

Budget and time constraints

- The project can be completed in 4-5 months with a budget of 1 million PKR.

1.12 References

√ https://www.mihaileric.com/posts setting-up-a-machine-learning-project/
√ https://towardsdatascience.com/data-visualization-for-machine-learning-and-data-science-a45178970be7
√ https://builtin.com/machine-learning/nlp-machine-learning
√ https://towardsdatascience.com/my-first-nlp-pipeline-99d24aafb773