Project Plan

Sicuro: Understanding Mental Health Better – A Data Driven Approach

Final Year Project

SUPERVISOR: Dr. Yu, Tao
Department of Computer Science

AUTHOR: Sahabudeen Asimudeen
Bachelor of Engineering (Computer Science)

October 2, 2022
Contents

LIST OF FIGURES .................................................................................. 3
LIST OF TABLES ................................................................................... 3
1. INTRODUCTION .................................................................................. 4
   1.1 BACKGROUND .............................................................................. 4
   1.2 PROBLEM STATEMENT – THE CHALLENGES OF UNDERSTANDING ONE’S MENTAL HEALTH .................................................. 4

2. OBJECTIVES ...................................................................................... 5
   2.1 EDUCATIONAL OBJECTIVE .......................................................... 5
   2.2 RESEARCH OBJECTIVE .................................................................. 5

3. RELATED WORK .................................................................................. 6
   3.1 METHODOLOGY OF RELATED WORK ............................................ 6

4. METHODOLOGY ................................................................................ 7
   4.1 PART 1: DATA COLLECTION ........................................................... 7
      4.1.1 Types of data to be collected ..................................................... 8
   4.2 PART 2: ANALYSIS ......................................................................... 8
   4.3 PART 3: DISPLAY ............................................................................ 10

5. SCHEDULE ....................................................................................... 10

6. REFERENCES .................................................................................... 11
List of Figures
Figure 1 Architecture diagram of project workflow (data flow perspective) 7

List of tables
Table 1 Objective and subjective types of data to be collected 8
Table 2 Project milestones 10
1. Introduction

1.1 Background

The advancement of mhealth (mobile health) applications regarding mental health was a significant turning point in how we interact with behavioral change and emotional well-being. Most mental health apps can provide information, monitor mood, and track behavioral patterns. These apps have been known to play a significant role in creating awareness and in building support for mental health [1].

Mental health is vital at every stage of life, from childhood and adolescence through adulthood. Our genes, life experiences, upbringing, and environment all affect our mental health and influence how we think and respond to situations. Being aware of these factors can make it easier to understand how we operate [2]. By monitoring and learning about one’s own emotions, it was reported that people felt more in control of their mood, made more informed decisions with the goal to become happier, and became more confident and positive in their emotional well-being [3]. In short, the first step to emotional well-being is self-awareness and self-understanding. Self-tracking mood helps users increase their awareness and proactive self-regulation of their emotional well-being. However, there exists a variety of challenges when understanding one’s mental health: multifaceted in nature, lack of data, and affordability.

The following is a project plan to build a data-driven mental health application to 1) aid users in understanding their mental health, by enabling users to be aware of what affects their emotions and contextualize data, and 2) to analyze how demographic factors affect mental health.

1.2 Problem statement – The challenges of understanding one’s mental health

Multi-faceted in nature – The environment, human behavior, and emotional well-being have a multitude of dimensions. We live in immensely sophisticated environments that consist of many social, environmental, cultural, and economic variables. This makes it difficult to construct associations and linkages between which specific variables affect our mood at a given timeframe.
Lack of data – Obtaining data on different variables from people for a sustained period poses a challenge – conducting large-scale studies across many geographical locations, would be time-consuming and costly.

Affordability – One of the best ways to understand one self’s mental health is to see a therapist. However, this is not an affordable option for many.

2. Objectives

This project aims to achieve several educational & research objectives in support of assisting the understanding of one self’s mental health and change for the better.

2.1 Educational objective

As understanding and dealing with mental health is not a linear process, the primary objective of this project is to allow users to keep track of their emotional well-being by emphasizing building a progressive healthy journey with them, through apprehending the environment they are in and how various variables affect their mood; and thus, presenting the findings clearly and understandably. The functions of the app aim to appeal to a wide audience who are intrigued to explore how their surroundings and personal variables affect their emotional well-being.

2.2 Research objective

The research objective is to experiment and design an algorithm that integrates different socio-economic variables and user logged emotional-wellbeing data to output visual graphs. The correlation between various demographic factors and emotional well-being will be measured and analyzed quantitatively by measuring the intensity of each variable against the frequency of occurrence. In addition, a consolidated database will be curated to find general trends and serve future research purposes.
3. Related work

There is a research paper from the International Journal of Mental Health Systems studying the relationship between depression and sociodemographic factors. This study used the Canadian Community Health Survey, Cycle 1.2 (CCHS-1.2) dataset [4]. Similarly, a study was done using the 2019 National health Interview Survey conducted by the Centers for Disease Control and Prevention (CDC), whereby it was studied if demographic information like income, education level, race, and household region correlated with chronic mental health conditions like anxiety and depression.

3.1 Methodology of related work

The dataset from the CDC consisted of the following fields: age, sex, education, sexual orientation, marital status, served in the military, citizenship, income, Body Mass Index (BMI), urban or rural types of living conditions, and anxiety and depression occurrence and frequency. The study built a model to quantify mental health by calculating a score for each data field:

\[ s = score \]
\[ a_t = anxiety\ level \]
\[ a_f = anxiety\ frequency \]
\[ d_t = depression\ level \]
\[ d_f = depression\ frequency \]

\[ s = a_t * a_f + d_t * d_f \]

The study then curated a correlation coefficient for the score in association with each variable and compared the associations using generated box-and-whisker diagrams [5].

This project will refer to the models used in the above projects and modify them to better fit this project.

The above projects have presented a model to quantify scores given a survey dataset at a given point in time. This project focuses on analyzing emotional well-being trends across a timeframe.
4. Methodology

The project is divided into three subparts. The first part focuses on data collection – the types of data to be collected, the method of data collection, data processing, and storing of data. The second part focuses on the analysis of data – types of analysis, levels of analysis, and methods and tools used. The third part focuses on the display of insights – the data to be displayed, the framework of app design, and visualization. The methodologies revolve around the first three laws of behavioral change — make it obvious, make it attractive, and make it easy. The figure below illustrates the three main steps of this project:

![Architecture diagram of project workflow (data flow perspective)](image)

Figure 1

4.1 Part 1: Data collection

In order to keep the project on track and manageable given the time constraint, test data sets will be created to test the model for the development stages.

The data collection process will be developed with the intention of removing as much friction as possible in the procedure, for example, rather than asking users to write how they feel, instead, they would be asked to express their emotions based on a Likert scale. This way users do not have to exert much effort into typing, but rather just simple taps.
4.1.1 Types of data to be collected

The data to be collected can be classified into two main categories: Subjective data and Objective data. Subjective data are information from the user’s point of view, including feelings, perceptions, and more. Objective data are factual variables of the users, including age, height, weight, and more.

The figure below includes more detail on the data to be collected:

<table>
<thead>
<tr>
<th>Objective data</th>
<th>Subjective data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, Height, Weight, Gender, Occupation, Salary range, Sleep hours, type of housing, district of residence, pre-diagnosed psychological/physical disorder, number of family members, insurance coverage, working industry, etc.</td>
<td>Stress levels, moods, social life, self-esteem, workload, sense of belonging, usual reasons for stress, etc.</td>
</tr>
</tbody>
</table>

*Table 1 Objective and subjective types of data to be collected*

The data will be stored in Google Firebase or MongoDB. The databases will be further evaluated in the later stages of the project.

4.2 Part 2: Analysis

A variety of Python libraries will be used for analysis, including but not limited to Pandas, Numpy and Scipy. There are two main levels of analysis: 1) **Individual level analysis** – refers to the analysis of data from each individual user and 2) **Group level analysis** – refers to separating data into multiple strata based on the objective data collected (i.e., age, height, weight, etc.)

In the individual level, the analysis is more personalized. The analysis will focus on finding correlation of how their logged in subjective data affects other subjective data. For example, how their social life (how many people they have met in a day/week) affects their stress levels, or how their workload affects their sense of belonging.

In the group level, the analysis is more generalized. The analysis will focus on finding correlation of how objective data affects the subjective data. For example, how education level affects stress levels, or how number of family members affect social life.
The correlation will be analyzed by calculating a quantifiable score based on the probabilities of occurrence of an objective parameter with subjective parameters against total number of users:

\[ s = score \]
\[ O = number \ of \ users \ with \ objective \ parameter \]
\[ x = subjective \ parameter \]
\[ t = total \ number \ of \ users \]

\[ s = \frac{O_x}{t} \]

For example, if the correlation between high education level and stress levels were to be found, the formula would be as follows:

\[ s = score \]
\[ O = number \ of \ users \ with \ high \ level \ of \ education \]
\[ x = with \ high \ stress \ levels \]
\[ t = total \ number \ of \ users \]

The higher the score the higher the correlation between the two specified parameters (i.e., the correlation between high stress levels with the level of education)

Subjective parameters like stress level will be obtained through the app based on users’ perception of stress. The values will be numerical on a Likert scale. For example, users will choose a value from a scale of one to five (one being the least stressful and five being the most stressful).

After a critical mass of data has been collected/curated, machine learning can be used to determine whether it is possible to predict mental health scores using demographic information. This will require both a regression and a classification task. Different methods of regressors and classifiers will be considered within the course of this project.
4.3 Part 3: Display

The main framework for front-end development will be react native. The wireframes and UI/IX design prototyping will be carried out on Miro and Figma respectively.

For the display of data in a clear and understandable manner, the Python libraries matplotlib and seaborn will be used to generate visuals for users. The best graphs and diagram structures for visualization will be determined based on the data types in the later stages of the project.

5. Schedule

The table below illustrates the schedule of this project:

<table>
<thead>
<tr>
<th>Milestones</th>
<th>Time period</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phase 1: Ideation and research</strong></td>
<td>August 2022 – September 2022</td>
</tr>
<tr>
<td>• Background research on existing models for analyzing mental health</td>
<td></td>
</tr>
<tr>
<td>• Background research on what helps people have control of their emotional well-being</td>
<td></td>
</tr>
<tr>
<td><strong>Deliverable:</strong> Project plan, project website</td>
<td>2nd October 2022</td>
</tr>
<tr>
<td><strong>Phase 2: Project development</strong></td>
<td>October 2022 – February 2023</td>
</tr>
<tr>
<td>• UI/UX Design</td>
<td></td>
</tr>
<tr>
<td>• Code implementation</td>
<td></td>
</tr>
<tr>
<td>• Documentation</td>
<td></td>
</tr>
<tr>
<td><strong>Deliverable:</strong> 1st Presentation</td>
<td>9th – 13th January 2023</td>
</tr>
<tr>
<td><strong>Deliverable:</strong> Preliminary implementation and detailed interim report</td>
<td>22nd January 2023</td>
</tr>
<tr>
<td><strong>Phase 3: Testing and refinement</strong></td>
<td>February 2023 – March 2023</td>
</tr>
<tr>
<td>• Modifying and fine-tuning code</td>
<td></td>
</tr>
<tr>
<td><strong>Deliverables:</strong> Finalized tested implementation and Final report.</td>
<td>18th April 2023</td>
</tr>
<tr>
<td><strong>Deliverable:</strong> Final presentation</td>
<td>17th April 2023 – 21st April 2023</td>
</tr>
<tr>
<td><strong>Project exhibition</strong></td>
<td>3rd May 2023</td>
</tr>
</tbody>
</table>

*Table 2 Project milestones*
6. References


