What is the problem?

Mental Health is a hot button issue within the United States, and in terms of big business, a generally untapped market. From climbing rates of substance use disorders, to people claiming the lack of a “third place’ to lingering ramifications of most people lacking genuine social interaction during the covid pandemic and never recovering. Mental health and how to help patients feel empowered and have the skills to improve their own life is important in today’s world.

Understanding how our lives in the technological age with a 40+ hour work week, social media and technology usage eating away at our days. Understanding how these factors correlate and how to effectively attack problems from multiple angles, statistically, could prove to be life altering.

**How does this support our customers**

Psychotherapy is often used to discover patterns in behavior and identify facts. These facts in terms of our consumers are data points such as how many hours a day are they on social media, how often do they preform exercise and how would they rate their mental health. Understanding how the participation or lack thereof in physical activity would affect mental health would allow us to assist end users in assessing which factors to increase or decrease in order to obtain higher feelings of good mental health.

Understanding the underlying factors such as total technology usage and the various other factors that are included in our study will provide a more holistic picture and allow us to make recommendations based on statistics and clinicians who use our product to make recommendations confidently based on a breadth of data.

**What is our Data product?**

Our data product is a turnkey solution for clinicians to analyze data. Given some factors about clients they can analyze where they would fall within a population mean and given their other factors surrounding technology usage and mental health, could predict and suggest which factors could be adjusted to seek improvement.

The project is a standalone application which trains on a data set and then makes predictions about omitted factors based on its training. This will allow users to take data sets and see if their patients are over or under a mean, how their other factors lead to their current result in on category and which factors have the strongest weight in their mental health. All this data in turn, can be used to create a care plan for that patient.

**Our Dataset**

The data set current used for training is collected from the open-source dataset website Kaggle. This open-source dataset is inclusive of 13 unique data points to analyze relationships for patients. The data set currently hold 10,000 records with HIPPA sensitive information pre-sanitized. This data will be fed to the machine learning algorithms created from the library produced by Sci-Kit learn. As our end users enter information, this can also used for training to tune the model and produce are more accurate and confident result.

**Objectives and Hypotheses**

The objective of the data product is to have an easily integrated comprehensive tool that allows clinicians to analyze their patients’ objective factors and see how they compare to other persons around the world and deviation from means. While this approach is reductionist in nature, it allows a quick and easy analysis of factors many people associate with causing poor mental health.

It is the hypothesis that the product will provide an accurate preduction and understanding of the strengths of correlations to clinicians who can then use it in their practice. While it is anticipated that the models training set and tuning will be frequent this accuracy will be improving consistently is a target within it’s development.

**Methodology**

Developing the model using an agile framework will allow for optimal tuning across the various models that are used to determine outputs. Allowing for frequent adjustments and modifications to the source code and data set as the design philosophy will support the implementation of minor tweaks and enhancements with seamless integration into the live product that users will be interacting with daily maintain a positive and up-to-date solution.

The model should also be developed with cloud integration in mind. While the product does have the initial training set integrated with it which would allow for offline operation, pushing new data sets to the users in the form of updates may not be the most optimal solution. The agile method plans on pushing frequent tweaks and patches as well as a consistently updated database for training, ensuring cloud integration is key to ensuring this methodology will be successful.

**Funding Requirements**

The estimated team of developers is 6 developers at an estimated $100,000 per annum rate. Over the 3-month development period this can be estimated to cost $150,000 as well as any technology cost if more computers need to be requisitioned. There should be consideration for server space by a cloud provider in order to facilitate development, testing and launching anticipated at $5,000 per year.

Depending on the model of employee desire for building determines the cost of the product. Ensuring that at least 1 developer remains as a permanent employee will allow for tuning and updates to be done in the future. The rest of the positions could be contracted out during the development cycle and during any other future overhauls that may potentially be needed.

**Stakeholder impact**

Stakeholders will see an impact very quickly as they use the product. Simply saying a person does not partake in physical activity or uses too much social media is a very easy thing to do. This product will add data to those statements and show how a deviation from the mean in those factors can lead to decreased mental health scores. This will allow clinicians to better understand which factors are majorly affecting their patients and make recommendations.

In terms of patients, as they are a stakeholder by proxy, they will see better outcomes. This product would allow clinicians to congregate data and better understand these relationships as more and more data is given to the model and it’s able to make better connections between the relationships within the model.

**Ethical and Legal**

There are currently no legal or ethical concerns for this project. Ethically, user data is submitted by a clinical who should receive authorization from the client as part of their practice’s paperwork, and the model does not support any identifying information. Ensuring sanitation of information attempting to be inserted to the database will keep private health information out of our hands.

There are no legal impacts forecasted. The libraries utilized are open source and do not require any direct forms of licensing or payment. Verification of the licenses and will occur to ensure the product is compliant but from the planning phase of development was determined to be a non-issue.

**Recommendations**

The development team lead for this project is relatively new to the field of computer science but has shown very promising work in other areas and within machine learning. Obtaining a Bachelor of Computer Science and being already familiar with the technologies we intend on using they have an involved understanding of what is needed for this project.