**The Opportunity**

Machine learning is a big component in emerging medicines practices. The ability for doctors to analyze and quickly discern patient conditions and outcomes is revolutionizing the field. This revolutionary change is also happening with psychology. Creating a database and machine learning model that can synthesize data for practitioners to be able to analyze data that would take them days to do by hand will create greatly improved understanding of psychological factors and improve patient outcomes.

**Our customers**

This product will serve for a non-profit psychology association that attempts to help clinicals determine effective treatment modalities. The model will provide the clinicians with an understanding of what the most pressing factors affecting patients are and how they can coach adjustments to improve outcomes.

While this target audience will be obtained with initial releases, the product and data base could be modified for non-clinical use for every day people to see correlations and ways to improve their mental health.

**Target Market**

This model will be an entirely new product and is currently in the form of producing a minimally viable product. Creating a model that will take inputs and give the desired format in output is the first objective. While the model should be complete and functional, it may suffer in accuracy which will be determined for resolution with a round of revisions.

While there are similar products, such as mental health chatbots and web-based clinicians, this product looks to bridge the gap in the numerical break down of patients. It explores the relationships, in a reductionist way, of the growth of technology, decrease in mental health and which factors could be modified to invert that relationship.

**The Data**

As users input data, the data could be scrubbed and ensure HIPPA compliance, then be used for further training of the model. While the initial data set included 10,000 data points, the inclusions of more, potentially diverse data points could help discover unrealized relationships that would lead to future better patient outcomes.

This data could be implanted in a few methods. Patches to the training database on scheduled timeframes will allow for any issues that arise from the new data to be patched with the update being pushed. However, if data is properly sanitized proper to being introduced to the model, the data could be added in real time to allow for the most accurate model at all times.

**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.

A preliminary training data set is the first deliverable. Understanding the categories of data and relationships to be explored are vital to developing the model and isolating the primary focuses of the study. This deliverable will be in the form of a study with numerical/enumerable data points and data categories we wish to isolate.

**Deliverables**

The most pertinent deliverable is the model and underlying documentation of it. Delivering the final software and all related information is the project's primary purpose and should be the final target in every situation.

A data set must be delivered early in the evaluation process by stakeholders. Ensuring that stakeholders are comfortable with the data that the model is being trained on will translate to their confidence in the results of the model. Ensuring ample time for them to review and raise questions about the data set is critical to this deliverable.

**Implementation plan**

The overview of this project is to use the Sci-Kit learn python library to develop Random Forest models that will learn from the data set and predict a chosen variable given an input set. This will be through developing a boiler plate model for Regression/Classification and implanting training for each desired parameter in separate modules. These modules will then allow the user to run a data set against its training to predict a given parameter.

The model and data to run the model with for prediction will be accessible from a GUI built inside of Tkinter. The user will be able to pick the model/parameter they wish to predict and enter data for the model to calculate an output. The first implemnation of this will again by minimally viable, lacking styling but fully functional.

**Validation, Testing and Acceptance**

The outcome of this product is the development of 7 machine learning models that use the same boiler plate and predict data points out of a given set of data. These models will train on the currently 10,000 data points currently obtained and predict data about a given subject in an attempt to predict factors surrounding their mental health and draw correlations on how to improve the subject’s mental health.

The first step to ensuring the product meets stakeholder requirements will be internal testing. Ensuring that the users are able to adequately input data, it returns timely, accurate information is the first step. Ensuring the product functions in a desirable method with acceptable lag times is important for ensuring it can be used. Secondly, user acceptance testing will be implemented. Ensuring that end users can adequately navigate and utilize the product interface and functionalities will validate that the product is complete. Failure at any section will allow for change to be planned and implemented within the agile framework.

**Development environment and costs**

The programming language utilized will be Python with Visual Studio Code being the recommended IDE but not a strict requirement. 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.

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 or server space by a cloud provider to facilitate development, testing and launching.

**Timeline**

10/11/2024-10/25/2024 – The initial proposal is submitted for stakeholders to review and add any requirements they feel are pertinent. These are reviewed with developers for feasibility and added if applicable.

10/28/2024-11/8/2024 – Stakeholders review data and ensure there are no anomalies, and that data is acceptable to all stakeholders; machine learning can begin on this data set.

11/8/2024-11/29/2024 – Initial implementation of machine learning algorithm; the algorithm is a minimally viable product, can make correct predictions but has less insight, and is not fully integrated with data.

11/29/2024-12/2/2024 – Review code at the approximate halfway point; all data is integrated, algorithms should be moderately refined, and when given data points, the algorithm should be able to extract related data points with 80% confidence.

12/2/2024-12/20/24 – Deliver final product to stakeholders, model works within a 95% confidence interval in predicting mental health scores based on other factors or other factors based on mental health scores.