AI-Based Precision Medicine Platform – Proposal
Capstone: The Art of Approximation

Ojas Chaturvedi, Zaheen Jamil, Saianshul Vishnubhaktula, Ritwik Jayaraman

November 21, 2023

1. **Language:** Python, a simple and popular language for machine learning and data science due to its extensive libraries and frameworks

2. **Objective:** To build an AI-powered platform to analyze symptoms, previous medical records, and research to provide chances of diseases and then give personalized treatment recommendations

3. **Implementation:**

   (a) **Overview of steps:**

      i. **Data Collection:** Collect previous medical records and research data from various sources and databases
         A. Ex: github.com/bruzwen/ddxplus
      ii. **Data Processing:** Process the data to extract relevant features, such as removing stop works and blank lines, etc.
         A. **Homomorphic Encryption** will protect sensitive health data instead of conforming to HIPAA and other health data protection regulations
      iii. **Model Training:** Train machine learning models to predict disease risk and treatment outcomes
      iv. **Model Deployment:** Deploy the models on a secure platform to be used by clinicians and patients

   (b) **Libraries:**

      i. **Pandas:** For data manipulation and analysis
      ii. **Matplotlib:** For visualizations
      iii. **NumPy:** For numerical computing and working with arrays
      iv. **Scikit-learn:** For data mining and analysis
      v. **TensorFlow:** For complex neural network modeling
      vi. **PyTorch:** For natural language processing
      vii. **NLTK/spaCy:** For human language data with symptom inputs
      viii. **Flask/Django:** For backend web development
      ix. **SQLAlchemy:** For SQL databases and Object-Relational Mapping

   (c) **Manual Work:**

      i. Building the machine learning model
      ii. Collection of datasets of diseases and percent chance of symptoms
         A. Will contact local hospitals for datasets for the latest outcomes
      iii. Homomorphic encryption implementation
      iv. GUI development
         A. Website or App
         B. Users can input their symptoms (and this will go into the dataset anonymously with homomorphic encryption)
         C. Users receive a percent chance of diagnosis based on these symptoms from the model we trained

4. **Jobs:**

    (a) 2 Machine Learning/Data Collections Specialist
        i. Make the Machine Learning/AI Model using Python and related libraries
        ii. Compile all data needed

    (b) Data Security Specialist
        i. Focuses on the implementation of homomorphic encryption

    (c) GUI Developer
        i. Makes the website/app and all of its functionality (UI)
        ii. Would work with a Data Security Expert for the implementation of homomorphic encryption