Project 4 Proposal

Anthony Costa, Jonathan Diaz, Sarah Kim, Aakash Nagalapura

**Purpose**:

The primary objective of this proposal is to implement a data model for disease prediction that leverages advanced analytics and machine learning techniques. By utilizing relevant health indicators, we aim to develop an accurate prediction model that can assist healthcare professionals in making informed decisions and improving patient outcomes.

**Data:**

<https://www.kaggle.com/datasets/kaushil268/disease-prediction-using-machine-learning?select=Testing.csv>

**Data Cleaning**:

Assign variables to each symptom. Clean to remove possible duplicate records and fill in missing values.

**Algorithm Development**:

Develop an algorithm that will accurately recognize symptoms used to predict diseases. Tune algorithm as needed.

**Model Evaluation**:

Evaluate the accuracy of the algorithm using clinical trials by comparing the generated predictions to the actual diagnosis.

>> Possible model: Decision Tree

**Limitations**:

This particular dataset is solely dependent on a patient’s symptoms. It does not take the following into account: a patient’s medical history or demographic, possible disease trends, or other factors that may also contribute to a health professional’s diagnosis.

**Internal Milestones**:

* Project Proposal - Thurs 6/1
* Data cleaning - Mon 6/5
* Analysis/Testing - Wed 6/7
* Creating documentation - Thurs 6/8
* Creating the presentation - Thurs 6/8

**>> PRESENTATION ON MONDAY, JUNE 12TH**