**PROJECT PROPOSAL**

**PROJECT 4 – GROUP 5**

**Group 5 Members:** Uma Selvaraj, Mason Seifaddini, Rekha Renukappa, Thanh Vinh Giang

**Github link:** [**https://github.com/alvin-giang/Disease\_Prediction\_Chatbot.git**](https://github.com/alvin-giang/Disease_Prediction_Chatbot.git)

**Dataset:**

[**https://www.kaggle.com/datasets/itachi9604/disease-symptom-description-dataset/data?select=dataset.csv**](https://www.kaggle.com/datasets/itachi9604/disease-symptom-description-dataset/data?select=dataset.csv)

**PROJECT OVERVIEW:**

**Train an algorithm to recognize disease symptoms and create a chatbot that will help interact with patients.**

**1. Introduction**

The aim of this project is to develop a machine learning model capable of predicting diseases based on symptoms using the "Disease Symptom Description" dataset available on Kaggle. This dataset contains a comprehensive list of diseases and their associated symptoms, which can be utilized to build an efficient and accurate prediction model.

**2. Objectives**

* Data Preprocessing: Clean and preprocess the dataset to handle missing values, outliers, and ensure that the data is in a suitable format for model training.
* Feature Engineering: Perform feature extraction and selection to identify the most significant symptoms that contribute to accurate disease prediction.
* Model Development: Build and train multiple machine learning models (e.g., Decision Trees, Random Forest, SVM, Neural Networks) to predict diseases based on symptoms.
* Model Evaluation: Evaluate the performance of the models using metrics such as accuracy, precision, recall, and F1-score. Compare the results to select the best-performing model.
* Deployment: Create a user-friendly interface for the prediction model, allowing users to input symptoms and receive potential diagnoses.
* Visualization: Create a chatbot that help interact with patients, provide diseases’ details and give some advice.

**3.** **Conclusion**

This project will leverage machine learning to develop a predictive model that can assist in diagnosing diseases based on symptoms, potentially aiding healthcare professionals and patients in early detection and treatment planning. The successful completion of this project could lead to a valuable tool in the healthcare domain, improving both the speed and accuracy of diagnoses.