Machine Learning Project Proposal

## Team Composition:

Khushnur Binte Jahangir

Adam Broniewski

## Dataset:

We will use the [UCI - Heart Disease Data Set](https://archive.ics.uci.edu/ml/datasets/Heart+Disease)

## Problem:

Predict whether a patient has heart disease.

## Why this dataset:

This dataset was chosen as it meets all the requirements of the project scope. This was the most important factor for us so that we can focus on learning and understanding fundamental machine learning requirements without getting stuck on unclean data or a difficult dataset.

Interestingly, there are 14 main attributes that are used in other papers that make use of the dataset, although there are 75 attributes that can be chosen from for prediction. This provides a lot of room for experimentation and improvement on past research.

## Dataset composition:

There are several datasets available from different locations (Cleveland, Hungary, Switzerland). The UCI description indicates the “Cleveland database is the only one that has been used by ML researchers to this date”. The Cleveland database is described as being “processed”; however it is not known to what extent or what exactly this processing includes.

|  |  |
| --- | --- |
| Number of instances | 303 |
| Number of attributes | 14-75 |
| Attribute types | |
| categorical |  |
| integer |  |
| real |  |
| Analysis type | classification |

## Other References:

There are 61 other relevant papers that make use of this dataset as identified by UCI website

## Project Title:

Heart Disease Machine Learning Project 2022