Team3\_H Project Overview and Breakdown.

Key term: Prognosis, forecasting the likelihood of the progression of an illness.

Problem statement: Build a machine learning model to forecast the likelihood of the progression of the diseases or illnesses identified in the dataset, based on each patient’s medical history.

Keep in Mind:

1. The problem is an unsupervised machine learning problem, this is because there is no target feature here to help with the model training. PS: I have never worked on an unsupervised ML project so y’all should do your research.
2. From what I have identified so far, we will be checking for the progression of the following illnesses in each patient (skin cancer, heart disease, depression, diabetes, arthritis, other cancer.

Algorithm to use: We will be using clustering algorithms so it is advisable for us to learn about them too.

Steps

Data preprocessing and EDA:

1. Data cleaning: check for null values and outliers.
2. data visualization to understand the distribution of each field in the data set.
3. conversion of text data to numeric data(encoding).
4. Dealing with missing data, outliers or unwanted data by imputing, deleting or any other way.
5. Checking for anomalies
6. Divide the data into training and test split.

Feature engineering:

1. Observing the features to see if a new feature can be formed from them, eg weight and height to form BMI, distance and time to form speed.
2. Feature selection, after performing feature engineering it is important to weigh the feature importance and select the features best for our model.

Model selection and training

1. Selection of models to use and training the model.
2. Evaluate the model based on the metric provided and test the generality of the model on the test set.
3. Tune the parameters and keep evaluating the model till you get the best results.