**Data Mining Final Project -** [**ISM6136.004F23**](https://usflearn.instructure.com/courses/1809618)

**Group Name:** Matrix Masters

**Group Members**:

* + - Venkata Dharma Teja Konakanchi
    - Pavan Venkata Vijaya Ajith Varma Datla
    - Shyam Prasad Nalasani
    - Sai Chowdary Penikalapati
    - Manikanta Yashwanth Jalagadugu
    - Sunil Porida

**Data used:**

We have used a heart disease dataset. The dataset contains a variety of demographics and health-related information for a group of patients. It includes a CLASS variable "num" that indicates the heart health of each of the patients. The "goal" field refers to the presence of heart disease in the patient. It is an integer value from 0 (no presence) to 4. Experiments with the dataset have concentrated on simply attempting to distinguish the presence (value 1,2,3,4) from the absence (value 0) of heart disease.

**Problem Statement and Analysis:**

“Matrix Masters” group was approached to work for a medical institution that wants to improve the heart health of its patients. We have obtained a dataset that contains a variety of demographics and health-related information for a group of patients. It also includes a CLASS variable "num" that indicates the heart health of each of the patients. The values are:

1. No heart health issues.
2. Slight heart health issues
3. Moderate heart health issues
4. High heart health issues
5. Extreme heart health issue

We have been asked to develop a classifier based on the dataset data, to predict the CLASS of new patients so they can be enrolled for treatments based on their demographic data.

The COSTs of the treatments are as follows, based on the predicted class of each patient.

0 – Checkup = 100 USD

1 – Minor Treatment = 200 USD

2 – Moderate Treatment = 300 USD

3 – Significant Treatment = 400 USD

4 – Extreme Treatment = 500 USD

TRUE\_CLASS = 0, 1, 2, 3, 4

The BENEFITs of the treatments are as follows:

* If the classification is correct and the correct treatment is given to the patient:

500 \* (TRUE\_CLASS + 1)

* If an incorrect classification is made and the wrong intervention is given to the patient:

0

We would like to find a classifier that maximizes the overall Net Benefit.

**NET\_BENEFIT = BENEFIT - COST**.

Therefore, a larger positive number is a good outcome.

So, for example:

* If a patient's true class is 4 and their predicted class is 0,

**NET\_BENEFIT = 0 - 100 = -100**

* If a patient's true class is 4 and they are correctly classified,

**NET\_BENEFIT is 500\*(4+1) - 500 = 2000**

The medical institution would like us to evaluate a few models. We agreed to do so and would like to try a few classifiers and find out which classifier would give the maximum benefit.

As of now, we plan to implement KNN, SVM, and Decision Tree classifiers. Final changes will be made after we start working on the dataset.