Authors :

* Mankirat Singh Bhamra MXB220061
* Harikrishna Dev HXD220000
* Krishnan Venkatesan KXV220007
* Medha Priyanga Saravanan MXS220057

1. Research question
   * Can we predict the probability of a patient being diagnosed with heart disease using his current health attributes?
   * Our audience is the medical instrumentation industry.
   * Using historical data, we can predict the potential diagnosis of the condition to help the doctor make better-informed decisions.
2. Data:
   * Our dataset is from Kaggle: [Link](https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction)
   * The period for this data is 30 May 1989 to 2 Dec 1996.
   * The level of data is an individual patient level.
   * We have instances where there are multiple entries for a patient and missing variables.
3. Models:
   * The outcome is whether the patient is likelier to have heart disease or no heart-related illness.
   * X – variables:
     1. Age: age of the patient [years]
     2. Sex: sex of the patient [M: Male, F: Female]
     3. ChestPainType: chest pain type [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic]
     4. RestingBP: resting blood pressure [mm Hg]
     5. Cholesterol: serum cholesterol [mm/dl]
     6. FastingBS: fasting blood sugar [1: if FastingBS > 120 mg/dl, 0: otherwise]
     7. RestingECG: relaxing electrocardiogram results [Normal: Normal, ST: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), LVH: showing probable or definite left ventricular hypertrophy by Estes' criteria]
     8. MaxHR: maximum heart rate achieved [Numeric value between 60 and 202]
     9. ExerciseAngina: exercise-induced angina [Y: Yes, N: No]
     10. Oldpeak: oldpeak = ST [Numeric value measured in depression]
     11. ST\_Slope: the slope of the peak exercise ST segment [Up: upsloping, Flat: flat, Down: downsloping]
   * Models:
     1. Logistic regression
     2. Support Vector Classifier
     3. KNN Classifier
     4. Decision tree Classifier
     5. Random Forest Classifier
     6. Cross-validation / Bootstrapping
4. Results:
   * We predict that our model will have an accuracy of ~75-85% and with cross validation will increase.