



**Ahmedabad  
University**

**CSE523 - Machine Learning**

**Section: 1**

## **Heart Disease Prediction - Report**

**Submitted to Prof. Mehul Raval**

**Group name: Bias\_Variance\_Dilemma**

<b>Enrollment Number</b>	<b>Name</b>
<b>AU1940110</b>	<b>Shrey Patel</b>
<b>AU1940155</b>	<b>Shubham Patel</b>
<b>AU1940207</b>	<b>Keyur Nagar</b>

### ❖ Things we have done:

1. Found out the data and the CSV file regarding that
2. We did the data set analysis that contains 14 attributes regarding the heart disease prediction model.
3. The final output depends on the 13 parameters, and the 14th one, i.e. target, defines whether the patient has a disease or not.
4. We also did a statistical analysis of the dataset.
5. We then divided the dataset into 80 % as a training dataset and 20 % as a test dataset.
6. Performed KNN for training the data set.
7. Derived the misclassification rate vs a number of neighbors (K) plot. Find the appropriate value of K, which doesn't overfit or underfit the data. Achieved an accuracy of 72.31 %.

### ❖ Things to be done:

- We will try to understand the 13 parameters more thoroughly using data analysis techniques and their effect on the outcome, i.e. our target parameter.
- We will try other algorithms (Logistic Regression, SVM, etc.) for our dataset and select the model with high accuracy.
- In the end, we will try to deploy our model on the front-end website where the user can enter their values for the 13 parameters, and the model will detect the disease and show relevant results. This will ensure our model is widely used.

### ❖ Data Set:

This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the data set combines four databases, namely Cleveland, Hungary, Switzerland, and the VA Long Beach. The "target" field refers to heart disease in the patient. It is integer-valued from 0 (no presence) to 4. Researchers have concentrated on simply attempting to distinguish presence (values 1,2,3,4) from absence (value 0). So in this data set, the value of the field "target" has only two integer values, 0 (no disease) and 1 (disease).