CSE523 - Machine Learning

Section: 1

Heart Disease Prediction - Report

Submitted to Prof. Mehul Raval

Group name: Bias_Variane_Dillema

Enrollment no	Name
AU1940155	Shubham Patel
AU1940110	Shrey Patel
AU1940207	Keyur Nagar

❖ Things we have done:

- ➜ As mentioned in the earlier report, we have started working on the new data set that consisted of 70000 instances and 12 attributes including the target variable.
- ➜ In this week we have tried to perform feature extraction and selection on our data set in order to lessen down the dimensionality and select only the best variables that incorporate the output(target)
- ➜ We had height and weight attributes in our dataset, so we thought that we can use BMI (Body Mass Index) that is derived using a formula that consists of height and weight features.
- ➜ By doing so we can lessen the feature by combining two features into a new feature. Also the originality of the dataset remains the same as the new feature (BMI) is derived from the original features (height and weight).
- ➜ The same thing we did on another parameters (features) i.e. ap_hi and ap_low. We found that we can find the BP by taking difference of these two parameters. BP = ap_hi – ap_low
- ➜ So we converged another two features into one without loosing the originality of the data.
- ➜ From the Exploratory Data Analysis, we found that the alcohol and smoking features don't affect the final output parameter i.e. target. So, we tried to remove these two parameters while training our model.
- ➜ We have reduced the attributes from 11 to 9 and tried to fit the model with Logistic Regression and KNN algorithm.
- ➜ Logistic Regression gave the accuracy of 68.77% and KNN gave the accuracy of 69% with neighbour = 80.

❖ Things to be done:

- ● We will try to incorporate other feature selection algorithms to further reduce our features and to increase our model accuracy.

- ● In the end, we will try to deploy our model on the front-end website where the users can enter their values for requested parameters and can get to know their results.