|  |  |
| --- | --- |
|  |  |

**ALY 6020: Predictive Analytics**

**CRN:** 81941

**Week 6 Final Assignment:**

**Submitted By:**

**Rishhikumar Janakiraman** [janakiraman.r@husky.neu.edu**]**

**Nikhil Sanjay Thorat [**thorat.n@husky.neu.edu**]**

**Megha Ravi [**ravi.m@husky.neu.edu**]**

**Introduction**

Cardiovascular diseases (Heart disease) are one of the top reasons for deaths worldwide claiming upto 17.9million lives each year. There are multiple types of heart diseases such as coronary heart disease, cerebrovascular disease, rheumatic heart disease with 4 people out of 5 dying due to strokes and heart attacks. The symptoms of individuals under a risk of any of these diseases are generally high blood pressure, variations in glucose levels, obesity. The one way to reduce premature deaths caused by these diseases is early accurate diagnosis, proper treatment and with the advent of technology, use of historic data can make this possible.

The project is aimed at understanding the various factors that influence a heart disease (age, sex,other medical conditions) and predict what combinations are more susceptible by creating test and train dataset, building models and performing analysis.

**Dataset**

The dataset is taken from Kaggle and has 14 variables namely sex, age, chest pain type, blood pressure level, serum cholestoral, fasting blood sugar, electrocardiographic results, maximum heart rate of 303 patients. Predictions will be done using attributes blood pressure, heart rate, cholesterol levels, age.

**Planned Work -**

The dataset contains various attributes related to heart and the planned work to be implemented includes basic data cleaning followed by the Exploratory Data Analysis (EDA). The main focus is to predict heart related diseases using various models and determine their accuracy. Machine learning has a huge impact especially in the field of medical and health science so we have planned to implement the following algorithms and machine learning techniques on the dataset to provide data driven insights :

1. Logistic Regression model
2. Classification model using Neural networks
3. Binary classification
4. Random forest

**Conclusion -** For this project we are going to use supervised and unsupervised learning in order to predict the accuracy of the results. The most accurate model will be taken into consideration for predicting the heart related diseases. For our project Neural networks and other similar algorithms will be used to predict heart diseases by taking in some attributes in consideration such as Blood pressure, heart rate, etc. Regularization and normalization will be performed for initial analysis.

References -

[1] Cardiovascular diseases. (n.d.). Retrieved from<https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1>

[2] Real Python. (2020, January 13). Logistic Regression in Python. Retrieved May 30, 2020, from <https://realpython.com/logistic-regression-python/>

[3] Koehrsen, W. (2018, January 17). Random Forest in Python. Retrieved May 29, 2020, from https://towardsdatascience.com/random-forest-in-python-24d0893d51c0