ENEL 682 - Project Proposal

Presented By:

Manav Pandya- 30169952
Hetalben Virani- 30183515
Pujan Patel- 30169747

Motivation behind choosing the topic:

Diabetes is a chronic condition that could lead to a catastrophe in the global health care system. Approximately 382 million individuals worldwide have diabetes, according to the International Diabetes Federation. This will almost double to 592 million by 2035. Moreover, as stated on the website of government of Canada, over 3 million people living in Canada, or 8.9% population have already been diagnosed with diabetes and the repots have stated that this is increasing at an alarming rate of 3.3 percent per year. Diabetes, also known as diabetes mellitus or just diabetes, is a condition brought on by elevated blood glucose levels. For the diagnosis of diabetes, several conventional techniques based on physical and chemical studies are available. Nevertheless, since diabetes affects human organs like the kidney, eye, heart, nerves, foot, and others, early diabetes prediction is a difficult job for medical professionals.

However, as the field of data science and machine learning has grown in leaps and bounds, it is comparatively easy to predict that an individual will become diabetic in the future or not. This type of prediction can be made by assessing various qualities of a person such as their age, blood pressure, thickness of their skin, body mass index, etc.

Who can use this model: This Machine Learning model can be used by the healthcare workers in order to track diabetes patients; in our case we opt to work with the Alberta Health Services or the healthcare department of any province of the country.

Plan of Attack: Our first and the foremost step would be to find a suitable data set and then after we will check all the attributes related to diabetes. Once, we have assimilated all the required data, we will remove all the other data using data cleaning. Furthermore, we will visualize each column for the better understanding of the data and for the prediction we will separate the data into test and train data. Lastly, using the different model we will classify the output and based on the accuracy we will choose our best model.

Important Data sets and Inputs:

Dataset: Diabetic Dataset, which has been retrieved from Kaggle.

Models: As advised, we will be primarily working with three machine learning models, one of them being Neural Network and the rest would be Logistic Regression and Support Vector Machine (SVM).

Frameworks: Data gathering, Preprocessing, Data Visualization, Feature Extraction as well as Classification.

Components: Data Cleaning, Data Integration, Data Transformation as well as Data Reduction.