EXECUTIVE SUMMARY

Diabetes is a chronic condition prevalent among more than 100 million people across the world. The chances of readmission of diabetic patients are quite high. Despite major advances in science and technology, diabetes continues to be a chronic disease, with a thirty-day readmission rate of around 20%, as compared to an average of 12% for the rest of the diseases. Additionally, readmissions cost hospitals a fair amount of money, so the end goal is to identify and reduce the possibility of a readmission. Prevention of patient readmission has been given a greater importance due to large cost involvement.

So in order to create a prediction system for readmission I studied US hospitals dataset.

The dataset represents 10 years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. It includes over 50 features representing patient and hospital outcomes.

The objective of this project is to predict whether the patient will be readmitted to the hospital or not. For predicting it most accurately I have used various prediction models like Logistic Regression, Decision Trees, Neural Networks, Random Forests, SVM and Naïve Bayes and seen their performance in order to get the best results.

The feature selection was done on the basis of PCA and the correlation analysis. Data was converted ad pre-processed by removing null values and changing categorical variables. It even involved changing categorical variables using domain knowledge for diagnosis codes from ICD-9 codes.

The EDA gave us some interesting insights as the readmission is more in female people and in the Caucasian race. The overall comparison of methods on the basis of their accuracy sensitivity and specificity was generated as the output.