Project Proposal

Project Title:
Using Machine Learning techniques to Predict Hospital Readmission for Diabetic Patients.

Project Description:
Predict if a patient with diabetes will be readmitted to the hospital within 30 days.

Background:
CMS, Center of Medicare and Medicaid Services has created many programs to improve the quality of patient care. One such program is called the Hospital Readmission Reduction Program. Under this program, the objective is to encourage hospitals to improve communication and care coordination to better engage patients and caregivers in discharge plans and, in turn, reduce avoidable readmissions. For those hospitals which are currently disciplined under this program, need to create interventions to provide additional assistance to patients with increased risk of readmission.

But how do we identify these patients? We can use predictive modeling from data science to help prioritize patients.

Importance:
One patient population that is at increased risk of hospitalization and readmission is that of diabetes. Diabetes is a medical condition that affects approximately 1 in 10 patients in the United States. Diabetic patients have almost double the chance of being hospitalized than the regular population. Therefore, in my project, I am planning to focus on predicting hospital readmission for patients with diabetes.

Source Data:
The data used in this project originally comes from the UCI machine learning repository (https://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008).
It consists of over 100000 hospital admissions from patients with diabetes from 130 US hospitals between 1999–2008.

Proposed Plan:
In this project I will demonstrate how to build a model predicting readmission using the following steps:

1) **Data Exploration:** Exploring the data is the first step for the analysis. It is important to validate the dataset before using it for our prediction model. Then an exploratory data analysis shall be performed to identify what are the most correlated features.
2) **Feature Engineering:** In this section, we will create features for our predictive model. For each section, we will add new variables to the data frame and the keep track of which columns of the data frame we want to use as part of the predictive model features.
3) **Splitting Training / validation / test samples:** The idea behind splitting the data is so that I can measure how well my model would perform on unseen data.
4) **Model Selection / Model Building:** In this section, I will train a few machine learning models and use a few techniques for optimizing them. I will then select the best model based on performance on the validation set.
5) **Model Evaluation:** Now that I have selected my best model. I will evaluate the performance of the test set.