

PREDICTING HOSPITAL READMISSION FOR PATIENTS WITH DIABETICS

A Project report submitted to

DEPARTMENT OF

COMPUTER SCIENCE AND ENGINEERING

In partial fulfillment of the requirements for the award of the degree of

Bachelor Of Technology In

Computer Science And Engineering



Presented by

CH.RENUKA	Y16CS1219
B.MOUNIKA	Y16CS1212
M.UNNATHA	Y16CS1263
B.GOPI KRISHNA	Y16CS1216

Under the esteemed guidance of

Dr.B.V.V.S.PRASAD,

Associate professor,

Dept. of CSE

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CHALAPATHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by A.I.C.T.E, Affiliated To Acharya Nagarjuna University)

GUNTUR-522034

2019-2020

Introduction

As the healthcare system moves toward value-based care, CMS has created many programs to improve the quality of care of patients. One of these programs is called the Hospital Readmission Reduction Program (HRRP), which reduces reimbursement to hospitals with above average re admissions. For those hospitals which are currently penalized under this program, one solution is 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.

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. According to Ostling et al, patients with diabetes have almost double the chance of being hospitalized than the general population (Ostling et al 2017). Therefore, in this article, I will focus on predicting hospital readmission for patients with diabetes.

In this project I will demonstrate how to build a model predicting readmission in Python using the following steps

- data exploration
- feature engineering
- building training/validation/test samples
- model selection
- model evaluation

You can follow along with the Jupyter Notebook provided on my github (https://github.com/andrewwlong/diabetes_readmission).

Data Exploration

The data that is used in this project originally comes from the UCI machine learning repository

Feature Engineering

In this section, we will create features for our predictive model. For each section, we will add new variables to the dataframe and then keep track of which columns of the dataframe we want to use as part of the predictive model features. We will break down this section into numerical features, categorical features and extra features.

Numerical Features

The easiest type of features to use is numerical features. These features do not need any modification.

Categorical Features

The next type of features we want to create are categorical variables. Categorical variables are non-numeric data such as race and gender. To turn these non-numerical data into variables, the simplest thing is to use a technique called one-hot encoding.

Extra Features

The last two columns we want to make features are `age` and `weight`. Typically, you would think of these as numerical data.

Model Selection: Baseline models

In this section, we will first compare the performance of the following 7 machine learning models using default hyperparameters:

- K-nearest neighbors
- Logistic regression
- Stochastic gradient descent
- Naive Bayes
- Decision tree
- Random forest
- Gradient boosting classifier

Conclusion

Through this project, we created a machine learning model that is able to predict the patients with diabetes with highest risk of being readmitted within 30 days. The best model was a gradient boosting classifier with optimized hyperparameters. The model was able to catch 58% of the readmissions and is about 1.5 times better than just randomly picking patients. Overall, I believe many healthcare data scientists are working on predictive models for hospital readmission.