CISC 451 – Assignment 2

Supervised Learning

October 25th, 2020

Gavin McClelland – 10211444

Marshall Cunningham - 20249991

# Introduction

A dataset was provided representing 10 years of clinical care at 130 hospitals in the United States from years (1999-2008). The train and test datasets include 50 features representing a total of 101766 individual diabetes patients and their corresponding outcomes. This dataset was also provided with the accompanying mapping for each feature. The objective of this assignment is to train a predictive model that would predict if a patient would be readmitted to the hospital again or not. Moreover, the ideal model would classify if a patient would be readmitted in fewer than or greater than 30 days. This paper discusses the tools used for analysis and the approach to analysis. This is followed by a discussion of some approaches that didn’t work, and an evaluation of results.

# Software Packages and Download Instructions

All code written for this assignment was written in the Python programming language. The main software packages used in this assignment were numpy, pandas, and Sci-Kit learn. An exhaustive list of dependencies can be downloaded by running the command “pip install -r requirements.txt” in the root directory of the submitted folder.

# Analytics Process

## 3.1 EDA

## 3.2 Preparation

## 3.3 Modeling

## 3.4 Other approaches explored

# 4.0 Results

# 5.0 Conclusion

# References