Predicting diabetes patient hospital readmission

Capstone 1 proposal
Pavel Zimin

Hospital readmission is a highly preventable cause for high healthcare costs. In fact, hospitals are financially punished if they have higher readmission rate than the “expected” level. It is important for the hospitals to be able to know which population of patients are at a higher risk for being readmitted. This knowledge will help prioritize patients that will benefit from hospital discharge follow up programs.

The outcome of this analysis will be helpful to the hospital healthcare teams with prioritizing patient support program. Perhaps most importantly this analysis will benefit patients who will receive improved health care, decreased chances of readmission while incurring smaller cost.

The data that will be used for this project has been published in UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets/diabetes+130-us+hospitals+for+years+1999-2008). The data has been collected for 10 years (1999-2008) from 130 hospitals located throughout the U.S. The data was initially obtained from a large clinical dataset which was filtered for such criteria as inpatient encounter (admitted to the hospital), diabetes diagnosis codes, the length of stay was in the range of 1-14 days, laboratory tests were performed during the encounter and medications were administered during the encounter (see the original publication by Strack et al. for further details). After applying these filters the dataset consists of 101,766 encounters (observations) with 55 attributes. Attributes include demographic data, medications, diagnostic results and information regarding their hospital stay. The outcome for this dataset is whether the patient was readmitted within 30 days from discharge.

This is a classification problem. I will try multiple algorithms and will select the one that more accurately predicts the hospital readmission.

At the conclusion of this project I plan to produce a Jupyter Notebook containing code, narrative explanation, visualization and model selection.