Diabetes Readmission Prediction for Patients

Submitted To:

Prof. Tristan Glatard

Submitted By: Team 7

Mo Rokibul Islam 40060110

Sakib Shuvo 40077755

Kirti Dhir 40094627

Introduction

Readmission: a patient who had been discharged from a hospital is **admitted again** within a specified time interval

Related to the safety of patient's health and live

Related to more use of hospital service and cost

Data science can help to solve the problem

Identify the patients who are at risk for readmission

Reduce readmission rate of patients

Data Set Description

A 10 years of clinical care diabetic dataset

Supervised learning

Target classes: readmit <30 days, readmit >30 days and never readmit

Samples: 101766, Features: 50
Features Type: Numerical and Categorical

Data Pre-processing

- Identifiers: encounter_id and patient_nbr
- Missing Values > 35 %: weight (97%), payer code (40%), and medical specialty (47%)
- Constant value: examide, citoglipton and glimepiride-pioglitazone
- Multiple inpatient visits
- Patients who expired
- Patients who discharge to a hospice

Data Pre-processing

- Label encoding for categorical features
- Binning samples
- Data scaling
- Data resampling

Model Selection

- Logistic Regression and Random Forest Classifiers
- Python, scikit-learn, pandas
- Data splitting: training : test = 70 : 30
- Model parameters are selected using k-fold cross validation
- Accuracy matric: accuracy, precision, recall, F1

Results and Discussions

Without resampling

Matric	Logistic Regression		Random Forest	
	Training	Test	Training	Test
accuracy score	0.602	0.600	0.852	0.609
precision score	0.602	0.600	0.852	0.609
recall score	0.545	0.546	0.880	0.574
f1_score	0.572	0.572	0.866	0.591

With resampling

Matric	Logistic Regression		Random Forest	
	Training	Test	Training	Test
accuracy score	0.430	0.430	0.965	0.815
precision score	0.430	0.430	0.965	0.815
recall score	0.425	0.425	0.966	0.815
f1_score	0.427	0.428	0.965	0.815

Future Work

Cluster-computing and Parallel-computing Framework: Apache, Dask

More accurate model like Neural Network

thank you!