**Diabetic Patients Readmission**

**Abstract:**

This project aims to use classification models to predict diabetic patients with high risk of readmission.

Early identiﬁcation of patients facing a high risk of readmission can enable healthcare providers to conduct additional investigations and possibly prevent future readmissions. This not only improves the quality of care but also reduces the medical expenses on readmission.

## Design:

## The main purpose of this research is to facilitate healthcare institutions in predicting readmission of a diabetic patient by allowing the model to learn the relation among features and their importance in determining whether the patient will be readmitted or not. This helps the hospitals in providing the best inpatient treatment and improve the cost eﬃciency of healthcare centers. At the same time, it is important to identify the key factors responsible for the readmission of a diabetic patient.

Pre-process, analyze, visualize, and conduct supervised learning on dataset.

## Data:

I used dataset represents 10 years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. The dataset included around 100,000 records and 50 features which representing patient and hospital outcomes. The data extracted from the database for encounters with certain criteria.

The target labels, indicating:

“<30” if the patient was readmitted in less than 30 days, Label **Yes.**

“>30” if the patient was readmitted in more than 30 days, Label **No.**

“No” for no record of readmission, Label **No.**

## Algorithms:

## Feature Engineering:

## The data contains such attributes as patient number, race, gender, age, admission type, time in hospital, medical specialty of admitting physician, number of lab test performed, HbA1c test result, diagnosis, number of medications, diabetic medications, and number of outpatients, inpatient, and emergency visits in the year before the hospitalization.

## After removing unnecessary fear. I’ve removed columns which contain more than 40% missing values, one category and admission ID and patient medical record number

## Models:

I used classification analysis in order to classify whither patients will readmitted based on the features. First, I’ve started with data pre-processing by exploring the data, managing missing values, and transforming variables. Following that, I’ve applied analysis by using Random forest, decision tree algorithm and Boosting Classifier.

## Model Evaluation and Selection:

## With comparing models accuracy, Boosting was the best in the evaluation.

## Evaluation:

## Accuracy:65

## F1:64

## Precision:67

## Recall:59

## Tools:

## Numpy and Pandas for EDA.

## SKlearn for modeling.

## Seboron and Matpolitlib for visualizing

## Communication:

## Time in hospital is important factor in predicting readmission.

Chart, bar chart, histogram

Description automatically generatedChart, histogram

Description automatically generated