

CS513-A (KDD) PROJECT PROPOSAL: (Heart Disease Detection)

Project Group No: 5

Problem Statement: (Description of the problem statement)

According to the Centers for Disease Control and Prevention (CDC), Diabetes is a chronic (long-lasting) health condition that affects how your body turns food into energy. With diabetes, your body doesn't make enough insulin or can't use it as well as it should. When there isn't enough insulin or cells stop responding to insulin, too much blood sugar stays in your bloodstream. Over time, that can cause serious health problems such as heart disease, vision loss, and kidney disease. There isn't a cure yet for diabetes, but losing weight, eating healthy food, and being active can help treat it.

Diabetes is not only one of the top ten leading causes of death in the world, but also the most expensive chronic disease in the US. Therefore, reducing readmission rates for diabetic patients has a great potential to reduce medical cost significantly.

Therefore, we want to classify a diabetic patient's hospital readmission status by using data mining techniques on a dataset that consists of 101,766 records of patients.

Dataset: (Description)

The dataset consists of 101,766 records and 49 features in form of columns.

Source of Dataset: <https://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008>

Implementation Strategy and algorithms Used: (List different models)

We have decided to implement and compare 8 different models. We have chosen a few models from our course and few from outside the course. The following are the models selected by us:

- a. Random Forest with Randomized Search CV
- b. Logistic Regression with Grid Search CV
- c. Support Vector Machine with Grid Search CV
- d. K-nearest neighbor with Grid Search CV
- e. Naive Bayes
- f. Decision Tree

Model metrics and Evaluation: AUC-ROC, Confusion Matrix , F1, Recall, Precision

Team Members Names: Group 5

- a. Aditi Gupta
- b. David Laziuk
- c. Michael Scoli