

Data Mining Project Proposal

Project Members: Raghavi Ravi
Surabhi Priya

Problem Statement:

One person making a generous donation can save life of many.

Blood is essential for treatments, chronic diseases, injuries.

"We all have it but might need it".

Taking into account the data sets and introducing few categorical attributes we will be determining the probability of a person donating blood or not in near future.

Data Set Name: National Blood bank data set

Data source: Blood Transfusion Service Center

<https://archive.ics.uci.edu/ml/datasets/Blood+Transfusion+Service+Center>

Procedures involved:

Data Cleaning

Converting data to appropriate format

Scaling the data sets

Removing Redundant data

Filling the null values

Finding Correlation

Pattern Evaluation:

Applying Machine learning patterns like Decision Tree, Deep Learning, Neural Networks etc.

Accuracy Computation: Using Logistic Regression, KNN, Bagging, Gradient Boosting etc.

Accuracy visualization and Comparison: based on the computational value we plot a bar graph to compare the data.