

The blood you donate gives someone
another chance at life.

GIVE THE GIFT OF LIFE
DONATE BLOOD



Project Report on Blood donation prediction

Presented by
Pratik Jadhav

Introduction

The donation of blood is important because most often people requiring blood do not receive it on time causing loss of life. Examples include severe accidents, patients suffering from dengue or malaria, or organ transplants. Extreme health conditions such as Leukemia and bone marrow cancer, where affected individuals experience sudden high blood loss and need an urgent supply of blood and do not have it can also lead to loss of life. Sound data-driven systems for tracking and predicting donations and supply needs can improve the entire supply chain, making sure that more patients get the blood transfusions they need, which can reduce mortality risk.

Problem Statement

To predict if a blood donor will donate within a given time window.

Why solve this problem

- ❖ To reduce the number the deaths due to lack of blood.
- ❖ To spread awareness among people about blood donation
- ❖ To predict whether the donar will donate in the given time span or not.

Data

The dataset used in our study is one used by others researchers studying the problem posted on the UCI Machine Learning Repository. The source data has been taken from blood donor database of the Blood Transfusion Service Center in Hsin-Chu City in Taiwan. 748 donors were randomly selected from the donor database for the study. The features measured include: R (Recency - months since last donation), F (Frequency - total number of donation), M (Monetary - total blood donated in c.c.), T (Time - months since first donation), and a binary variable representing whether the donor donated blood in March 2007 (1 stands for donating blood; 0 stands for not donating blood).

Evaluation Metrics

- ❖ The Evaluation metric for this project is AUC Score.
- ❖ **AUC** represents the probability that a random positive (green) example is positioned to the right of a random negative (red) example. **AUC** ranges in value from 0 to 1.

First Steps EDA

The information which we got from the dataset is as under: There were no null values in the dataset

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 748 entries, 0 to 747
Data columns (total 5 columns):
Recency (months)                748 non-null int64
Frequency (times)              748 non-null int64
Monetary (c.c. blood)          748 non-null int64
Time (months)                  748 non-null int64
whether he/she donated blood in March 2007  748 non-null int64
dtypes: int64(5)
memory usage: 29.3 KB
```


Data Modelling

We have used logistic regression and Tpot classifier the two models as this problem is a classification problem.

The TPOT Classifier performs an intelligent search over machine learning pipelines that can contain supervised classification models, preprocessors, feature selection techniques, and any other estimator or transformer that follows the [scikit-learn API](#). The TPOTClassifier will also search over the hyperparameters of all objects in the pipeline.

By default, TPOTClassifier will search over a broad range of supervised classification algorithms, transformers, and their parameters. However, the algorithms, transformers, and hyperparameters that the TPOTClassifier searches over can be fully customized using the `config_dict` parameter.

AUC Score

Model	AUC Score
Logistic Regression	0.7890972663699937
Tpot Classifier	0.7852828989192625