**Give Life: Predict Blood Donations**

Build a binary classifier to predict if a blood donor is likely to donate again.

#### Project Description

"Blood is the most precious gift that anyone can give to another person — the gift of life." ~ [World Health Organization](http://www.who.int/features/qa/61/en/)

Forecasting blood supply is a serious and recurrent problem for blood collection managers: in January 2019, "Nationwide, the [Red Cross](https://www.kjrh.com/news/local-news/red-cross-in-blood-donation-crisis) saw 27,000 fewer blood donations over the holidays than they see at other times of the year." Machine learning can be used to learn the patterns in the data to help to predict future blood donations and therefore save more lives.

In this Project, you will work with data collected from the donor database of Blood Transfusion Service Center in Hsin-Chu City in Taiwan. The center passes its blood transfusion service bus to one university in Hsin-Chu City to gather blood donated about every three months. The dataset, obtained from the [UCI Machine Learning Repository](https://archive.ics.uci.edu/ml/datasets/Blood+Transfusion+Service+Center), consists of a random sample of 748 donors. Your task will be to predict if a blood donor will donate within a given time window. You will look at the full model-building process: from inspecting the dataset to using the [tpot](https://www.datacamp.com/community/tutorials/tpot-machine-learning-python) library to automate your Machine Learning pipeline.

To complete this Project, you need to know some Python, pandas, and logistic regression. We recommend one is familiar with the content in DataCamp's [Manipulating DataFrames with pandas](https://www.datacamp.com/courses/manipulating-dataframes-with-pandas), [Preprocessing for Machine Learning in Python](https://www.datacamp.com/courses/preprocessing-for-machine-learning-in-python/), and [Foundations of Predictive Analytics in Python (Part 1)](https://www.datacamp.com/courses/foundations-of-predictive-analytics-in-python-part-1) courses.

#### Project Tasks

* 1 Inspecting transfusion.data file
* 2 Loading the blood donations data
* 3 Inspecting transfusion DataFrame
* 4 Creating target column
* 5 Checking target incidence
* 6 Splitting transfusion into train and test datasets
* 7 Selecting model using TPOT
* 8 Checking the variance
* 9 Log normalization
* 10 Training the linear regression model
* 11 Conclusion