Predict Liver Diseases Patient :

The dataset ILPD (Indian Liver Patient Dataset) [1] comprises 583 instances with each having 10 features and 1 target variable. The dataset is used to classify if a patient, given a feature vector, has the Liver Disease or not (binary classification).

The predictive algorithms Random Forest ,SVM, KNeighbors and Logistic Regression are chosen for this task.

The task is made possible thanks to Python, and especially Scikit-Learn/Pandas libraries. Indeed, I used Anaconda3 [2] for "all-in-one" installation.

GridSearchCV is used to automatically search for optimal parameters in Random Forest.

In Python for Data Science,matrixplot This library is used to make Feature Scatter plot. ,NumPy:It provides some advance math functionalities to python.

The algorithms' performance is compared using pd.plotting.scatter_matrix.

The quickest way to reproduce this report is to run "python.exe ClassificationTask.py" with "DataPreparation.py" and "Indian Liver Patient Dataset (ILPD).csv" put in the same place.

Software Environment:

Anaconda3 v4.0.0 64bit (Python v3.5.2)

PyCharm 2016.1.3

IPython Notebook

Reference:

ILPD (Indian Liver Patient Dataset) Data Set

https://www.kaggle.com/uciml/indian-liver-patient-records