Project ID: PW20MG02

Project Type: MACHINE LEARNING

Project Title: Prediction Of Hepatitis and Liver Damage

Team Members: M SUMUKHA (01FB16ECS185)

M PRADEEP KUMAR(01FB16ECS183)

Sreerama Priyanka (01FB16ECS395)

Project Guide: Prof. MAHITHA.G

Project Abstract:

1. Hepatitis B:

Dataset of Hepatitis B contains text yes/no representing presence of an enzyme or symptom and values were missing in few attributes. So imputer was used to replace missing values with mean value and yes/no values are converted into numerical value.

After pre=processing Random forest, gradient Boosting and Decision Tree algorithms were used to build models to predict hepatitis. PCA is used to fetch 15 attributes out 20 attributes and above models are built. Which increased the prediction accuracy of Hepatitis B.

1. Liver Damage:

There are several reasons for Liver damage it might be due to virus or lack of self-care. Dataset contains 10 attributes all of them found to be necessary for prediction. Missing values are replaced by mean using imputer. To increase the size of the dataset SMOTE Over Sampling technique is used. Prediction models are trained before and after oversampling.

1. Hepatitis C:

Hepatitis C dataset contains ALT and RNA, enzyme and virus count which has been tested at regular weeks. As there is no missing values or textual values in our dataset, machine learning models were directly applied. This gave us low accuracy.

Due to this the attributes were categorized based on clinical ranges (e.g. age: 0-32 is grouped under single category ). Now machine learning models are applied, by doing this accuracy didn’t vary much. Then PCA is applied, even after this accuracy didn’t increase. It has been noticed that ALT/AST enzyme count is normal for Hepatitis C infected patients in one third of cases. So, ALT/AST columns were dropped. SMOTE Oversampling increased not more than 10% of data. Even after doing this accuracy didn’t increase more than 32%.

Having lesser number of rows in the dataset gave low accuracy.

Code Execution:

For Code Analysis:

1) Open jupyter notebook

2) install the required modules- sklearn, numpy, pandas, and matplotlib or execute this command “pip install -r requirement.txt”

3) navigate to “Liver/ Hepatitis B/ Hepatitis C” directory and run files in jupyter notebook.

To run Application:

1) run “python api.py” in the root directory.

2) navigate to “./UI” directory and open index.html in chrome.