Statistical Machine Learning Approaches to Liver Disease Prediction

Ideation:

To predict that the person having liver disease or not using various machine learning algorithms,We must need previous patient records . For that the data set is taken from kaggle. This data set contains 416 liver patient records and 167 non liver patient records collected from North East of Andhra Pradesh, India. The "Dataset" column is a class label used to divide groups into liver patient (liver disease) or not (no disease). This data set contains 441 male patient records and 142 female patient records**.**

Any patient whose age exceeded 89 is listed as being of age "90".

Columns:

- Age of the patient
- Gender of the patient
- Total Bilirubin
- Direct Bilirubin
- Alkaline Phosphotase
- Alamine Aminotransferase
- Aspartate Aminotransferase
- Total Protiens
- Albumin
- Albumin and Globulin Ratio
- Dataset: field used to split the data into two sets (patient with liver disease, or no disease)

Now ,the user can predict the disease by entering above parameters like(Age of the patient, Gender of the patient, Total Bilirubin,Direct Bilirubin, Alkaline Phosphotase, etc)in the web application. The first step is Preprocessing of data which includes cleaning of data ,removing noise from data and partitioning dataset for Training and Testing. Here we take 80% for Training and 20% for Testing because how much you test that,

much we got accuracy. After partitioning we use different machine learning algorithms to train the model and predict the accuracy by testing. Comparing those accuracy and comes out with the maximum accuracy.