

Statistical Machine Learning Approaches to Liver Disease Prediction

Team ID: PNT2022TMID26382

Team Leader : Madhulika I

Team Member : Rithika AM

Team Member : Kavya A P

Team Member : Kirthiga S

Reading the Dataset

- You might have your data in .csv files, Excel files or .tsv files or something else. But the goal is the same in all cases. If you want to analyse that data using pandas, the first step will be to read it into a data structure that's compatible with pandas.
- Let's load a .csv data file into pandas. There is a function for it, called `pd.read_csv()`. We will need to locate the directory of the CSV file at first (it's more efficient to keep the dataset in the same directory as your program).

```
In [2]: data = pd.read_csv('indian_liver_patient.csv')
```

Content

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 a liver patient (liver disease) or not (no disease). This data set contains 441 male patient records and 142 female patient records.

Note: We have not started any data analysis yet, this is just to show you all the authenticity of the dataset.