SRIHARSHA DAPARTI

UX46295

DATA 606 – CAPSTONE IN DATA SCIENCE

A VIEW ON HCV DATA

AIM

The aim of the project is to delve into the Hepatitis C Virus (HCV) for Egyptian patients dataset thoroughly and gain as much information as possible. Along this process, I expect the data to throw me problems which I intend to solve in the span of this semester. The goal is to develop a predictive model that predicts the progression of the disease in a given patient.

DATASET

Hepatitis C Virus (HCV) for Egyptian patients. This data was obtained from UCI Machine Learning Repository. Citation: Dua, D. and Graff, C. (2019). The dataset has two files. The first is the dataset itself which shows the anonymous records of Egyptian patients who underwent treatment dosages for HCV about 18 months and the second file contains the discretization parameters for each and every attribute in the first file.

The dataset contains about 1000 patient records with 29 attributes for each record explaining the treatment .The attributes of the patient records are:

- Age
- BMI
- Nausea/Vomiting
- Diarrhea
- Jaundice
- WBC Count – White Blood Cell Count
- HGB - Haemoglobin
- AST 1 – Aspartate Transaminase ratio
- ALT 4 – ALT Week 4
- ALT 24 – ALT Week 24
- ALT 48 – ALT Week 48
- RNA Base
- RNA 12
- RNA EF – RNA Elongation Factor
- Baseline Histological Staging
- Gender
- Fever
- Headache
- Fatigue & Generalized bone ache
- Epigastric Pain
- RBC Count - Red Blood Cell Count
- Platelets Count
- ALT 1 – Alanine Transaminase ratio Week 1
- ALT 12 – ALT Week 12
- ALT 36 – ALT Week 36
- ALT after 24 w – ALT after 24 weeks
- RNA 4
- RNA EOT – RNA at End Of Treatment
- Baseline Histological Grading

Data Link: Hepatitis C Virus (HCV) for Egyptian patients.