DTSC 5502 – Project

Liver Disorder prediction

**Description:**

This project is a prediction model used to predict the liver disorder. In this project the model is trained by taking the input from the dataset. And the model predicts the remaining dataset. And the model’s prediction is purely based on its accuracy. To get more accuracy we must train the model with maximum input.

In this project we have used two datasets one from the Kaggle website and the other one is from the data.world. To make the model more accurate we are merging the datasets.

**Variables:**

There are seven variables in the dataset. And these variables are used to determine the condition of the liver.

* **mcv** mean corpuscular volume
* **alkphos** alkaline phosphotase
* **sgpt** alamine aminotransferase
* **sgot** aspartate aminotransferase
* **gammagt** gamma-glutamyl transpeptidase
* drinks number of half-pint equivalents of alcoholic beverages drunk per day
* selector field used to split data into two sets.

Structure of the dataset is 7 attributes and 345 instances.