

PROBLEM STATEMENT: Liver Disease Prediction

This Machine Learning Challenge requires participants to create predictive models to predict the stage of liver Cirrhosis using 18 clinical features. Cirrhosis damages the liver from a variety of causes leading to scarring and liver failure.

Hepatitis and chronic alcohol abuse are frequent causes of the disease. Liver damage caused by cirrhosis can't be undone, but further damage can be limited. Treatments focus on the underlying cause. In advanced cases, a liver transplant may be required. Predicting the stage of cirrhosis and beginning the treatment before it's too late can prevent the fatal consequences of the disease.

ABOUT DATASET:

- Train Dataset - It consists of a total of 6801 datapoints.
- Test Dataset - You must predict the stage of cirrhosis of 3201 datapoints.

- 1) ID: Unique Identifier
- 2) N_Days: number of days between registration and the earlier of death, transplantation, or study analysis time.
- 3) Status: status of the patient C (censored), CL (censored due to liver tx), or D (death)
- 4) Drug: type of drug. D-penicillamine or placebo
- 5) Age: age in [days]
- 6) Sex: M (male) or F (female)
- 7) Ascites: presence of ascites N (No) or Y (Yes)
- 8) Hepatomegaly: the presence of hepatomegaly N (No) or Y (Yes)
- 9) Spiders: the presence of spiders N (No) or Y (Yes)
- 10) Edema: the presence of edema N (no edema and no diuretic therapy for edema), S (edema present without diuretics, or edema resolved by diuretics), or Y (edema despite diuretic therapy)

- 11) Bilirubin: serum bilirubin in [mg/dl]
- 12) Cholesterol: serum cholesterol in [mg/dl]
- 13) Albumin: albumin in [gm/dl]
- 14) Copper: urine copper in [ug/day]
- 15) Alk_Phos: alkaline phosphatase in [U/liter]
- 16) SGOT: SGOT in [U/ml]
- 17) Triglycerides: triglycerides in [mg/dl]
- 18) Platelets: platelets per cubic [ml/1000]
- 19) Prothrombin: prothrombin time in seconds [s]
- 20) Stage: histologic stage of disease (1, 2, 3, or 4)