**Classification Project: Predict Liver Disease**

**Business Objective**: The variable to be predicted is categorical (no disease, suspect disease, hepatitis c, fibrosis, cirrhosis). Therefore, this is a classification project.

**Data Set Details:** The number of instances (rows) in the data set is 615, and the number of variables (columns) is 13. Nearly all input variables are numeric-valued except one, Sex, which is binary, and most of them represent measurements from blood and urine analysis. The following list summarizes the variables information:

**category(diagnose):** No disease, suspect disease, hepatitis c, fibrosis or cirrhosis.

**age: (0-100).** The normal ranges change depending on the age.

**sex: (f or m).** The normal ranges change depending on the sex.

**albumin:** (normal range 34-54g/L). A blood albumin level below normal may be a sign of kidney disease or liver disease as Hepatisis o Cirrhosis.

**alkaline\_phosphatase:** (40-129 U/L). It is an alkaline phosphatase test to check for liver disease or liver damage. The main causes or conditions that can cause normal levels to rise is liver disease.

**alanine\_aminotransferase:** (7-55 U/L). Alanine aminotransferase or ALT is an enzyme found primarily in the liver. An elevated amount indicates liver damage from hepatitis, infection, cirrhosis, liver cancer, or other liver diseases.

**aspartate\_aminotransferase:** (8-48 U/L). It is an enzyme found primarily in the liver, but also in the muscles. Elevated levels of AST in the blood may indicate hepatitis, cirrhosis, mononucleosis, or other liver diseases.

**bilirubin:** (1-12 mg/L). It is a yellowish substance that forms during the normal process of breaking down red blood cells in the body. Lower-than-normal bilirubin levels are generally not a concern. Elevated levels may indicate liver disease or damage.

**cholinesterase:** (8-18 U/L). It is a blood test that studies the levels of 2 substances that help the nervous system to function properly. Cholinesterase deficiency causes liver disease, renal disease...

**cholesterol:** (Less than 5,2 mmol/L). Cholesterol is a waxy, fat-like substance found in every cell in your body. A high level could cause carotid artery disease, stroke, and peripheral artery disease...

**creatinina:** (61.9-114.9 ??mol/L for m and 53-97.2 ??mol/L for f). It measures the level of creatinine in the blood. High level of creatinine in the blood and low level in the urine indicate kidney disease or affecting the function of the kidneys

**gamma\_glutamyl\_transferase :** (from 0 to 30-50 IU/L.). It is an enzyme that indicates cholestasis. The higher the level of GGT, the greater the level of liver damage.

**protein:** (Less than 80 mg). It measures the amount of protein in the urine. If protein levels in the urine are elevated, this could indicate kidney damage or another medical problem.

**Acceptance Criterion:**

Need to deploy the end results using Streamlit etc.

**Milestones:**

**30 days to complete the Project**

|  |  |  |
| --- | --- | --- |
| **Milestone** | **Duration** | **Task start - End Date** |
| Kick off and Business Objective discussion | 1 day | 24-Sep-2024 |
| Data set Details | 2 to 3 days | 26-Sep-2024 |
| EDA | 1 Weeks – 1 ½ week | 01-Sep-2024 |
| Model Building | 1 Week – 1 ½ week | 08-Sep-2024 |
| Model Evaluation | 1 week | 15-Sep-2024 |
| Feedback | 15-Sep-2024 |
| Deployment | 1 Week | 22-Sep -2024 |
| Final presentation | 1 day | Will decide according to status of all teams |

Protocols:

1. All participants should adhere to agreed timelines and timelines will not be extended.
2. All the documentation – Final presentation and python code to be submitted before the final presentation day.
3. All the participants must attend review meetings.