Prediction of Liver Disease

**Algorithms Used:**

* Logistic Regression
* K-Nearest Neighbors
* Support Vector Classifier
* Decision Tree Classifier
* Random Forest Classifier
* Gradient Boosting Classifier
* XGBoost Classifier

**Colab Notebook:** <https://colab.research.google.com/drive/1NbEry2bAfACZbrFudTFu7AOmQKzm-IIW?usp=sharing>

**Source:**

<https://www.kaggle.com/code/chanchal24/liver-disease-prediction-using-7-models/notebook>

**Dataset:**

<https://www.kaggle.com/code/resolut/eda-indian-liver-patient-records/input>

**Rows, columns:** (583, 11)

**About the dataset**

The dataset used for liver disease prediction consists of medical records from patients, including various health indicators and demographic information. These parameters offer a comprehensive view of an individual's liver function and can be instrumental in predicting the presence or absence of liver disease. The dataset comprises 10 key features along with a label indicating whether or not the patient has a liver disease.

Here is a breakdown of the attributes used in this prediction model:

* **Age**: The age of the patient in years. Age is a critical factor, as liver disease prevalence tends to increase with age.
* **Gender**: Gender is a binary indicator (Male/Female) which may have implications on liver disease risk and progression. Certain liver diseases can show a varying prevalence between genders.
* **Total Bilirubin**: Bilirubin is a yellow compound that occurs in the blood as a result of the breakdown of red blood cells. High levels of total bilirubin can indicate liver dysfunction or other related issues.
* **Direct Bilirubin**: A portion of total bilirubin that is directly processed by the liver. An elevated direct bilirubin level can signal liver damage or bile duct obstruction.
* **Alkaline Phosphatase (ALP)**: An enzyme found in the liver and bones. Increased levels of ALP are associated with liver damage or bone disorders, making it an important marker in this dataset.
* **Alanine Aminotransferase (ALT)**: An enzyme predominantly found in the liver. High ALT levels indicate liver cell damage, which is typical in conditions like hepatitis or fatty liver disease.
* **Aspartate Aminotransferase (AST)**: Another enzyme that is present in the liver and heart. When liver cells are damaged, AST levels in the blood increase.
* **Total Proteins**: Total proteins include albumin and globulin in the blood. Lower levels of total proteins may suggest liver disease since the liver plays a crucial role in protein synthesis.
* **Albumin**: A protein made by the liver, albumin helps keep blood vessels stable and prevents fluid from leaking into other tissues. Low albumin levels may indicate poor liver function.
* **Albumin and Globulin Ratio**: This ratio measures the balance between albumin and globulin proteins. An abnormal ratio can reflect liver damage or other health issues related to the liver.
* **Dataset**: This column specifies whether the individual is classified as having liver disease (1) or not (0), providing the label for the prediction task.

**Algorithms Model and Accuracy**

|  |  |
| --- | --- |
| **Models** | **Scores** |
| Logistic Regression | 0.7699 |
| Random Forest Classifier | 0.7257 |
| Support Vector Classifier | 0.7168 |
| Gradient Boosting Classifier | 0.7168 |
| XGBoost | 0.7168 |
| Decision Tree Classifier | 0.6991 |
| K-Nearest Neighbors | 0.6637 |

Best: **Logistic Regression**