

An anatomical illustration of the human digestive system, including the liver, stomach, and intestines, rendered in a semi-transparent blue style. The liver is highlighted in a solid orange color, positioned centrally in the upper abdominal region. The background is a dark blue gradient.

Liver Cirrhosis Stage Prediction

- Kathula Shalom Ramesh

TABLE OF CONTENT

❖ Introduction and Objective of project

❖ Problem Statement

❖ Dataset Overview

❖ Data Characteristics

❖ Conclusion

INTRODUCTION

- ❖ Liver cirrhosis is a severe, progressive condition where healthy liver tissue is replaced by scar tissue (fibrosis), leading to a gradual loss of liver function. The liver plays a crucial role in detoxifying harmful substances, producing bile, and regulating metabolism. As cirrhosis progresses, the liver's ability to perform these vital functions diminishes, potentially leading to life-threatening complications such as liver failure, portal hypertension, and liver cancer.

OBJECTIVE:

- ❖ The objective of this project is to develop an advanced machine learning model that can accurately predict the stage of liver cirrhosis using patient data, including clinical test results, demographics, and lifestyle factors. This model will be integrated into [Your Company Name]'s existing healthcare platform, enabling healthcare providers to make informed decisions about patient care and treatment.

PROBLEM STATEMENT:

- ❖ Prato is a leading healthcare technology company focused on developing innovative solutions for chronic disease management.
- ❖ As a company dedicated to improving healthcare delivery, Prato recognizes the need for a reliable, scalable solution to predict the stage of liver cirrhosis in patients. Accurately identifying the stage of cirrhosis is critical for timely interventions, reducing healthcare costs, and improving patient quality of life.

Dataset Overview

❖ Key Features:

Total Records: Number of rows - 25000 rows

Total Features: Number of columns- 19 columns

❖ Features:

❖ Categorical Features:

- List of categorical features : ['Status', 'Drug', 'Sex', 'Ascites', 'Hepatomegaly', 'Spiders', 'Edema']

❖ Numerical Features:

- List of numerical features :
['N_Days', 'Age', 'Bilirubin', 'Cholesterol', 'Albumin', 'Copper', 'Alk_Phos', 'S GOT', 'Tryglicerides', 'Platelets', 'Prothrombin', 'Stage']

Data Characteristics

❖ Data Cleaning Process :

Handling Null Values : Filled missing values with appropriate substitutes.

Removing Duplicates : Removed duplicate records to ensure each entry is unique.

Categorical Encoding : Converted categorical variables into numerical format.

Applied techniques such as One-Hot Encoding or Label Encoding.

❖ Cleaned Data Summary:

Description: Summarized the state of the cleaned data.

Characteristics:

- **No Missing Values:** All null values have been addressed.
- **No Duplicates:** Duplicate records have been removed.
- **Transformed Features:** Data is normalized/standardized and encoded appropriately.

Model Selection :

- **Model Selection:**
- **Decision Tree:**
 - **Description:** A decision tree model was considered for classification.
 - **Advantages:** Easy to interpret, visual representation of decision rules.
 - **Limitations:** Prone to overfitting, can be less accurate with complex data.
- **Random Forest:**
 - **Description:** Chosen as the final model for classification.
 - **Advantages:** Combines multiple decision trees to improve accuracy and robustness.
 - **Features:** Reduces overfitting, handles large datasets well.

Model Performance:

- **Random Forest Accuracy:**
- **Description:** Achieved high accuracy in predicting liver cirrhosis stages.
- **Performance Metrics:**
 - **Accuracy Score:** 0.9530666666666666
 - **Benefits:** Provides reliable predictions and handles feature interactions effectively

Comparison:

- **Decision Tree vs. Random Forest:**
 - **Decision Tree:** Lower accuracy and higher susceptibility to overfitting.
 - **Random Forest:** Superior performance with better generalization.

Suggestions And Recommended :

- **Clinical Integration:**
- **Collaboration with Experts:** Work with medical professionals to validate predictions and integrate the model into clinical decision-making processes.
- **User Training:** Ensure that end-users understand the model's predictions and limitations for effective application in practice.
- **Documentation and Reporting:**
- **Detailed Documentation:** Maintain comprehensive documentation of model development, including data preprocessing, model parameters, and performance metrics.
- **Clear Reporting:** Communicate findings and recommendations clearly to stakeholders and decision-makers.

Conclusion

- **Chosen Model: Random Forest**
- **Reason for Selection:**
 - **Superior Accuracy:** Demonstrated high accuracy in predicting liver cirrhosis stages.
 - **Robustness:** Handles large datasets and complex feature interactions effectively.
 - **Reduced Overfitting:** Combines multiple decision trees to provide reliable predictions.

Outcome:

- **Model Performance:** Random Forest outperformed other models in accuracy and reliability for the task of predicting liver cirrhosis stages.

**Thank
You**

