SPRINGBOARD

PATTERNS IN LIVER PANEL TESTS FOR DIAGNOSING AND STAGING

HEP C PREDICTION

Naomi Lopez

AUGUST 2024

Data Science Method

Define.Clean. Explore. Model. Conclude

- Problem identification
- Hypothesis formation

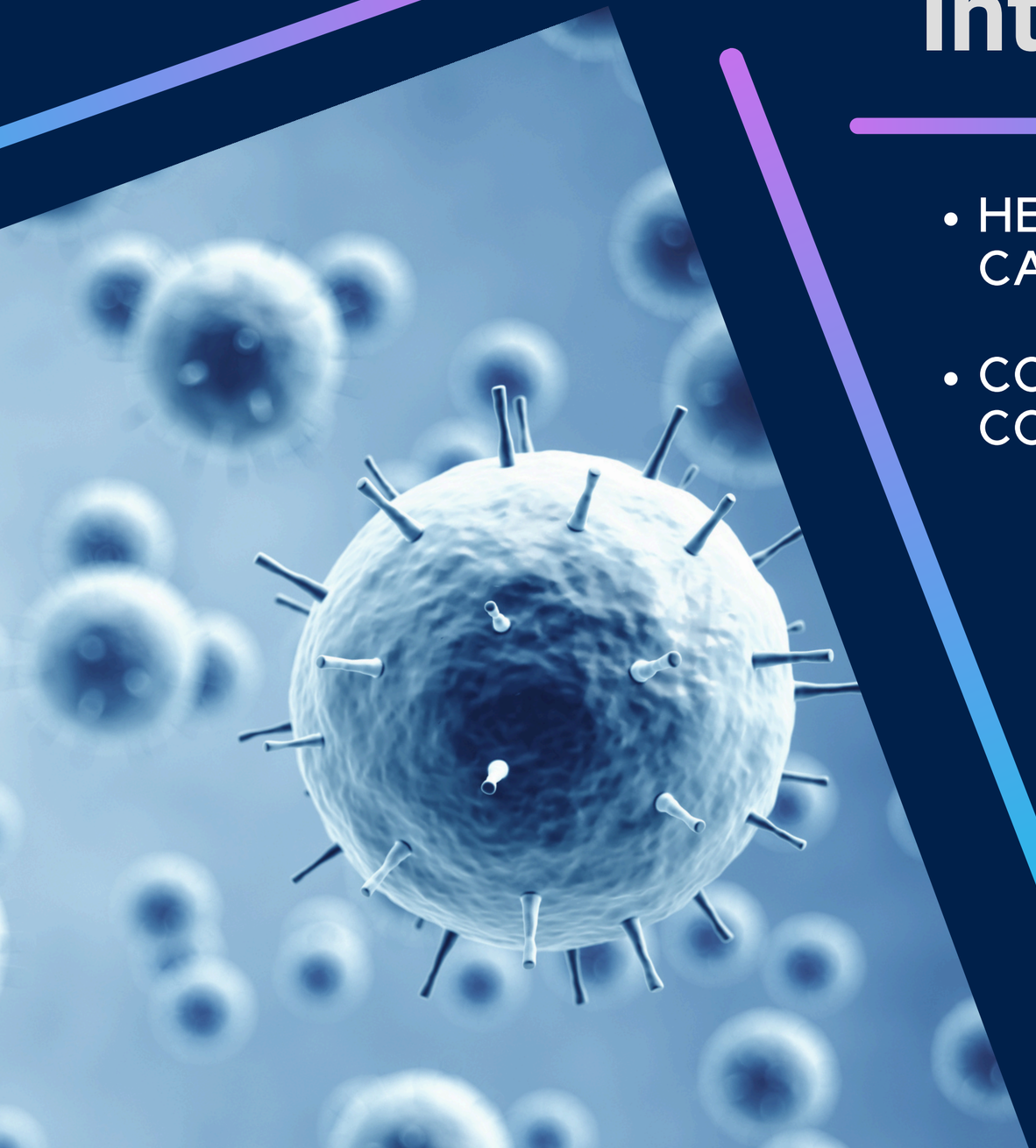
- Gather data
- Assess missing values
- Ensure Consistency

- Visualize trends
- Notice correlations
- Identify outliers

- Deploy Models
- Explore parameters
- Make Adjustments

- Make connections
- Insights
- Improvement

Define



Introduction

- HEPATITIS C CAUSES LIVER DAMAGE, CANCER, AND IRREVERSIBLE LIVER FAILURE.
- COMMON TRANSMISSION ROUTES INCLUDE CONTACT WITH INFECTED BLOOD.

Objective

- CLASSIFICATION MODEL
- DISTINGUISH HEP C STATUS & SEVERITY
- LIVER FUNCTION PANEL DATA

Impact

DOCTORS

- Liver test abnormalities could trigger Hep C test
- More frequent screening
- CDC's current recommendation is once in a lifetime

PATIENTS

- Reduces the risk of undiagnosed Hepatitis C
- Earlier intervention and better health outcomes

HOSPITALS

- Accurate staging through blood work could reduce the need for expensive diagnostic tests like MRIs and ultrasounds

Cleaning

01 **Inconsistent Data Types and Redundancies**

- “Category” converted from string to numerical
- Redundant default index removed

02 **Handling Null Values**

- Columns ALP , CHOL dropped due to high null values.
- Rows with missing values in PROT, ALB, and ALT dropped

03 **Data Distribution and Patient Count**

- Dataset was imbalanced
- 75 out of 615 patients positive for Hep C
- To balance dataset, majority class was under-sampled

Clean

Processing

01

One Hot Encoding

"Sex" column converted to Dummy Variables

02

Numerical Category

Hep C - = 0
Hep C + early stage = 1
Hep C + advanced stage = 2
Hep C + Cirrhosis = 3

03

Feature Scaling

SKLearn's StandardScaler ensure features have mean of 0 and a standard deviation of 1

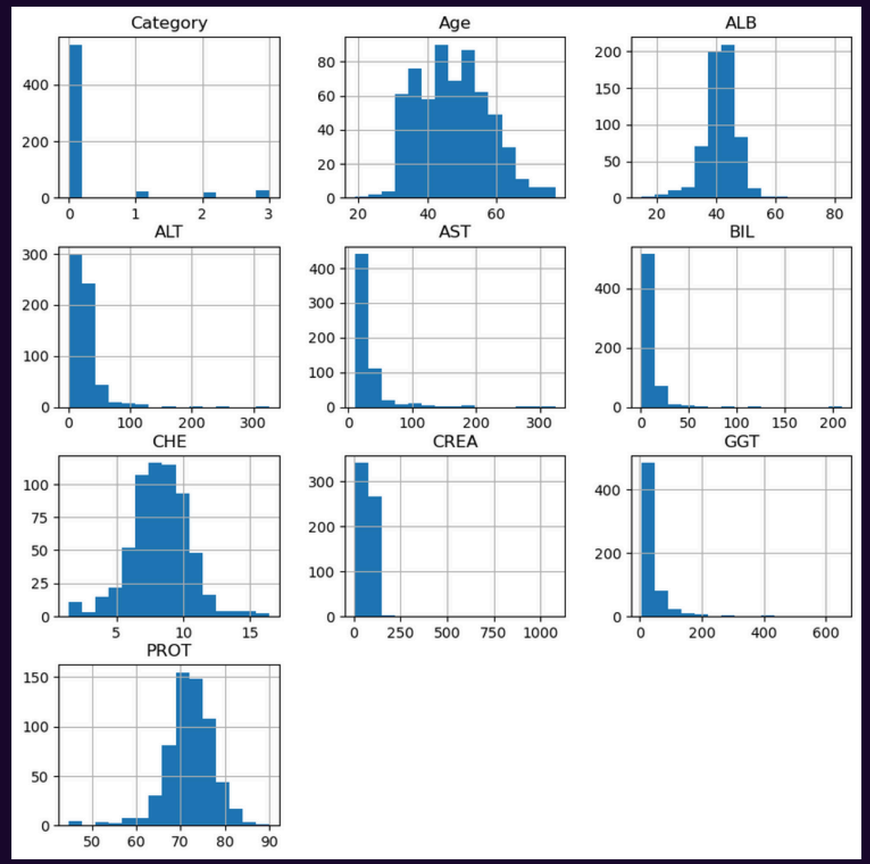
04

Split Data

80% Training Data (80 samples)
20% Testing Data (20 samples)

Explore

Data Analysis



Features that are skewed peak interest

ALT. AST. BIL. CREA. GGT

Liver enzyme levels increase during instances of liver damage

Explore

Data Analysis

- **The heatmap depicts how features relate to each other.**
- **Can help identify patterns in the dataset**
- **Strong relationship between AST and GGT: potential Hepatitis C**

Feature Correlation

MODEL

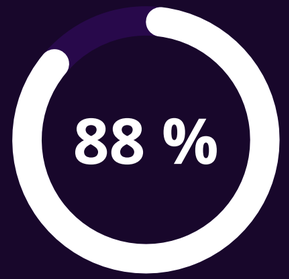
Modeling

Gini Impurity

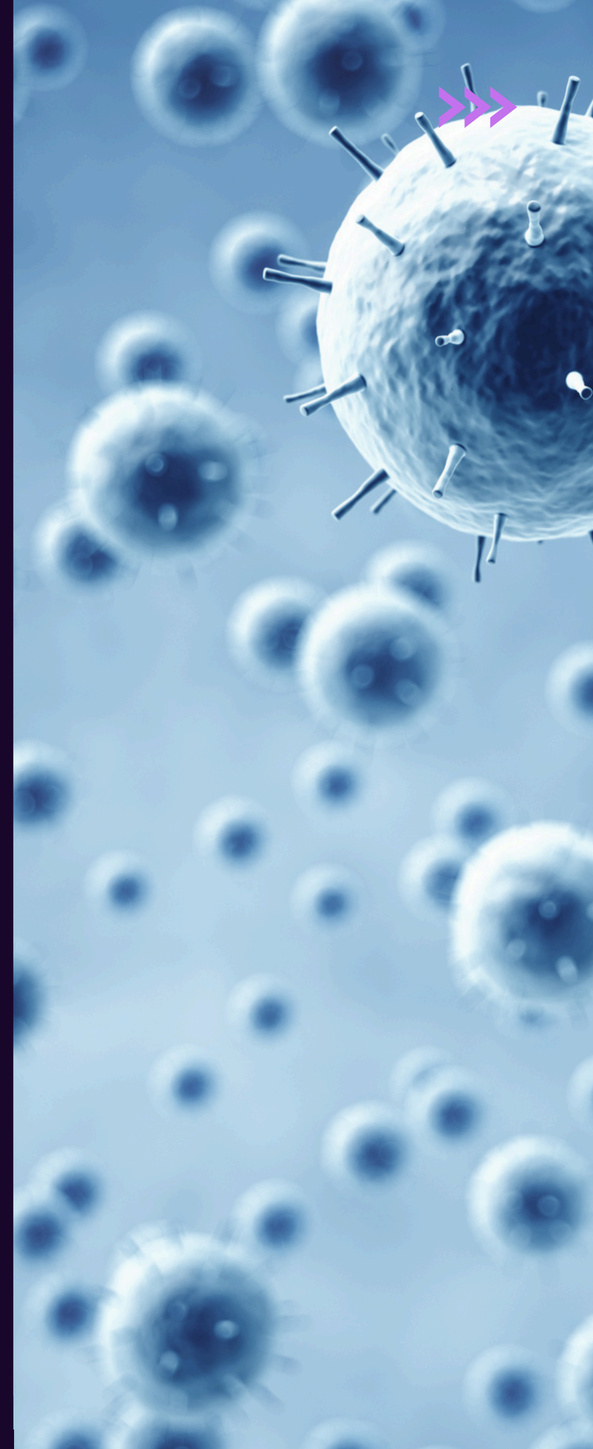
Entropy

Precision

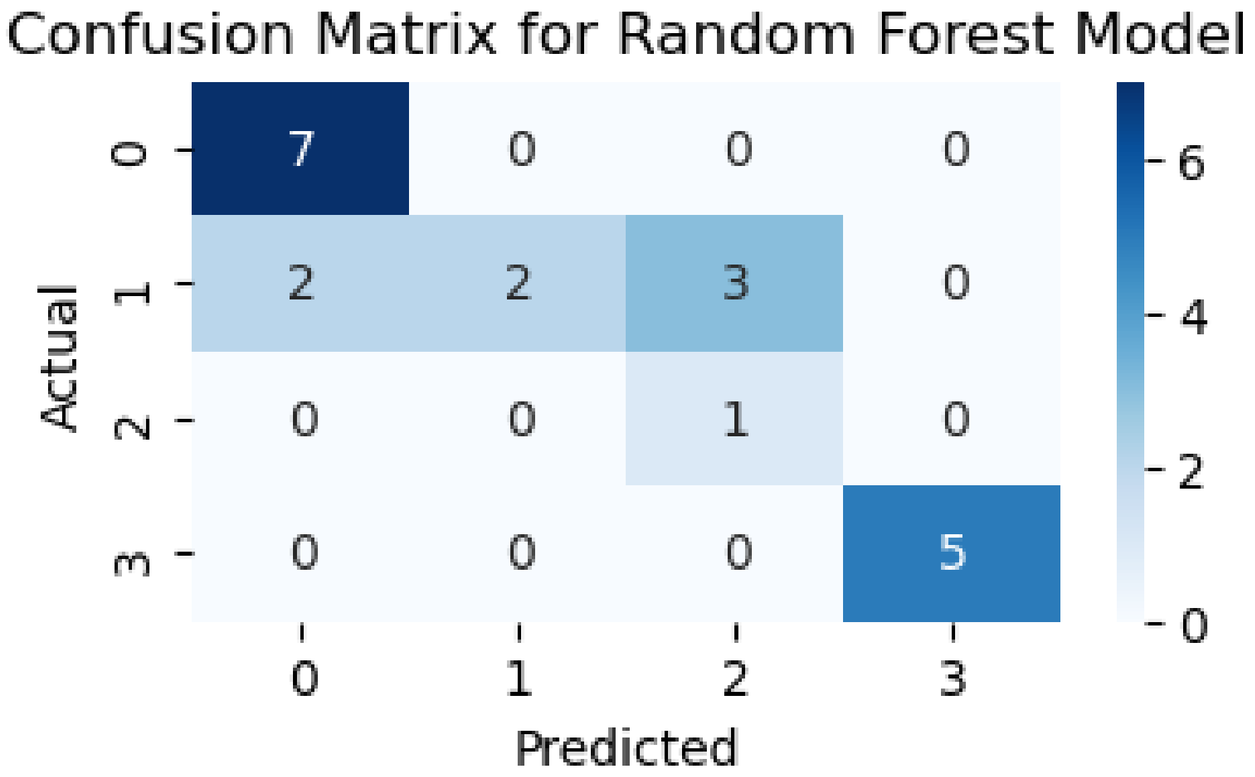
Random Forest



Support Vector Machine



Model Performance



Random Forest

Confusion Matrix

- **Highest Accuracy, Balanced Accuracy, Precision, Recall, and F1 Score**
- **Class 0: Highest accuracy**
- **Class 3: Perfectly classified**

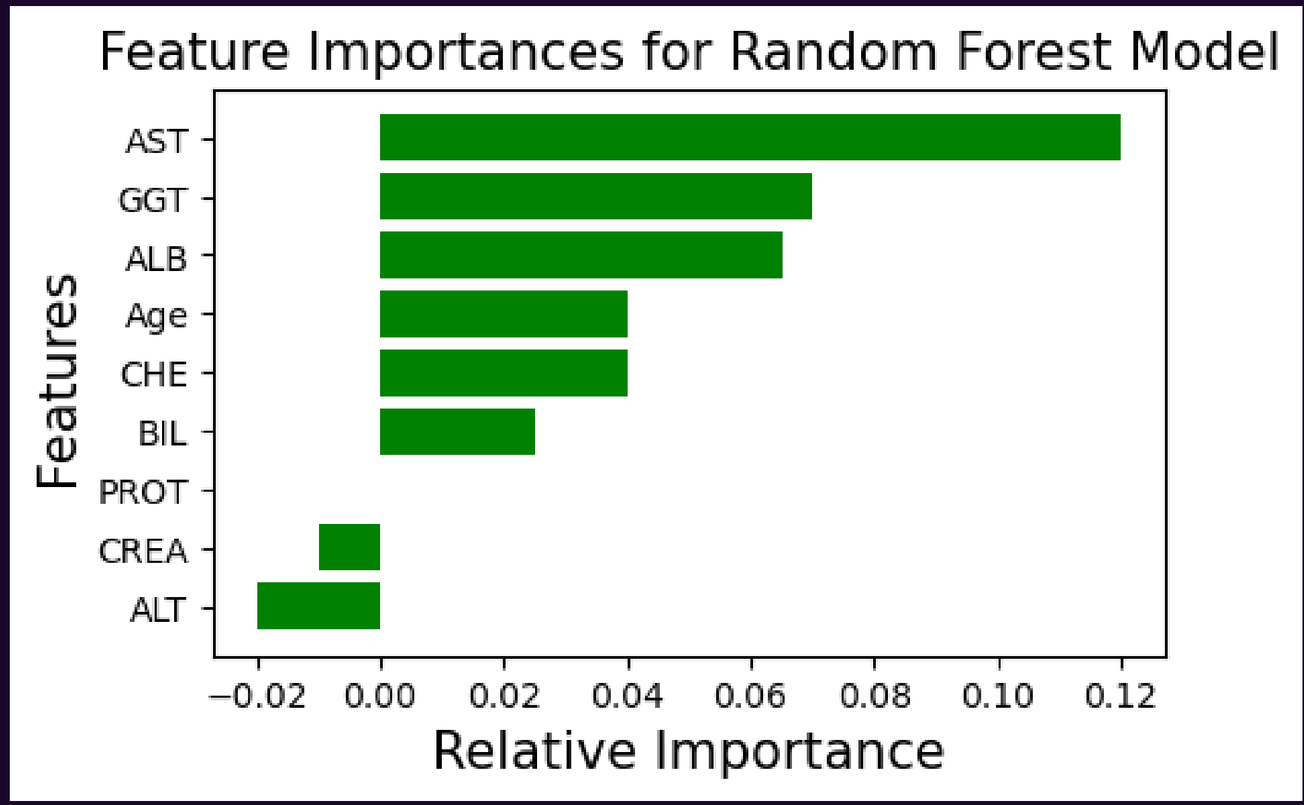
Cross Validation

- **Range: 0.50 to 0.80**
- **Mean Score: 0.68**

Conclude

Insights

- **Elevated AST/ GGT indicate liver damage**
- **Consistent with Hep C impact on liver**
- **Highly correlated in heatmap skewed histograms (outliers)**



ROC/AUC Performance

- **Class 0: 0.99**
- **Class 1 & 2: 0.95**
- **Class 3: 1.00**

Conclude

Future Improvement

- **Enhanced Data Balancing:**
  - Implement SMOTE to improve class balance and address model sensitivity
- **Feature Engineering:**
  - Explore additional relevant features and interactions to boost model performance
- **Advanced Modeling:**
  - Consider integrating Gradient Boosting or ensemble methods
- **Cross-Validation:**
  - Employ k-fold cross-validation to reduce overfitting
- **Binary Classification Exploration:**
  - Investigate binary classification models to potentially improve differentiation between Hepatitis C positive and negative cases

CONCLUSION

- **Successful Model Development:**
  - Model effectively predicts the presence and stages of Hep C
- **Clinical Implications:**
  - Identified key liver panel tests that can aid in early detection and staging, potentially guiding more targeted screenings
- **Impact Potential:**
  - Streamlines the diagnostic process, saving time and money for both medical providers and patients
- **Path Forward:**
- Future improvements and additional data could further enhance model performance and clinical utility

www.github.com/Naomi-Lopez

THANK YOU