

# Machine Learning Methods to Identify Hepatitis C Risk Factors in Blood Donors Based on Biochemical Indicators



Yu Chieh Cheng, Chien Hsiao, He Jin Chu - American University STAT 627 Final Project

### Introduction / Context:

- Early detection of Hepatitis C in blood donors through data analysis is crucial for identifying health risks.
- Accurate predictions enhance blood donation safety and reduce infection risks.
- Guaranteeing risk-free blood tested in secure conditions is essential for public health.

### Approach:

- Explored five different methods for classification, utilizing cross-validation techniques to identify the most suitable model.
- Feature selection was employed using statistical correlation to predict biochemical indicators, ensuring a targeted analytical approach.
- The predictive performance of the model was assessed using a series of metrics, including accuracy.

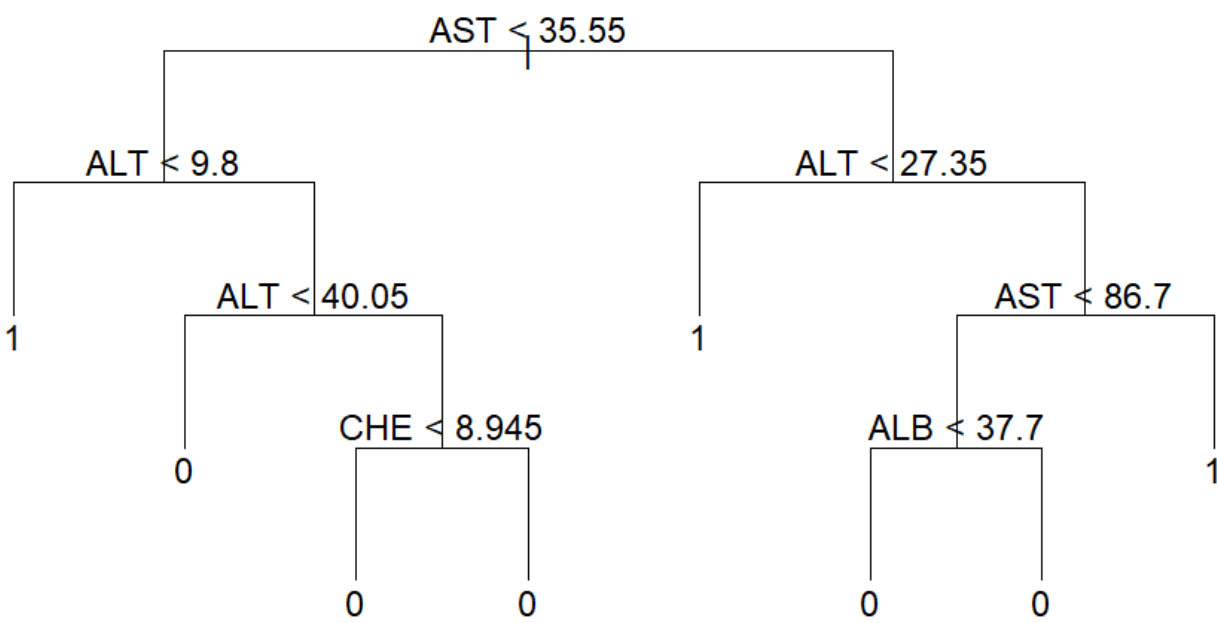
### Reference:

Lichtinghagen, R., Klawonn, F., & Hoffmann, G. (2020). HCV data [Data set]. UCI Machine Learning Repository. <https://doi.org/10.24432/C5D612>

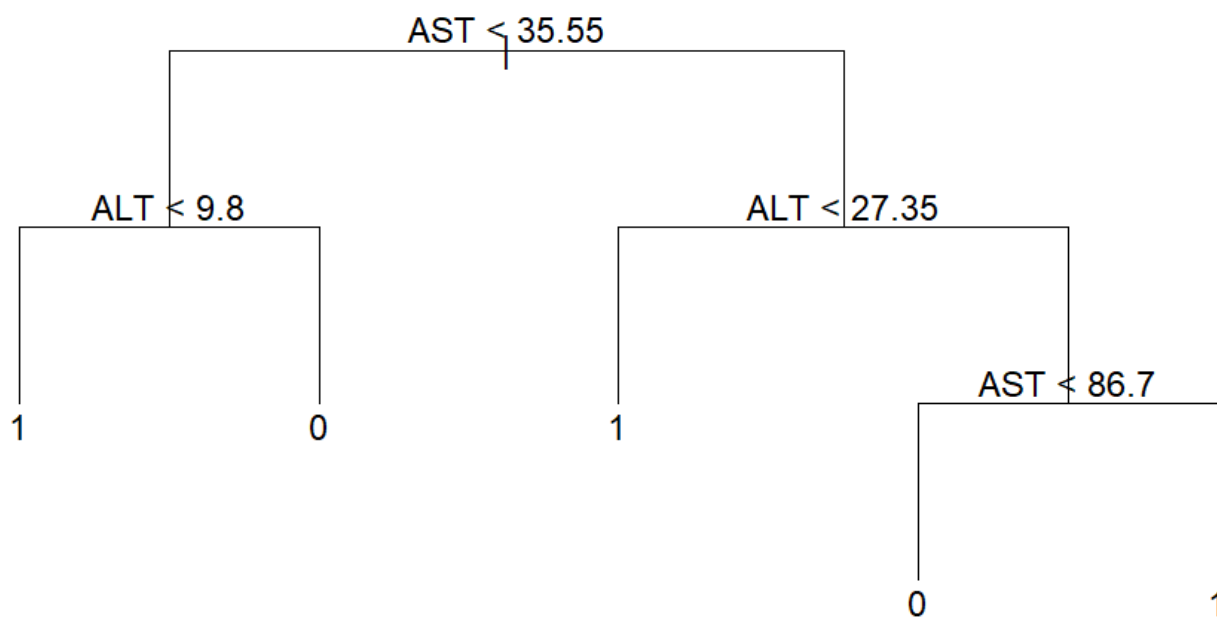
### Data:

- Source: UC Irvine Machine Learning Repository
- Size/Scale: 615\*14

	Methods	Accuracy Rate	Parameters
Status	Logistic Regression	93.22%	8
	LDA	95.42%	13
	Classification Tree	97.03%	2
	Random Forest	97.88%	3
	Support Vector Machine	98.98%	13



# Before Prune



# After Prune

### Results:

- The SVM method is the best fit for HCV data.
- SVM accuracy rate is the highest among all of the methods.
- Finding more data to enhance the reliability of the results.

### Assumptions/Limitations/Challenges:

- Case-wise deletion for handling missing values may lead to the loss of important information and potentially introduce non-random bias.
- The assumption is that the dataset has no systematic errors or biases, ensuring the biochemical data reliably represents the donors' health.

### Ethical Implications:

- Case-wise deletion removes any record with missing data, which can introduce bias and affect result reliability.

### Stakeholders:

- Blood Donors
- Donation Centers
- Public Health Authorities