# Hackatón BDM 2020

Cavasin, Rapaport

## Dataset elegido: HCV data

- Dataset con valores de laboratorio de 615 análisis de sangre
- La muestra se compone de donantes de sangre (y donantes sospechados) y pacientes con hepatitis C (solo hepatitis, Fibrosis, Cirrosis).
- Variable objetivo: "Categoría"

# Objetivo de KDD

 Dado un análisis de sangre con las variables del dataset, determinar si un paciente es donante, o una persona enferma de Hepatitis C en cualquiera de sus formas.

# 1° Paso: Preprocesamiento

- Eliminación de la columna 'unnamed: 0' por tratarse de un índice innecesario.
- 2. Determinar y especificar la cantidad de valores faltantes.
  - a. Faltantes: ['ALB', 'ALP', 'ALT', 'CHOL', 'PROT']
  - b. Imputación por hot-deck en dichas variables mediante
    KNNImputer
- 3. Cambiar la variable "sex" a numérica.

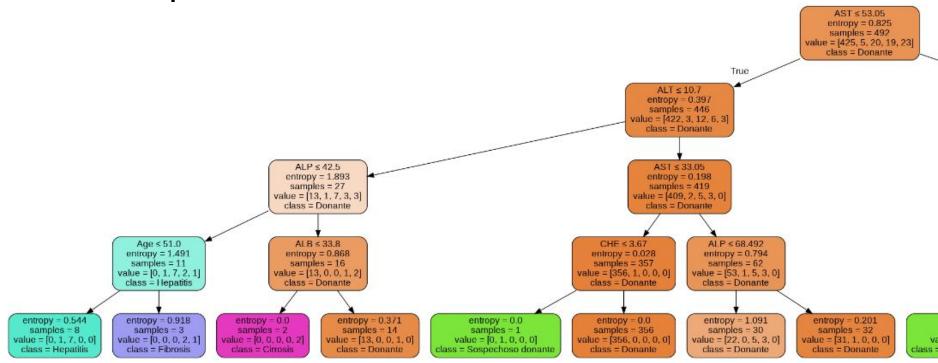
## 2° Construcción del árbol

- 1. División del dataset en datos de training y testing. Proporción 80-20.
- 2. Entrenamiento del árbol con datos de training

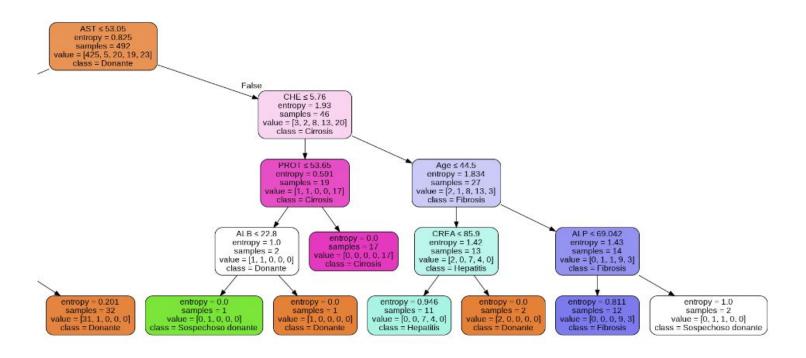
## 3° Evaluación del modelo

1. Evaluamos la precisión en las predicciones con SKLearn

## Rama izquierda del árbol



#### Rama derecha del árbol



#### Conclusiones

- Obtuvimos un árbol con 4 niveles con una precisión que ronda el 91%.
- Un paciente con AST menor a 53.05 y ALT mayor a 10.7 es, con una alta probabilidad, un donante de sangre.
  - Si su ALT es menor que 10,7 pero su ALP es mayor a 42.5, también tiene altas probabilidades de ser donante. Sin embargo, si su ALB está por debajo de 33.8, es probable que tenga cirrosis.
- Pacientes con AST menor a 53.05, ALT menor a 10.7, y ALP inferior a 42.5, son personas con Hepatitis C. Si son jóvenes, sólo tienen Hepatitis. Los mayores, han desarrollado Fibrosis.

### Conclusiones

- Para las observaciones con AST mayor a 53.05, tenemos que:
  - Tienen muchas posibilidades de tener hepatitis C en alguna de sus formas.
  - A excepción de los que poseen un PROT menor a 53.65 y un ALB mayor a 22.8, casos muy raros.
  - Los menores a 44 años y con CREA superior a 85.9, suelen ser donantes.