

**Universidad Tecnológica Centroamericana**

**UNITEC**

“Mini Proyecto#2 ”

Sistemas Inteligentes

Catedrático:  Kenny Mauricio Dávila Castellanos

Alumnos:

Martín José Pérez Gálvez 11711267

Bryan Manrique Amador Mena 11711211

12 de diciembre de 2020

# Introducción

Un árbol de decisión es un modelo de predicción que dado un conjunto de datos se crean construcciones lógicas muy similares a los sistemas de predicción basados en reglas, que ayudan a categorizar una serie de condiciones para la resolución de un problema.

En este reporte se muestran implementación, resultados y conclusiones del mini proyecto Arboles de Decisión que consiste en primero en crear un árbol de decisión a partir de datos entrenamiento, después crear una versión podada del árbol generado por medio del entrenamiento y por último usar unos de los 2 árboles para generar una matriz de confusión y el Accuracy a partir de datos de prueba.

# Implementación

## Librerías

1. pandas
2. numpy
3. matplotlib.pyplot
4. seaborn
5. sklearn.model\_selection (KFold)
6. sklearn.preprocessing (StandardScaler)
7. sklearn.metrics (f1\_score, confusion\_matrix)
8. sklearn.svm (SVC)
9. sklearn.ensemble (RandomForestClassifier)
10. sklearn.naive\_bayes (GaussianNB, BernoulliNB, CategoricalNB)
11. sys
12. pickle

## Scripts ejecución

1. mp\_2\_parte\_1.py dataset
2. mp\_2\_parte\_2\_cross.py dataset
3. mp\_2\_parte\_2\_train.py dataset archivosalida
4. mp\_2\_parte\_2\_test.py dataset archivoclasificador.pickle
5. mp\_2\_parte\_3\_cross.py dataset
6. mp\_2\_parte\_3\_train.py dataset archivosalida
7. mp\_2\_parte\_3\_test.py dataset archivoclasificador.pickle
8. mp\_2\_parte\_4\_cross.py dataset
9. mp\_2\_parte\_4\_ train.py dataset tipoclasificador archivosalida
10. mp\_2\_parte\_2\_test.py dataset archivoclasificador.pickle

# Resultados y Estadísticas Parte 1

## Datos Discretos

1. sexo



1. dias\_fiebre



1. dias\_ultima\_fiebre



1. nauseas



1. vomitos



1. rash



1. mialgas



1. artralgias



1. prueba\_torniquete



1. dolor\_abdominal



1. acumulación\_fluidos



1. sangrado\_mucosas



1. hemorragia



1. shock



1. letargia



1. irritabilidad

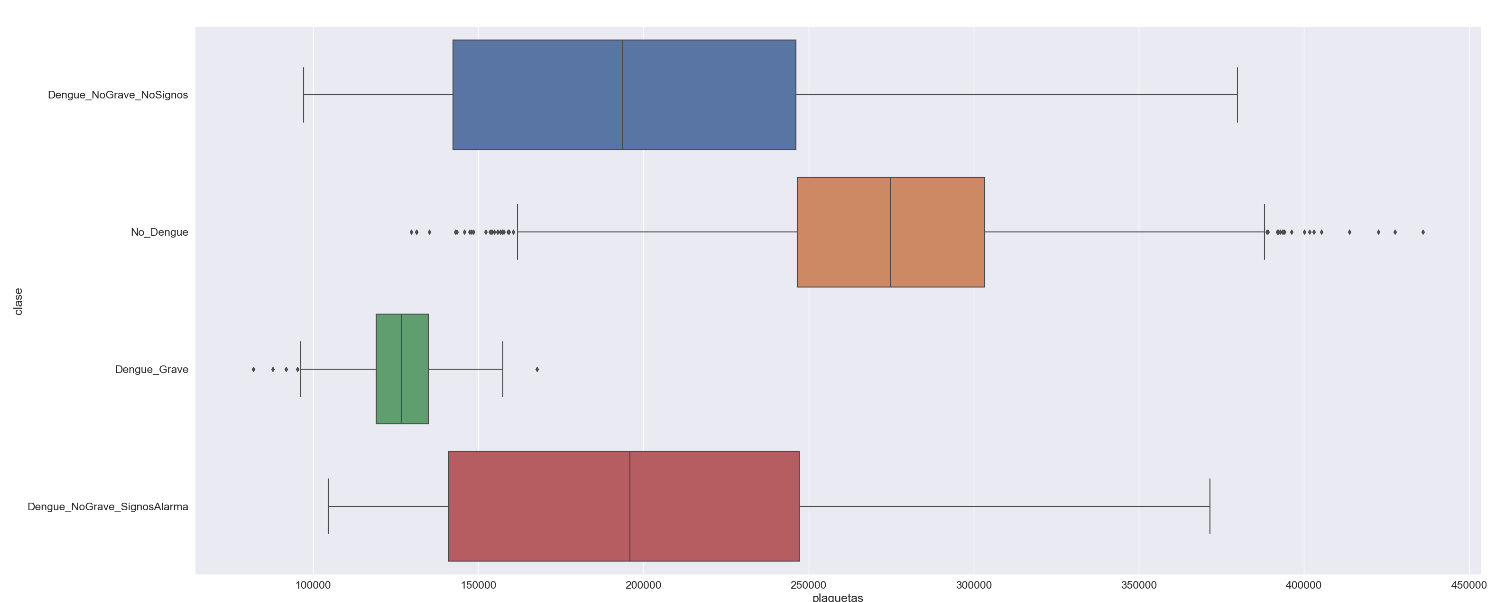


1. hepatomegalia

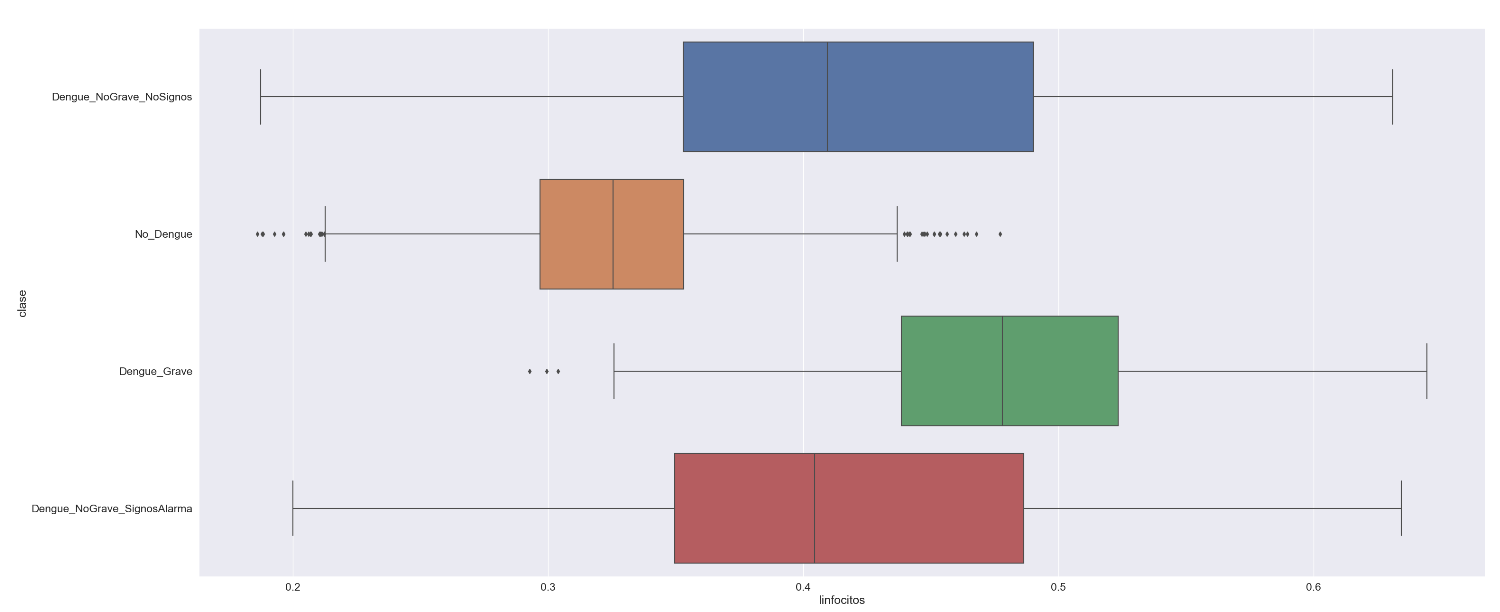


## Datos Continuos

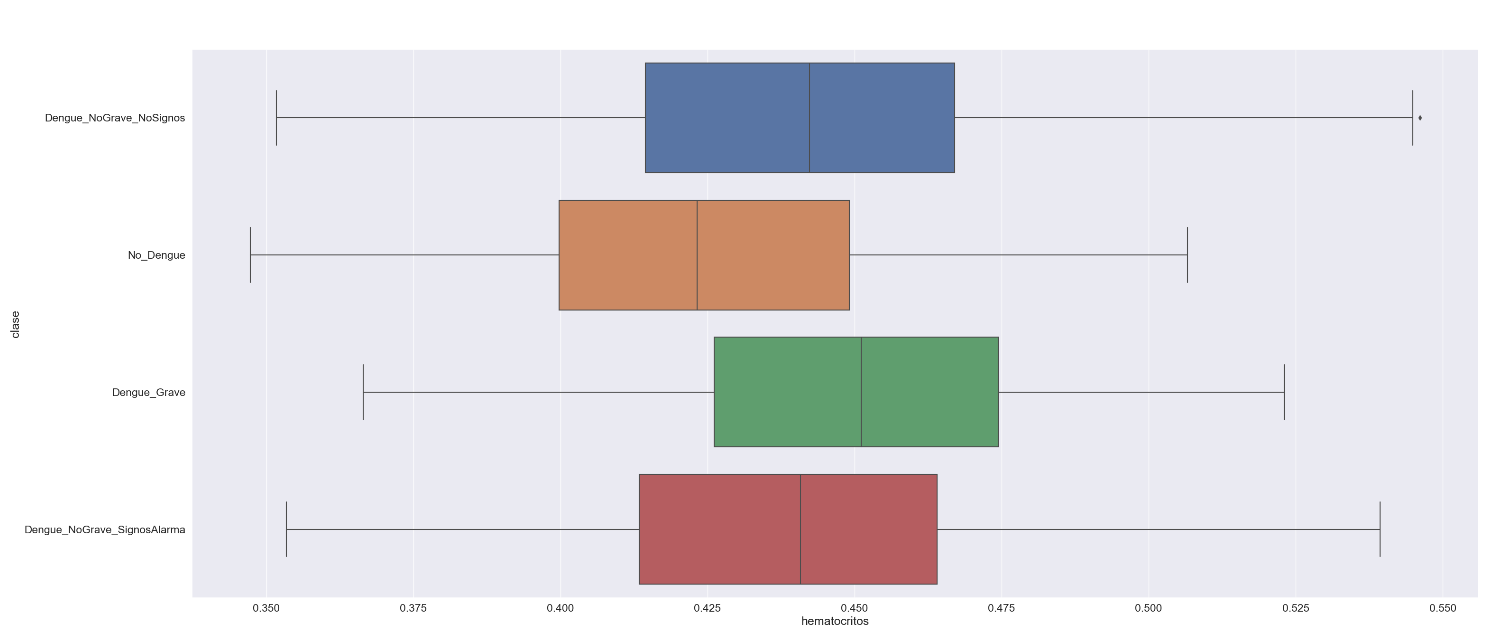
1. Plaquetas



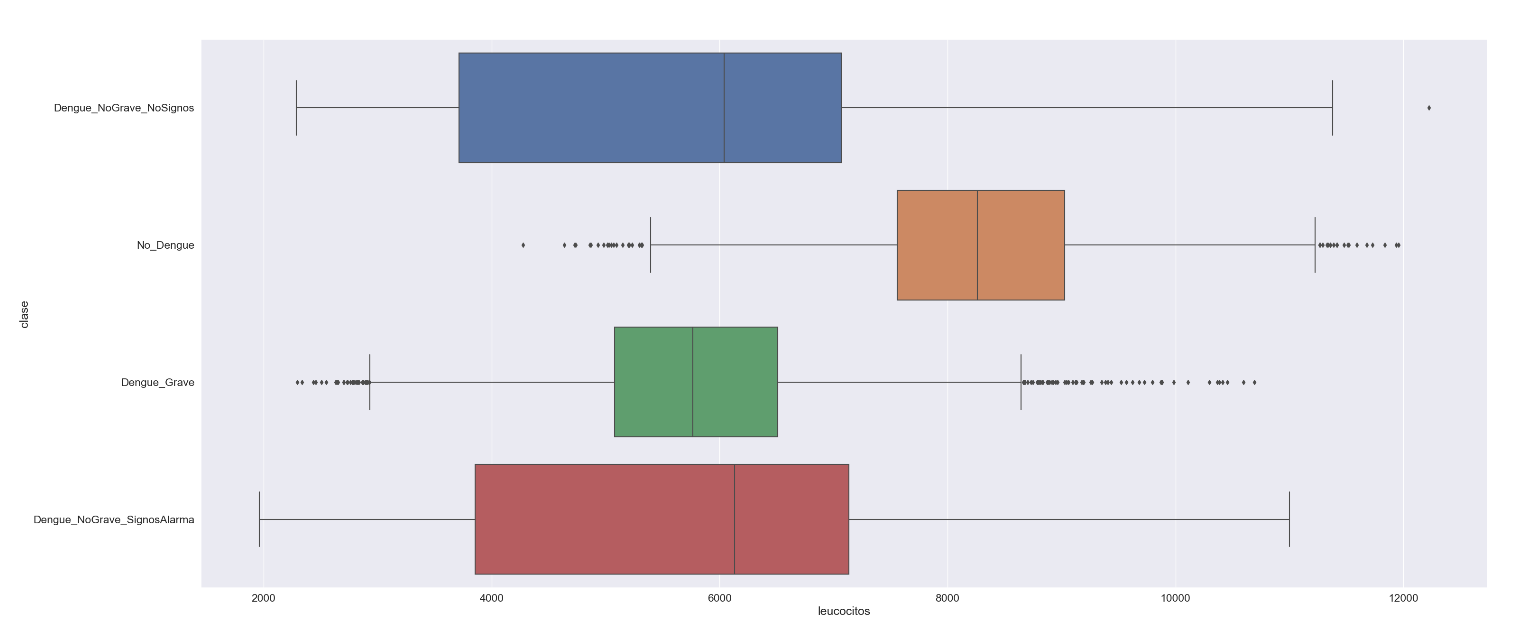
1. Linfocitos



1. Hematocritos



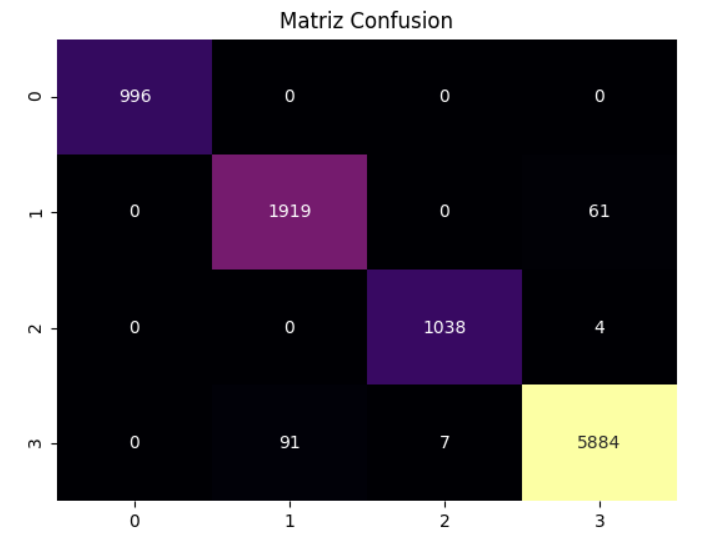
1. Leucocitos



## Resultados y Estadísticas Parte 2

1. **Clínica**





F1 1 : 1.0

F1 2 : 0.9619047619047619

F1 3 : 0.9947292764734068

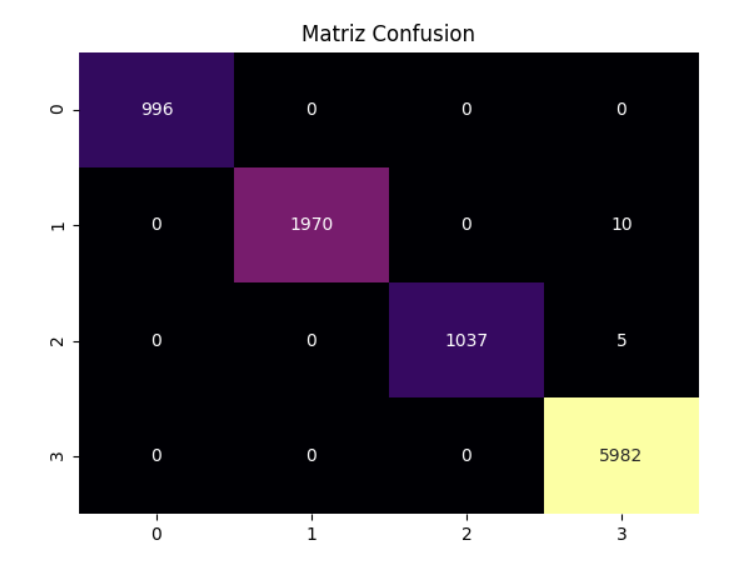
F1 4 : 0.9863381108037884

F1 total: 0.9857430372954893

Efectividad: 98.37%

1. **Laboratorio**





F1 1 : 1.0

F1 2 : 0.9974683544303797

F1 3 : 0.9975949975949976

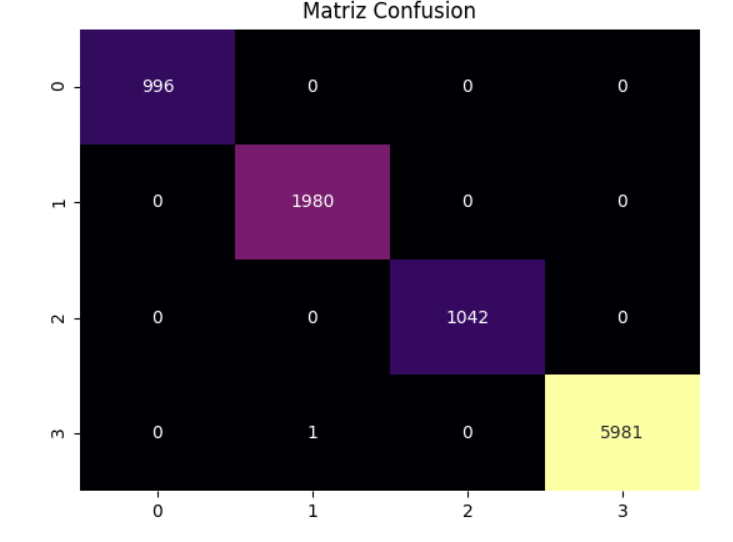
F1 4 : 0.998747808665164

F1 total: 0.9984527901726353

Efectividad: 99.85000000000001%

1. **Completo**





F1 1 : 1.0

F1 2 : 0.9997475385003786

F1 3 : 1.0

F1 4 : 0.9999164089275265

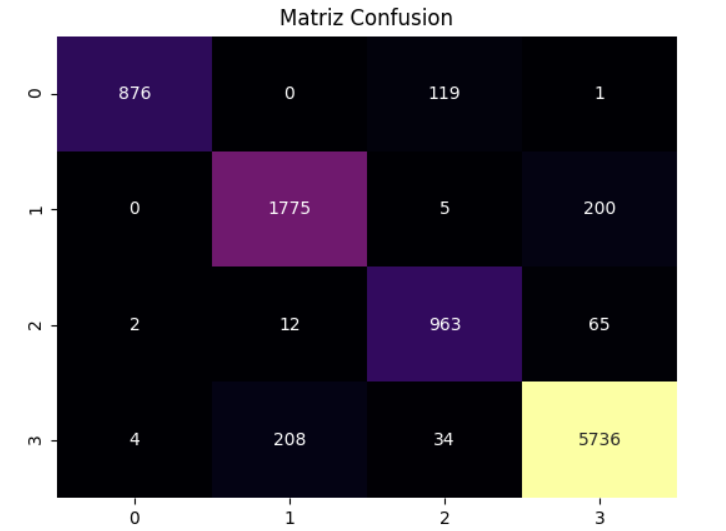
F1 total: 0.9999159868569762

Efectividad: 99.99%

## Resultados y Estadísticas Parte 3

## Resultados y Estadísticas Parte 4

Completo



## Análisis resumido

## Dificultades Encontradas

## Conclusiones