

**Sebastián Barrera A00358271**

**Santiago Hurtado A00362570**

**Miguel Sarasti A00364978**

**Sebastián Morales A00365920**

A continuación, se encontrará con una breve introducción al experimento sobre árboles de decisión.

**Árboles de decisión:** Un árbol de decisión **es un modelo predictivo** que divide el espacio de los predictores agrupando observaciones con valores similares para la variable respuesta o dependiente.

Para dividir el espacio muestral en sub-regiones es preciso aplicar una serie de reglas o decisiones, para que cada sub-región contenga la mayor proporción posible de individuos de una de las poblaciones.

Si una sub-región contiene datos de diferentes clases, se subdivide en regiones más pequeñas hasta fragmentar el espacio en sub-regiones menores que integran datos de la misma clase.

## **Desarrollo del experimento**

**Objetivo del experimento:** Determinar cual implementación entre un árbol de decisión desarrollado por nosotros y el árbol de decisión de la librería de **Accord.NET**

**Framework** tiene mayor precisión para predecir la variable objetivo del dataset.

**Unidad Experimental:** La unidad experimental de este experimento son las estructuras de los árboles de decisión tanto el implementado por nosotros y la librería utilizada. Estos serán desarrollados en C# para evaluar su rendimiento, junto a sus pruebas que tendrán factores como el tamaño del dataset que se va a predecir y el accuracy. Con el fin de determinar cuál tiene mejor desempeño.

**Variable(s) de Respuesta:** La variable de respuesta que determinamos para realizar el experimento es el porcentaje de predicción de los datos por cada algoritmo al ejecutarse, cuando se ingrese el dataset.

**Factores controlables:** Son variables de proceso o características de los materiales experimentales que se pueden fijar en un nivel dado.

- El tipo de árbol de decisión, tanto el de la librería o el implementado por nosotros.
- Cantidad de datos del dataset.
- Datos del dataset de:
  - ❖ **Validación**
  - ❖ **Testeo**
  - ❖ **Entrenamiento**
- La RAM del computador donde se ejecuta el algoritmo (2GB, 4GB, 8GB, 16GB, etc.) es un factor que podemos controlar porque podemos escoger computadores con un tamaño de RAM específico o incluso limitar el uso de esta, para el proceso del algoritmo.
- Procesador del computador donde se ejecuta el algoritmo.
- Sistema Operativo.
- Cantidad de aplicaciones que se están ejecutando en el computador mientras se ejecuta el algoritmo.

**Factores no controlables:** Son factores que no podemos controlar en el experimento.

- La implementación del algoritmo de la librería utilizada.
- La cantidad de procesos que se están ejecutando en el computador mientras se ejecuta el algoritmo.
- El tiempo de entrenamiento del algoritmo.
- La cantidad de RAM que utiliza los algoritmos en su ejecución.

**Factores estudiados:**

Los factores que vamos a estudiar en este experimento son los que influyen directamente en la precisión de la variable objetivo de las estructuras de árboles de decisión utilizadas.

- La implementación del árbol de decisión a utilizar para predecir los datos.
- Cantidad de datos del dataset

**Niveles****Tipo de implementación:**

1. Nuestro árbol de decisión
2. Librería Accord.NET arbol de decision

**Cantidad:**

1. 100 datos
2. 200 datos
3. 300 datos

Tipo de implementación	Nivel del tipo de implementación
Nuestro árbol de decisión	1
Librería Accord.NET árbol de decisión	2

Cantidad	Nivel de cantidad
100	1
200	2
300	3

## Tratamientos

<b>Tipo de implementación</b>	<b>Cantidad de datos</b>	<b>Tratamiento</b>	<b>Repetición</b>	<b>% Precisión Promedio</b>
Nuestro árbol de decisión	<b>100</b>	<b>1</b>	<b>50</b>	<b>74 %</b>
Nuestro árbol de decisión	<b>200</b>	<b>2</b>	<b>50</b>	<b>79 %</b>
Nuestro árbol de decisión	<b>300</b>	<b>3</b>	<b>50</b>	<b>76 %</b>
Librería Accord.NET árbol de decisión	<b>100</b>	<b>4</b>	<b>50</b>	<b>74 %</b>
Librería Accord.NET árbol de decisión	<b>200</b>	<b>5</b>	<b>50</b>	<b>74 %</b>
Librería Accord.NET árbol de decisión	<b>300</b>	<b>6</b>	<b>50</b>	<b>70 %</b>

## Conclusiones

Las conclusiones que llegamos después de realizar un análisis estadístico de los datos son las siguientes:

Para realizar el experimento tuvimos como variables de estudio la cantidad de datos para predecir y el tipo de árbol de decisión usado, comparando la precisión de cada repetición, en base a esto con los resultados que nos arrojó el experimento y usando la técnica de ANOVA, nos dio como resultado que para una gran cantidad de datos a predecir es más preciso el árbol de decisión implementado por nosotros que el de la librería, no obstante esto ocurrió cuando evaluamos una entrada pequeña, mediana y todo el dataset.

Otra conclusión que observamos es que el árbol de decisión de la librería siempre tiene el mismo porcentaje de precisión para todos sus intentos, volviéndose constante.

La varianza del árbol de decisión de la librería es mucho menor a la del árbol implementado por nosotros, haciendo que su porcentaje de acierto sea más constante, pero nuestro árbol, aún teniendo una mayor varianza, su precisión/acuracy es más favorable que el de la librería.

Observamos que cuando el algoritmo de la librería realiza la predicción sobre una gran cantidad de datos, el porcentaje de precisión tiende a disminuir, en comparación con menores cantidades de datos.

Existe un experimento atípico, porque normalmente el árbol de decisión nuestro tiende a mejorar su precisión cuando aumenta la cantidad de datos en el experimento, pero cuando hay una cantidad de datos mediana la precisión/accuracy tiene un pico, dado que llega a una precisión promedio de casi un 80%.

Analizamos que el árbol de decisión implementado por nosotros y el árbol de la librería, cuando realizan una predicción a una pequeña cantidad de datos, tienen medias iguales.