*Simple Random Forest Algorithm*

Tabla de contenido

[**1.** **Resumen** 2](#_Toc12363303)

[2. Introducción 3](#_Toc12363304)

[Referencias 4](#_Toc12363305)

[Figuras 4](#_Toc12363306)

|  |  |
| --- | --- |
| Raimundo Flores Cabra  *dpto. Ciencias de la Computación e Inteligencia Artificial* *Universidad de Sevilla* Sevilla, España Correo electrónico UVUS: [raiflocab@alum.us.es](mailto:raiflocab@alum.us.es) Correo electrónico de contacto: [raimundokarate98@gmail.com](mailto:raimundokarate98@gmail.com) | Vicente López Vázquez *dpto. Ciencias de la Computación e Inteligencia Artificial* *Universidad de Sevilla* Sevilla, España Correo electrónico UVUS: [viclopvaz1@alum.us.es](mailto:viclopvaz1@alum.us.es) Correo electrónico de contacto: [vlopezvazquez3@gmail.com](mailto:vlopezvazquez3@gmail.com) |

# **Resumen**

*El objetivo de nuestro trabajo ha sido construir una versión simplificada del algoritmo de Bosques Aleatorios. Pero ¿en qué consiste este algoritmo? Un bosque aleatorio es un conjunto de árboles de decisión entrenados a partir de un conjunto de datos de entrenamiento con un conjunto aleatorio de atributos seleccionados para cada uno y tras la aplicación de Bootstrapping como técnica de muestreo. Esto nos permite construir un conjunto de árboles de decisión partiendo de un solo conjunto de datos de entrenamiento, ya que dichos datos son difíciles de conseguir y no muy abundantes. Por esta razón, se aplican técnicas de muestreo, en este caso Bootstrapping con reemplazo, para dar lugar a un nuevo conjunto de datos para cada árbol de decisión.*

*Esta técnica facilita el desarrollo de sistemas de predicción fiables y con mayor capacidad de predicción que un solo árbol de decisión. Debemos destacar que, a lo largo de este proyecto hemos comprendido que los datos con los que trabajamos no siempre son suficientes y tenemos que adaptarnos a ellos; que las técnicas de muestreo nos facilitan el trabajo con conjuntos de datos pequeños y con la construcción de varios árboles de decisión; y que un solo árbol no es tan eficaz como un conjunto promediado de ellos.*

# Introducción

La Inteligencia Artificial [1] tiene numerosos usos actualmente, desde sistemas de planificación automática o sistemas de decisión, a reconocimiento de la escritura o del habla. Nuestro caso trata sobre los árboles de decisión [2], un ejemplo de Aprendizaje Supervisado [3].

Un árbol de decisión es un modelo de predicción construido a partir de un conjunto de datos, de donde se extraen una serie de reglas, que sirven para representar y categorizar una serie de condiciones que ocurren de manera sucesiva, para la resolución del problema. En la siguiente figura podemos ver la estructura de un árbol de decisión:

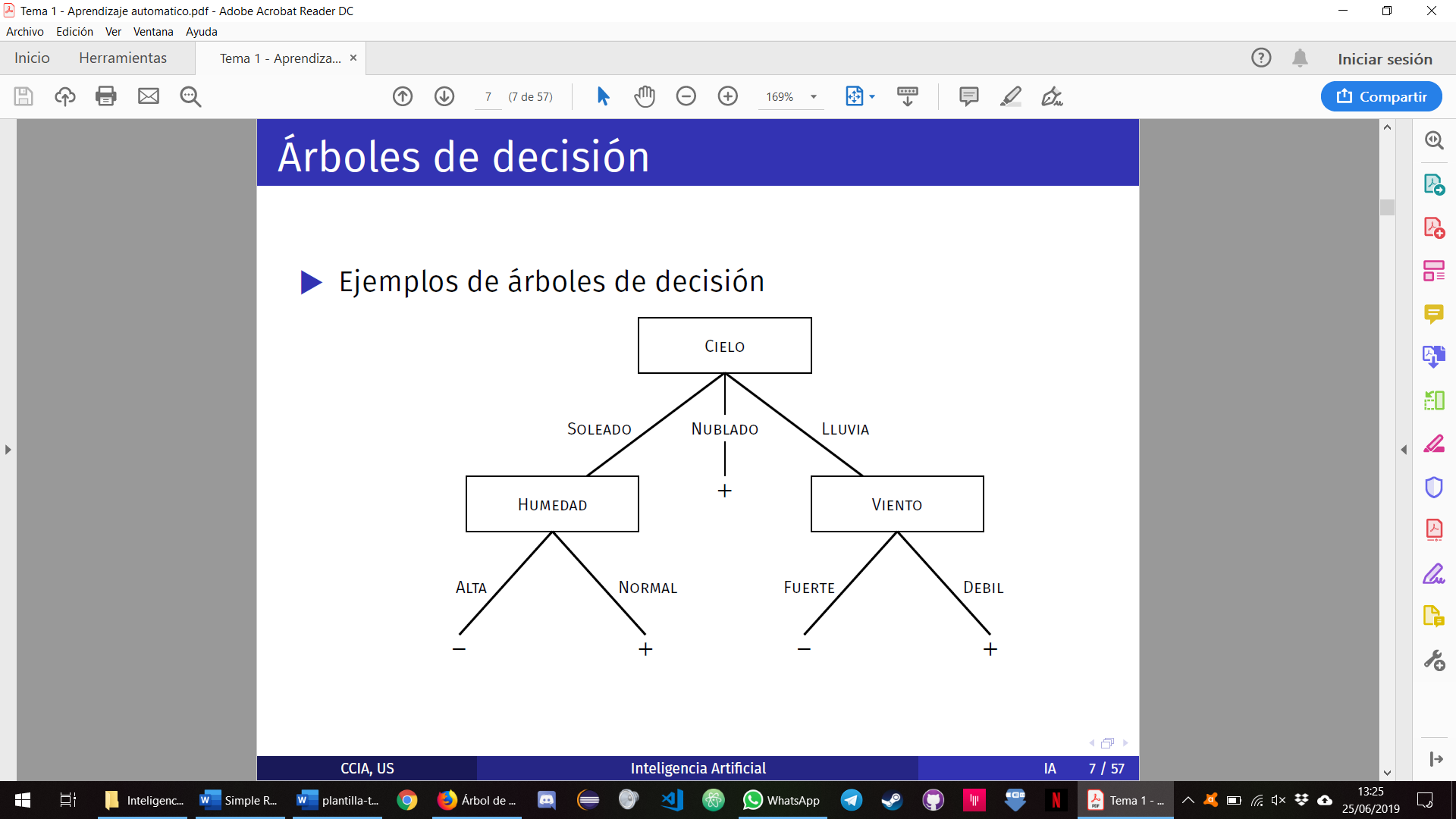
Podemos ver que el árbol clasificará los distintos ejemplos teniendo en cuenta como primera forma de clasificación el atributo “Cielo”. Dependiendo del valor de dicho atributo, el siguiente atributo usado para clasificar el ejemplo será “Humedad” o “Viento” o en caso de que el valor de “Cielo” sea nublado la clasificación es directa.

Figura 1. Árbol de decisión con 3 nodos de decisión.

A grandes rasgos, nuestro proyecto trabaja con dichos árboles de decisión. Pero se complica al introducir técnicas de muestreo como Bootstrapping [4] y Random Subspace Method [5]. Además, trabajaremos con varios árboles con el objetivo de aumentar la precisión calculando la salida promedia. La aplicación de técnicas de muestreo para obtener nuevos conjuntos de datos a partir de uno y el uso de varios árboles entrenados con diferentes conjuntos de datos da lugar a una técnica llamada Random Forest (Bosque Aleatorio) [6].

El uso de Random Forests aumenta la precisión que el uso de un solo árbol de decisión proporciona y facilita el trabajo con conjuntos de datos pequeños o difíciles de conseguir. Por eso su uso está muy extendido y aceptado.

Como conclusión, nuestro proyecto ha consistido en diseñar y construir una forma simple del algoritmo del Random Forest, compararlo con el uso de árboles individuales y profundizar en el uso de las técnicas de muestreo arriba citadas.

# Referencias

1. <https://es.wikipedia.org/wiki/Inteligencia_artificial>
2. <https://es.wikipedia.org/wiki/%C3%81rbol_de_decisi%C3%B3n>
3. <https://es.wikipedia.org/wiki/Aprendizaje_supervisado>
4. <https://es.wikipedia.org/wiki/Bootstrapping_(estad%C3%ADstica)>
5. <https://en.wikipedia.org/wiki/Random_subspace_method>

1. <https://es.wikipedia.org/wiki/Random_forest>

# Figuras

1. *Tema 1 – Aprendizaje automático. dpto. Ciencias de la Computación e Inteligencia Artificial Universidad de Sevilla. Sevilla, España.*