

## TÍTULO (DESCRIPCIÓN CORTA DEL PROYECTO. ENTRE 8 Y 12 PALABRAS)

Juan José Gabriel Cañón Díaz  
Universidad Eafit  
Colombia  
jjcanond@eafit.edu.co

Mauricio Toro  
Universidad Eafit  
Colombia  
mtorobe@eafit.edu.co

### 1. RESUMEN

El objetivo de esta entrega es encontrar las posibles problemáticas del proyecto el cual es predecir si la persona tendrá o no éxito académicamente a nivel superior; una vez encontradas las problemáticas buscar las posibles soluciones mas eficientes o mas factibles, ya que la idea principal del proyecto es predecir con mayor precisión posible mediante una serie de datos que recibirá la máquina, y así, en base al historial, descubra cual es el factor mas grande que influye si la persona tendrá éxito o no.

### 2. Palabras clave

Información

### 3. Palabras clave de la clasificación de la ACM

### 4. INTRODUCCIÓN

En Colombia las pruebas Icfes miden el nivel académico de cada estudiante de bachillerato que llega al grado 11, recolectando así información social, económica, demográfica y familiar de todos los estudiantes del país que las presentan, de esta forma con la basta información recolectada se puede hacer infinidad de predicciones respecto al futuro académico de cada estudiante.

En este proyecto pretendemos predecir si el resultado que el estudiante obtendrá en las pruebas Saber Pro aplicada en los estudiantes universitarios del país con el mismo fin de evaluar conocimientos, estará por encima o por debajo del promedio.

Cabe resaltar que los estudiantes que están por encima del promedio (en ambas pruebas) ocasionalmente reciben becas o ayudas de diferentes entidades del estado o privadas, asegurando una mejor calidad de vida.

### 5. PROBLEMA

Lograr predecir cual es la probabilidad de éxito de una persona en base a sus datos y descubrir cual es el factor que mas afecta este resultado de si tendrá éxito conforme aprende la maquina.

### 6. PROBLEMAS RELACIONADOS

1. Problema overfitting y underfitting: Este problema consiste que el programa es entrenado con un número de datos limitado, el cual le ayuda a aprender conforme va leyendo los datos, sin embargo, la máquina solo aprende en base a los datos con los que ya ha interactuado y al recibir un dato diferente, no lo reconoce, por ejemplo,

- i. • Underfitting: se entrena la máquina con un solo tipo de planta, al recibir otro tipo de planta, no la reconocerá como planta.
- ii. • Overfitting: se entrena con varios tipos de planta con cualidades semejantes, al recibir un tipo de planta que no tenga ningún atributo semejante a las que ya reconoce, no la reconocerá como una planta [1]

2. problema de predicción de series temporales: este problema se basa en que las respuestas se basan en los datos "x" sin embargo, puede haber datos independientes que no tengan relación con el resultado, un ejemplo claro de este sería el salario de un trabajador y el peso de una persona. sin embargo, el programa lee toda la información teniendo en cuenta estos datos independientes, generando un error de predicción conforme a la relación de los datos ya que este tiene en cuenta todos los datos entregados Solución: usar el método de clasificación, el cual consiste en leer todos los datos y seleccionar los datos no independientes enviarlos al programa de predicción para obtener un resultado más cercano al deseado.[2]

## 7. DISEÑO DE LA ESTRUCTURA DE DATOS

1.

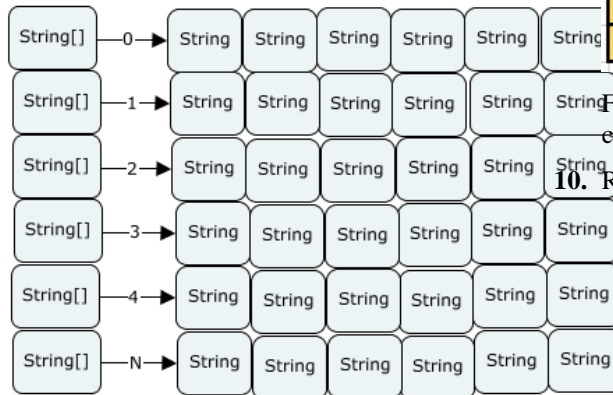


Figura 1: Estructura de la matriz que guarda los datos que son extraídos del archivo cvs.

2.

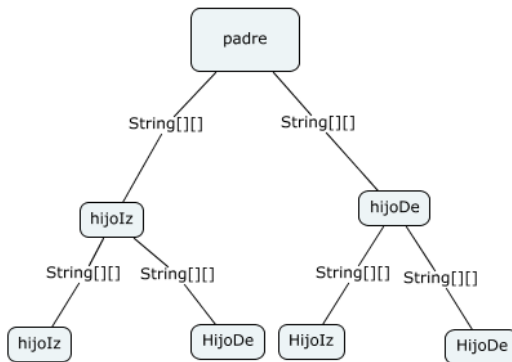


Figura 2: Estructura del árbol que ayuda al aprendizaje de los datos.

## 8. MEMORIA

	MEMORIA(MB)
conjunto de datos 1 (5000)	70
conjunto de datos 2 (15000)	245
conjunto de datos 3 (25000)	323

Figura 3: tabla de memoria consumida por tres diferentes conjuntos de datos

## 9. TIEMPO

	tiempo de ejecucion (s)
conjunto de datos 1 (5000)	228
conjunto de datos 2 (15000)	1524
conjunto de datos 3 (25000)	2584

Figura 4: tabla de tiempo que dura al correr el código con sus respectivos datos.

## 10. REFERENCIAS

1. Na8, Que es overfitting y underfitting y cómo solucionarlo, diciembre 12 de 2017, Aprende Machine Learning antes de que sea demasiado tarde, <https://www.aprendemachinlearning.com/que-esoverfitting-y-underfitting-y-como-solucionarlo/>
2. Álvaro G., Problemas Comunes En Aprendizaje automático, 10 de Mayo de 2018, <https://machinelearningparatodos.com/problemas-comunes-en-aprendizaje-automatico/>