

## Sprint 11 – Trabajo Final

Allison Garces Castaño.

### **Presentación del conjunto de datos escogido:**

La base de datos elegida se basa en una prueba de estado, la intención con esta es procesar los datos y obtener información sobre como se agrupan las diferentes variables y determinar si un estudiante puede aprobar o no, en base a las características informadas por el.

También, estudiar los diferentes resultados obtenidos en general.

### **Características generales:**

Instituto Colombiano para la Evaluación de la Educación - ICFES (2023).

Datalcfes. Saber 11. <https://www.icfes.gov.co/web/guest/data-icfes>

La siguiente es una base de datos tomada del ICFES. En específico la prueba saber 11, que se realiza a estudiantes del grado 11. Que ya van a finalizar sus estudios de bachillerato, es un instrumento de evaluación estandarizada que mide oficialmente la calidad de la educación formal impartida a quienes terminan el nivel de educación media.

Esta base de datos incluye datos como resultados individuales por estudiante, puntaje global de la prueba y puntaje individual de inglés, matemáticas, competencias ciudadanas y lenguaje. A su vez también ofrece características demográficas, culturales y económicas de cada estudiante. Como estrato de la vivienda, trabajo de los padres, entre otros.

### **Definición de las variables:**

El conjunto de datos posee (20049, 82) siendo 82 las columnas y 20049 las filas. De las cuales, se eliminar aproximadamente la mitad de estas, ya que no realizan ningún aporte significativo al estudio.

La mayoría de las variables son categóricas, así que he procesado previamente volviendo la mayoría en datos numéricos, de la siguiente manera:

FAMI_ESTRATOVIVIENDA  Sin Estrato = 0 Estrato 1 = 1 Estrato 2 = 2 Estrato 3 = 3 Estrato 4 = 4 Estrato 5 = 5 Estrato 6 = 6	FAMI_PERSONASHOGAR  1 a 2 = 1 3 a 4 = 2 5 a 6 = 3 7 a 8 = 4 9 o más = 5	FAMI_CUARTOSHOGAR  Uno = 1 Dos = 2 Tres = 3 Cuatro = 4 Cinco = 5 Seis o mas = 6
FAMI_NUMLIBROS  0 A 10 LIBROS = 0 11 A 25 LIBROS = 1 26 A 100 LIBROS = 2 MAS DE 100 LIBROS = 3	FAMI_SITUACIONECONOMICA  Peor = 0 Igual = 1 Mejor = 2	ESTU_DEDICACIONLECTURADIARIA 30 minutos o menos = 0 Entre 30 y 60 minutos = 1 Entre 1 y 2 horas = 2 Más de 2 horas = 3 No leo por entretenimiento = 4
ESTU_DEDICACIONINTERNET 30 minutos o menos = 0 Entre 30 y 60 minutos = 1 Entre 1 y 3 horas = 2 Más de 3 horas = 3 No Navega Internet = 4	COLE_NATURALEZA  NO OFICIAL = 1 OFICIAL = 2	COLE_BILINGUE  N = 0 S = 1
COLE_CARACTER  NO APLICA = 0 ACADÉMICO = 1 TÉCNICO = 2 TÉCNICO/ACADÉMICO = 3	COLE_AREA_UBICACION  RURAL = 1 URBANO = 2	COLE_JORNADA  UNICA = 0 MAÑANA = 1 TARDE = 2 NOCHE = 3 SABATINA = 4 COMPLETA = 5
FAMI_TIENEINTERNET FAMI_TIENESERVICIOTV FAMI_TIENECOMPUTADOR FAMI_TIENELAVADORA FAMI_TIENEHORNOMICROOGAS FAMI_TIENEAUTOMOVIL FAMI_TIENEMOTOCICLETA FAMI_TIENECONSOLAVIDEOJUEGOS ESTU_TIENEETNIA  NO = 0 SI = 1	CALIFICACION GANO = 0 PERDIO = 1 REGULAR = 2	

Presentación de los objetivos:

La finalidad a la que quiero llegar es poder explorar los datos, comprender un poco como son los resultados finalidad de la prueba en base a las características explicadas anteriormente y si es posible llegar a determinar el resultado de un estudiante, si por ejemplo tiene un estrato social bajo o si no influye en nada para el resultado de la prueba.