

¿LA TASA DE DECERSION ES AFECTADA POR LA REPROBACION DE CURSOS?

Juan Felipe Roa Teuta

Facultad ingeniería de sistemas, Universidad cooperativa de Colombia, Colombia

Recibido: 10 de noviembre de 2023

RESUMEN

Se utilizó un conjunto de datos en el cual se trata distintos temas como lo es la deserción estudiantil, y la repetencia de cursos en los distintos departamentos de Colombia. Colombia tiene una tasa de deserción estudiantil del 10.68%, la cual hace varios años llegaba a ser superior alcanzando unos valores del 16.72%. para tratar de predecir la tasa de deserción en el futuro, se utilizó un modelo de machine learning basado en regresión lineal, el cual tendrá en cuenta un variable acerca de los estudiantes que reprobaron el año escolar. Con base a los resultados se podría decir que hay factores externos que ocasionan que los estudiantes se retiren de los colegios.

ABSTRACT

A data set is used that addresses different topics such as student dropout and repetition of courses in the different departments of Colombia. Colombia has a student dropout rate of 10.68%, which several years ago was higher, reaching values of 16.72%. To try to predict the dropout rate in the future, a machine learning model based on linear regression was used, which will consider a variable about students who failed the school year. Based on the results, it could be said that there are external factors that cause students to withdraw from schools.

1. INTRODUCCION

La deserción estudiantil representa un desafío significativo en el panorama educativo colombiano, afectando no solo la trayectoria académica individual, sino también teniendo repercusiones de largo alcance en el desarrollo social y económico del país. En este artículo científico, nos sumergimos en un análisis exhaustivo de los factores subyacentes que contribuyen a la deserción estudiantil en Colombia.

La integración de técnicas avanzadas de machine learning nos permite examinar patrones complejos y no lineales en grandes conjuntos de datos, desentrañando todos

Aquellos elementos educativos que influyen en las decisiones de los estudiantes a desertar sus estudios.

A través de este enfoque innovador, buscamos no solo comprender profundamente los desafíos específicos que enfrenta el sistema educativo colombiano, sino también proponer soluciones basadas en datos que pueden adaptarse dinámicamente a las cambiantes dinámicas educativas.

Para el desarrollo de este estudio se utilizará una metodología de minería de datos conocida como CRISP -DM. para así cumplir y verificar cada uno de los objetivos propuestos con un enfoque cuantitativo usando un diseño

correlacional.

2. METODOLOGIA

Se utiliza una metodología conocida como CRISP -DM, con la cual se busca realizar la comprensión del proyecto, y gracias a ello, se establecerán los objetivos del proyecto, además de que se evaluará la situación, de esa manera se podrán establecer los objetivos de la minería de datos. Gracias a lo anterior mencionado, se podrá obtener un conjunto de datos, en el cual se deberá realizar una limpieza para poder trabajar con los datos. Seguidamente se seleccionará alguna técnica de modelado en la que se deberá diseñar una evaluación del modelo y de esa manera se procederá a la construcción del modelo. Y, por último, se determinará si los resultados coinciden con los objetivos planteados inicialmente.

La investigación desarrollada se realizó con un enfoque cuantitativo debido a que no se basa en datos empíricos o revisiones documentales, sino más bien establece precedentes, compara grupos, categorías o clases, determina causas y efectos a partir de un estudio el cual permite desarrollar tecnología e innovaciones. Por esto, según Hernández Sampieri [1], La investigación cuantitativa se vincula a conteos numéricos, el cual se representa en un conjunto de procesos organizando de una manera secuencial para comprobar ciertas suposiciones. En esta investigación se realizará un diseño correlacional identificando las variables, las cuales son el sistema de detección de intrusos y las anomalías en la red.

3. ANALISIS DE DATOS

En el siguiente apartado se encontrará la descripción del conjunto de datos además del método que se empleo para realizar el análisis de los datos.

3.1. DESCRIPCION DE LOS DATOS

Para la realización de la investigación se empleará un conjunto de datos llamado “MEN_ESTADISTICAS_EN_EDUCACION_EN_P REESCOLAR, BÁSICA Y MEDIA_POR_DEPARTAMENTO”, el cual brinda una visión detallada de diversas métricas educativas en Colombia, centrándose en la población de estudiantes de 5 a 16 años. A través de los diferentes campos, se pueden analizar aspectos clave del sistema educativo en términos de matriculación, cobertura, deserción y rendimiento académico en diferentes niveles educativos y en distintas regiones del país.

Cabe recalcar que el conjunto de datos abarca varios años, lo que permite realizar análisis a lo largo del tiempo y evaluar posibles tendencias o cambios en el sistema educativo colombiano. Además, La información está desglosada por departamentos, lo que facilita el análisis de disparidades regionales en términos de acceso a la educación y desempeño académico. Por otra parte, podemos observar de manera general la población estudiantil entre los 5 y 16 años. Además de que nos permite conocer ciertos aspectos como la población estudiantil que está inscrita en instituciones educativas, Las tasas de deserción, aprobación, reprobación y repetición de cursos en los diferentes niveles educativos.

Considero, que Este conjunto de datos es valioso para los investigadores y responsables de políticas educativas, ya que les brinda la oportunidad de identificar patrones, evaluar el impacto de las intervenciones y diseñar estrategias más efectivas para abordar los desafíos educativos en Colombia.

3.2. DESCRIPCION DEL METODO USADO PARA EL ANALISIS DE LOS DATOS

Para realizar el análisis de los datos se empleó un método de machine learning conocido como regresión lineal el cual se utiliza para analizar la relación entre una variable dependiente (la variable que queremos predecir) y una o más variables independientes (las variables que se utilizan para hacer la predicción).

El método de regresión lineal se basa en la siguiente ecuación:

$$Y=\beta_0+\beta_1X+\varepsilon$$

- Y: Variable dependiente (la que queremos predecir).
- X: Variable independiente (la que usamos para hacer la predicción).
- β_0 : Término de intercepción (el valor de Y cuando X es cero).
- β_1 : Coeficiente de pendiente (representa la relación entre X e Y).
- ε : Término de error (representa la variabilidad no explicada).

El objetivo de la ecuación es minimizar la suma de los cuadrados de los errores (ε) para encontrar los valores óptimos de β_0 y β_1 .

4. RESULTADOS

En la figura 1 podemos observar en como afecta la reprobación del grado escolar con la deserción académica de los estudiantes y uno de los hechos que se pueden observar y es que en cierta parte al reprobar los grados académicos los estudiantes suelen desertar sus estudios

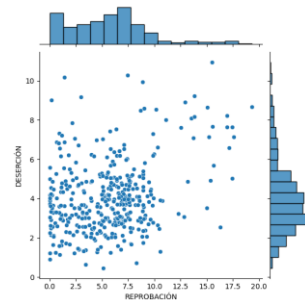


Figura 1 reprobación vs deserción

En la figura 2 podemos observar cómo afecta la reprobación de cursos en la repetición de estos, y gracias a ello podemos deducir que entre menos estudiantes reprueban sus grados académicos, estos suelen optar por repetir

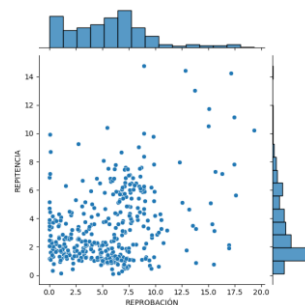


Figura 2 reprobación vs repetición

En la figura 3 podemos observar en como el departamento afecta la deserción de estudiantes

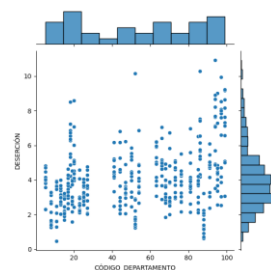


Figura 3 departamento vs repetición

Una vez entrenado el modelo con nuestro modelo de regresión lineal podemos observar como entre mas estudiantes pierden sus grados académicos, mayor es la tasa de deserción, sin embargo, gracias a esta técnica podemos conocer que la precisión de la predicción es del 16%, y gracias a este hecho se podría considerar que hay otros factores externos que pueden llegar a afectar la deserción de estudiantes.

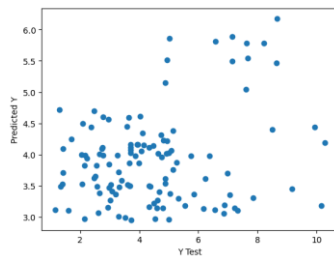


Figura 4: predicción modelo de machine learning

5. CONCLUSIONES

Al realizar el estudio pudimos darnos cuenta en como hay distintos factores que pueden afectar la deserción estudiantil como lo son la ubicación de las instituciones educativas, el estado de la conexión a internet, la cantidad de estudiantes que estas poseen.

Por otra parte, se puede decir que hay factores que no poseía el dataset los cuales tienen mas peso en la deserción de estudiante, esto se debe al poco porcentaje de correlación entre las distintas variables estudiadas.

Para próximas investigaciones recomendaría tener en cuenta otros factores como lo son el estrato al que pertenece el estudiante, su situación como persona vulnerable o desplazada, si tiene alguna condición física que le haya impedido continuar con sus estudios e incluso se podría llegar a estudiar de manera mas especifica cada uno de los

municipios, para de esta manera poder llegar a conocer algunos aspectos como si esta llega a tener algún tipo de conflicto armado y saber si el estudiante es víctima de conflicto.

6. REFERENCIAS

[1] R. Hernández Sampieri and C. P. Mendoza Torres, *Metodología de la investigación : las rutas cuantitativa, cualitativa y mixta*.