



Modelos de minería de datos aplicados al rendimiento académico universitario: Educación virtual durante pandemia COVID-19

Data mining models applied to university academic performance: Virtual education during COVID-19 pandemic

Jesús Eduardo Gamboa Unsihuay¹*, Arturo Zuñiga Blanco¹*

* Autor de correspondencia: jgamboa@lamolina.edu.pe

RESUMEN

En el 2020 se suscitó la pandemia por COVID-19, la cual trajo consigo un cambio en las rutinas de las personas hacia la virtualidad. La educación no ha sido ajena a estos cambios, los cuales han venido acompañados de un deterioro en la salud mental, manifestándose a través de la depresión, ansiedad, estrés, entre otros. Esta investigación buscó caracterizar a los estudiantes de cursos básicos de estadística en su primer ciclo de clases virtuales, así como establecer qué variables influyeron en su rendimiento académico. Se encontraron 3 conglomerados de estudiantes, diferenciados principalmente por la edad, historial académico y puntajes de depresión, ansiedad y estrés. Por otro lado, los atributos que mayor ganancia de información presentan al predecir el rendimiento académico fueron el promedio ponderado semestral del ciclo anterior, la asignatura en la que se matriculó, la cantidad de créditos matriculados, la edad, el estrés y la ansiedad.

Palabras clave: estudiantes universitarios, síntomas emocionales, análisis de conglomerados, árboles de clasificación, software R.

ABSTRACT

In 2020, the COVID-19 pandemic began, which brought with it a change in people's routines towards virtuality. Education has not been the exception to these changes, which have been accompanied by a deterioration in mental health, manifested through depression, anxiety, stress, among others. This research sought to characterize the students of basic statistics courses in their first cycle of virtual classes, as well as to establish which variables influenced their academic performance. Three clusters of students were found, differentiated mainly by age, academic record, and depression, anxiety, and stress scores. On the other hand, the attributes with the greatest gain

Forma de citar el artículo (Formato APA):

Gamboa, J., & Zuñiga, A. (2021). Modelos de minería de datos aplicados al rendimiento académico universitario: Educación virtual durante pandemia COVID-19. *Tierra Nuestra*. 15(1), 18-28. <http://dx.doi.org/10.21704/rtn.v15i1.1812>

¹ Universidad Nacional Agraria La Molina, 15024, Lima, Perú. jgamboa@lamolina.edu.pe; azuniga@lamolina.edu.pe

in information when predicting the academic performance were the weighted average of the previous semester, the enrolled course, the number of credits enrolled, age, stress and anxiety.

Keywords: university students, emotional symptoms, cluster analysis, classification trees, R software.

Introducción

En el año 2020 se suscitó la pandemia de la COVID – 19, declarada como tal el día 11 de marzo del año en mención (Organización Mundial de la Salud [OMS], 2020). Para entonces, los casos confirmados alrededor del mundo excedían apenas los 100 mil contagios y las muertes eran 4 mil (John Hopkins University, s.f.). No obstante, el SARS-CoV-2, un nuevo tipo de coronavirus causante de la COVID - 19, se esparció rápidamente a casi todos los países del mundo, lo cual empujó a los gobiernos a tomar diversas medidas, tales como el cierre de escuelas, universidades y centros de labores, y en casos más extremos los toques de queda y cuarentenas. Este nuevo contexto transformó la rutina de millones de personas de un modo presencial a uno remoto o virtual, siendo las actividades laborales y educativas dos de los ejes principales de cambio.

En cuanto a la realidad educativa en Perú, de acuerdo con lo señalado en la Ley Universitaria 30220, el régimen de estudios se puede dar en la modalidad presencial, semipresencial o a distancia. En particular, para los estudios de pregrado, esta última modalidad no puede ser mayor al 50% del total de créditos (Superintendencia Nacional de Educación Universitaria [SUNEDU], 2014). Es por este motivo que muchas instituciones educativas no se encuentran preparadas para el cambio en el modelo de enseñanza – aprendizaje presencial a uno totalmente virtual y a distancia, lo cual implica un gran reto para las autoridades universitarias, docentes y estudiantes pues cada uno de ellos tiene un rol importante en la adaptación al cambio.

En primer lugar, las autoridades son las encargadas de brindar las plataformas de aprendizaje sincrónico y asincrónico, así como de velar por el acceso masivo a estos sistemas por parte de la comunidad universitaria, mientras que los docentes tienen que adaptar sus procesos pedagógicos a entornos no presenciales, lo cual implica una constante capacitación en nuevas herramientas tecnológicas y técnicas pedagógicas, así

como una gestión más flexible del tiempo, distribuido entre sesiones sincrónicas y asincrónicas. Finalmente, los estudiantes vienen asumiendo el compromiso de una mayor autonomía (Vlachopoulos & Makri, 2019) y responsabilidad en su proceso de aprendizaje, en muchos casos a pesar de las limitaciones tecnológicas (Huanca et al., 2020). Es así que este contexto de pandemia puede mellar su rendimiento académico por diversos factores que inciden en la salud mental (Lozano Díaz et al., 2020).

El presente estudio busca identificar perfiles en estudiantes universitarios de cursos básicos de estadística en función a su rendimiento en dichos cursos, su historial académico, así como su nivel de depresión, ansiedad y estrés al inicio del primer semestre académico en modalidad virtual en el contexto de la pandemia de COVID-19, haciendo uso de modelos de minería de datos, tales como los árboles de clasificación y el análisis de conglomerados.

Marco teórico

Rendimiento académico en tiempos de pandemia

Al realizar estudios de rendimiento académico universitario, la carrera del estudiante es una variable que debe ser considerada en el análisis (Porcel et al., 2010). Esta afirmación también es respaldada por (Dicovski Riobóo & Pedroza, 2017), quienes añaden el sexo como una característica relevante.

En el contexto de la pandemia de COVID-19, (Al-kandari et al., 2021) señalaron que en diversos estudios se reporta estrés, miedo, ansiedad y depresión como síntomas de deterioro en la salud mental, las cuales podrían deberse al sistema de salud de cada país y también al sesgo de selección de los participantes voluntarios en los estudios. Por otro lado, (Lozano Díaz et al., 2020) mencionaron que el confinamiento ha causado un fuerte impacto en el ámbito psicológico y académico de los estudiantes universitarios y proponen la incorporación de la formación en resiliencia. Por su

parte, (Haider & Al-Salman, 2020) recopilaron datos en estudiantes universitarios de Jordania, los cuales revelaron que al menos el 80% de estos padecieron de estrés, frustración y depresión a causa de las medidas de confinamiento y cuarentena. Asimismo, de acuerdo con (Montalvo Romero & Montiel Rosales, 2020) el estrés en los alumnos universitarios se ha incrementado debido al COVID-19, lo cual se constituye en un factor externo y no controlable por los estudiantes, y que, ante ello, el personal educativo (en el cual se incluye a los docentes) debe apoyarlos para dar continuidad a su formación profesional.

En Perú, (Huarcaya-Victoria, 2020) hace mención al incremento de problemas psicológicos y mentales, tales como la depresión, la ansiedad y el estrés, debido al confinamiento. Asimismo, cita a (Qiu et al., 2020) y (Wang et al., 2020) quienes afirmaron que los estudiantes universitarios, junto a las mujeres y adultos mayores son grupos de alto riesgo de afectación a la salud mental. (Kaparounaki et al., 2020) señalaron también un aumento en el número de horas de sueño, pero con disminución de la calidad del mismo. También reportaron un incremento de pensamientos e intentos de suicidio.

Por otro lado, (Wang et al., 2020) afirmaron que las estudiantes con síntomas físicos (mialgia, mareos, catarro) tuvieron un mayor impacto psicológico debido al brote de COVID-19 generando altos niveles de estrés, ansiedad y depresión, siendo esta una evidencia de que el sexo es un factor relevante al momento de determinar el impacto de la COVID-19. De acuerdo con (Cao et al., 2020), la estabilidad económica familiar, vivir en un área urbana y con familiares disminuyen el riesgo de ansiedad en estudiantes universitarios de medicina. Sin embargo, si algún conocido o uno de los miembros de la familia contrae COVID-19, el riesgo de ansiedad se incrementa.

Respecto a los alumnos de la Universidad Nacional Agraria La Molina (UNALM) matriculados en el curso Estadística General, (Salazar Vega, 2019) encontró que el rendimiento académico pasado es el mejor predictor del actual. Por otro lado, (Soto & Zuñiga Blanco, 2020) afirmaron que el estilo de aprendizaje (visual, auditivo o kinestésico) predominante del alumno no tiene efectos en su rendimiento académico.

Síntomas emocionales

Para el Ministerio de Salud (2007), la depresión es denominada como un trastorno mental, el cual principalmente se manifiesta con humor depresivo, pérdida del interés y disfrute de los aspectos positivos de la vida, además cansancio o fatiga, los cuales empobrecen la calidad de vida y generan adversidades en el entorno familiar, laboral o académico y social de quienes la padecen.

En cuanto a la ansiedad, se dice que es una respuesta anticipatoria a una amenaza futura, está más asociada con la tensión muscular, vigilancia en relación a un peligro futuro y comportamientos cautelosos o evitativos, además es normal ver crisis de pánico como un tipo particular de respuesta al miedo, pero aclarar que no es propio de la ansiedad, sino de varios trastornos mentales (Asociación Americana de Psiquiatría, 2014).

Por otro lado, existe dificultad en definir el estrés, debido a las distintas realidades que existen en el mundo. El estrés generalmente se interpreta en referencia a una amplia gama de experiencias, entre las que figuran el nerviosismo, la tensión, el cansancio, el agobio, la inquietud y otras sensaciones similares, como la constante presión escolar, laboral o de otra índole. También se atribuye a situaciones de miedo, temor, angustia, pánico, afán por cumplir, vacío existencial, premura por desear cumplir metas y propósitos en un tiempo corto (Martínez y Díaz, 2007).

Una manera de medir estas variables de afectación emocional es a través del cuestionario DASS (escalas de depresión, ansiedad y estrés, por sus siglas en inglés), cuya versión original consta de 42 preguntas, sin embargo, estas no manejan adecuadamente la correlación entre las afecciones emocionales en estudio, puesto que la depresión y la ansiedad comparten diversas características. Sin embargo, tienen algunas que los diferencian como los bajos niveles de afecto positivo, los cuales son exclusivos de la depresión, y por otro lado la hiperactividad, que es causada solamente por la ansiedad. Ante ello, (Antony et al., 1998) propusieron una versión resumida de 21 ítems que lidia mejor con las correlaciones entre factores, y probaron estas escalas en pacientes no clínicos, así como en otros que padecían ciertos desórdenes mentales como

trastorno obsesivo – compulsivo, fobia social, entre otros.

A su vez, (Antúnez & Vinet, 2012) validaron la escala DASS-21 en estudiantes universitarios chilenos, mientras que (Román Mella et al., 2014) realizaron lo propio en estudiantes chilenos de educación secundaria. Posterior a ello, (Tijerina et al., 2018) aplicaron este cuestionario en estudiantes que inician la carrera profesional de nutrición. A nivel local, (Polo Martínez, 2017) validó las escalas en estudiantes universitarios de Chimbote. También en Perú, (Contreras-Mendoza & Olivas-Ugarte, 2020) analizaron la validez, fiabilidad y equidad de las DASS-21 en adolescentes peruanos de 14 a 17 años.

Metodología

La investigación se llevó a cabo con un enfoque del tipo cuantitativo, debido a que las variables fueron objeto de medición, de tipo descriptivo-correlacional de corte transversal. El diseño de investigación fue no experimental, puesto que no se controlaron las variables (Hernández et al., 2014).

Los datos fueron obtenidos de distintas fuentes: la primera de ellas corresponde a la Oficina de Estudios y Registros Académicos (OERA) de la UNALM, quien proporcionó los datos del historial académico de los estudiantes, mientras que el nivel de estrés, ansiedad y depresión se recogió a través del cuestionario DASS-21, el cual fue aplicado de manera virtual, voluntaria e informada a los estudiantes matriculados en los cursos Estadística General y Métodos Estadísticos de la UNALM.

La población de estudio estuvo compuesta por 958 estudiantes matriculados en los cursos de Estadística general y Métodos estadísticos para la investigación de la UNALM en el semestre 2020 - I. Para el análisis de datos, se consideró una muestra de estudiantes de tamaño 358 estudiantes de los cuales 220 (61.5%) fueron mujeres con edad promedio de 21 años y 138 (38.5%) varones con edad promedio 21.3 años.

El análisis de los datos consideró dos etapas: la primera fue descriptiva, en la que se realizaron exploraciones univariadas y bivariadas de las variables en estudio a fin de identificar posibles patrones y asociaciones

relevantes entre estas. Las variables del estudio se listan en la tabla 1.

Tabla 1: Variables del conjunto de datos

Variables	Descripción
Sexo	Sexo del alumno/a
Carrera	Carrera que estudia
Situación académica	Un(a) estudiante puede estar en situación académica “Normal” o “Riesgo” (cualquier situación distinta a Normal)
Créditos matriculados actuales	Número de créditos matriculados en el ciclo 2020 - I
Créditos aprobados en ciclo 1	Número de créditos aprobados en su primer ciclo de estudios
Años desde el ingreso	Número de años transcurridos desde que ingresó a la universidad
Promedio ponderado semestral (PPS)	Promedio ponderado del último semestre cursado, previo al ciclo 2020-1
Número de retiros	Número de veces que se retiró de algún curso (valor máximo según Reglamento General: 6)
Curso	Curso básico de estadística en el que se matriculó: Estadística general o Métodos estadísticos para la investigación I
Turno	Turno de la sección del curso matriculado: mañana, tarde o fin de semana
Nota	Nota obtenida en el curso al finalizar el semestre académico, la cual se establece en escala vigesimal.
Nota categorizada	Baja (nota menor o igual a 13), media (nota mayor a 13 y menor a 16) y alta (nota de 16 a más)
Ansiedad (A)	Puntaje total obtenido con 7 ítems de la escala Dass 21 (ítems: 2, 4, 7, 9, 15, 19 y 20)
Estrés (S)	Puntaje total obtenido con 7 ítems de la escala Dass 21. (ítems: 1, 6, 8, 11, 12, 14 y 18)
Depresión (D)	Puntaje total obtenido con 7 ítems de la escala Dass 21. (ítems: 3, 5, 10, 13, 16, 17 y 21)

En una segunda etapa se consideraron los análisis de agrupación y clasificación. Para el primer caso, se

utilizó el análisis de conglomerados, también conocido como análisis cluster, el cual consiste en el descubrimiento de grupos de unidades elementales, basado en medidas de distancia o similitud entre estos (Everitt et al., 2011). La viabilidad del uso de esta técnica fue evaluada mediante el estadístico de Hopkins, mientras que el número de grupos considerados se obtuvo mediante la técnica de la silueta. Posterior a ello, el algoritmo de agrupamiento empleado fue el de k-medias, el cual hizo uso de las variables cuantitativas del conjunto de datos. Finalmente, la interpretación de los clusters encontrados se realizó con el conjunto de variables completo.

Por otro lado, para el análisis de clasificación, la variable objetivo fue el rendimiento académico (nota final del curso) categorizado en 3 clases, tomando como referencia los percentiles 33 y 66 (tabla 1), para luego aplicar un algoritmo de árbol de clasificación, el cual es una técnica de minería de datos que basa su aprendizaje en similitudes encontrando relaciones no lineales entre las distintas características de una población. (Orea et al., 2015). Con este algoritmo se pretendió buscar los factores o indicadores más importantes que estén asociados al rendimiento académico, a fin de caracterizar y clasificar a los estudiantes en grupos diferenciados.

Resultados

Un resumen descriptivo de las variables cuantitativas puede ser revisado en la tabla 2, a partir del cual cabe resaltar que todos los alumnos eran mayores de edad y que ya había transcurrido al menos un año desde que

ingresaron a la universidad. Asimismo, que por lo menos la mitad de los estudiantes no había realizado retiro de cursos previo al ciclo 2020-1. Por otro lado, respecto a las variables cualitativas, la situación académica del 78% de los encuestados era normal, el 45% estaba matriculado en el curso Métodos estadísticos para la investigación y el 55% restante en Estadística general.

La figura 1 muestra las correlaciones bivariadas entre cada par de variables cuantitativas. Así, se encontró que la edad estaba asociada de manera directa con los años desde el ingreso y el número de retiros efectuados, siendo esta una asociación natural dado que están influenciadas por el paso del tiempo. Por otro lado, los puntajes de depresión, ansiedad y estrés también mostraron una relación alta y directa. La variable que presentó mayor asociación lineal con la nota obtenida en el curso fue el promedio ponderado del último ciclo de estudios. No se evidenció la presencia de correlaciones negativas de alta magnitud, siendo la existente entre la cantidad de años desde el ingreso y el número de créditos matriculados en el ciclo verano 2020 la más resaltante ($r=-0.28$).

Análisis de conglomerados

Se verificó la viabilidad del análisis de conglomerados mediante el estadístico de Hopkins, cuyo valor fue de 0.73, indicando la factibilidad de encontrar clusters en el conjunto de datos. Luego, el análisis de la silueta sugirió el uso de 3 conglomerados usando la técnica de k-medias.

Tabla 2: Estadística descriptiva de las variables cuantitativas

Variable	Media	Desv. estándar	Mínimo	Mediana	Máximo
Edad	21.1	2.5	18	21	32
Años desde el ingreso	2.8	1.8	1	2.5	13
Créditos aprobados en su primer ciclo	15.3	3.6	4	16	21
Número de retiros de cursos	0.6	1.1	0	0	9
Promedio ponderado semestral (PPS)	13.6	2.5	0	13.5	20
Créditos matriculados en el ciclo 2020-I	17.1	3.3	4	17	26
Nota en el curso	14.2	2.1	4	14	19
Depresión (D)	7.5	5.0	0	7	21
Ansiedad (A)	5.5	4.1	0	5	19
Estrés (S)	8.9	4.4	0	9	21

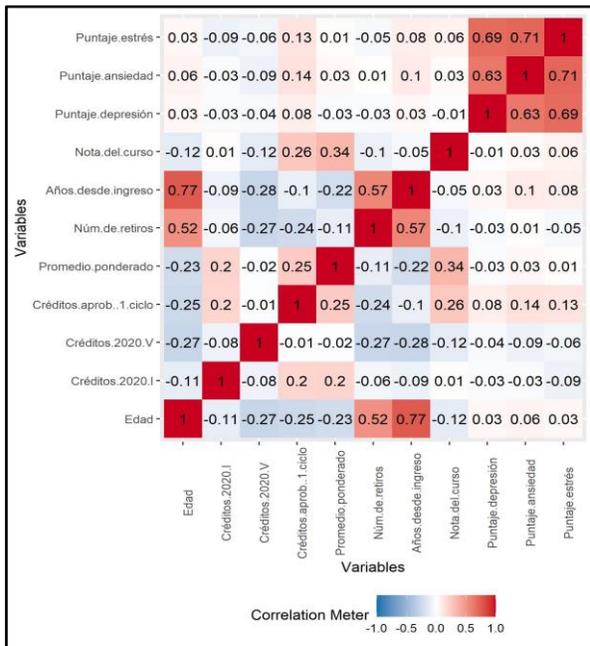


Figura 1: Correlaciones bivariadas entre las variables cuantitativas

El sexo no fue un factor relevante en la segmentación, así como la nota obtenida al finalizar el curso. A continuación, la descripción de cada uno de los clusters encontrados:

Cluster 1: Conformado por el 20% de los estudiantes, cuyo perfil se ajustó a las siguientes características:

- Presentaron la mayor edad promedio (24.3 años), como se aprecia en la densidad de la figura 3. Además, su ingreso a la universidad se dio hace 5 años en promedio, lo cual indica que deberían estar por culminar sus estudios. Por otro lado, en su primer ciclo aprobaron menos créditos (12) que en los otros clusters.
- Presentaron la mayor cantidad media de retiros de cursos (2), más de la mitad (62.3%) no se encontraba en situación académica normal (figura 2) y la media de su promedio ponderado semestral (12) era la más baja entre los clusters.
- Pocos cursaron el ciclo verano 2020, meses antes al inicio del ciclo virtual, ya que el número promedio de créditos matriculados en dicho ciclo fue apenas 0.71.

- En el ciclo 2020-1, la cantidad media de créditos matriculados (15.9) fue la menor entre los clusters. Además, el 77% se encontraba cursando Métodos estadísticos para la investigación.
- Sus puntajes de depresión, ansiedad y estrés se situaron en una posición intermedia al ser comparada con los otros conglomerados, tal como se muestra en la figura 4.

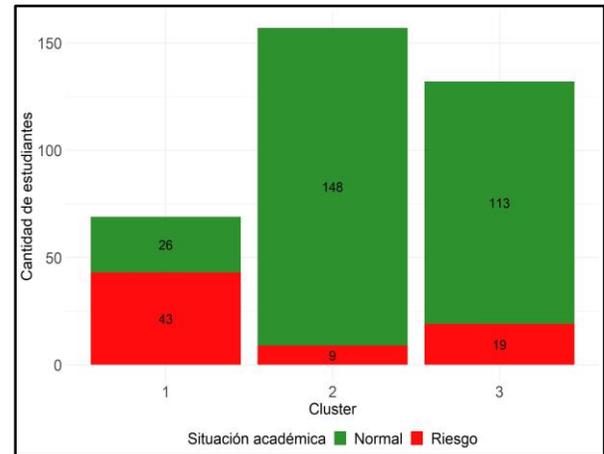


Figura 2: Distribución de la situación académica por cluster

Cluster 2: Conformado por el 44% de los estudiantes, los cuales presentan siguientes características:

- Presentaron la menor edad promedio (19.9 años), como se observa en la densidad de la figura 3. Asimismo, su ingreso a la universidad se dio hace 1.8 años en promedio, es decir estaban cursando el cuarto semestre.
- Presentaron la menor cantidad media de retiro de cursos (0.2) y la gran mayoría (95%) se encontraba en situación académica normal (figura 2).
- Cursaron en promedio 2.8 créditos en ciclo verano 2020, cantidad mayor respecto a los otros dos clusters.
- En el ciclo 2020-1, presentaron la mayor cantidad de créditos matriculados (17.4), junto con los integrantes del cluster 3. Además, el 77% se encontraba cursando Estadística general (59% en la modalidad de tres créditos y el 18% en la de cuatro).

- Se obtuvieron los menores puntajes promedio de depresión (4.2), ansiedad (3.1) y estrés (6.0), como se puede apreciar en la figura 4.

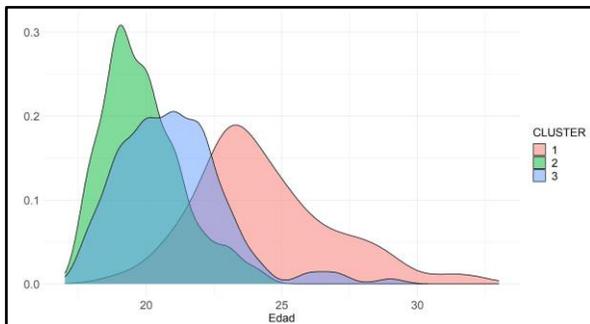


Figura 3: Distribución de la edad por cluster

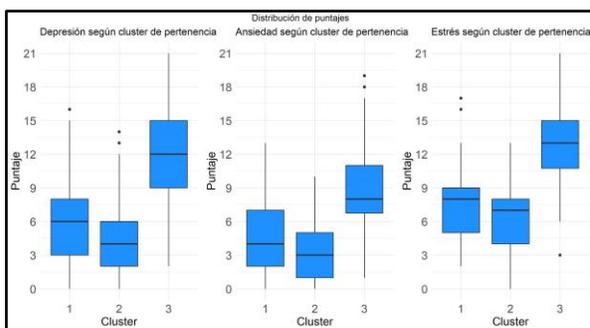


Figura 4: Distribución de puntajes de depresión, ansiedad y estrés por cluster

Cluster 3: Conformado por el 36% de los estudiantes, los cuales poseen el siguiente perfil:

- Presentaron una edad promedio (20.9) no tan alta como en el primer cluster ni tan baja como en el segundo (figura 3). Además, ingresaron a la universidad hace 2.7 años en promedio, por lo que deberían encontrarse en la mitad de su carrera. Por otro lado, en su primer ciclo aprobaron, en promedio, 16.5 créditos, la cantidad más alta respecto a los otros conglomerados.
- La mayoría (85.6%) se encontraba en situación académica normal (figura 2).
- El 55% se encontraba cursando Métodos estadísticos para la investigación.
- Presentaron los mayores puntajes promedio de depresión (12), ansiedad (8.9) y estrés (12.8), como se puede ver en la figura 4.

Análisis de clasificación mediante árboles

En el programa R se ejecutó el algoritmo Árboles de clasificación con la función Rpart del paquete rpart, la función mencionada usa la medida de Gini para la división de los nodos. Luego, para evitar tener un árbol muy profundo que sobreestime (overfitting), se buscó el parámetro de regularización y optimización que nos permitió identificar el mejor punto para realizar la poda, dicho parámetro se llama costo de complejidad (cp), el cual penaliza a modelos con mayor número de nodos hoja en el árbol; es así que el cp elegido fue de 0.014286 quien minimizó el error relativo de predicción.

Con los criterios antes explicados se obtuvo un árbol de clasificación con 10 hojas o nodos terminales (figura 5). El algoritmo consideró las variables que mejor discriminaron las categorías de la nota final (rendimiento académico), las cuales fueron el promedio ponderado semestral del último ciclo, la asignatura de estadística en la que se matriculó, la cantidad de créditos matriculados en el ciclo 2020-1, la edad, la ansiedad y el estrés, obteniéndose reglas de clasificación, de las cuales se explican a continuación las más importantes por abarcar una mayor cantidad de muestra y tener una mejor tasa de predicción:

- Regla 1: Clasificar a un estudiante en la categoría 3 (nota final mayor igual a 16), si su promedio semestral pasado es mayor o igual a 15.46. Esto nos indica que un estudiante que tiene un récord de notas alto en ciclos pasados, es de esperar que mantenga notas altas en siguiente semestre.
- Regla 2. Clasificar a un estudiante en la categoría 2 (nota final mayor a 13 y menor a 16), si su último promedio semestral es menor a 15.46, si lleva el curso de Métodos estadísticos para la investigación y si su cantidad de créditos matriculados en el semestre evaluado es menor a 21.
- Regla 3. Clasificar a un estudiante en la categoría 1 (Nota final menor igual a 13), si su promedio semestral anterior es menor a 15.46, si está matriculado en cursos de estadística general de 3 o 4 créditos y además si:
 - Si está cursando menos de 13 créditos y tiene 21 o más años de edad.
 - Si está cursando 13 o más créditos y: a) su último promedio semestral es menor o igual a 11, o b) su

último promedio semestral es mayor o igual a 12 y su puntaje de estrés es menor a 4.5.

Observando los nodos terminales u hojas del árbol del lado izquierdo (figura 5), se puede notar que los alumnos son clasificados en su mayoría en categorías 1 y 2, donde se evidencia la presencia de ansiedad y estrés en estudiantes cuyo último promedio ponderado se encuentra entre 12 y 15, este es un dato relevante a tomar en cuenta para realizar diagnósticos más profundos a estudiantes con este rendimiento.

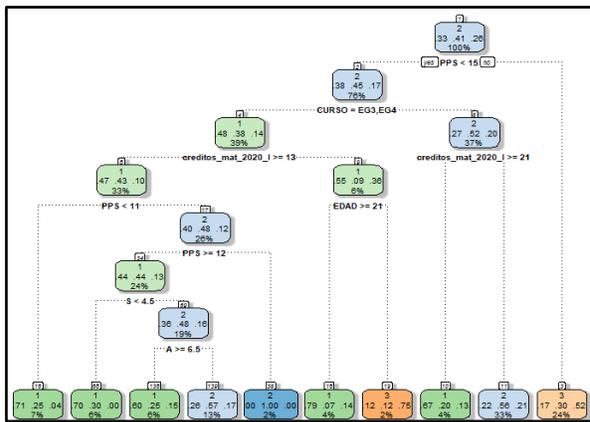


Figura 5: Árbol de decisión

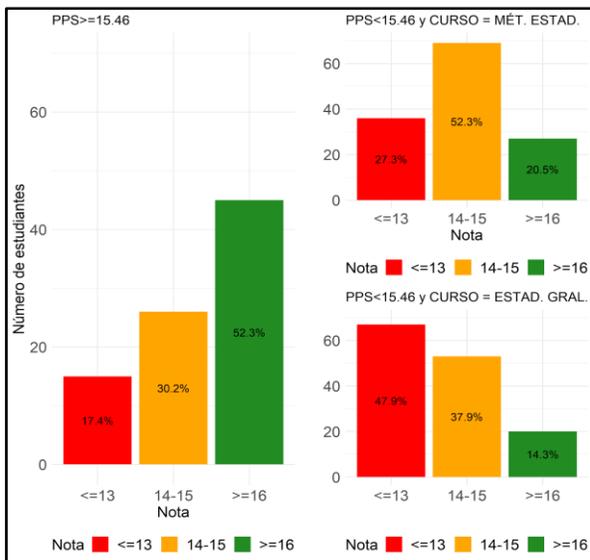


Figura 6: Distribución de estudiantes según curso, nota y promedio ponderado

Se puede observar que si consideramos las dos variables más importantes, el promedio ponderado semestral y el curso en que se matriculó, podemos obtener tres grupos

de estudiantes que se encuentran detallados en la figura 6. Un primer grupo de estudiantes está conformado principalmente por aquellos cuya nota del curso fue mayor o igual a 16. Un segundo grupo de estudiantes está conformado por estudiantes que el ciclo anterior obtuvo un promedio ponderado semestral menor a 15.46 y llevó el curso de métodos estadísticos. En este grupo, más de la mitad (52.3%) de los estudiantes obtuvo una nota final de 14 a 15.9. Mientras que un tercer grupo de estudiantes está caracterizado por un 47.9% que obtuvo notas finales menores iguales a 13, y su característica es que el ciclo anterior obtuvo PPS menores a 15.46 y lleva en el semestre de estudio cursos de estadística general de 3 y 4 créditos.

Discusión

En concordancia al hallazgo de Salazar Vega (2019), el rendimiento académico actual mostró ser influenciado por el desempeño pasado, sin embargo, en esta investigación no se considera el promedio ponderado acumulado sino el obtenido en el ciclo previo. Por otro lado, a diferencia de lo señalado por Dicovski y Riobóo & Pedroza, (2017), la variable sexo no tuvo aporte significativo en los modelos construidos.

Uno de los conglomerados está conformado por estudiantes jóvenes con menos de 2 años de estudio en la universidad y con los menores puntajes de ansiedad, así como depresión y estrés. Esto se podría deber a que están más adaptados al cambio debido a que hace poco pasaron por una transición del colegio a la universidad. Un resultado similar es obtenido por (Mamani Yto, 2020), cuya investigación señaló que el 96% de los estudiantes ingresantes universitarios no presentaban ansiedad y el 84% no presentaban depresión.

Al pasar el tiempo en las aulas universitarias, (Islam et al., 2020) mencionan que la edad y la ansiedad presentan una asociación positiva, es decir que los estudiantes de mayor edad presentaban mayores niveles de ansiedad. En nuestra investigación, aquellos de mayor edad no fueron quienes presentaron mayores niveles de afección en su salud mental, sino los que se encontraban en la mitad de sus estudios. Una posible explicación para esto puede estar relacionada a la disponibilidad limitada de tecnologías de información y comunicación (TIC) de los estudiantes (Huanca et al., 2020), así como a las características de su núcleo

familiar (Pérez López et al., 2020), variables que no fueron consideradas en el presente estudio.

Para (Timarán-Pereira et al., 2019), quienes utilizaron Árboles de decisión, afirman que los atributos asociados al buen desempeño académico son el estrato socioeconómico medio o alto, la jornada de estudio en la mañana o completa, el índice TIC regular y la edad. Asimismo, entre los atributos asociados a un bajo desempeño académico, se puede listar el estrato socioeconómico bajo, el índice TIC bajo y el nivel SISBEN 1 (Sistema de Identificación de Potenciales Beneficiarios de Programas Sociales), mientras que nosotros encontramos que el PPS previo alto es un factor asociado a altos rendimientos académicos, mientras que tener menores PPS en el ciclo anterior, matricularse en el curso de estadística general, la abundante cantidad de créditos matriculados, además de la presencia de ansiedad y estrés son atributos ligados al bajo rendimiento académico.

Conclusiones

Los conglomerados obtenidos presentan la siguiente caracterización: i) estudiantes con mayor tiempo de estudios desde el ingreso a la universidad, pero con un rendimiento inferior, ii) estudiantes jóvenes con menor depresión, ansiedad y estrés, y iii) estudiantes que se encuentran en la mitad de carrera con un rendimiento aceptable pero con niveles altos de depresión, ansiedad y estrés.

Los atributos que mayor ganancia de información presentan para predecir un alto, mediano o bajo rendimiento académico de los estudiantes universitarios, son el promedio ponderado semestral del ciclo anterior, la asignatura en la que se matriculó, la cantidad de créditos matriculados, la edad, el estrés y la ansiedad; Obtenido mediante el modelo de minería de datos árboles de decisión.

El algoritmo de árbol de decisión clasifica a un estudiante con una nota mayor o igual a 16, si su último promedio semestral es mayor o igual a 15.46. Además, un estudiante es clasificado con nota final mayor a 13 pero menor a 16, si su promedio semestral previo fue menor a 15.46, se matriculó en el curso de métodos estadísticos y llevó menos de 21 créditos. Por último, un estudiante es clasificado con una nota final menor a

13, si su último promedio semestral fue menor a 15.46 y se matriculó en cursos de estadística general, en algunos casos de este grupo se encontró presencia de estrés y ansiedad como variables relacionadas al rendimiento académico.

Los resultados de esta investigación procuran ser de gran utilidad para las oficinas correspondientes encargadas del bienestar y la salud mental de los estudiantes de la UNALM, así como de su desempeño académico. Recomendamos tener en cuenta el presente estudio para la previsión del regreso a clases presenciales en el mediano plazo, pues existe la posibilidad de que los alumnos que comenzaron sus estudios universitarios junto con el brote de la pandemia, para entonces ya hayan recorrido un buen trecho de su carrera en la modalidad virtual y les cueste adaptarse a una modalidad presencial, es decir un fenómeno similar a uno de los conglomerados encontrados.

Conflictos de intereses

Los autores firmantes del presente trabajo de investigación declaran no tener ningún potencial conflicto de interés personal o económico con otras personas u organizaciones que puedan influir indebidamente con el presente manuscrito.

Contribuciones de los autores

Preparación y ejecución: GJ, ZA; Desarrollo de la metodología: GJ, ZA; Concepción y diseño: GJ, ZA; Edición del artículo: GJ, ZA; y, Supervisión del estudio: GJ, ZA.

Referencias bibliográficas

- Al-kandari, A., Law, J., Alhashmi, H., Alshammari, O., & Bhandari, P. (2021). Staying (mentally) healthy – the impact of COVID19 on personal and professional lives. *Techniques and Innovations in Gastrointestinal Endoscopy*. <https://doi.org/10.1016/j.tige.2021.01.003>
- Antony, M., Bieling, P., Cox, B., Enns, M., & Swinson, R. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment*, 10, 176-181. <https://doi.org/10.1037/1040-3590.10.2.176>

- Antúnez, Z., & Vinet, E. V. (2012). Escalas de Depresión, Ansiedad y Estrés (DASS - 21): Validación de la Versión abreviada en Estudiantes Universitarios Chilenos. *Terapia psicológica*, 30(3), 49-55. <https://doi.org/10.4067/S0718-48082012000300005>
- Asociación Americana de Psiquiatría. (2014) Manual diagnóstico y estadístico de los trastornos mentales (DSM-5), 5 a Ed. Arlington, VA.
- Cao, W., Fang, Z., Hou, G., Han, M., Xu, X., Dong, J., & Zheng, J. (2020). The psychological impact of the COVID-19 epidemic on college students in China. *Psychiatry Research*, 287, 112934. <https://doi.org/10.1016/j.psychres.2020.112934>
- Contreras-Mendoza, I., & Olivas-Ugarte, L. O. (2020). Escalas abreviadas de Depresión, Ansiedad y Estrés (DASS-21): Validez, fiabilidad y equidad en adolescentes peruanos. *Revista de Psicología Clínica en Niños y Adolescentes*, 7.
- Dicovski, L. M., & Pedroza, M. E. (2017). Minería de datos, una innovación de los métodos cuantitativos de investigación, en la medición del rendimiento académico universitario. *Revista Científica de FAREM-Estefí*, 24, 143-152. <https://doi.org/10.5377/farem.v0i24.5557>
- Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis* (5.a ed.). Wiley.
- Haider, A. S., & Al-Salman, S. (2020). Dataset of Jordanian university students' psychological health impacted by using e-learning tools during COVID-19. *Data in Brief*, 32, 106104. <https://doi.org/10.1016/j.dib.2020.106104>
- Hernández, R., Fernández, C., & Baptista, M. (2014). *Metodología de la investigación* (6ta ed.). McGRAW-HILL education.
- Huanca, J., Supo, F., Sucari, R., & Supo, L. (2020). El problema social de la educación virtual universitaria en tiempos de pandemia, Perú | *Innovaciones Educativas*. <https://revistas.uned.ac.cr/index.php/innovaciones/article/view/3218>
- Huarcaya-Victoria, J. (2020). Consideraciones sobre la salud mental en la pandemia de COVID-19. *Revista Peruana de Medicina Experimental y Salud Pública*, 37(2), 327-334. <https://doi.org/10.17843/rpmesp.2020.372.5419>
- Islam, M. A., Barna, S. D., Raihan, H., Khan, M. N. A., & Hossain, M. T. (2020). Depression and anxiety among university students during the COVID-19 pandemic in Bangladesh: A web-based cross-sectional survey. *PLOS ONE*, 15(8), e0238162. <https://doi.org/10.1371/journal.pone.0238162>
- Kaparounaki, C. K., Patsali, M. E., Mousa, D.-P. V., Papadopoulou, E. V. K., Papadopoulou, K. K., & Fountoulakis, K. N. (2020). University students' mental health amidst the COVID-19 quarantine in Greece. *Psychiatry Research*, 290, 113111. <https://doi.org/10.1016/j.psychres.2020.113111>
- Johns Hopkins University (s.f.). COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Recuperado el 09 de marzo de 2021 de <https://coronavirus.jhu.edu/map.html>
- Lozano Díaz, A., Prados, J. S. F., Canos, V. F., & Martínez, A. M. M. (2020). Impactos del confinamiento por el COVID-19 entre universitarios: Satisfacción Vital, Resiliencia y Capital Social Online. *RISE*, 9(Extra 1), 79-104.
- Mamani Yto, G. D. (2020). Evaluación de ansiedad, depresión y dinámica familiar en ingresantes a la U.C.S.M. entre los años, 2018-2019, Arequipa 2020.
- Martínez Díaz, E. S., & Díaz Gómez, D. A. (2007). A psychosocial approximation to school stress. 13. Ministerio de Salud. (2007) *Guía de Práctica Clínica en Trastornos Mentales y del Comportamiento debido al consumo de sustancias psicotropas*. Lima: Dirección General de Salud de las Personas. Dirección de Salud Mental. [http://bvs.minsa.gob.pe/local/minsa/2907.PDF %20](http://bvs.minsa.gob.pe/local/minsa/2907.PDF%20)
- Montalvo Romero, N., & Montiel Rosales, A. (2020). Impacto del COVID-19 en el estrés de universitarios. <http://www.itchetumal.edu.mx/avacient/index.php/revista/issue/view/15/9>
- Orea, S. V., Vargas, A. S., & Alonso, M. G. (2005). Minería de datos: Predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos. http://fcaenlinea.unam.mx/anexos/1566/1566_u6_act1b.pdf
- Organización Mundial de la Salud (11 de marzo de 2020). *Alocución de apertura del Director*

- General de la OMS en la rueda de prensa sobre la COVID-19 celebrada el 11 de marzo de 2020. <https://www.who.int/es/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>
- Pérez López, E., Vázquez Atochero, A., & Cambero Rivero, S. (2020). Educación a distancia en tiempos de COVID-19: Análisis desde la perspectiva de los estudiantes universitarios. *RIED. Revista Iberoamericana de Educación a Distancia*, 24(1), 331. <https://doi.org/10.5944/ried.24.1.27855>
- Polo, R. (2017). Propiedades Psicométricas de la Escala de Depresión, Ansiedad Y Estrés (DASS-21) en estudiantes universitarios de Chimbote. 77.
- Porcel, E. A., Dapozo, G. N., & López, M. V. (2010). Predicción del rendimiento académico de alumnos de primer año de la FACENA (UNNE) en función de su caracterización socioeducativa. *Revista Electrónica de Investigación Educativa*, 12(2), 22.
- Qiu, J., Shen, B., Zhao, M., Wang, Z., Xie, B., & Xu, Y. (2020). A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: Implications and policy recommendations. *General Psychiatry*, 33(2), e100213. <https://doi.org/10.1136/gpsych-2020-100213>
- Román Mella, F., Vinet, E. V., & Muñoz, A. M. A. (2014). Escalas de Depresión, Ansiedad y Estrés (DASS-21): Adaptación y propiedades psicométricas en estudiantes secundarios de temuco. *Revista Argentina de Clínica Psicológica*, XXIII (2), 179-190.
- Salazar, R. J. (2019). Modelo de ecuación estructural explicativo del rendimiento académico de los estudiantes del curso de estadística general en la UNALM. 107.
- Soto, I., & Zuñiga, A. (2020). Los Estilos de aprendizaje bajo un modelo de programación neurolingüística y rendimiento académico en estudiantes universitarios. *Espí-ritu Emprendedor TES*, 4(3), 77-91. <https://doi.org/10.33970/eetes.v4.n3.2020.217>
- Superintendencia Nacional de Educación Universitaria. (2014). *Ley Universitaria*. http://www.minedu.gob.pe/reforma-universitaria/pdf/ley_universitaria.pdf
- Timarán-Pereira, R., Caicedo-Zambrano, J., & Hidalgo-Troya, A. (2019). Árboles de decisión para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas Saber 11°. *Revista de Investigación Desarrollo e Innovación, RIDI*, 9(2), 12. <https://doi.org/10.19053/20278306.v9.n2.2019.9184>
- Tijerina, L., González, E., Gómez, M., Cisneros, M. A., Rodríguez, K. Y., & Ramos, E. G. (2018). Depresión, ansiedad y estrés en estudiantes de nuevo ingreso a la educación superior. *RESPYN Revista de Salud Pública y Nutrición*, 17(4), 41-47. <https://doi.org/10.29105/respyn17.4-5>
- Vlachopoulos, D., & Makri, A. (2019). Online communication and interaction in distance higher education: A framework study of good practice. *International Review of Education*, 65(4), 605-632. <https://doi.org/10.1007/s11159-019-09792-3>
- Wang, C., Pan, R., Wan, X., Tan, Y., Xu, L., Ho, C. S., & Ho, R. C. (2020). Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. *International Journal of Environmental Research and Public Health*, 17(5). <https://doi.org/10.3390/ijerph17051729>
- Sánchez, A. (2009). Caso Perú. Cambios demográficos y movilidad laboral en la región Asia- Pacífico: 2007-2008, OIM-UNFPA, Lima.
- Tamagno, C. (2003). Entre acá y allá. Vidas transnacionales y desarrollo. Peruanos entre Italia y Perú. Lima.
- Torres, J. (2006). Remesas en el Perú, Centrumun, Lima.