

Revista de Investigación Científica y Tecnológica

Llamkasun

Trabajemos



Características para un modelo de predicción de la deserción académica universitaria. Caso Universidad Nacional de Santa



Characteristics for a predictive model of university academic dropout. Case of the National University of Santa Fe.



Características para um modelo preditivo de abandono académico universitário. O caso da Universidade Nacional de Santa Cruz.

<https://doi.org/10.47797/llamkasun.v2i4.61>



UNIVERSIDAD NACIONAL AUTÓNOMA DE
TAYACAJA DANIEL HERNÁNDEZ MORILLO

VICEPRESIDENCIA DE INVESTIGACIÓN
PAMPAS TAYACAJA - HUANCAVELICA

ISSN: 2709-2275

Características para un modelo de predicción de la deserción académica universitaria. Caso Universidad Nacional del Santa

Characteristics for a prediction model of university academic desertion. Case: Universidad Nacional del Santa

Características para um modelo preditivo de deserção académica universitária. O caso da Universidade Nacional del Santa

Hugo Esteban Caselli Gismondi 
Universidad Nacional del Santa

Luis Vladimir Urrelo Huiman 
Universidad Privada Antenor Orrego

RESUMEN

Un modelo predictivo puede contribuir a mitigar la deserción académica de estudiantes universitarios, elemento de preocupación durante el seguimiento académico en varios países, en Perú la Universidad Nacional de Santa muestra un comportamiento desfavorable. Consecuentemente se declara como objetivo: proponer las características que deberán formar parte de un modelo de predicción basado en Machine Learning que contribuya a la adopción de medidas oportunas durante el seguimiento académico. Esta investigación responde a un caso de estudio y forma parte de los resultados del proceso de generación del modelo predictivo. Como materiales y métodos se emplearon: la revisión documental y métodos estadísticos. Se utilizaron los softwares libres: Python y Pandas. Se destacan como resultados: el análisis de las variables más frecuentemente empleadas en los estudios de predicción, la data maestra de la Universidad depurada y la propuesta de características que deberán integrar el modelo. El banco mundial concluye: en Perú sólo el 65% de los estudiantes universitarios logran el grado de bachiller; las características halladas durante el estudio, así como los resultados de la data de esta institución constituyeron los referentes para la propuesta. Estas formarán parte del modelo predictivo que posibilitará predecir quienes se quedan como estudiantes, egresados, bachiller o titulado.

Palabras clave: deserción académica, características predictivas, seguimiento académico.

RECIBIDO : 21-08-2021
ACEPTADO : 28-10-2021

DOI:

<https://doi.org/10.47797/llamkasun.v2i4.61>



ABSTRACT

A predictive model can contribute to mitigate the academic desertion of university students, an element of concern during the academic follow-up in several countries, in Peru the National University of Santa shows an unfavorable behavior. Consequently, the objective is stated as: to propose the characteristics that should be part of a predictive model based on Machine Learning that contributes to the adoption of timely measures during the academic follow-up. This research responds to a case study and is part of the results of the predictive model generation process. As materials and methods were used: documentary review and statistical methods. Free software was used: Python and Pandas. The following results stand out: the analysis of the variables most frequently used in prediction studies, the refined master data of the University and the proposal of characteristics that should integrate the model. The world bank concludes: in Peru only 65% of university students achieve a bachelor's degree; the characteristics found during the study, as well as the results of the data of this institution constituted the referents for the proposal. These will be part of the predictive model that will make it possible to predict who will remain as students, graduates, bachelor's degree holders or graduates.

Keywords: academic desertion, predictive characteristics, academic tracking.

RESUMO

Um modelo de previsão pode contribuir para mitigar a deserção acadêmica dos estudantes universitários, elemento de preocupação durante o acompanhamento acadêmico em vários países, no Peru a Universidade Nacional do Pai Natal mostra um comportamento desfavorável. Consequentemente, o objectivo é propor as características que devem fazer parte de um modelo preditivo baseado na Aprendizagem Automática que contribua para a adopção de medidas atempadas durante o acompanhamento académico. Esta investigação responde a um estudo de caso e faz parte dos resultados do processo de geração do modelo preditivo. Os materiais e métodos utilizados foram: revisão documental e métodos estatísticos. Foi utilizado software livre: Python e Pandas. Destacam-se os seguintes resultados: a análise das variáveis mais frequentemente utilizadas nos estudos de previsão, os dados de mestrado refinados da Universidade e a proposta de características que devem ser incluídas no modelo. O Banco Mundial conclui: no Peru apenas 65% dos estudantes universitários atingem o grau de bacharel; as características encontradas durante o estudo, bem como os resultados dos dados desta

instituição, constituíram as referências para a proposta. Estes farão parte do modelo de previsão que permitirá prever quem permanecerá como estudante, licenciado, bacharelato ou licenciado.
Palavras-chave: abandono escolar, características predictivas, acompanhamento académico.

INTRODUCCIÓN

El seguimiento académico del estudiante universitario ha sido objeto de estudio en nacionalidades como Australia, Chile, Colombia, México, Paraguay y Costa Rica, y en este sentido autores como Scott, Shah, Grebennikov, & Singh (2008), Donoso & Arias R. (2010), Puchi, Moraga, & Villagran (2016), Torres Guevara (2012), Hernández Herrera (2016) y Fernández (2018) lo han investigado desde la perspectiva de la retención, mientras que otros como Abarca Rodríguez & Sánchez Vindas (2005), Castaño, Gallón, Gómez, & Vásquez (2004), Hellas et al. (2018), Spady (1970), Fishbein & Ajzen (1975), CLABES (2019) Morí Sánchez (2012) y Ethington (1990), han centrado su atención en el abandono o deserción.

Desde el siglo pasado Salvador-Blanco & Garcia-Valcarcel Muñoz-Repiso (1989) reconocieron la estrecha relación entre el seguimiento académico y el desempeño académico, medido éste a través de los índices de promoción por asignaturas, la desaprobación de alguna asignatura y/o el nivel de deserción.

Según estudios realizados por Ferreyra, Avitabile, Álvarez, Paz, & Urzúa (2017) en relación con el comportamiento de la deserción académica de jóvenes entre 25 y 29 años en América Latina, Bolivia destaca como el país de menor número de graduados y la mayor cantidad de desertores con más del 40%, Honduras y Uruguay muestran las cifras más altas de matriculados sin haberse graduado oscilando en torno al 50%, mientras que México y Perú muestran los mejores resultados pero sin alcanzar el 70% de graduados de la educación superior.

En Perú, el “II Informe bienal sobre la realidad universitaria en el Perú” (SUNEDU II, 2020) revela que en el 2018 sólo el 55% de los jóvenes entre 25-29 años lograron egresar, el 27,4% aún estudian y el 17,6% interrumpieron sus estudios. Similar comportamiento se apreció en la Universidad Nacional de Santa (UNS) hasta el 2018 (Compendios Estadísticos UNS 1987-2018), donde la Escuela Profesional de Ingeniería de Sistemas e Informática (EPISI) y en la de Energía y Física (EPIE) no lograron alcanzar el 50% de graduados y titulados y ninguna de las cuatro escuelas

sobrepasó el 70% en ninguna de estas dos categorías.

Los elevados porcentajes de estudiantes universitarios que no han logrado graduarse y titularse en el Perú devino en la incorporación en este propio informe de aspectos novedosos para el seguimiento al desempeño de los estudiantes como los factores que se asocian a la interrupción o abandono de los estudios universitarios lo que es muestra de la importancia que se le presta a esta problemática en el país. Entretanto, para el mundo académico, este tipo de problemática identificada en varios países deberá constituir un objeto a investigar tal y como defienden Kostopoulos, Lipitakis, Kotsiantis, & Gravvanis (2017), quienes además reconocen la necesidad de realizar estudios de predicción para el seguimiento académico dada su complejidad y las múltiples variables que intervienen.

La predicción contribuye, entre otros, a determinar con mayor certeza el nivel de permanencia y/o promoción de las asignaturas en las que se matriculó el estudiante, así como quiénes tenderán al abandono, de allí que aparezcan reportados diferentes propuestas de modelos de estudio que explican el comportamiento de grupos determinados de estudiantes que están inmersos en el sistema universitario desde

perspectivas psicológicas y causales, entre los que se encuentran Spady (1970), Tinto (1975), Bean (1985) y Ethington (1990).

Estos modelos incluyen las categorías o variables que pueden explicar la ocurrencia de cada caso de allí que en este sentido exigen de la determinación previa de las variables independientes que definen el comportamiento de las variables respuesta, las cuales deberán definirse en correspondencia con las características del contexto universitario objeto de estudio, aspecto éste que unido al comportamiento de la UNS condujo a plantearse como objetivo de este trabajo: proponer las características que deberán formar parte de un modelo de predicción basado en Machine Learning que contribuya a la adopción de medidas oportunas durante el seguimiento académico.

METODOLOGÍA

Este trabajo responde a un caso de estudio que forma parte de los resultados de una investigación dirigida a la generación de un modelo para predecir la deserción de estudiantes universitarios en la UNS utilizando el Machine Learning. En correspondencia se dio cumplimiento a las fases 1 y 2 de la metodología para desplegar el proyecto de analítica predictiva de

deserción de estudiantes de la UNS que aparece en la figura 1. En él se exponen los antecedentes que ponen luz roja en la deserción académica universitaria y los hallazgos derivados del análisis de las variables más frecuentemente empleadas en los estudios de predicción, como resultados de la fase 1 Entendimiento del entorno universitario y las características que

pueden influir en la deserción universitaria; también se exponen los resultados de la ingesta y preprocesamiento de los datos, para así dar cumplimiento a la fase 2 Adquisición y preprocesamiento de datos históricos de los semestres 2004-2018 de la Facultad de Ingeniería de la UNS. Por último, se proponen las características que deberán integrar el referido modelo.

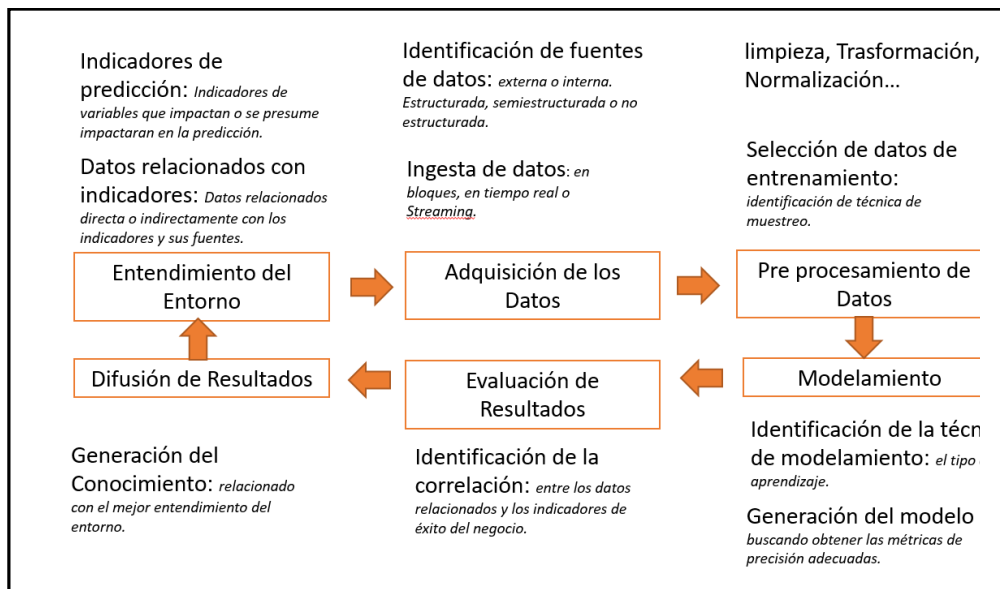


Figura 1. Modelo para desplegar el Proyecto de analítica predictiva de deserción académica en la UNS

Fuente: Elaboración propia

Se identificó como población a todos los estudiantes de la UNS, mientras se intencionó una muestra que abarcó a los estudiantes de las cuatro escuelas profesionales de la Facultad de Ingeniería de esta institución en los semestres de los años entre el 2004 y 2018. La selección de esta muestra se fundamentó tomando en cuenta que esta Facultad resulta la de más

baja proporción de bachilleres vs ingresantes (39,13%) y titulados vs ingresantes (20,06%) con respecto a las otras dos Facultades de la UNS.

Se aplicaron el análisis-síntesis y la inducción-deducción para comprender el fenómeno de la deserción académica en las universidades; identificar las variables ya empleadas en este tipo de investigación y

arribar a generalizaciones en torno a las más frecuentemente empleadas para su selección y posterior uso en el modelo de predicción de la deserción académica en la UNS.

Para analizar la tendencia de la deserción académica en la UNS se determinó el periodo comprendido entre 1993 y el 2018 para lo cual se recolectaron, procesaron y analizaron datos registrados en sistema de gestión académica de la UNS (SIIGAA).

El análisis de las variables más frecuentemente empleadas se realizó a partir de la sistematización exhaustiva de las fuentes consultadas, lo que permitió distinguir las más utilizadas ante los diferentes escenarios objeto de estudio. Durante este análisis se lograron varios hallazgos.

Se descubrió que Spady (1970), Fishbein & Ajzen (1975), Tinto (1975), Bean (1985), Ethington (1990) coincidieron al identificar variables que incidían en la persistencia o abandono estudiantil, siendo estas: antecedentes familiares, desempeño académico, expectativa de éxito, integración social, rendimiento académico previo, apoyo de profesores y el compromiso institucional.

También se identificaron como elementos que influyen en la retención: los programas de orientación en el primer año, los

programas de apoyo por pares, grupos de estudio y tutoría por pares, estos declarados por Scott, Shah, Grebennikov, & Singh (2008).

Torres Guevara (2012) y Donoso & Arias (2010) coincidieron en el peso de programas de apoyo académico y financiero para estimular la retención estudiantil universitaria, mientras que estos tipos de programas en lo psicológico y desde la gestión universitaria fueron defendidos sólo por los primeros. Dresel & Rindermann (2011) consideraron que la consejería desarrollada después de las evaluaciones permitirían asegurar y aumentar la calidad de la enseñanza en la educación superior, Arco-Tirado, Fernández-Martín, & Fernández-Balboa (2011) recomendaron la tutoría, Colvin (2015) defendió la mentoría por parte de compañeros como aspecto que influye en la retención, mejora el desempeño académico y favorece la parte afectiva del trabajo no personalizado en el aula y Whitlock (2018) reconoció cuatro factores asociados al logro del estudiante de graduarse: características institucionales, académicas, aptitud estudiantil y la comunidad de estudiantes.

Asimismo, Castaño, et al. (2004) declararon un conjunto de variables que influían en la decisión de los estudiantes de permanecer o abandonar su estudio, algunas de las cuales

resultaron coincidentes con las de sus antecesores. Estas variables fueron tomadas en la Guía de Gestión de permanencia estudiantil emitida por el Ministerio de Educación Nacional Colombia (2015) y

agrupadas por éste según su naturaleza en: individuales, académicas, institucionales y socioeconómicas (ver figura 2).



Figura 2. Variables que influyen en la permanencia o abandono de estudiantes universitarios

Fuente: Adaptado de Castaño, Gallón, Gómez, & Vásquez (2004).

En cuanto a las variables utilizadas específicamente en la predicción de la deserción académica, se descubrió que Hellas y sus colaboradores (2018), habían identificado las diez más empleadas. Por orden de mayor a menor según la cantidad de veces que fueron empleadas la variable relacionada con la calificación de las asignaturas fue utilizada 88 veces representando el 24,4% del total, le siguieron la relacionada con la puntuación en los exámenes/post-prueba de grado con 14,7%, rango de calificación en el curso con 13,6%, programa de graduación/retención

con 13,4%, desempeño y el promedio de calificaciones con 12,2% cada una y rendimiento de la tarea con 11,4%. Los tres restantes no superaron el 10%.

En este sentido Shahiri, Husain, & Rashid (2015) identificaron que los atributos mayormente utilizados eran: las actividades extracurriculares, los antecedentes de secundaria y la red de interacción social, mientras Hernández (2016) consideró tres factores como esenciales en el desempeño académico: la evaluación, la motivación del estudiante y el docente, este último también

destacado por Fernández (2018) junto con las calificaciones.

Hellas et al. (2018) también establecieron las relaciones entre las variables predictivas (variables independientes) y las variables pronosticadas (variables dependientes). Definieron que para el caso del pronóstico de retención/abandono, sus variables predictoras más comunes resultaron: el desempeño en las pre-asignaturas, en las asignaturas y la escuela secundaria, así como los datos demográficos, genero, edad, familia y los datos de personalidad. Estos resultaron de gran valor para la propuesta que se realiza.

Específicamente para la predicción aplicando Machine Learning, se apreció que las variables más empleadas en los últimos años eran muy heterogénea, no obstante se identificó que: género del estudiante, promedio de secundaria, programa y plan de estudio, utilizados por Bendangnuksung & Prabu (2018), Forero Zea, Piñeros, & Rodríguez (2019), Jia & Mareboyana (2014), y Oancea, Dragoescu, & Ciucu (2013) resultaron las de mayor coincidencia; mientras edad del estudiante y diferencia años entre secundaria-universidad fueron sólo utilizadas por estos últimos; financiamiento y retención por Jia & Mareboyana (2014) y estrato socio

económico, temas aprobados, temas no aprobados por Forero et al. (2019).

Para la ingesta y preprocesamiento de los datos se tomó como principal fuente el SIIGAA. Para obtener la data socio-económica se utilizó el sistema de bienestar universitario y para la data de graduados y titulados la base de datos de la Oficina de grados y títulos de la UNS, ambas exportadas a Excel.

Los datos fueron cargados y preprocesados con Python y la librería Pandas. De acuerdo a Müller & Guido (2016) Python es la lengua franca para las aplicaciones de la ciencia de datos, ya que tiene librerías para carga de data, visualización, estadísticas y más, asociada con la biblioteca scikit-learn, el cual es un proyecto de código abierto dependiente de los paquetes Python: NumPy y SciPy, los cuales deben ser asociados con Pandas, Matplotlib y Jupyter Notebook, todos estos trabajan en la distribución de Python: Anaconda.

Con el objetivo de mejorar la calidad de los datos las datas fueron sometidas a un proceso de depuración, descartando las variables que ofrecían sesgo elevado, así como las que mostraban exceso de nulos y/o faltantes. Para evitar el sesgo de las variables que contenían valores muy alto con respecto a otras, se procedió a su

regularización de tal manera que todas quedaron en el rango de 0 a 1.

Toda vez depuradas las datas socio económica, la de graduados y titulados, así como los promedios de notas en todos los semestres matriculados y la cantidad de semestres matriculados se procedió a efectuar un merge de estas fuentes de datos con el software Pandas a través del identificador código estandarizado como clave primaria en todas las datas.

Por último, se realizó la propuesta de las variables a utilizar en el modelo de predicción en la UNS.

RESULTADOS Y DISCUSIÓN

Entendimiento del entorno universitario y las características que pueden influir en la deserción universitaria

Antecedentes que ponen luz roja en la deserción académica en la UNS

En Perú la deserción académica se ha convertido en una preocupación de las autoridades del país. De acuerdo a los datos más recientemente publicados por el SUNEDU II (2020), en las cuatro zonas del país el por ciento de graduados es similar oscilando entre el 51,0% y el 55,9%, mientras que las interrupciones académicas llegan a alcanzar más un 24,0% en la Costa y en la Selva, con mejores comportamientos en la zona capitalina (ver figura 3).

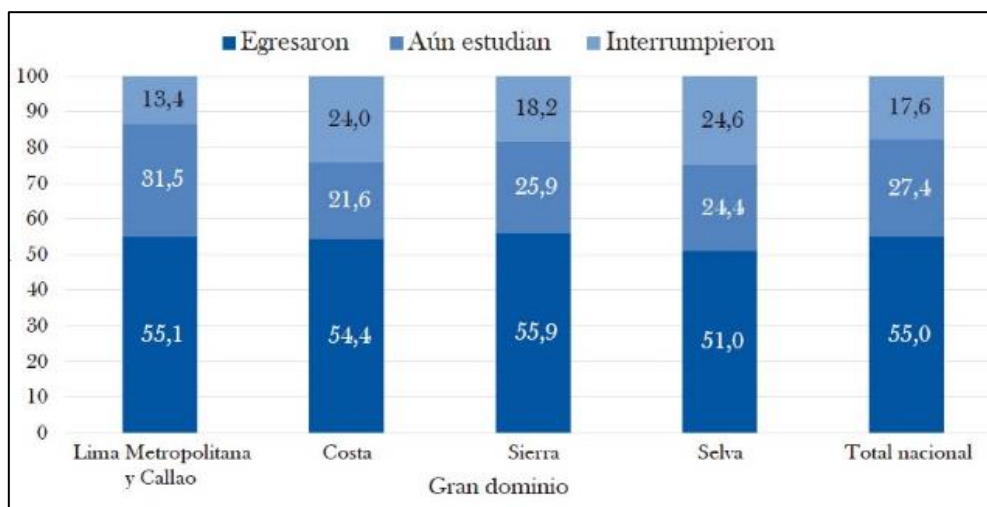


Figura 3. Porcentaje de jóvenes que egresaron, interrumpieron o aun estudian en la UNS

Fuente: (SUNEDU II, 2020)

El propio informe aportó datos que permitieron obtener el porcentaje de jóvenes que interrumpieron sus estudios y analizar esta interrupción a través de cuatro variables (referidas como características socioeconómicas): máximo nivel educativo de los padres, quintil de gasto, lengua materna y miembro del hogar dependientes.

Este análisis permitió identificar que el 47,7% de los jóvenes que interrumpieron sus estudios provenían de hogares con el nivel económico más bajo (Quintil 1 del Quintil de gastos) y que el 22,1% que lo hacía procedía de padres sin estudios superiores (ver figura 4).

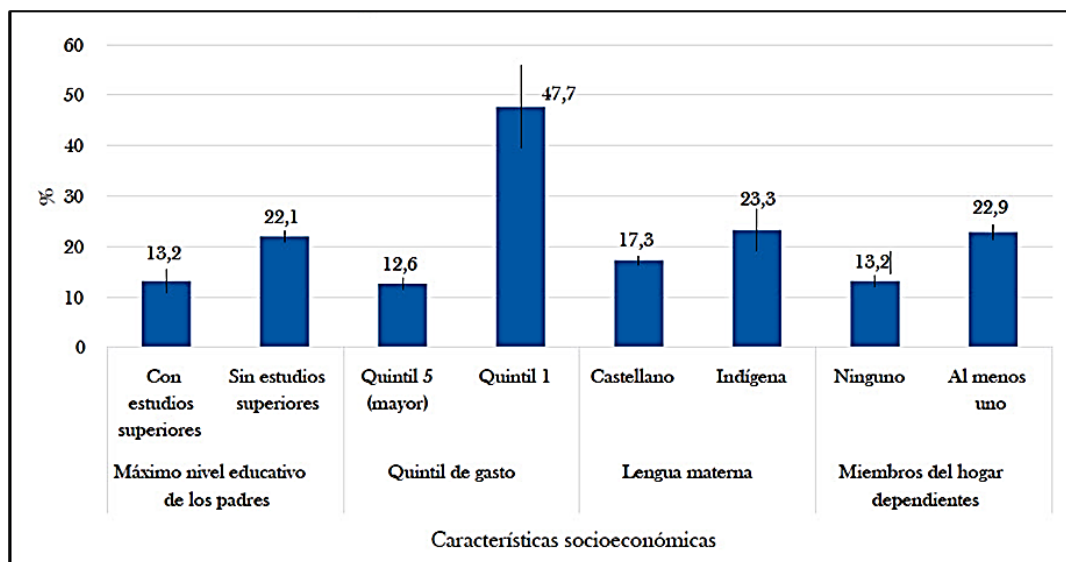


Figura 4. Comportamiento de los jóvenes que interrumpieron estudio según variables socioeconómicas
Fuente (SUNEDU II, 2020)

También resultaban alarmantes la calidad de notas que se publicaban en los diversos boletines estadísticos de la UNS. Durante los periodos 2011-2017, el promedio global de las asignaturas de todas las carreras de la Universidad sólo alcanzó 12.72 en la escala vigesimal. Estos bajos promedios provocan que muchos de sus egresados vean limitadas sus posibilidades de postularse a becas de estudios de posgrado.

Igualmente, preocupante resultaron los niveles de abandono luego de haber culminado los estudios, pero sin obtener el Grado académico y el Título profesional que ocurrieron en la UNS, específicamente en las escuelas profesionales de la Facultad de Ingeniería. La figura 5 contiene gráficas comparativas independientes que muestran el comportamiento de los Ingresantes en

relación con los Graduados y con los Titulados para cada una de estas escuelas. En este sentido se puede apreciar que entre 1993 y el 2018 hubo un comportamiento en zig zag con caídas y subidas abruptas de los porcentajes tanto de graduados como de titulados y que, de manera general, en toda la facultad estos números estuvieron muy distante de los porcentajes de ingresantes. Asimismo, quedó evidente que la EPISI y la

EPIE obtuvieron los porcentajes más bajos de la facultad al no alcanzar el 45% ni de graduados ni de titulados. También se observa que, pese a no sobrepasar el 70% en ninguno de los años, la EPIE en los titulados y la EPIA en los graduados sí lograron alcanzar este porcentaje en el 2018, siendo éstos los dos mejores resultados alcanzados en la facultad durante el periodo estudiado.

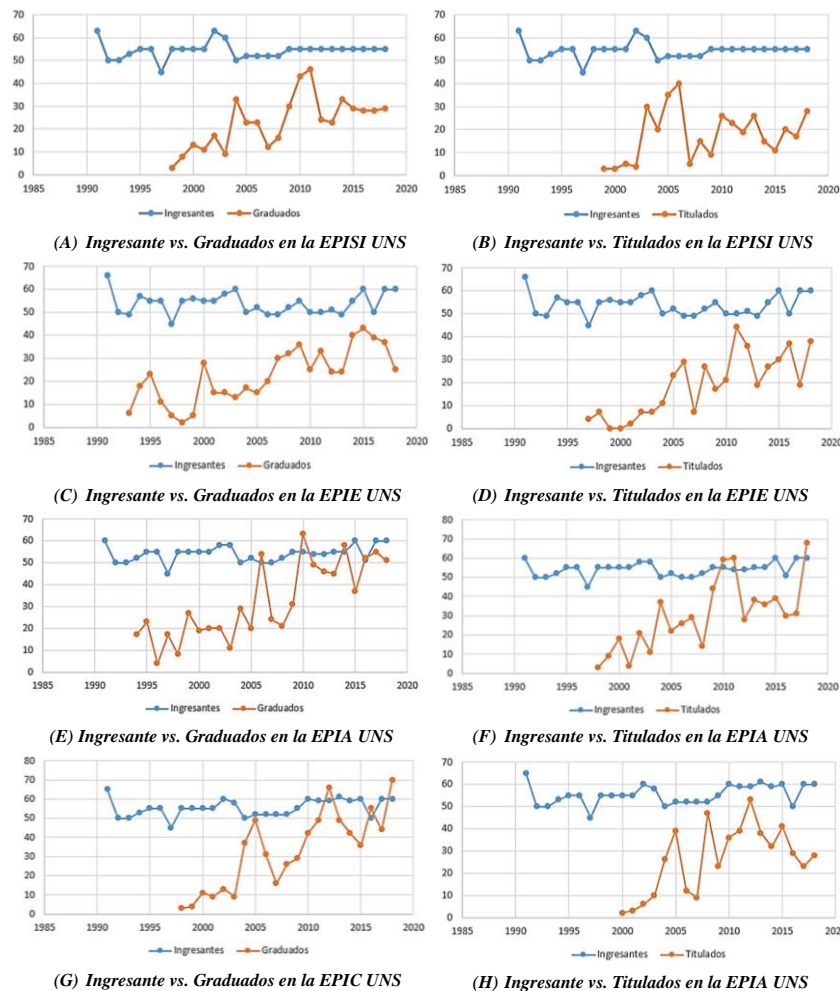


Figura 5. Comportamiento de ingresantes, graduados y titulados en: (A) y (B) EPISI, (C) y (D) EPIE, (E) y (F) EPIA y (G) y (H) EPIC de la UNS

Fuente: Compendios Estadísticos UNS 1987-2018

Estos resultados de los abandonos académicos, en particular en la UNS, constituyen la luz roja que indicó el curso de esta investigación.

Variables independientes más frecuentemente empleadas para la predicción de la deserción académica.

El análisis riguroso de las fuentes consultadas que reportaron estudios de la deserción o abandono académico permitió arribar a la identificación de las variables más frecuentemente empleadas en los diferentes contextos y países. También se apreció que éstas aparecen referidas de manera indistinta como factores, variables, características, sin embargo, independientemente de la manera que se mencionen se trata de las variables que determinan en mayor o menor medida el comportamiento del abandono o deserción académica y sobre las cuáles deberá accionarse con efectividad para buscar su disminución.

Se descubrió, de esta manera, que las variables más empleadas respondían a cuatro grandes grupos: individuales, académicas, institucionales y socioeconómicas. Igualmente se pudo identificar cuáles de las variables de estos grupos resultaban las más estudiadas.

Entre las variables individuales destacaban: la edad y género del estudiante, los antecedentes familiares, el apoyo de la familia, la integración a la sociedad y sus expectativas futuras.

En el grupo de las variables académicas se encontraron las relacionadas con: el rendimiento académico, la calidad del programa de estudio, el tipo de colegio, las ausencias diarias y los métodos de estudio. En relación con las variables institucionales se identificaron de peso por su aplicación las que tenían que ver con el compromiso institucional expresado esencialmente en: el apoyo académico, con recursos materiales y financieros, así como la interacción estudiante-docente.

Por último, las variables de tipo socioeconómico más relevantes dado por la frecuencia con que fueron estudiadas se relacionaban con: la dependencia económica del estudiante, su situación laboral y la de sus padres, así como las personas a cargo económicamente.

Muchas de estas variables fueron tomadas en cuenta para la propuesta resultante en este caso de estudio.

Del preprocesamiento de los datos de la Facultad de Ingeniería de la UNS

Se obtuvieron las datas del sistema de bienestar universitario conformada por 15

características según se muestra en la tabla 1 y la de Graduados y titulados con tres

características relacionadas al estado de graduación (ver tabla 2).

Tabla 1

Lista de características de la data del sistema de bienestar universitario

Número	Característica	Registros	Tipo de dato
1	Código	767	Int64
2	Fecha de Nacimiento	767	string
3	Lugar de Nacimiento	767	string
4	Sexo	767	string
5	Estado civil	767	string
6	Celular	767	string
7	Dependencia económica	767	string
8	Condición de trabajo responsable de familia	767	string
9	Alumno trabaja	767	string
10	Ingreso total familiar	767	float64
11	Lugar de procedencia	767	string
12	Material de vivienda	767	string
13	Servicios	767	string
14	Tipo de colegio	767	string
15	Veces que se postuló	767	Int64

Tabla 2

Lista de características de la data Graduados y titulados

Número	Característica	Registros	Tipo de dato
1	Código	2873	Int64
2	Escuela profesional	2873	String
3	Fecha diploma	2873	String

Como resultado del proceso de análisis de la calidad de la data del sistema de bienestar universitario generado por el Python en un mapa de calor en cuanto a nulos y faltantes

(ver figura 6) se descubrió que la característica Alumno trabaja y Tipo de colegio, le faltaban más del 90 % de sus valores.

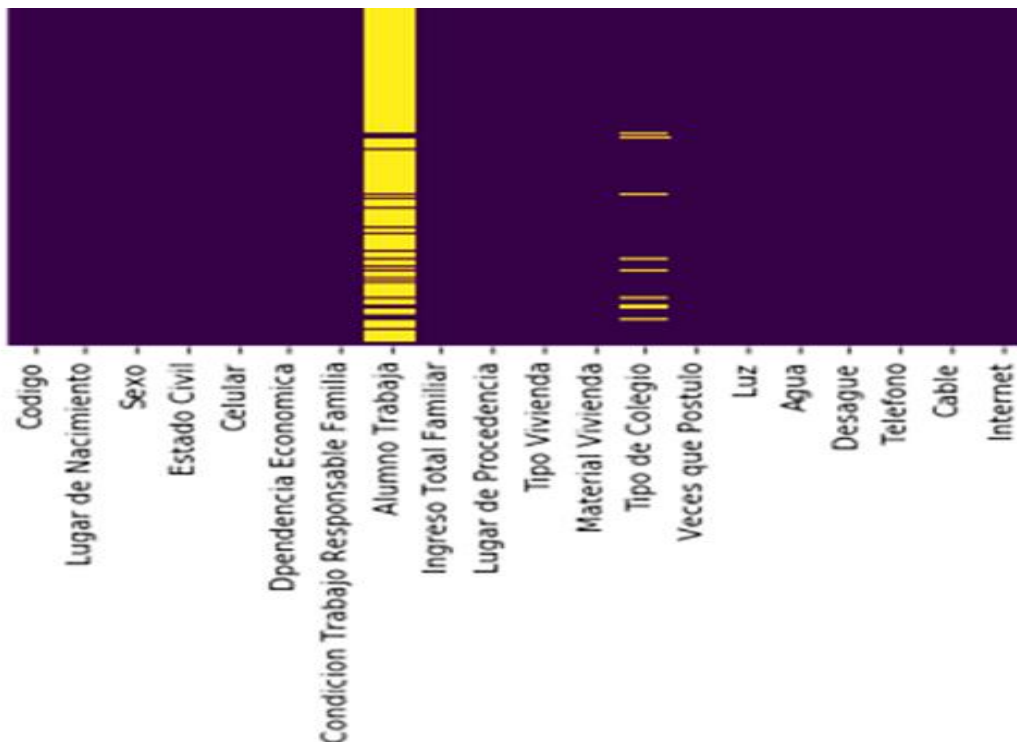


Figura 6. Mapa de calor para evaluar la calidad de la data del sistema de bienestar universitario

Fuente: Software Python

Tanto las características con exceso de valores nulos

como: Lugar de nacimiento, Estado civil, Tipo de vivienda, Veces que postuló y Material de la vivienda fueron descartadas al considerarse con demasiado sesgo.

El proceso de depuración de la data del sistema de bienestar universitario arrojó otras deficiencias. Se apreció un mal diseño de las bases de datos transaccionales lo que provocaba la presencia de campos multivaluados como el caso de la característica Servicio, también se descubrió la existencia de identificadores de

los registros con diferente formato y tipo de dato.

Para mejorar la calidad de los datos se unieron las datas a través de los identificadores de registros ya estandarizados, se eliminaron los campos sin valor y se categorizaron aquellos que se encontraban muy dispersos. También se modificó la característica Servicios al estar multivaluada con los elementos: luz, agua, desagüe, teléfono, cable e internet. Esta característica quedó separada de manera independiente en cada uno de estos seis servicios para proponer las variables predictoras.

Además, se reajustaron a valores booleanos (0 y 1) las características Sexo y Celular (si tenía o no). Las características: Dependencia económica, Condición de trabajo, responsable de familia y Lugar de procedencia fueron normalizados a valores numéricos.

También se descartó la característica Fecha de nacimiento por no ser relevante para el estudio.

Como consecuencia de la depuración de la data de Graduados y titulados sólo se consideró de valor predictivo la característica Código pues resultaba indicativa del estudiante graduado.

Se obtuvo un dataset maestro (ver estructura en tabla 3) resultante del merge de las datas depuradas.

Tabla 3

Dataset maestro

Número	Característica	Registros	Tipo de dato
1 (*)	Código	767	String
2	Sexo	767	Int64
3	Celular	767	Int64
4	Dependencia económica	767	Int64
5	Condición de trabajo responsable familia	767	Int64
6	Ingreso total familiar	767	category
7	Lugar de procedencia	767	Int64
8	Luz	767	Int64
9	Agua	767	Int64
10	Desagüe	767	Int64
11	Teléfono	767	Int64
12	Cable	767	Int64
13	Internet	767	Int64
14	Promedios	767	Float64
15	Número de semestres	767	Float64
16	Estudiante	767	Int64
17	Egresado	767	Int64
18	Graduado	767	Int64
19	Titulado	767	Int64

Leyenda: (*) El código se excluye para efectos de evaluación

Propuesta de las características que deberán formar parte del modelo predictivo

Como resultado de todo el proceso de depuración de las datas y la evaluación final de la calidad a la dataset de entrenamiento maestro se realizó la propuesta de las

características que deberán formar parte del modelo predictivo, éstas agrupadas en académicas y no académicas según muestra la tabla 4. Las características de la 2 a las 14 se corresponden con las variables predictivas (variables independientes), mientras que las 4 últimas conforman la función objetivo.

Tabla 4

Características seleccionadas para el modelo de deserción académica

Número	Categoría	Característica	Valores
1		Sexo	M/F
2		Celular	V/F
3		Dependencia económica	Rango
4		Condición de trabajo responsable de familia	Rango
5		Ingreso total familiar	Rango
6	Características no académicas (Demográficas)	Lugar de procedencia	0-9
7		Luz	V/F
8		Agua	V/F
9		Desagüe	V/F
10		Teléfono	V/F
11		Cable	V/F
12		Internet	V/F
13			Promedios
14		Número de semestres	Rango
15	Características académicas	Estudiante	V/F
16		Egresado	V/F
17		Graduado	V/F
18		Titulado	V/F

Fuente: elaboración propia

Teniendo en cuenta la problemática de este caso de estudio, estas 18 características a las que se arribaron deberán formar parte del

modelo de predicción multiclase para la deserción académica, el que posibilitará proyectar, a partir de las características

académicas y socioeconómicas (variables dependientes del modelo) cuántos de los estudiantes y egresados lograrán alcanzar el grado de bachiller o titularse.

CONCLUSIONES

La deserción académica constituye una problemática que afecta a varios países del mundo con significativa presencia en países Latinoamérica, y pese a que Perú aparece entre los de mejores resultados de la región sólo el 65% de los que ingresan a la educación superior logran el grado de bachiller.

Los resultados de abandono académico sin haber obtenido Grado académico ni Título profesional que muestra la UNS y la calidad del promedio global de las asignaturas en todas sus carreras ponen luz roja a esta problemática evidenciando la necesidad de su estudio.

Las variables de tipo individual, académica, institucional y socioeconómica halladas durante el análisis de la literatura, así como los resultados del procesamiento y depuración de las datas de la UNS constituyeron los referentes para la propuesta de las variables a emplear en el modelo de predicción académica.

Las 18 características a las que se arribaron deberán formar parte del modelo de

predicción multiclase para la deserción académica, el que posibilitará proyectar cuántos de los estudiantes y egresados lograrán alcanzar el grado de bachiller o titularse profesionalmente.

REFERENCIAS BIBLIOGRÁFICAS

- Abarca Rodríguez, A., & Sánchez Vindas, A. (2005). La deserción estudiantil en la educación superior: El caso de la Universidad de Costa Rica. *Revista Electrónica Actualidades Investigativas En Educación*, 5. <https://doi.org/10.15517/aie.v5i4.9186>
- Arco-Tirado, J. L., Fernández-Martín, F. D., & Fernández-Balboa, J.-M. (2011). The impact of a peer-tutoring program on quality standards in higher education. *Higher Education*, 62(6), 773–788. <https://doi.org/10.1007/s10734-011-9419-x>
- Bean, J. P. (1985). Interaction Effects Based on Class Level in an Explanatory Model of College Student Dropout Syndrome. *American Educational Research Journal*, 22(1), 35–64. <https://doi.org/10.3102/00028312022001035>

- Bendangnuksung, & Prabu, D. (2018). Students' Performance Prediction Using Deep Neural Network. 1171–1176.
- Castaño, E., Gallón, S., Gómez, K., & Vásquez, J. (2004). Deserción estudiantil universitaria: una aplicación de modelos de duración. *Lecturas de Economía*, (60), 39–65. Retrieved from <https://www.redalyc.org/articulo.oa?id=155217798002>
- CLABES. (2019). Noveno Congreso Latinoamericano sobre el abandono en la educación superior. Retrieved from <https://www.urosario.edu.co/CLABES/S/inicio/>
- Colvin, J. W. (2015). Peer Mentoring and Tutoring in Higher Education BT - Exploring Learning & Teaching in Higher Education (M. Li & Y. Zhao, Eds.). https://doi.org/10.1007/978-3-642-55352-3_9
- Donoso Díaz, S., & Arias R., Ó. (2010). Retención de estudiantes y éxito académico en la educación superior análisis de buenas prácticas. Retrieved from https://www.google.com.pe/url?sa=t&rct=j&q=&esrc=s&source=web&cad=rja&uact=8&ved=2ahUKEwipy6ipgovwAhX7TjABHeJoBnUQFjABegQIBhAD&url=https%3A%2F%2Fwww.cned.cl%2Ffile%2F1927%2Fdownload%3Ftoken%3Dug9_b9uA&usg=AOvVaw2DCa_n1O4ht95y1UD6Jo_M
- Dresel, M., & Rindermann, H. (2011). Counseling University Instructors Based on Student Evaluations of Their Teaching Effectiveness: A Multilevel Test of its Effectiveness Under Consideration of Bias and Unfairness Variables. *Research in Higher Education*, 52(7), 717–737. <https://doi.org/10.1007/s11162-011-9214-7>
- Ethington, C. A. (1990). A psychological model of student persistence. *Research in Higher Education*, 31(3), 279–293. <https://doi.org/10.1007/BF00992313>
- Fernández, S. G. (2018). Rendimiento Académico en Educación Superior: Desafíos para el Docente y Compromiso del Estudiante. *Revista Científica de La UCSA*, 5, 55–63. Retrieved from http://scielo.iics.una.py/scielo.php?script=sci_arttext&pid=S2409-87522018000300055&nrm=iso
- Ferreira, M. M., Avitabile, C., Álvarez, J., Paz, F. H., & Urzúa, S. (2017).

- Momento Decisivo: La Educación Superior en América Latina y el Caribe.
- Fishbein, M. A., & Ajzen, I. (1975). Belief, attitude, intention and behaviour: An introduction to theory and research (Vol. 27).
- Forero, L. D., Piñeros, Y. F., & Rodríguez, J. I. (2019). Machine Learning for the Identification of Students at Risk of Academic Desertion BT - Learning Technology for Education Challenges (L. Uden, D. Liberona, G. Sanchez, & S. Rodríguez-González, Eds.). Cham: Springer International Publishing.
- Hellas, A., Ihantola, P., Petersen, A., Ajanovski, V. V, Gutica, M., Hynninen, T., ... Liao, S. N. (2018). Predicting Academic Performance: A Systematic Literature Review. Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education, 175–199.
<https://doi.org/10.1145/3293881.3295783>
- Hernández, C. A. (2016). Diagnóstico del rendimiento académico de estudiantes de una escuela de educación superior en México. *Revista Complutense de Educación*, 27(3) SE-Artículos).
https://doi.org/10.5209/rev_RCED.2016.v27.n3.48551
- Jia, J.-W., & Mareboyana, M. (2014). Predictive Models for Undergraduate Student Retention Using Machine Learning Algorithms BT - Transactions on Engineering Technologies (H. K. Kim, S.-I. Ao, & M. A. Amouzegar, Eds.). Dordrecht: Springer Netherlands.
- Kostopoulos, G., Lipitakis, A.-D., Kotsiantis, S., & Gravvanis, G. (2017). Predicting Student Performance in Distance Higher Education Using Active Learning BT - Engineering Applications of Neural Networks (G. Boracchi, L. Iliadis, C. Jayne, & A. Likas, Eds.). Cham: Springer International Publishing.
- Ministerio de Educación Nacional Colombia. (2015). Guía para la implementación del Modelo de Gestión de permanencia y graduación estudiantil en Instituciones de Educación Superior. Imprenta Nacional de Colombia.
- Morí Sánchez, M. del P. (2012). DESERCIÓN UNIVERSITARIA EN ESTUDIANTES DE UNA UNIVERSIDAD PRIVADA DE IQUITOS. *Revista Digital de*

- Investigación En Docencia Universitaria, 6(1 SE-Artículos de investigación).
- <https://doi.org/10.19083/ridu.6.42>
- Müller, A. C., & Guido, S. (2016). Introduction to Machine Learning with Python. O'Reilly Media, Inc.
- Oancea, B., Dragoescu, R., & Ciucu, S. (2013, April 26). Predicting students' results in higher education using neural networks. 190–193. Jelgava, Latvia.
- Puchi, R., Moraga, A., & Villagran, W. (2016). Modelo De Seguimiento De La Retención Y El Rendimiento De Los Estudiantes De Pregrado De Primer Año En La Universidad De La Frontera.
- Salvador-Blanco, L., & Garcia-Valcarcel Muñoz-Repiso, A. (1989). El rendimiento académico en la universidad de Cantabria: abandono y retraso en los estudios (M. de E. y Ciencia, Ed.). Retrieved from <https://sede.educacion.gob.es/publive/nta/el-rendimiento-academico-en-la-universidad-de-cantabria-abandono-y-retraso-en-los-estudios/investigacion-educativa/1300>
- Scott, G., Shah, M., Grebennikov, L., & Singh, H. (2008). Improving Student Retention A University of Western Sydney Case Study. Journal of Institutional Research, 14(1), 9–23.
- Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A Review on Predicting Student's Performance Using Data Mining Techniques. Procedia Computer Science, 72, 414–422. <https://doi.org/https://doi.org/10.1016/j.procs.2015.12.157>
- Spady, W. G. (1970). Dropouts from higher education: An interdisciplinary review and synthesis. Interchange, 1(1), 64–85. <https://doi.org/10.1007/BF02214313>
- SUNEDU I. (2018). Informe Bienal sobre la realidad universitaria peruana. Retrieved from <https://www.gob.pe/institucion/sunedu/informes-publicaciones/606251-informe-bienal-sobre-la-realidad-universitaria-2018>
- SUNEDU II. (2020). II informe bienal sobre la realidad universitaria en el Perú. Retrieved from <https://www.gob.pe/institucion/sunedu/informes-publicaciones/1093280-ii-informe-bienal-sobre-la-realidad-universitaria-en-el-peru>
- Tinto, V. (1975). Drop-Outs From Higher Education: A Theoretical Synthesis of Recent Research. Review of

Educational Research, 45, 89–125.

<https://doi.org/10.2307/1170024>

Torres Guevara, L. E. (2012). Retención estudiantil en la educación superior. Revisión de la literatura y elementos de un modelo para el contexto colombiano (Editorial Pontificia Universidad Javeriana, Ed.). Bogotá.

Whitlock, J. L. (2018). Using Data Science and Predictive Analytics to Understand 4-Year University Student Churn.

Contacto

Hugo Esteban Caselli Gismondi

hcasellig@uns.edu.pe