

UNIVERSIDAD PERUANA UNIÓN

SEGUNDA ESPECIALIDAD EN INGENIERÍA

Segunda Especialidad en Estadística Aplicada para
Investigación



**Modelo de Machine Learning para predecir la deserción
universitaria a partir de factores socioeconómicos y
académicos**

Tesis para obtener el Título de Segunda Especialidad en Estadística Aplicada
para Investigación

Autores:

Sandra Sofia Izquierdo Marin
William Robert Torres Navarrete

Asesor:

Dr. Juan Jesús Soria Quijaite

Lima, abril 2024

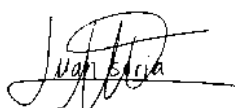
DECLARACIÓN JURADA DE ORIGINALIDAD DE TESIS

Yo, Juan Jesús Soria Quijaite, docente de la Segunda Especialidad en Estadística Aplicada para Investigación de la Universidad Peruana Unión.

DECLARO:

Que la presente investigación titulada: “**Modelo de Machine Learning para Predecir la Deserción Universitaria a Partir de Factores Socioeconómicos y Académicos**” de los autores Sandra Sofia Izquierdo Marin y William Robert Torres Navarrete tiene un índice de similitud de 6% verificable en el informe del programa Turnitin, y fue realizada en la Universidad Peruana Unión bajo mi dirección.

En tal sentido asumo la responsabilidad que corresponde ante cualquier falsedad u omisión de los documentos como de la información aportada, firmo la presente declaración en la ciudad de Lima, a los 02 días del mes de abril del año 2024.



Juan Jesús Soria Quijaite

ACTA DE SUSTENTACIÓN

En Lima, Ñaña, Villa unión a 2 días del mes de abril del año 2024, siendo las 10:20 horas, se reunieron de forma online sincrónica, bajo la dirección del presidente del jurado Dra. Ethel Altez Ortiz, el secretario Mg. Lizeth Huanca López; los demás miembros: y el PhD. Javier Linkolk López Gonzales y el Mg. Nemias Saboya Rios y el asesor Dr. Juan Jesús Soria Quijaite con el propósito de administrar el acto académico de sustentación de Tesis de la Segunda Especialidad titulada “*Modelo de Machine Learning para predecir la deserción universitaria a partir de factores socioeconómicos y académicos*”, conducente a la obtención del Título de Segunda Especialidad Profesional de ingeniería: Estadística Aplicada para Investigación.

El presidente inició el acto académico de sustentación invitando al candidato hacer uso del tiempo determinado para su exposición. Concluido la exposición, el Presidente invitó a los demás miembros del Jurado a efectuar las preguntas, cuestionamientos y aclaraciones pertinentes, los cuales fueron absueltos por el candidato. Luego se produjo un receso para las deliberaciones y la emisión del dictaminador del Jurado.

Posteriormente, el jurado procedió a dejar constancia escrita sobre la evaluación en la presente acta, con el dictamen siguiente:

Candidatos: William Robert Torres Navarrete
Sandra Sofia Izquierdo Marin

CALIFICACIÓN	ESCALAS			Mérito
	Vigesimal	Literal	Cualitativa	
Aprobado	15	B-	Con nominación de Bueno	Muy bueno

(*) Ver parte posterior

Finalmente, el Presidente del Jurado invitó al candidato a ponerse de pie, para recibir la evaluación final. Además, el Presidente del Jurado concluyó el acto académico de sustentación, procediéndose a registrar a registrar las firmas respectivas.

Presidente

Secretario

Asesor(a)

Miembro

Miembro

Candidato(a)

AGRADECIMIENTOS

A nuestro asesor, el Dr. Juan Soria Quijaite por el apoyo y motivación en el proceso de la realización de la presente tesis.

A la Universidad Peruana Unión, a través de su coordinado de Segunda Especialidad en Estadística Aplicada para Estadística, Dr. Javier López Gonzales.

Los autores

DEDICATORIA

A Dios, padre todo poderoso, quien me ha dado la fuerza y motivación para forjar mi vida.

A mi madre, Gloria, quien, con su ejemplo, me ha enseñado a luchar por mis sueños.

A mi esposo, Jorge, e hijos, Adrián, Benjamín y Gabriel, quienes son mi motor diario, mis alegrías, mis orgullos. Los amo con todo mi ser.

Sandra Sofia Izquierdo Marin

Dedico esta investigación a nuestro a dios por ser el inspirador para cada uno de mis pasos dados en mi convivir diario; a mi madre por ser los guías en el sendero de cada acto que realizo hoy, mañana y siempre; a mis hijos, por ser el incentivo para seguir adelante con este objetivo, a mi esposa Linda por darme su amor y apoyo constante en cada proyecto que emprendo.

William Robert Torres Navarrete

ÍNDICE

DECLARACIÓN JURADA DE ORIGINALIDAD DE TESIS	ii
ACTA DE SUSTENTACIÓN	iii
AGRADECIMIENTOS.....	iv
DEDICATORIA.....	v
ÍNDICE	vi
Resumen.....	1
Abstract.....	2
INTRODUCCIÓN.....	3
MÉTODOLOGÍA	6
RESULTADOS	8
CONCLUSIONES.....	20
REFERENCIAS	21
ANEXOS	30

Modelo de Machine Learning para Predecir la Deserción Universitaria a Partir de Factores Socioeconómicos y Académicos

Machine Learning Model to Predict University Dropout from Socioeconomic and Academic Factors

Resumen

El estudio se centró en generar un modelo de Machine Learning que pueda predecir la deserción de los estudiantes de la Universidad Católica de Trujillo en el periodo 2021-2022, a partir de factores socioeconómicos y académicos, cuya efectividad sea superior al 75%. El diseño metodológico de este estudio fue mixto, ya que no solo utilizó un método estadístico, sino también el análisis documental de casos, los cuales están recolectado de forma numérica y categórica para determinar la relación entre factores socioeconómicos y académicos y la deserción universitaria de los estudiantes, así como explicar y predecir a través de ellos escenarios futuros de la variable dependiente. La población estudiada consistió en dos cohortes (2021 y 2022) de 11 carreras de nivel pregrado dictadas en la Universidad Católica de Trujillo; totalizando 3156 alumnos, de los cuales 789 han abandonado los estudios (índice de deserción del orden del 25%). En conclusión, el modelo propuesto es significativo (p -value: 0) en base a cuatro factores predictores (lugar de procedencia, hermanos, institución educativa de egreso y cantidad de cursos desaprobados). Además, los factores socioeconómicos abordados si se relacionan de forma significativa con la deserción universitaria, a excepción del género (masculino y femenino), cuya relación no es significativa ($\chi^2=,338$). Además, los resultados obtenidos mediante el análisis de chi- cuadrado indica que todos los factores académicos están relacionados de forma significativa. Finalmente, el modelo es capaz de clasificar correctamente 98.7% de las observaciones de entrenamiento.

Palabras claves: Machine Learning, deserción, factores socioeconómicos y académicos.

Abstract

The study focused on generating a Machine Learning model that can predict the dropout of students at the Catholic University of Trujillo in the period 2021-2022, based on socioeconomic and academic factors, whose effectiveness is greater than 75%. The methodological design of this study was mixed, since it not only used a statistical method, but also the documentary analysis of cases, which are collected numerically and categorically to determine the relationship between socioeconomic and academic factors and the university dropout of the students. students, as well as explain and predict future scenarios of the dependent variable through them. The studied population consisted of two cohorts (2021 and 2022) of 11 undergraduate courses taught at the Catholic University of Trujillo; totaling 3,156 students, of which 789 have dropped out (dropout rate of around

25%). In conclusion, the proposed model is significant (p-value: 0) based on four predictor factors (place of origin, siblings, educational institution of graduation and number of failed courses). Furthermore, the socioeconomic factors addressed are significantly related to university dropout, with the exception of gender (male and female), whose relationship is not significant ($\chi^2=.338$). Furthermore, the results obtained through the chi-square analysis indicate that all academic factors are significantly related. Finally, the model is able to correctly classify 98.7% of the training observations.

Keywords: Machine Learning, dropout, socioeconomic and academic factors.

INTRODUCCIÓN

La deserción universitaria, al ser un proceso de abandono de la formación profesional, representa un serio inconveniente que impacta a numerosas instituciones de educación superior globalmente, particularmente en los primeros años de estudio universitario; el cual trae como consecuencias, un gran impacto en la economía y desarrollo del país por cuanto se minimiza la mano de obra calificada y amplia los índices de desigualdad de oportunidad laboral (Zárate y Mantilla, 2014).

La deserción universitaria tiene una serie de resultados negativos para estudiantes e instituciones de educación superior. Los estudiantes que abandonan la universidad se enfrentan a mayores dificultades económicas y sociales y tienen menos probabilidades de obtener empleos bien remunerados y satisfactorios; además, las dichas instituciones enfrentan una pérdida de recursos, una disminución de la retención de estudiantes y una disminución en la calidad educativa (Quintero, 2016).

Aunque las instituciones efectúan esfuerzos para abordar el problema de la deserción universitaria, sigue siendo un desafío importante. Una de las razones de esto es que la deserción universitaria es un problema complejo y multifacético que se ve influenciado por una serie de factores; por ejemplo, en países de Asia, uno de los principales factores que contribuyen a la deserción universitaria es la falta de recursos económicos tanto para el ingreso a la universidad, como para el sustento a lo largo de la carrera universitaria, entendiéndose no solo la carrera en sí, sino los materiales para las diversas practicas pre profesionales (Niyogisubizo et al., 2022) y la falta de apoyo familiar al considerar que dicha carrera no es lucrativa, por ende, no debería ser estudiada (Shuja et al., 2016). Otro factor importante en la deserción es la falta de motivación, ya sea en la propia carrera -tomada como una obligación en la vida adulta- o bien por las expectativas negativas de conseguir empleo posterior al estudio (Li y Xue, 2023).

En Europa, la problemática de la deserción se genera por los altos costos de la propia vida universitaria, donde la mayoría de las estudiantes consideran

conveniente una educación particular sobre la educación estatal, por considerarlas de mejor calidad (Contini y Zotti,

2020), así mismo, el fracaso académico o decepción hacia la propia carrera -al ver que no cubre las expectativas o agrado propio- también genera deserción universitaria (Zajac y Komendant, 2019).

En Norte América también se aprecia que la falta de economía para mantener una vida universitaria es otro indicador de deserción (Goldrick et al., 2016), así como la falta de habilidades y técnicas de estudio (Mintz, 2019). En América del Sur, la pobreza y falta de economía impulsa el abandono universitario en los primeros ciclos de la formación profesional (Améstica et al., 2020) y (Carneiro y Sampaio, 2014).

En nuestro país, la alta tasa de pobreza y el difícil acceso a la educación universitaria es uno de las principales causas de deserción académica (Lopez, 2022), teniéndose en cuenta además, que la gran mayoría de colegios de la zona sierra y selva del Perú no cuentan con el nivel de preparación y exigencia para la vida universitaria, incluso la Evaluación Censal de Estudiantes (ECE) reportó que estudiantes que completaban su educación secundaria, no sabían leer ni escribir correctamente, lo cual les dificulta su correcta inserción y adaptación a las exigencias de la vida universitaria (Lopez, 2022).

Como se puede apreciar, existe una multicausalidad en la deserción universitaria, donde cada factor juega un rol crucial para el aumento de la probabilidad de esta problemática, más aún cuando el principal sector de riesgo son los estudiantes de los primeros ciclos. Entre los principales factores observados en las diversas partes del mundo, se aprecia la recurrencia de los factores económicos y académicos derivados de la baja calidad educacional en niveles previos a la universidad (Lopez, 2022). Estos factores interactúan de maneras complejas y no lineales, por lo que su abordaje en nuestra sociedad y principalmente en la localidad de Trujillo resulta de vital importancia a fin de poder predecir escenarios futuros y, por consiguiente, poder optar por las mejores alternativas para evitar que este inadecuado desenlace llegue a concretarse; optando en esta ocasión por analizar y pronosticar esta situación

de manera más precisa mediante herramientas basadas en machine learning (Valero et al., 2022).

Actualmente, los modelos utilizados para predecir la deserción universitaria están enfocados en la caracterización multidimensional de las mismas ya sea utilizando características académicas o socioeconómicas, sin embargo, estos métodos no proporcionan herramientas para realizar una predicción que permita generar una alerta temprana en la deserción; ay que un problema de los modelos estadísticos actuales es que necesitan de supuestos rigurosos para ajustarse a mejor resultados (Valero et al., 2022).

Razón por la cual, el presente estudio se basa en el uso del machine learning, toda vez que tiene la ventaja de ajustar modelos a una amplia variedad de datos, además que posee la capacidad de ajustar modelos con alta precisión, lo que se traduce en predicciones más exactas y, por lo tanto, en mejores resultados. Investigadores como Barreno (2012), Prieto (2015), Arismendy y Morales (2018) indican que uno de los modelos más factibles a emplear para la deserción universitaria, es el modelo de regresión logística binaria, que por su naturaleza permite indicar si un estudiante se encuentra en riesgo de deserción o no; es así que, Arismendy et al. (2018), observó que su modelo de regresión logística planteado en base a sexo, cursos repetidos y convivencia con ambos padres logró predecir y acertar el

76% de casos de deserción universitaria. El estudio de Morocho (2020), planteó un modelo de regresión logística en base a edad, estado civil, carrera profesional, cantidad de hermanos y la presencia de hijos, que logró una predicción del 82.72% de casos de deserción, además, la investigación de Solís et al. (2022), logró que, mediante la edad, estado civil, situación laboral y salario, se puedan predecir el 99,06% de los casos de deserción universitaria.

Estos estudios demuestran, no solo la eficacia de la regresión logística para poder trabajar con datos cuantitativos y cualitativos, sino también permite focalizar adecuadamente cuales son las variables que poseen un mayor peso explicativo en el modelo, permitiendo así, identificar las principales causas de la

deserción universitaria y, por ende, poder tomar las acciones preventivas o correctivas correspondientes.

Es así que la investigación se encuentra delimitada en tiempo y espacio a los estudiantes registrados en los semestres académicos 2020 y 2021 de la Universidad Católica de Trujillo, quienes cursaron el primer ciclo de formación profesional y de quienes se analizarán sus factores socioeconómicos y académicos, los cuales podrían estar asociados a la deserción universitaria.

En cuanto al objetivo general, se buscó Generar un modelo de Machine Learning que pueda predecir la deserción de los estudiantes de la Universidad Católica de Trujillo en el periodo 2021-2022, a partir de factores socioeconómicos y académicos, cuya efectividad sea superior al 75%, en tanto que los objetivos específicos fueron: Describir los factores socioeconómicos y académicos de los estudiantes de la Universidad Católica de Trujillo en el periodo 2021-2022, Identificar los factores socioeconómicos y académicos que más se relacionan con la deserción en los estudiantes de la Universidad Católica de Trujillo en el periodo 2021-2022 y determinar la relación que existe entre los factores socioeconómicos y académicos en los estudiantes de la Universidad Católica de Trujillo en el periodo 2021- 2022.

MÉTODOLOGÍA

El diseño es no experimental, toda vez que no se manipulan las variables independientes o dependiente, sino que se basan en sus comportamiento y patrones observados para poder realizar proyecciones futuras en base a modelos estadísticos (Zacarías y Supo, 2020).

El diseño metodológico de este estudio fue mixto, ya que no solo utilizó un método estadístico, sino también el análisis documental de casos, los cuales están recolectado de forma numérica y categórica para determinar la relación entre factores socioeconómicos y académicos y la deserción universitaria de los estudiantes, así como explicar y predecir a través de ellos escenarios futuros de la variable dependiente (Hernández & Mendoza, 2018). Además, se trata de una

investigación aplicada y tecnología porque en base a una investigación básica se formulan hipótesis de trabajo para resolver la deserción universitaria; además es tecnológica, pues se basa en la comprobación del beneficio de las tecnologías basadas en inteligencia artificial (modelo de Machine Learning) (CONCYTEC, 2018).

El nivel de investigación fue descriptivo, correlacional, explicativa y predictiva; descriptivo porque el objetivo es reconocer y detallar los elementos socioeconómicos y académicos que contribuyen al abandono universitario, además de analizar la correlación existente entre dichas variables (Hernández & Mendoza, 2018). Además, es explicativa porque analiza las causas o relaciones entre las variables, es decir, por qué ocurren ciertos fenómenos o comportamientos, elaborando un modelo teórico que permita entender los elementos que inciden en el abandono de estudios superiores; y predictiva porque tiene como objetivo hacer predicciones sobre la deserción académica, basándose en datos previos y en un modelo de Machine Learning (Hair Jr. et al., 2014). Este estudio se basa en un diseño transversal o de corte transversal, ya que se recopilan los datos en un solo punto en el tiempo para analizar la relación entre las variables (Hernández & Mendoza, 2018).

La población estudiada consistió en dos cohortes (2021 y 2022) de 11 carreras de nivel pregrado dictadas en la Universidad Católica de Trujillo; totalizando 3156 alumnos, de los cuales 789 han abandonado los estudios (índice de deserción del orden del 25%).

En cuanto a la Técnica, esta fue la revisión documental (registro administrativo, base de datos secundario), y sobre el Instrumento, este fue la Base de datos académica (interna) de la universidad católica de Trujillo del 2021-2022 en relación con las variables. La plantilla contendrá 7 campos: 6 columnas para las variables independientes, 1 columna para el nombre de los estudiantes y 2367 filas correspondientes a los estudiantes de la UCT periodo

2021-2022. La información para la plantilla será descargada del Software ERP University.

RESULTADOS

Tabla 1

Modelo de regresión logística para predecir la deserción universitaria

Coefficients				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.3600	0.7150	4.699	2.61e-06***
Procedencia.T	-2.1104	0.4529	-4.660	3.16e-06***
Procedencia.CH	2.7962	0.4119	6.789	1.13e-11***
Hermanos2	-0.3328	0.4405	-0.756	0.44989
Hermanos3	-0.9653	0.4774	-2.022	0.04316 *
Hermanos4	-23.0871	2807.0025	-0.008	0.99344
Hermanos5	-23.0709	2276.7585	-0.010	0.99192
IE_egreso.P	3.9835	0.3219	12.375	< 2e-16 ***
Cursos_desaprobados1	-2.2353	0.7257	-3.080	0.00207 **
Cursos_desaprobados2	-3.7301	0.7249	-5.146	2.66e-07***
Cursos_desaprobados3	-26.5408	2576.3405	-0.010	0.99178
Cursos_desaprobados4	-26.6861	1648.9741	-0.016	0.98709

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken

to be 1) **Null deviance:** 3549.46 on 3155

degrees of freedom **Residual deviance:**

226.88 on 3144 degrees of freedom

AIC: 250.88

Tabla 2*Intervalos de confianza del modelo*

	2.5 %	97.5 %
(Intercept)	2.085119	4.92099482
Procedencia.T	-2.987029	-1.19592998
Procedencia.CH	2.010840	3.63669096
Hermanos2	-1.211392	0.52556704
Hermanos3	-1.917121	-0.03423378
Hermanos4	-921.337716	101.70594114
Hermanos5	-793.030114	70.59820357
IE_egreso.P	3.392140	4.66481691
Cursos_desaprobados1	-3.824784	-0.94639505
Cursos_desaprobados2	-5.335010	-2.46128815
Cursos_desaprobados3	-1004.642576	59.96458595
Cursos_desaprobados4	-668.733568	25.75903654

Diferencia de residuos: 3322.581**Grados libertad: 11****p-value: 0**

Acorde a la tabla 1, el modelo propuesto es significativo (p-value: 0) en base a cuatro factores predictores (lugar de procedencia, hermanos, institución educativa de egreso y cantidad de cursos desaprobados).

Tabla 3

Factores socioeconómicos y académicos de los estudiantes de la Universidad Católica de Trujillo en el periodo 2021-2022.

Factores		Fr.	%	Media	Mediana	Moda	D.S.	
FACTORES SOCIALES	Género	Varón	137	4.3				
		Mujer	3019	95.7				
	Procedencia	Trujillo	2612	82.8				
		Chiclayo	369	11.7				
		Chimbote	175	5.5				
	Número de hermanos	1 - 2	2119	67.1	2.16	2.00	1	1.133
	3 - 4	834	26.4					
	5 a más	203	6.5					
FACTORES ECONÓMICOS	Nivel socioeconómico	NSE E	422	13.4				
		NSE D	677	21.5				
		NSE C	1589	50.3				
		NSE B	467	14.8				
		NSE A	1	0.0				
	Condición laboral padres	Trabajo informal	681	21.6				
		Trabajo formal	2475	78.4				
	Tipo de vivienda	Arrendada	13	0.4				
Propia		3143	99.6					
FACTORES ACADÉMICOS	Institución educativa de procedencia	Pública	688	21.8				
		Privada	2468	78.2				
	Repitencia académica en la secundaria	Si	641	20.3				
		No	2515	79.7				
	Récord de asistencia	75% a menos	105	3.3	88.25	90.00	90	7.292
		76% a	636	20.1				
		80% a	939	29.8				
		90%						
		91% -	1476	46.8				
	Cursos desaprobados	Ninguno	958	30.4	1.42	1.00	0	1.330
1 - 2		1593	50.5					
3 - 4		605	19.1					

Como se aprecia en la tabla 3, la mayor parte de los participantes del estudio son mujeres (95.66%), mientras que solo una minoría son varones (4.34%). El 82.76% de estudiantes de la Universidad católica de Trujillo, provienen de la propia ciudad de Trujillo, observándose que la segunda proporción de estudiantes (11.69%) son provenientes de la ciudad de Chiclayo y solo el 5.54% de ellos provienen del distrito de Chimbote. Se observa que la mayor

parte de estudiantes tienen entre uno a dos hermanos, siendo muy pocos los que poseen entre 4 a 5 hermanos. El 13.37% de estudiantes poseen un nivel socioeconómico E, es decir poseen en sus hogares un ingreso mensual promedio de s/1300.00 soles, mientras que el 21.45% de ellos poseen un nivel socioeconómico D, es decir registran un ingreso mensual promedio de s/2480.00 soles. El 50.35% de estudiantes poseen un nivel socioeconómico C, demostrando un ingreso mensual de s/3970.00 soles, el 14.80% de los estudiantes que conforman la muestra poseen nivel socioeconómico B, es decir, ostentan un ingreso mensual promedio de s/7020.00 soles y solo el 0.03% de ellos registran nivel socioeconómico A, es decir poseen un ingreso mensual promedio de s/12660.00 soles.

El 78.42% de los padres de los integrantes de la muestra poseen empleo formal, ya sea de forma dependiente o independiente, pero que cuentan con acceso a servicios de salud y aseguramiento de un ingreso constante; mientras que solo el 21.58% de ellos presentan trabajo informal, es decir no cuentan con todos los beneficios sociales. El 99.59% de los estudiantes de la UCT poseen vivienda propia, mientras que solo el 0.41% de ellos poseen vivienda en situación de alquiler. Además, el 78.20% de estudiantes proceden de instituciones educativas privadas, mientras que el 21.80% de ellos han egresado de las instituciones educativas del estado. El 79.69% de estudiantes de la UCT no registran ninguna repitencia académica durante su formación en nivel secundaria, mientras que el 20.231% de ellos poseen el antecedente de al menos una repitencia de grado durante la educación secundaria. La mayor parte de estudiantes de la UCT posee un registro de asistencia superior al 90% de sus clases y la mayor parte de estudiantes poseen solo entre cero a un curso desaprobado (véase tabla 4).

Tabla 4

Deserción universitaria en los estudiantes de la Universidad Católica de Trujillo en el periodo 2021-2022.

Deserción universitaria			
		Frecuencia	Porcentaje
Válido	Si	789	25,0
	No	2367	75,0
	Total	3156	100,0

La tabla N°4 muestra la cantidad de estudiantes que han desertado de la educación universitaria, apreciándose que el 25% de ellos han abandonado sus estudios, mientras que el 75% continúan de forma satisfactoria su formación profesional.

Tabla 5

Análisis de relación entre los factores socioeconómicos y académicos con la deserción universitaria.

Factores			Deserción		Prueba de Valor	Chi-cuadrado Sig. A. (b)
			Si	No		
FACTORES SOCIALES	Género	Varón	39	98	,918	,338
		Mujer	750	2269		
	Procedencia	Trujillo	338	2274	1191.759	<,001
		Chiclayo	325	44		
		Chimbote	126	49		
	Numero de hermanos	1 - 2	235	1884	1249,646	<,001
3 - 4		351	483			
5 a más		203	0			
FACTORES ECONÓMICOS	Nivel socioeconómico	NSE E	422	0	1635,860	<,001
		NSE D	217	460		
		NSE C	139	1450		
		NSE B	11	456		
		NSE A	0	1		
	Condición laboral padres	Trabajo informal	681	0	2605,135	<,001
		Trabajo formal	108	2367		
	Tipo de vivienda	Arrendada	12	1	31,540	<,001
		Propia	777	2366		
	FACTORES ACADÉMICOS	Institución educativa de procedencia	Pública	647	41	2236,605
Privada			142	2326		
Repitencia académica en la secundaria		Si	555	86	1626,987	<,001
		No	234	2281		
Récord de asistencia		75% a menos	86	19	2208,056	<,001
		76% a 80%	493	143		
		81% a 90%	210	729		
		91% - 100%	0	1476		
Cursos desaprobados		Ninguno	6	952	2322,867	<,001
		1 - 2	178	1415		
	3 - 4	605	0			

Los resultados plasmados en la tabla 5 demuestran que los factores socioeconómicos abordados si se relacionan de forma significativa con la deserción universitaria, a excepción del género (masculino y femenino), cuya relación no es significativa ($\chi^2=,338$). Además, los resultados obtenidos mediante el análisis de chi-cuadrado indica que todos los factores académicos están relacionados de forma significativa.

Tabla 6

Modelo propuesto en base a los 10 factores del estudio

Coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.835e+01	2.391e+05	0.000	1.000
Genero.M	-2.865e+00	2.981e+04	0.000	1.000
Procedencia.T	-1.289e+01	1.974e+04	-0.001	0.999
Procedencia.CH	3.626e+00	4.711e+04	0.000	1.000
Hermanos2	-9.056e-02	8.841e+03	0.000	1.000
Hermanos3	-1.012e+00	9.445e+03	0.000	1.000
Hermanos4	-2.430e+01	3.509e+04	-0.001	0.999
Hermanos5	-1.958e+01	4.480e+04	0.000	1.000
Nivel_socioeconomico.E	-2.631e+01	2.360e+05	0.000	1.000
Nivel_socioeconomico.D	-4.666e+01	1.936e+05	0.000	1.000
Nivel_socioeconomico.C	-1.430e+01	1.186e+05	0.000	1.000
Nivel_socioeconomico.B	-9.983e+00	4.459e+04	0.000	1.000
Trabajo.I	3.067e+01	2.871e+04	0.001	0.999
Vivienda.A	-2.010e+00	6.999e+04	0.000	1.000
IE_egreso.P	2.874e+01	4.930e+03	0.006	0.995
Repitencia.N	1.239e+01	2.002e+04	0.001	1.000
Asistencia75	-6.370e+01	2.121e+05	0.000	1.000
Asistencia76	-6.283e+01	2.116e+05	0.000	1.000
Asistencia77	-5.115e+01	2.420e+05	0.000	1.000
Asistencia78	-6.318e+01	2.116e+05	0.000	1.000
Asistencia79	-4.087e+01	1.964e+05	0.000	1.000
Asistencia80	-4.414e+01	2.043e+05	0.000	1.000
Asistencia81	-4.306e+01	2.056e+05	0.000	1.000
Asistencia82	-4.529e+01	2.098e+05	0.000	1.000

Asistencia83	-4.821e+01	2.840e+05	0.000	1.000
Asistencia84	-4.848e+01	2.771e+05	0.000	1.000
Asistencia85	-5.875e+01	2.449e+05	0.000	1.000
Asistencia86	-4.205e+01	2.594e+05	0.000	1.000
Asistencia87	-4.418e+01	2.282e+05	0.000	1.000
Asistencia88	-5.992e+01	2.511e+05	0.000	1.000
Asistencia89	-2.218e+01	2.287e+05	0.000	1.000
Asistencia90	-2.350e+01	2.120e+05	0.000	1.000
Asistencia91	-1.241e+01	2.120e+05	0.000	1.000
Asistencia92	-2.346e+01	2.125e+05	0.000	1.000
Asistencia93	-5.921e+01	3.193e+05	0.000	1.000
Asistencia94	-2.311e+01	2.124e+05	0.000	1.000
Asistencia95	-2.253e+01	2.130e+05	0.000	1.000
Asistencia96	-4.289e+01	2.210e+05	0.000	1.000
Asistencia97	-2.464e+01	2.129e+05	0.000	1.000
Asistencia98	-2.183e+01	2.176e+05	0.000	1.000
Asistencia99	-4.151e+01	2.074e+05	0.000	1.000
Asistencia100	-4.587e+01	2.219e+05	0.000	1.000
Cursos_desaprobados1	-1.744e+01	2.561e+04	-0.001	0.999
Cursos_desaprobados2	-1.898e+01	2.475e+04	-0.001	0.999
Cursos_desaprobados3	-4.173e+01	3.159e+04	-0.001	0.999
Cursos_desaprobados4	-2.556e+01	5.997e+04	0.000	1.000

La tabla 6 demuestra que un modelo de predicción basada en 10 factores predictores no es viable, por lo que, para poder implementar un modelo adecuado para la predicción de deserción universitaria, se debe realizar un filtrado de los modelos más relevantes, en este caso, se optó por realizar tres modelos individuales, el primero basado solo en factores sociales, el segundo basado solo en factores económicos y el tercero basado solo en factores académicos tal como detalla la siguiente tabla:

Tabla 7*Reconocimiento y depuración de factores para implementación del modelo*

Modelo basado en los 3 factores sociales propuestos				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.84997	0.20929	4.061	4.88e-05 ***
Genero.M	-0.20074	0.24438	-0.821	0.411
Procedencia.T	-2.15375	0.15134	-14.231	< 2e-16 ***
Procedencia.CH	2.11056	0.16589	12.722	< 2e-16 ***
Hermanos2	-0.09351	0.17616	-0.531	0.596
Hermanos3	-1.57628	0.16816	-9.374	< 2e-16 ***
Hermanos4	-20.40291	450.38950	-0.045	0.964
Hermanos5	-20.51166	369.39343	-0.056	0.956
Modelo basado en los 3 factores económicos propuestos				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-5.573	9786.770	-0.001	1.000
Nivel_socioeconomico.E	32.885	30505.456	0.001	0.999
Nivel_socioeconomico.D	-7.736	25781.816	0.000	1.000
Nivel_socioeconomico.C	1.865	15267.754	0.000	1.000
Nivel_socioeconomico.B	-7.642	5788.805	-0.001	0.999
Trabajo.I	27.435	1141.775	0.024	0.981
Vivienda.A	-10.748	2380.685	-0.005	0.996
Modelo basado en los 4 factores económicos propuestos				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.971e+01	7.579e+04	0.000	0.999793
IE_egreso.P	5.055e+00	8.713e-01	5.802	6.55e-09 ***
Repitencia.N	1.625e+01	2.005e+03	0.008	0.993533
Asistencia75	-2.747e+01	7.583e+04	0.000	0.999711
Asistencia76	-2.568e+01	7.583e+04	0.000	0.999730
Asistencia77	-2.927e+01	7.583e+04	0.000	0.999692
Asistencia78	-2.789e+01	7.583e+04	0.000	0.999707
Asistencia79	-2.748e+01	7.583e+04	0.000	0.999711
Asistencia80	-2.761e+01	7.583e+04	0.000	0.999709
Asistencia81	-2.744e+01	7.583e+04	0.000	0.999711

Asistencia82	-2.689e+01	7.583e+04	0.000	0.999717
Asistencia83	-6.243e+00	9.316e+04	0.000	0.999747
Asistencia84	-7.688e+00	9.131e+04	0.000	0.999933
Asistencia85	-5.609e+00	8.783e+04	0.000	0.999949
Asistencia86	-6.243e+00	9.316e+04	0.000	0.999947
Asistencia87	-7.864e+00	8.095e+04	0.000	0.999922
Asistencia88	-6.327e+00	9.073e+04	0.000	0.999944
Asistencia89	-2.618e+00	8.209e+04	0.000	0.999975
Asistencia90	-4.104e+00	7.594e+04	0.000	0.999957
Asistencia91	-3.230e+00	7.602e+04	0.000	0.999866
Asistencia92	1.524e+01	7.607e+04	0.000	0.999940
Asistencia93	-5.508e+00	1.147e+05	0.000	0.999962
Asistencia94	1.326e+01	7.598e+04	0.000	0.999861
Asistencia95	-2.506e+00	7.645e+04	0.000	0.999974
Asistencia96	1.502e+01	7.633e+04	0.000	0.999843
Asistencia97	-4.639e+00	7.619e+04	0.000	0.999951
Asistencia98	-1.281e+00	7.841e+04	0.000	0.999987
Asistencia99	-5.977e+00	7.725e+04	0.000	0.999938
Asistencia100	-6.190e+00	7.695e+04	0.000	0.999936
Cursos_desaprobados1	-3.274e+00	1.522e+00	-2.151	0.031498 *
Cursos_desaprobados2	-5.535e+00	1.574e+00	-3.517	0.000437***
Cursos_desaprobados3	-2.857e+01	6.914e+03	-0.004	0.996703
Cursos_desaprobados4	-2.921e+01	4.509e+03	-0.006	0.994831

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'

0.1 ' ' 1

Acorde a la tabla 7, el modelo en basado en los factores sociales es significativo y acorde a los p-values mostrados, la contribución al modelo es significativa para los predictores de lugar de procedencia y número de hermanos; así mismo, dicho modelo es capaz de clasificar correctamente 91.06% de las observaciones de entrenamiento. Por otro lado, se aprecia que el modelo no es significativo debido a factores económicos. Finalmente, en la tabla 8 se observa que, para los factores académicos, el el modelo en conjunto

sí es significativo y acorde a los p-values, los predictores significativos son la I.E. de egreso y cantidad de cursos desaprobados, dicho modelo es capaz de clasificar correctamente 97.81% de las observaciones de entrenamiento.

En síntesis, para implementar un modelo predictivo funcional, la regresión logística debe abarcar únicamente los factores sociales y académicos significativos, dichos factores brindan un aporte adecuado para modelar la deserción universitaria.

Tabla 8

Evaluación de las predicciones en el modelo propuesto

Observaciones	0	1
Si	764	25
No	16	2351

La tabla 8 muestra el funcionamiento predictivo del modelo final propuesto, que está compuesto por los factores sociales: lugar de procedencia y número de hermanos, así como los factores académicos de institución de egreso y cantidad de cursos desaprobados, siendo así, un total de cuatro factores que contribuyen a modelar la deserción universitaria.

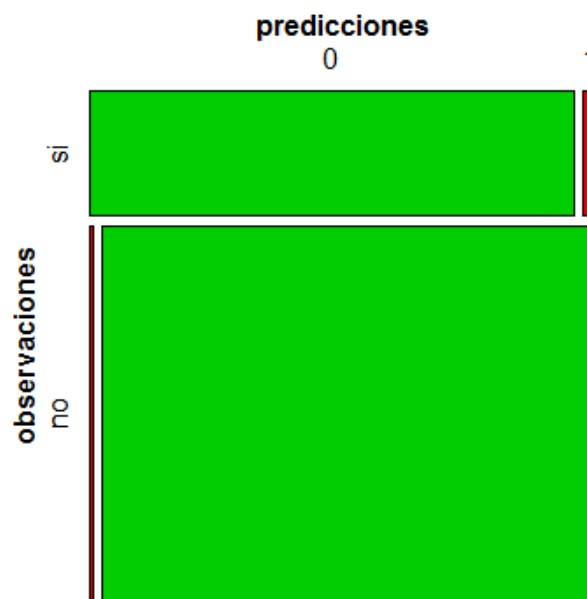


Figura 1

Mosaico de la distribución de falsos positivos y falsos negativos en la predicción de la deserción universitaria.

A razón de la tabla 8, la proporción de clasificación de observaciones se puede cuantificar de la siguiente forma: $(764+2351)/(764+2351+16+25) = 0.9870089$. Es decir, el modelo es capaz de clasificar correctamente 98.7% de las observaciones de entrenamiento.

CONCLUSIONES

- El modelo propuesto es significativo (p-value: 0) en base a cuatro factores predictores (lugar de procedencia, hermanos, institución educativa de egreso y cantidad de cursos desaprobados)
- Los factores socioeconómicos abordados si se relacionan de forma significativa con la deserción universitaria, a excepción del género (masculino y femenino), cuya relación no es significativa ($\chi^2=,338$). Además, los resultados obtenidos mediante el análisis de chi-cuadrado indica que todos los factores académicos están relacionados de forma significativa.
- Se optó por realizar un modelo de predicción de tres modelos individuales, el primero basado solo en factores sociales, el segundo basado solo en factores económicos y el tercero basado solo en factores.
- Para implementar un modelo predictivo funcional, la regresión logística debe abarcar únicamente los factores sociales y académicos significativos, dichos factores brindan un aporte adecuado para modelar la deserción universitaria.
- El modelo es capaz de clasificar correctamente 98.7% de las observaciones de entrenamiento.

REFERENCIAS

- Ajzen, I., & Cote, N. G. (2008). *Attitudes and the prediction of behavior: Attitudes and attitude change*. In W.D. Crano & R. Prislin (Eds.), pág. 289-311. <https://psycnet.apa.org/record/2008-09973-013>
- Albán, J., y Calero, J. (2017). *El rendimiento académico: aproximación necesaria a un problema pedagógico actual*. Revista Conrado, 13(58), 213-220. <http://conrado.ucf.edu.cu/index.php/conrado>
- Álvarez, P., Cabrera, L., González, M. & Bethencourt, J. (2006). *Causas del abandono y prolongación de los estudios universitarios*. Revista Paradigma, vol. 27, número 1, pág. 349 - 363. http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1011-22512006000100002&lng=es&tlng=es
- Amat, J. (2017). *Árboles de decisión, random forest, gradient boosting y C5.0*. International (CC BY 4.0). [Artículo web]. https://www.cienciadedatos.net/documentos/33_arboles_decision_random_forest_gradient_boosting_C50.html
- Améstica, L., King, A., Sanhueza, D. & Ramírez, V. (2020). *Efectos económicos de la deserción en la gestión universitaria: el caso de una universidad pública chilena*. Revista Hallazgos, 18(35), 209-231. <https://doi.org/10.15332/2422409X.5772>
- Arismendy, C. & Morales, N. (2018). *Modelo de regresión logística como alternativa para medir la probabilidad de deserción temprana en la Universidad de los Llanos periodo 2015-2 - 2018-1*. [Tesis de licenciatura, Universidad de los Llanos]. <https://repositorio.unillanos.edu.co/handle/001/1101>
- Ariza, S. & Marín, D. (2009). *Factores intervinientes en la deserción escolar de la Facultad de Psicología, Fundación Universitaria Los Libertadores*.

Revista Universitaria de Tesis Psicológica, número 4, pág. 72-85.
<https://www.redalyc.org/articulo.oa?id=139013586006>

Barreno, E. (2012). *Análisis Comparativo de modelos de clasificación en el estudio de la deserción universitaria*. Revista INTERFASES, vol. 5, número 5, pág. 45-82.
<https://revistas.ulima.edu.pe/index.php/Interfases/article/view/149>

Basualdo, M. (2005). *Deserción y repitencia en la educación superior universitaria en Paraguay*.
<https://projectes.uab.cat/accedes/content/deserc%C3%B3n-y-repitencia-en-la-educaci%C3%B3n-superior-universitaria-en-paraguay.html>

Berens, J., Schneider, K., Gortz, S., Oster, S., & Burghoff, J. (2019). *Early Detection of Students at Risk - Predicting Student Dropouts Using Administrative Student Data from German Universities and Machine Learning Methods*. Journal of Educational Data Mining, 11(3), 1-41.
<https://doi.org/10.5281/zenodo.3594771>

Bernal, A. (2005). *La familia como ámbito Educativo*. Barcelona: Rialp.
<https://revistas.unav.edu/index.php/estudios-sobre-educacion/article/view/25792>

Brea, M. (2005). *Deserción en la educación superior pública en República Dominicana*. <https://www.psicologiacientifica.com/desercion-educacion-superior-publica-republica-dominicana/>

Carneiro, A. & Sampaio, H. (2014). *Access to higher education in Brazil*. Widening Participation and Lifelong Learning. Volumen 16, Number 1.
https://www.academia.edu/32128570/Access_to_higher_education_in_Brazil

Caselli, H. & Urrelo, L. (2021). *Características para un modelo de predicción de la deserción académica universitaria: Caso Universidad Nacional de Santa*. Revista de Investigación Científica y Tecnológica Llamkasun

Trabajemos, vol. 2, Número 4, pág. 2-22.
<https://lamkasun.unat.edu.pe/index.php/revista/article/view/61>

Cevallos, M & Barahona, C. (2021). *Modelo para automatizar el proceso de predicción de la deserción en estudiantes universitarios en el primer año de estudio*. [Tesis de grado, Universidad Peruana de Ciencias Aplicadas (UPC)]. Repositorio Académico UPC.
<http://hdl.handle.net/10757/656740>

Chong, E. (2017). *Factores que inciden en el rendimiento académico de los estudiantes de la Universidad Politécnica del Valle de Toluca*. Revista Latinoamericana de Estudios Educativos (México), vol. XLVII, núm. 1, pág. 91-108. <https://www.redalyc.org/journal/270/27050422005/html/>

Contini, D., Zotti, R. (2022). *Do Financial Conditions Play a Role in University Dropout? New Evidence from Administrative Data*. In: Checchi, D., Jappelli, T., Uricchio, A. (eds) Teaching, Research and Academic Careers, vol 1, pág. 39-70.
https://link.springer.com/chapter/10.1007/978-3-031-07438-7_3

Cruz, E., González, M., & Rangel, J. (2022). *Técnicas de machine learning aplicadas a la evaluación del rendimiento y a la predicción de la deserción de estudiantes universitarios, una revisión*. Prisma Tecnológico, 13(1), 77-87. <https://doi.org/10.33412/pri.v13.1.3039>

Cuevas, M., & Estévez, M. (2017). *Técnicas de análisis para la mejora y*. Universidad Complutense de Madrid, Ingeniería Informática, Madrid.
<https://eprints.ucm.es/id/eprint/44411/1/MemoriaTFG.pdf>

De La Cruz, V. (2019). *Diseño de un modelo predictivo basado en machine learning para el control de la deserción de estudiantes en la Universidad Ricardo Palma*. [Tesis de grado, Universidad Nacional Tecnológica de Lima Sur]. Repositorio Digital Institucional la Universidad Nacional Tecnológica de Lima Sur.
<https://repositorio.untels.edu.pe/xmlui/handle/123456789/282>

- Espinoza, A. y Oré, E. (2017). *Principales factores socio- económicos que influyen en la calidad de vida de los jóvenes venezolanos inmigrantes de 18 - 25 años de la Organización No Gubernamental Unión Venezolana en la ciudad de Lima - Perú, 2017*. [Tesis de Maestría, Universidad Nacional Mayor de San Marcos, Lima, Perú]. <http://repositorio.unsa.edu.pe/bitstream/handle/UNSA/5084/TSescaae.pdf>
- Garbanzo, G. (2007). *Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública*. Revista educación, vol. 3, numero 1, pág. 43 - 63. <https://www.redalyc.org/articulo.oa?id=44031103>
- Goldrick, S., Broton, K., & Eisenberg, D. (2016). *Hungry to learn: Addressing food and housing insecurity among undergraduates*. Wisconsin HOPE Lab. <https://www.acct.org/product/hungry-learn-addressing-food-and-housing-insecurity-among-undergraduates-2016>
- Gonzales, L. (2023). *Regresión Logística - Teoría*. [Artículo web]. <https://aprendeia.com/algoritmo-regresion-logistica-machine-learning-teoria/>
- Gonzales, J. & Peñaloza, M. (2021). *Identificación y predicción de estudiantes en riesgo de deserción académica por medio de modelos basados en machine learning*. [Tesis de licenciatura, Fundacion Universitaria Los Libertadores]. Repositorio Insitucional. <https://repository.libertadores.edu.co/handle/11371/4188>
- Grosso, P. (2020). *Rendimiento académico: un recorrido conceptual que aproxima a una definición unificada para el ámbito superior*. Revista de Educación, 11(20), 87-102. https://fh.mdp.edu.ar/revistas/index.php/r_educ/article/view/4165
- Hair Jr, J., Black, W., Babin, B., & Anderson, R. (2014). *Multivariate data analysis* (8va ed.). Harlow, Reino Unido: Pearson.

[https://www.scirp.org/\(S\(lz5mqp453ed%20snp55rrgjct55\)\)/reference/referencespapers.aspx?referenceid=2965867](https://www.scirp.org/(S(lz5mqp453ed%20snp55rrgjct55))/reference/referencespapers.aspx?referenceid=2965867)

Hernández Sampieri, R., & Mendoza Torres, C. (2018). *Metodología de la investigación : Las rutas cuantitativa, cualitativa y mixta*. México: McGraw-Hill Education.

<https://virtual.cuautitlan.unam.mx/rudics/?p=2612>

Li, J. & Xue, E. (2023). *Dynamic Interaction between Student Learning Behaviour and Learning Environment: Meta-Analysis of Student Engagement and Its Influencing Factors*. Behavioral Sciences. 2023; vol. 13, número 1, pág. 59. <https://doi.org/10.3390/bs13010059>

Lopez, R. (2022). *Factores económicos y académicos de deserción universitaria en Colombia, Chile, Perú y México, periodo 2011-2021*. [Tesis de Licenciatura, Pontificia Universidad Católica del Perú]. Repositorio Digital Institucional.

<https://tesis.pucp.edu.pe/repositorio/handle/20.500.12404/21993>

Marquín, M. J. (2017). *Predicción del Rendimiento Académico Mediante Técnicas del Análisis Multivariado*. [Tesis de Maestría, Universidad Tecnología De Pereira UTP]. Repositorio institucional UTP Colombia. <https://hdl.handle.net/11059/8356>

Martínez, J., Pérez, E., Ferrás, Y. & Bermúdez, L. (2021). *Análisis predictivo de la deserción estudiantil en la carrera de Medicina*. Revista Edumecentro, vol. 13, número 3, pág. 217-236. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2077-28742021000300217

Ministerio de Educación de Argentina (2021). *Fuerte impulso del Gobierno a las Becas Progresar*. <https://www.argentina.gob.ar/noticias/fuerte-impulso-del-gobierno-las-becas-progresar>

Ministerio de Educación de Chile. (2023). *JUNAEB: Junta Nacional de Auxilio Escolar y nBecas*. <https://www.chileatiende.gob.cl/instituciones/AJ009>

Ministerio de Educación del Perú [MINEDU] (2022). *El impacto del COVID 19 en la salud mental y deserción universitaria: respuestas para la continuidad.*

<https://www.minedu.gob.pe/conectados/pdf/cursos/webinars/2021/24-de-agosoto-2021-el-impacto-del-covid-19-salud%20mental-desercion-universitaria-respuestas-para-la-continuidad.pdf>.

Ministerio Federal de Educación e Investigación (s/f). *Deutsch land stip endium*

https://www.deutschlandstipendium.de/deutschlandstipendium/de/home/home_nod_e.html

Mintz S. (2019). *The 8 Steps Institutions Need to Take to Improve Student Success.* <https://www.insidehighered.com/blogs/higher-ed-gamma/8-steps-institutions-need-to-improve-student-success>

Montalvo, J. & Sánchez, N. (2023). Modelo de predicción de deserción: Un estudio de caso de estudiantes de la Universidad Politécnica Estatal del Carchi. *Revista Prometeo: Conocimiento científico*, vol. 3 número 2, pág. 32-34.

<https://prometeojournal.com.ar/index.php/prometeo/article/view/34>

Morocho, K. (2020). *Modelo de regresión logística de la deserción estudiantil en un Instituto Tecnológico en el Cantón Daule.* [Tesis de maestría, Escuela Superior Politécnica del Litoral]. DSpace en ESPOL.

<https://www.dspace.espol.edu.ec/handle/123456789/51577>

Mosquera, R., Castrillón, O., & Parra, L. (2018). *Máquinas de Soporte Vectorial, Clasificador Naïve Bayes y Algoritmos Genéticos para la Predicción de Riesgos Psicosociales en Docentes de Colegios Públicos Colombianos.* *Revista Información Tecnológica*, vol. 29, número 6, pág. 153 - 162. <https://dx.doi.org/10.4067/S0718-07642018000600153>

Niyogisubizo J., Liao, L., Nziyumva E., Murwanashyaka E. & Nshimyumukiza, P. (2022).

Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization. Journal Computers and Education: Artificial Intelligence, Volumen 3. <https://doi.org/10.1016/j.caeai.2022.100066>

Orellana, J. (2019). *¿Qué es machine learning y por qué es tan popular?* <https://www2.ucuenca.edu.ec/254-espanol/investigacion/blog-de-ciencia/ano-2019/julio-2019/1222-machine-learning>

Perez, C. & Rojas, L. (2020). *Diseño de un sistema para predecir la deserción de los alumnos mediante Machine learning en la Universidad Tecnológica del Perú.* [Tesis de grado, Universidad Tecnológica del Perú]. Repositorio Digital Institucional. <https://repositorio.utp.edu.pe/handle/20.500.12867/3843>

Perez, E., & Cupani, M. (2005). *Predictores de rendimiento académico en la escuela.* Redalyc. <https://www.redalyc.org/pdf/3350/335027178002.pdf>

Prieto, C. (2015). *Uso de Regresión Logística para predecir deserción estudiantil temprana.* [Tesis de licenciatura, Universidad de los Andes]. Repositorio Institucional. <https://repositorio.uniandes.edu.co/bitstream/handle/1992/18302/u721906.pdf>

Programas para el Bienestar (2022). *Beca para el Bienestar Benito Juárez de Educación Superior.* <https://programasparaelbienestar.gob.mx/beca-bienestar-benito-juarez-educacion-superior/>

Quintero, I. (2016). *Análisis de las causas de deserción universitaria.* [Artículo de Revisión]. Repositorio de la Universidad Nacional Abierta y a Distancia UNAD. <https://repository.unad.edu.co/bitstream/10596/6253/1/23783211.pdf>

Rivera, K. (2021). *Modelo predictivo para la detección temprana de estudiantes con alto riesgo de deserción académica.* Revista Innovación y Software,

vol. 2, número 2, pág. 6-13.

<https://www.redalyc.org/journal/6738/673870839001/html/>

Rodriguez, J. (2015). *El rendimiento escolar y intervención del trabajo social*. Cuenca: Universidad de Cuenca

RUSA (2017). *Revolutionizing higher education in india*. <http://rusa.nic.in/>

Segura, M., Mello, J. & Hernández, A. (2022) *Machine Learning Prediction of University Student Dropout: Does Preference Play a Key Role?* Mathematics, 10, 3359. <https://doi.org/10.3390/math10183359>

Shuja, A., Ali, A., Khan, Burki, S. B., & Bilal, S. (2022). *Perspectives on the Factors Affecting Students' Dropout Rate During COVID-19: A Case Study From Pakistan*. SAGE Open, 12(2). <https://doi.org/10.1177/21582440221097378>

Skills Future (2022). *What Is Skillsfuture?* <https://www.skillsfuture.gov.sg/AboutSkillsFuture>

Solís, J., Quiroz, S. & Fosado, O. (2022). *Modelo de regresión logística para la estimación de la deserción escolar del posgrado en la Universidad Técnica de Manabí, Ecuador*. Revista Bases de la Ciencia, vol. 7, número 3, pág. 1-14. <https://revistas.utm.edu.ec/index.php/Basedelaciencia/article/view/5197>

TeleSurTV. (2018). *Educación para todos: Los logros de Lula y Haddad en Brasil*. [Reseña periodística]. <https://www.telesurtv.net/news/brasil-lula-haddad-logros-educacion--20180912-0020.html>

Top Global University Japan. (2023). *Reading the data: the present and future of the Top Global University Project*. [Artículo Web]. <https://tgu.mext.go.jp/en/data/index.html>

Valero, J., Navarro, Á., Larios, A. & Julca, J. (2022). *Deserción universitaria: Evaluación de diferentes algoritmos de Machine Learning para su predicción*. Revista de Ciencias Sociales (Ve), Volumen XXVIII,

numero 3, pág. 362 - 375.

<https://www.redalyc.org/journal/280/28071865024/html/>

Vanderkooy, A., Regier, E. & Lilly, M. (2019). *Investing in inclusive growth: A systematic review of the role of financial incentives to promote lifelong learning*. Educational Research Review, Volume 27, pages 176-190, ISSN 1747-938X.

<https://www.sciencedirect.com/science/article/abs/pii/S1747938X1830407X>

Viera, D., Flores, M. & Pariachi, E. (2020).

Wild, S., Rahn, S. & Meyer, T. (2023). *The relevance of basic psychological needs and subject interest as explanatory variables for student dropout in higher education*. Eur J Psychol Educ. <https://doi.org/10.1007/s10212-022-00671-4>

Xu, L., Wang, H., Chen, J., Zhang, Y., Huang, Z., Yu, C. (2022). *English Learning Stress, Self-Efficacy, and Burnout among Undergraduate Students: The Moderating Effect of Mindfulness and Gender*. Int. J. Environ. Res. Public Health 2022, 19, 15819. <https://doi.org/10.3390/ijerph192315819>

Zacarías, H. & Supo, J. (2020). *Metodología de la Investigación Científica: Para las Ciencias de la Salud y las Ciencias Sociales*. Independently published (24 junio 2020). ISBN-13: 979-8656825252

Zajac, T. & Komendant, A. (2019). *Premeditated, dismissed and disenchanting: higher education dropouts in Poland*. Tert Educ Manag número 25, pág. 1-16. <https://doi.org/10.1007/s11233-018-09010-z>

Zárate, R. & Mantilla, E. (2014). *La deserción estudiantil UIS, una mirada desde la responsabilidad social universitaria*. Revista Zona Próxima, número 21, pág. 121 -134. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S2145-94442014000200010

ANEXOS

Evidencia de sumisión o publicación del artículo en una revista de prestigio

[Applied Sciences] Manuscript ID: applsci-3003293 - Submission Received

EO susy@mdpi.com en nombre de Editorial Office <applsci@mdpi.com>
Para: Sandra Izquierdo
CC: William Robert Torres Navarrete; Juan Jesús Soria Quijaite
Mar 23/04/2024 15:12

Responder Responder a todos Reenviar

Este mensaje está en Inglés Traducir a Español No traducir nunca de Inglés

Dear Dr. Izquierdo-Marin,

Thank you very much for uploading the following manuscript to the MDPI submission system. One of our editors will be in touch with you soon.

Journal name: Applied Sciences
Manuscript ID: applsci-3003293
Type of manuscript: Article
Title: Machine Learning Model to Predict University Dropout from Socioeconomic and Academic Factors
Authors: Sandra Sofía Izquierdo-Marin *, William Robert Torres Navarrete *, Juan Jesús Soria Quijaite
Received: 23 Apr 2024
E-mails: sandra.izquierdo@upeu.edu.pe, robert.torres@upeu.edu.pe, jesussoria@upeu.edu.pe