

# Revista de Ciencias Sociales



# Deserción universitaria: Evaluación de diferentes algoritmos de Machine Learning para su predicción

Valero Cajahuanca, Julio Elvis\*  
Navarro Raymundo, Ángel Fernando\*\*  
Larios Franco, Alfredo César\*\*\*  
Julca Flores, Janett Deisy\*\*\*\*

## Resumen

La deserción universitaria se ha incrementado significativamente en Perú antes y aún más después de la pandemia de COVID-19, es por ello, que las universidades públicas necesitan identificar e implementar programas para disminuirla. El trabajo tiene como propósito determinar el algoritmo de Machine Learning que tiene mejor desempeño para detectar la deserción universitaria. Este análisis se basó en el estudio de la deserción universitaria en Perú entre 2018 y 2021. La población está compuesta de 652 estudiantes, para los datos de entrenamiento se utilizó el 30% y para los de prueba el 70% de una data set de 106 datos válidos, para el desarrollo de los modelos de clasificación se utilizó el lenguaje Python de Anaconda a través de sus distintas librerías, el tipo investigación es aplicada y diseño descriptivo. Se obtuvo como resultado que el algoritmo K-Nearest-Neighbor con una precisión de 0.91, tiene mejor desempeño para pronosticar la deserción universitaria con las variables académicas y socioeconómicas de los estudiantes. En conclusión, el modelo obtenido puede ayudar a predecir en los primeros ciclos de estudios, los alumnos más probables en abandonar sus estudios, así como, alertar a la oficina de bienestar, la necesidad y atención de tutorías individuales y grupales.

**Palabras clave:** Deserción universitaria; modelo de predicción; regresión logística; árbol de clasificación; K-Nearest-Neighbor.

---

\* Doctor en Ingeniería de Sistemas. Docente en la Universidad Nacional de Lima Sur, Perú. E-mail: [jvalero@untels.edu.pe](mailto:jvalero@untels.edu.pe) ORCID: <https://orcid.org/0000-0002-8522-6249>

\*\* Doctor en Ingeniería de Sistemas. Docente en la Universidad Nacional de Lima Sur, Perú. E-mail: [anavarro@untels.edu.pe](mailto:anavarro@untels.edu.pe) ORCID: <https://orcid.org/0000-0002-9385-6187>

\*\*\* Doctor en Administración. Docente en la Universidad Nacional de Lima Sur, Perú. E-mail: [alarios@untels.edu.pe](mailto:alarios@untels.edu.pe) ORCID: <https://orcid.org/0000-0002-4258-8549>

\*\*\*\* Maestra en Ciencias de la Computación. Docente en la Universidad Nacional de Lima Sur, Perú. E-mail: [jjulca@untels.edu.pe](mailto:jjulca@untels.edu.pe) ORCID: <https://orcid.org/0000-0002-2354-3388>

Recibido: 2022-03-11 · Aceptado: 2022-05-30

# University desertion: Evaluation of different Machine Learning algorithms for its prediction

## Abstract

University dropout has increased significantly in Peru before and even more so after the COVID-19 pandemic, which is why public universities need to identify and implement programs to reduce it. The purpose of the work is to determine the Machine Learning algorithm that has the best performance to detect university dropout. This analysis was based on the study of university dropouts in Peru between 2018 and 2021. The population is made up of 652 students, 30% were used for training data and 70% for test data from a data set of 106 valid data, for the development of the classification models, the Anaconda Python language was used through its different libraries, the type of research is applied and descriptive design. It was obtained as a result that the K-Nearest-Neighbor algorithm with an accuracy of 0.91, has better performance to predict university dropout with the academic and socioeconomic variables of the students. In conclusion, the model obtained can help predict, in the first cycles of studies, the students most likely to drop out of their studies, as well as alert the welfare office, the need and attention of individual and group tutoring.

**Keywords:** University dropout; prediction model; logistic regression; classification tree; K-Nearest-Neighbor.

## Introducción

Basado en los reportes del año 2018 al 2020, se puede evidenciar el incremento en la deserción de estudiantes; en este sentido, el Banco Mundial el 2018, informó que aproximadamente el 30% de los estudiantes universitarios interrumpieron sus estudios debido a distintas causas. Además, según publicación en el diario El Comercio (Alayo, 2020) unos 174.000 (18,6%) estudiantes universitarios dejaron sus estudios en este año 2020.

Por otra parte, también es importante mencionar el contexto en el Perú y en el mundo con la pandemia del Covid-19. Es evidente que, al estar las personas en confinamiento, esto produjo una crisis económica en la mayoría de las familias y consecuentemente en la capacidad de solventar los estudios universitarios en muchos casos. Cuando una persona abandona sus estudios, marca un indicador preocupante en las instituciones universitarias, no sólo se trata de la frustración de los alumnos, sino también de encontrar las

causas, identificar los factores que hacen que se materialice este abandono.

Para Denle (2010), una variable muy importante a manejar en las universidades es la retención universitaria; esto debido al prestigio y reputación de la marca de una organización, por lo cual, es significativo monitorear, así como controlar los índices de deserción. También menciona que muchas veces los *rankings* de universidades consideran este factor para su clasificación; finalmente, menciona que la gestión de las matrículas universitarias debería considerar esta variable.

A partir de esto se plantea la pregunta: ¿Cuál de los modelos de *machine learning* será más eficiente para predecir la deserción universitaria? *Machine Learning* es parte de la ciencia de los datos que contiene distintos tipos de algoritmos de inteligencia artificial; por otra parte, existen dos tipos de aprendizajes los supervisados y los no supervisados, esto es importante para el aprendizaje de la máquina, es por ello que el objetivo del estudio se orienta a obtener un modelo matemático que ayude a predecir la deserción universitaria.

En los centros universitarios se tiene información del perfil de los estudiantes desde el momento de su inscripción y se cuenta con información de interés, además en el transcurso de los semestres académicos también van registrando más datos importantes para el análisis, como datos económicos, académicos y de salud.

## 1. Fundamentación teórica

### 1.1. Machine Learning

Lugo (2020), sostiene que *Machine Learning* se refiere a la ciencia que permite que las computadoras aprendan y actúen como lo hacen los humanos, mejorando su aprendizaje a lo largo del tiempo de una forma autónoma, alimentándolas con datos e información en forma de observaciones e interacciones con el mundo real. Al respecto, Fagella (2020), manifiesta que una de las formas en las que se utiliza la tecnología de inteligencia artificial es en el comercio minorista, donde se está aplicando como innovación en todo el ciclo de productos y servicios.

Por otra parte, *Machine Learning* utiliza algoritmos de Inteligencia Artificial (IA), la cual, tal como lo señalan Arbeláez-Campillo, Villasmil y Rojas-Bahamón (2021), de contar con el apoyo de información cualitativa y cuantifica eficiente, pueden llegar a alcanzar muchos objetivos propuestos; en este caso, la finalidad de estos algoritmos es determinar los patrones tomando como fuente los datos que se obtengan para luego a través de un modelo hacer predicciones, para ello se pueden tomar datos estructurados o no (Digital55, 2020).

El lenguaje de máquinas es entonces un mecanismo que permite a la misma, entrenarse con datos de un hecho pasado y del

cual se tiene muchos registros. Es importante conocer el alcance de estas herramientas que en la actualidad se utilizan básicamente para la predicción a través del aprendizaje supervisado, y, por otra parte, para la clusterización a través de aprendizaje no supervisados.

### 1.2. Modelos de clasificación de machine learning

Son algoritmos de inteligencia artificial, que, a través de aprendizajes supervisados y no supervisados, ayudan a aprender a una máquina, estos producen modelos matemáticos que son capaces de generalizar comportamientos. La presente investigación va a determinar el más efectivo.

En ese sentido, existen muchos modelos de *Machine Learning*, como, por ejemplo: Regresión logística; Árboles de decisión, y *K-Means*. Así, la regresión logística, es una herramienta muy versátil para realizar una clasificación de varias clases. Para graficar la regresión se tiene una curva con forma de S, esto permite dividir los datos en grupos (Microsoft Azure, 2022).

Greenwood (1978), menciona que la toma de decisiones es una resolución de problemas, y como tal, se debe buscar la mejor alternativa, aplicar un diagnóstico adecuado, así como realizar una búsqueda minuciosa, para elegir la mejor alternativa basados en un proceso de ramificación. Este proceso es el árbol de decisiones, que permite estimar alternativas ante la solución de problemas. Por otra parte, es necesario definir operacionalmente o dimensionar la variable de estudio modelos de clasificación de *Machine Learning* de acuerdo a las subvariables, que se pueden ver en el siguiente Cuadro 1.

### Cuadro 1 Dimensiones de la variable independiente

Variable Independiente	Dimensiones
Modelos de clasificación de <i>Machine Learning</i>	Precisión
	Sensibilidad
	Especificidad

Fuente: Elaboración propia, 2021.

Barreno-Vereau (2012), en su trabajo: “Análisis comparativo de modelos de clasificación en el estudio de la deserción universitaria”, realizó la comparación de dos modelos de clasificación. Como conclusión arribó que utilizando los mismos datos de origen y de evaluación se obtuvo los mismos resultados.

### 1.3. Deserción universitaria

Para poder definir la deserción universitaria, es necesario entender que es un tema complejo debido a los diversos factores que intervienen; por ejemplo, se define a la deserción de forma cuantificada sobre la cantidad de la población universitaria que dejan de estudiar; también se da el caso de reservas y cambios de carrera universitaria. Existen varios factores identificados por diversos estudios, como se muestra en UPlanner (2017), donde destacan los problemas financieros, pobre preparación escolar, la carrera no convence al estudiante, conflicto entre el estudio, la familia y el trabajo, entre otros; sin embargo, se cree que, si bien hay factores comunes, cada sector académico tiene sus propias características como es el caso de las universidades públicas.

Asimismo, Duche et al. (2020) manifiestan que todos los años en Perú, aumentan los egresados de la escuela secundaria que ingresan a la universidad; no obstante, en los primeros años de estudios la deserción, así como el abandono estudiantil es cada vez mayor. Según los autores, el problema surge cuando estudiantes

con características sociales, educativas, psicológicas, determinadas entran a formar parte de un nuevo contexto completamente distinto y diverso al de la escuela, al cual estuvieron acostumbrados durante más de 10 años. Y tal como lo señala Llanos (2019), ese paso a la universidad puede representar una vida totalmente nueva en otro lugar, así como separarse de familias y amigos, que, entre otras circunstancias, podrían acarrear alguna inestabilidad emocional, conllevándoles a la dimisión de tales estudios.

Por su parte, Chalpartar et al. (2022) sostienen que la deserción estudiantil se entiende como la suspensión de las actividades académicas; y de acuerdo con Rico (2006), puede ser el resultado de la interacción de una serie de características o variables, tales como el contexto, factores económicos, sociales, demográficos, familiares, individuales, entre otros; por lo cual, se considera un fenómeno dinámico y multicausal. De allí, que según Ferreyra et al. (2017), la posibilidad que los estudiantes no se gradúen revela que la inversión que se realiza en educación superior conlleva riesgos que afectan a unos estudiantes más que a otros, puesto que algunos están menos preparados académicamente para la educación superior y por tanto resulta más probable que abandonen.

En este sentido, fue necesario dimensionar la variable deserción universitaria, que para este estudio se entiende, cuando se interrumpe de forma temporal o definitiva la educación en el sistema universitario. Por lo tanto, en el Cuadro 2 se puede observar las dimensiones de la variable deserción universitaria.

**Cuadro 2**  
**Definición operacional Deserción universitaria**

Variable Dependiente	Dimensiones
Deserción universitaria	Factor económico
	Factor académico
	Factor familiar
	(Salud)

Fuente: Elaboración propia, 2021.

Asimismo, para realizar el procesamiento de datos, se utilizó un programa estadístico y realizaron las siguientes actividades; primero, la recolección de datos (limpieza de datos); luego la tabulación de resultados (*data set*); después el procesamiento de la información (Librerías de Anaconda); posteriormente, la presentación, así como publicación de resultados; y finalmente, los aspectos éticos y regulatorios. El trabajo está sujeto al código de ética del Colegio Nacional de Ingenieros del Perú.

**2. Metodología**

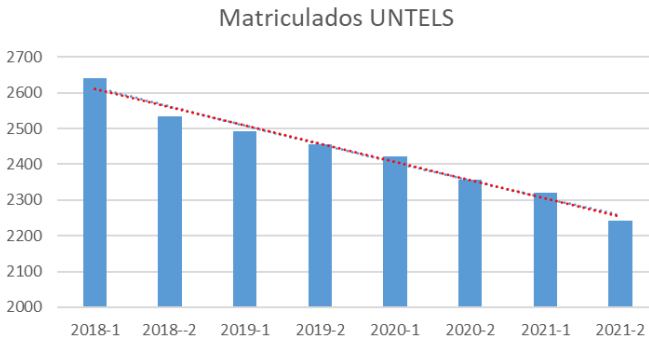
El trabajo de investigación es aplicado y tiene un alcance descriptivo, puesto que explica el comportamiento de una variable. Según Hernández, Fernandez y Baptista (2014), la investigación cuasi experimental mantiene la rigurosidad científica, el tipo de estudio también es transaccional descriptivo.

La población fue de 652 estudiantes, de la Universidad Nacional Tecnológica de Lima Sur (UNTELS) en Perú, para lo cual se tomó muestra de 106 datos válidos de estudiantes, a través del muestro aleatorio simple; y como técnica de recolección de datos, se utilizó la observación y la encuesta, a través de instrumentos como fichas de observación y los cuestionarios, respectivamente.

**4. Resultados y discusión**

**4.1. Deserción estudiantil**

En el Gráfico I, se muestran los datos estadísticos sobre la deserción universitaria desde el año 2018 hasta el año 2021, en la Universidad Nacional Tecnológica de Lima Sur (UNTELS) en Perú, encontrándose que, para el primer semestre del 2018, se contaba con más de 2.600 estudiantes en toda la universidad, llegando a contar con poco más de 2.200 en el segundo semestre del 2021, con un retiro, abandono o deserción de más de 400 estudiantes durante ese periodo.



Fuente: Elaboración propia, 2021 a partir de datos de la Oficina de Tecnología de la Información y Comunicaciones (OTIC) - Registros Académicos de la UNTELS.

**Gráfico I: Matriculados UNTELS (2018-2021)**

En este sentido, el presente trabajo revisa los estudios contemporáneos, analiza las nuevas técnicas de procesamiento de datos, selecciona los mejores algoritmos y determina un modelo que será de gran ayuda para predecir situaciones futuras. Por lo cual, se hace una revisión de las técnicas, algoritmos y modelos descritos, así como se busca confirmar los conocimientos aprendidos en el estado del arte. La comparación de estas técnicas de inteligencia artificial, permitirán determinar el de mejor rendimiento para aplicar al tema de deserción universitaria.

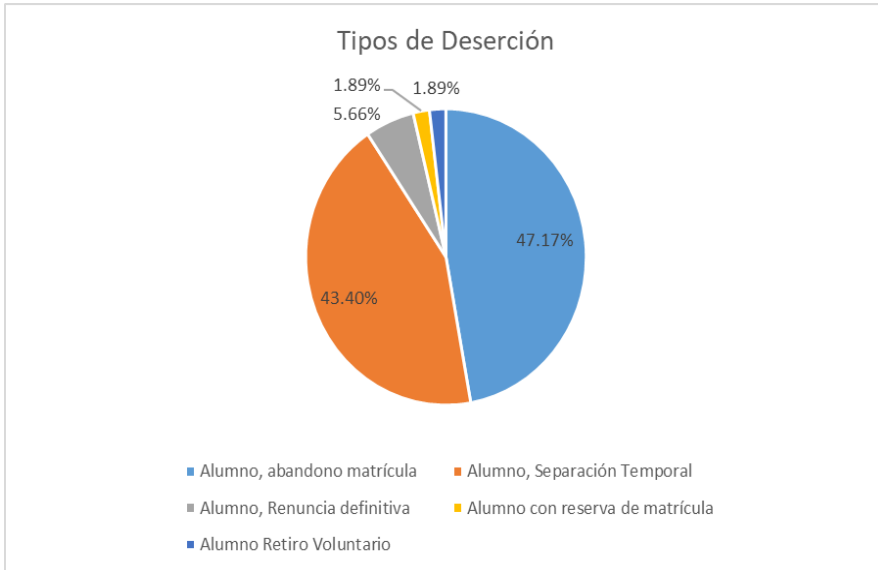
La investigación es importante porque busca generar valor al uso de las Tecnologías de Información (TI) en las organizaciones, teniendo una mejor capacidad de respuesta y eficiente toma de decisiones. Además, existe una justificación económica presupuestal, puesto que la Universidad podrá mantener mayor cantidad de estudiantes en ella; por cuanto, en las universidades públicas se considera el porcentaje de población estudiantil para la distribución de los presupuestos del sector.

De igual manera, la investigación aporta un modelo de *Machine Learning* entrenado y con buen *performance* en exactitud, sensibilidad y especificidad. Además, que este modelo necesita ser validado para determinar su fiabilidad y confianza. Este modelo matemático ayuda a poder evaluar los posibles escenarios futuros, y finalmente ayuda a la toma de decisiones. No obstante,

existen limitaciones respecto al acceso a la información, durante estos años se ha almacenado los datos de los estudiantes en distintos formatos y orden; por lo que, se tiene que realizar un trabajo de sistematización, digitalización y limpieza de los mismos.

En ese sentido, es necesario realizar un proceso de extracción de distintas fuentes, transformación a un solo formato plano, así como cargar a una herramienta de licencia gratuita. Los programas sin licencia como *WEKA* son limitados en cuanto al manejo de variables; por lo cual, no todos los datos se podrán recuperar después de la limpieza de datos, pero se trabajará con una cantidad significativa de ellos. Así, el estudio busca saber cuál de los algoritmos de clasificación de *machine learning* con aprendizaje supervisado tienen mejor desempeño para detectar la deserción universitaria.

De acuerdo a los datos recogidos de las fichas socioeconómicas, además de los datos académicos de los estudiantes, se puede obtener los siguientes resultados que se muestran en el Gráfico II, donde se observa que el abandono de matrícula (47,17%) es el tipo de deserción con mayor ocurrencia, esto sucede debido principalmente por las notas desaprobatorias obtenidas en el semestre anterior, coincidiendo con Gallegos y Campos (2019), quienes en su investigación concluyen que en el rendimiento académico destacan los factores personales sobre los sociales e institucionales.



Fuente: elaboración propia, 2021.

**Gráfico II: Tipos de deserción universitaria (2018-2021)**

En el Gráfico II mostrado anteriormente, se puede apreciar que la mayor cantidad de alumnos abandona la universidad primero de forma temporal (43,40%), esto ocurre cuando el alumno deja de estudiar un semestre y reserva su matrícula, después ya no se matriculan en los siguientes semestres académicos, pasando de un abandono temporal a un abandono definitivo (47,17%).

En ese sentido, Lizares (2017) en la tesis titulada: “Comparación de modelos de clasificación: Regresión logística y árboles de clasificación para evaluar el rendimiento académico” mencionan que las técnicas de minería de datos han funcionado muy bien para obtener modelos eficientes que ayuden a predecir el rendimiento académico, afirman que resulta ser herramienta eficaz para ser utilizado por los directivos de instituciones educativas.

Por su parte, Menacho (2017) en su artículo: “Predicción del rendimiento

académico aplicando técnicas de minería de datos”, hace una comparación entre tres modelos de inteligencia artificial en el ámbito educativo; llegando a la conclusión que a través de la matriz de confusión se puede realizar una evaluación precisa, así como comparar los clasificadores. De acuerdo a sus resultados, en las redes Naive de Bayes se presentó una tasa de aproximadamente 71,0% en lo que corresponde a la clasificación.

Es por ello, que en este estudio resultó conveniente seleccionar los aspectos socioeconómicos y académicos que tienen mayor relación con el abandono de los estudiantes. Para lo cual se ha codificado las variables que conforman cada uno de estos aspectos en el Cuadro 3. Por otra parte, se ha elaborado el siguiente cuadro de operacionalización de variables, donde se ha codificado los valores esperados por cada una de las variables.



**Cuadro 3**  
**Variables de estudio deserción universitaria**

FACTORES	VARIABLES	VALORES
SOCIO ECONÓMICO	Alumno_ Trabajador (AT)	1 0
	Tipo de Colegio de Procedencia (TCP)	1 0
	Preparación Pre Universitaria (PPU)	1 0
	Ingreso Familiar (IF)	1 2 3
	Efectividad Examen de Admisión (EEA) $EEA = \frac{\text{Puntaje obtenido Ex. Admisión}}{\text{Máximo puntaje posible}}$	[0;1]
ACADÉMICO	Promedio Ponderado Acumulado (PPA)	[0;20]
	Proporción de Créditos Aprobados (PCA) $PCA = \frac{\text{Total créditos aprobados}}{\text{Total de créditos matriculado}}$	[0;1]
	Presenta Repitencia de Asignaturas (PRA)	2 1 0

Fuente: Elaboración propia, 2021.

**a. Alumno Trabajador (AT):** Esta variable da una referencia sobre el tiempo de dedicación del estudiante en su educación, siendo 1 si el alumno trabaja o 0 en caso contrario.

**b. Tipo de Colegio de Procedencia (TCP):** Esta variable da una referencia sobre el tipo de procedencia del colegio del estudiante, siendo 1 si proviene de colegio particular y 0 en caso contrario.

**c. Preparación Pre Universitaria (PPU):** Esta variable da referencia si el estudiante tuvo preparación universitaria, siendo 1 si recibió preparación pre universitaria y 0 en caso contrario.

**d. Ingreso Familiar (IF):** Esta variable define la situación económica de la familia, siendo 1 el ingreso menor a S/.1000, 2 ingresos entre S/.1000 – S/.3000, y 3 los ingresos de más de 3000.

**e. Efectividad del Examen de Admisión (EEA):** Esta variable se obtiene de la división del puntaje obtenido en el examen entre el máximo puntaje posible. Los valores mínimos y máximos son [0;1].

**f. Promedio Ponderado Acumulado (PPA):** Es la media ponderada según el valor de créditos académicos de cada asignatura, de todas las notas obtenida por el estudiante en las asignaturas que haya cursado hasta el periodo de análisis 2018-2021. Los valores mínimos y máximos son [0;20] en la escala vigesimal.

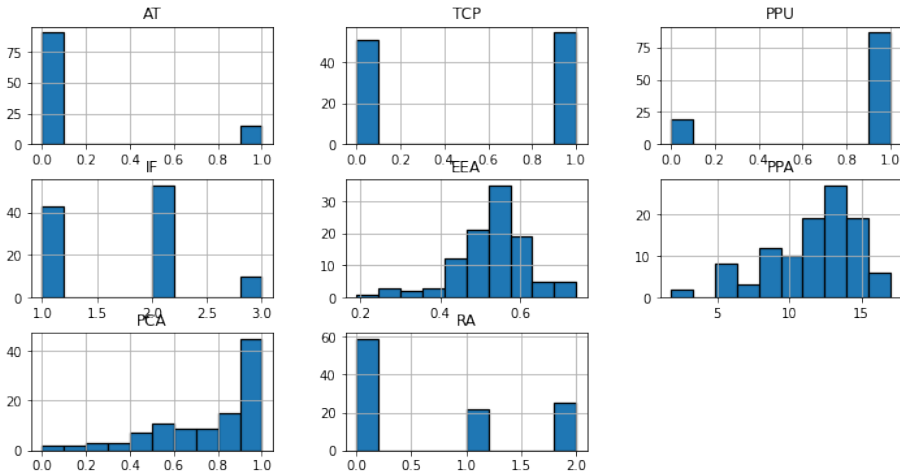
**g. Proporción de Créditos Aprobados (PCA):** Se obtiene de la división del total de créditos aprobados y total de créditos matriculados. Los valores mínimos y máximos son [0;1].

**h. Presenta Repitencia de Asignaturas (PRA):** Esta variable determina aquellos alumnos con más de 2 repitencias en una

misma asignatura. En caso que sea 2, significa tercera repitencia; en caso que sea 1, representa segunda repitencia; y 0 en caso contrario.

El siguiente paso fue un trabajo de recolección y tabulación de los datos para

elaborar la *data set* del proyecto. Una vez elaborada la misma, se verificó la consistencia de los datos, además de la normalidad en el caso de los valores, los cuales se muestran en la Figura I.

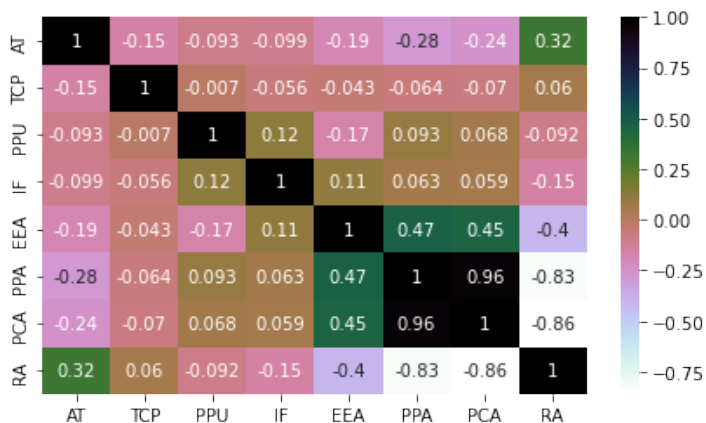


Fuente: Elaboración propia, 2021.

**Figura I: Histogramas para revisar la normalidad de los datos**

También se buscó saber si existe correlación entre las variables de estudio, los factores socioeconómicos y la deserción universitaria, para ello se muestra la matriz de correlación a continuación en la Figura II,

donde se puede apreciar una fuerte correlación entre las variables de la dimensión académica del estudio, promedio ponderado acumulado (PPA) y proporción de créditos aprobados (PCA).

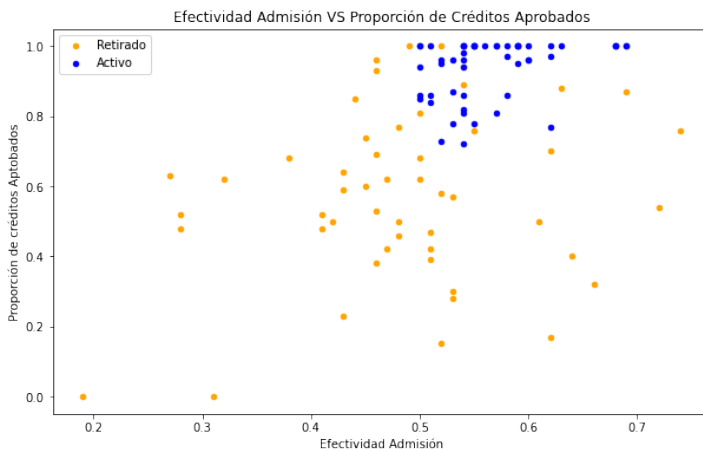


Fuente: Elaboración propia, 2021.

**Figura II: Matriz de correlación a través del mapa de calor**

Por otra parte, se graficó la dispersión entre las variables proporción de créditos aprobados (PCA) y efectividad en el examen de admisión (EEA), con la finalidad de explicar el factor académico del estudio, tal como se muestra en la Figura III. Por lo tanto,

permite visualizar la dispersión y ayuda a categorizar mejor a un estudiante como alumno regular o aquel estudiante que interrumpe sus estudios por los factores socioeconómicos y académicos.



Fuente: Elaboración propia, 2021.

**Figura III: Dispersión de variables**

Por su parte, en la Figura IV se muestran los resultados obtenidos de los algoritmos, que luego de entrenar al modelo generado a través del programa *Python* para la predicción universitaria, y someterlo a los datos de prueba

para su validación, evidencian el algoritmo KNN con la mejor precisión con 0.90625; seguido del algoritmo SVM con una precisión de 0.875.

```
The accuracy of:
❑ the SVM is: 0.875
❑ the Logistic Regression is 0.8125
❑ the Decision Tree is 0.8125
❑ the KNN is 0.90625
```

Fuente: Elaboración propia, 2021.

**Figura IV: Precisión de los algoritmos**

Finalmente, se muestran en la Tabla 1 los resultados obtenidos de precisión por tipo de variable; por ejemplo, si sólo se evaluara

las variables académicas, el algoritmo KNN tendría mejor *performance* 0,875 comparado con los otros algoritmos.

**Tabla 1**  
**Precisión de los algoritmos por tipo de variables**

Algoritmos	Variables		
	Total	Académico	Socioeconómico
SVM	0,875	0,8125	0,65625
Logistics Regression	0,8125	0,8125	0,6875
Decision Tree	0,8125	0,8438	0,59375
KNN	<b>0,90625</b>	<b>0,875</b>	0,59375

Fuente: Elaboración propia, 2021.

Con el modelo que obtuvo mejor precisión, el KNN, se realizó una prueba para poder verificar la validez del algoritmo, se ingresaron distintos [x1,x2,x3,x4,x5,x6,x7,x8] valores correspondientes a: Alumno Trabajador [AT], Tipo de Colegio de Procedencia [TCP], Preparación Pre Universitaria [PPU], Ingreso Familiar [IF], Efectividad del Examen de Admisión [EEA], Promedio Ponderado

Acumulado [PPA], Proporción de Créditos Aprobados [PCA], y Presenta Repitencia de Asignaturas [PRA], de las variables académicas y socioeconómicas, obteniendo el resultado que el estudiante con los datos ingresados de las variables tiene un 90,62% de probabilidad de deserción (retiro) del centro de estudios como se muestra en la Figura V.

```
model=KNeighborsClassifier(n_neighbors=3) #this examines 3 neighbours for putting
model.fit(train_X,train_y)
prediction=model.predict(test_X)
ejemplo=model.predict([[1,0,0,1,0.5,10.0,0.4,2]])
print (ejemplo)
print('The accuracy of the KNN is',metrics.accuracy_score(prediction,test_y))
['retirado']
The accuracy of the KNN is 0.90625
```

Fuente: Elaboración propia, 2021.

### Figura V: Ejemplo de predicción

Los resultados de la investigación, con respecto a 0,875 de precisión con variables académicas, están de acuerdo a lo mencionado por Lizares (2017) donde menciona que las técnicas de inteligencia artificial funcionan bien para el pronóstico de aspectos académicos. Sin embargo, los resultados evidencian que, considerando variables económicas y socioeconómicas tiene mejor precisión con el algoritmo KNN 90,6% en contraste a lo expresado por Menacho (2017), que obtuvo 71% de precisión con el algoritmo de Redes Naive de Bayes.

De igual manera, los hallazgos encontrados coinciden con Witten y Frank (2011), quienes sostienen que este algoritmo KNN, trata de una estructura de datos llamada árbol kD porque almacena un conjunto de puntos en un espacio k-dimensional, siendo k el número de atributos, permitiendo encontrar el vecino más cercano; por lo tanto, resulta mucho más rápido que examinar todos los puntos para encontrar ese vecino más cercano al objetivo, ayudando a lograr una mejor predicción en este caso de la deserción universitaria.

### Conclusiones

El algoritmo K-Nearest-Neighbor con una precisión de 0,91 tiene mejor desempeño para pronosticar la deserción

universitaria con las variables académicas y socioeconómicas de los estudiantes. Si sólo se evaluara las variables académicas y no las socioeconómicas el algoritmo KNN tendría una precisión de aproximadamente 0,88, esto es el mejor desempeño para pronosticar la deserción universitaria sólo con las variables académicas.

Asimismo, se evidenció que uno de los factores de riesgo más recurrentes entre los estudiantes que abandonaron sus estudios, es el bajo rendimiento académico 43,4%, durante su permanencia en la universidad. Por lo cual, el modelo obtenido puede ayudar a predecir en los primeros ciclos de estudio, qué alumnos son más probables de abandonar sus estudios, al mismo tiempo, alertar a la oficina de bienestar, la necesidad y atención de tutorías individuales, así como grupales.

### Referencias bibliográficas

- Alayo, F. (28 de Setiembre de 2020). Unos 174.000 estudiantes peruanos dejaron la universidad en lo que va del 2020. *El Comercio*. <https://elcomercio.pe/lima/sucesos/unos-174000-estudiantes-peruanos-dejaron-la-universidad-en-lo-que-va-del-2020-noticia/>
- Arbeláez-Campillo, D. F., Villasmil, J. J., y Rojas-Bahamón, M. J. (2021).

- Inteligencia artificial y condición humana: ¿Entidades contrapuestas o fuerzas complementarias? *Revista de Ciencias Sociales (Ve)*, XXVII(2), 502-513. <https://doi.org/10.31876/rcs.v27i2.35937>
- Barreno-Vereau, E. (2012). Análisis comparativo de modelos de clasificación en el estudio de la deserción universitaria. *Interfases*, (005), 45-82. <https://doi.org/10.26439/interfases2012.n005.149>
- Chalpartar, L. T. M., Fernández, A. M., Betancourth, S., y Gómez, Y. A. (2022). Deserción en la población estudiantil universitaria durante la pandemia, una mirada cualitativa. *Revista Virtual Universidad Católica Del Norte*, (66), 37-62. <https://doi.org/10.35575/rvucn.n66a3>
- Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support System*, 49(4), 498-506. <https://doi.org/10.1016/j.dss.2010.06.003>
- Digital55 (01 de Junio de 2020). Qué es Machine Learning: casos de éxito en empresas. *Digital55*. <https://www.digital55.com/innovacion/ques-machine-learning-casos-exito-empresas>
- Duche, A. B., Paredes, F. M., Gutiérrez, O. A., y Carcausto, L. C. (2020). Transición secundaria-universidad y la adaptación a la vida universitaria. *Revista de Ciencias Sociales (Ve)*, XXVI(3), 244-258. <https://doi.org/10.31876/rcs.v26i3.33245>
- Fagella, D. (March 4, 2020). Artificial Intelligence in retail – 10 present and future use cases. *Emerj*. <https://emerj.com/ai-sector-overviews/artificial-intelligence-retail/>
- Ferreyra, M. M., Avitabile, C., Botero, J., Haimovich, F., y Urzúa, S. (2017). *Momento decisivo: La educación superior en América Latina y el Caribe*. Grupo Banco Mundial. <https://openknowledge.worldbank.org/bitstream/handle/10986/26489/211014ovSP.pdf?sequence=5&isAllowed>
- Gallegos, J. A., y Campos, N. A. (2019). Determinantes del rendimiento académico estudiantil: Caso Universidad Católica de la Santísima Concepción. *Revista de Ciencias Sociales (Ve)*, XXV(2), 163-177.
- Greenwood, W. T. (1978). *Teoría de decisiones y sistemas de información*. Trillas.
- Hernández, R., Fernandez, C., y Baptista, M. D. P. (2014). *Metodología de la investigación*. McGraw-Hill / Interamericana de Editores, S.A. de C.V.
- Lizares, M. (2017). *Comparación de modelos de clasificación: Regresión logística y árboles de clasificación para evaluar el rendimiento académico* [Tesis de pregrado, Universidad nacional Mayor de San Marcos]. <https://cybertesis.unmsm.edu.pe/handle/20.500.12672/7122?show=full>
- Llanos, M. G. (2019). Percepciones de los hijos únicos sobre sus vivencias en la interacción universitaria. *Revista de Ciencias Sociales (Ve)*, XXV(E-1), 400-414. <https://doi.org/10.31876/rcs.v25i1.29630>
- Lugo, A. (2020). ¿Qué es el Machine Learning? *INVID*. <https://invidgroup.com/es/machine-learning-metodos/>
- Menacho, C. H. (2017). Predicción del rendimiento académico aplicando técnicas de minería de datos. *Anales Científicos*, 78(1), 26-33. <https://doi.org/10.21704/ac.v78i1.811>
- Microsoft Azure (03 de Enero de 2022). How to select algorithms for Azure

- Machine Learning. *Microsoft Azure*. <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-select-algorithms>
- Rico, D. A. (2006). *Caracterización de la deserción estudiantil en la Universidad Nacional de Colombia sede Medellín*. Universidad Nacional de Colombia.
- UPlanner (27 de Marzo de 2017). 8 causas de deserción estudiantil en la educación superior. *uPlanner* <https://uplanner.com/es/8-causas-de-desercion-estudiantil-en-la-educacion-superior-2/>
- Witten, I. H., Frank, E., y Hall, M. A. (2011). *Data mining: Practical Machine Learning tools and techniques*. Morgan Kaufmann Publishers.