

# **UN MODELO DE DECISIÓN PARA UNA INSTITUCIÓN DE EDUCACIÓN SUPERIOR DE PERÚ**

## **Resumen**

En la investigación se desarrolla un modelo de árbol de decisiones que permite identificar las variables que impacten significativamente en el bajo desempeño de los estudiantes de educación superior. Para la construcción del modelo se obtuvieron muestras de alumnos que pertenecen a instituciones de educación superior de tres ciudades del Perú, además se agregó una nueva variable nivel de pobreza y se realizaron muchas transformaciones convirtiendo la mayoría de las en categóricas para disminuir la dispersión de la variable. El modelo identifica aquellas variables que impactan más a la variable bajo desempeño. Las instituciones de educación superior pueden intervenir en aquellas variables que le competen para el desarrollo de la planificación estratégica en instituciones de la región y propone mecanismos que conllevan a ser más eficientes las organizaciones educativas y la sociedad.

Palabras claves: planificación estratégica, educación superior, árbol de decisiones

## **Introducción**

El estudio se evalúa la situación actual de las instituciones educativas de educación superior en el Perú, para ello, se cuenta con el censo de tres universidades. La primera de ellas tiene sede en Lima; la segunda en Trujillo y la tercera en Cajamarca. La base de datos consolida información de ocho semestres consecutivos (cuatro cursos) y las principales variables que influyen en su rendimiento académico.

El problema de la investigación es el diseño de un modelo de gestión del desempeño académico basado en árbol de decisiones en la educación superior, como resultado de su aplicación, se ofrece información relevante para la planificación estratégica en instituciones de la región y presenta la realidad actual. Además, se propone mecanismos que conllevan al logro de ser más eficientes, identificando las variables donde puede intervenir las instituciones educativas, planificando y desarrollando nuevas estrategias metodológicas que ayudarán a los estudiantes a alcanzar las competencias que requieren en su formación superior.

### *Objetivo:*

Desarrollar un modelo de decisión para una institución de educación superior de Perú.

## **Revisión bibliográfica.**

El trabajo desarrollado con las variables que disponemos inicia al asociar el colegio de procedencia de los alumnos según el precio de la matrícula con un nivel de pensión que pagaban en el colegio

de procedencia el cual se clasifico en cinco categorías. También se asignó un nivel de pobreza según el lugar de residencia a nivel de departamento, provincia y distrito donde residen los padres o tutores. Posteriormente, se observó el comportamiento de las otras variables que se tienen en la base de datos. Asimismo, se analizará cómo se asocian con el bajo desempeño académico.

#### *Deserción problema de alcance internacional en América Latina.*

La deserción universitaria es un problema tanto social, como económico para toda Latinoamérica, el Instituto para la Educación Superior en América Latina y el Caribe (IESALC) desarrolló un estudio sobre la “repetencia”, entendida como “la acción de cursar reiteradamente una actividad docente, sea por mal rendimiento del estudiante o por causas ajen (González, 2006).

Identificar las variables e implementar medidas remediales, es de suma importancia para quienes lideran las organizaciones educativas y además, para aquellos que tiene la responsabilidad de desarrollar normas para regular las universidades. En el Perú, la labor recae en la Superintendencia Nacional de Educación Superior Universitaria (SUNEDU). La UNESCO indica que invertir en educación es invertir en el desarrollo de una nación (Unesco. Organización de las Naciones Unidas para la Educación, 2009).

#### *Trabajo de transformación de variables.*

A partir de la variable “promedio ponderado”, se creó una nueva variable dicotómica “Desempeño Académico” con dos categorías: éxito: -bajo desempeño académico o desaprobado: cuando el promedio ponderado sea menor de 13- y fracaso: -aprobado, cuando el promedio ponderado es 13 o más-, la nueva variable -“Desempeño Académico”- será la variable respuesta para el diseño y evaluación de los modelos de clasificación, a partir de un conjunto de variables independientes o “predictoras”, entre las que se encuentran: colegio de procedencia, categoría de pago, carrera, tipo de ingreso, modalidad de ingreso, distrito de residencia, entre otras (Educación superior en Iberoamerica: Informe 2011; Sabino, 2014)

Se analizan dos métodos equivalentes para identificar las relaciones existentes entre la variable dependiente -variable respuesta-, y el conjunto de variables independientes: árbol de decisión y regresión logística.

Con la metodología del modelo del árbol de decisión, se busca clasificar casos en grupos y también, pronosticar valores de la variable dependiente (López López, 2009). Además, este procedimiento proporciona herramientas de validación para análisis de clasificación exploratorios y confirmatorios.

Como resultado del análisis, el modelo estimado se utiliza para identificar la asociación de las variables independientes y el “desempeño académico”. Adicionalmente, el análisis también genera

reglas, las mismas que pueden ser utilizadas para predecir eventos futuros, como la verosimilitud (veracidad, certidumbre) de que un alumno en su desempeño académico está aprobado o desaprobado a partir de las variables que permanecen en el modelo (Berlanga Silvente, Rubio Hurtado, & Vilà Baños, 2013)

Se recodifican las variables cuantitativas en variables categóricas, con excepción de la cuota de pago mensual según la carrera en la que está matriculado el alumno. Esta variable es constante para cada carrera; por lo tanto, no distorsiona al modelo.

Existen cuatro métodos disponibles en el módulo de redes de IBM SPSS (figura 1); sin embargo, como la variable respuesta es dicotómica, el método seleccionado para estimar el modelo de clasificación fue el CRT (*Classification and regression trees*). Este método consiste en un algoritmo de árbol binario completo que hace particiones de los datos y genera subconjuntos homogéneos respecto a la variable dependiente (Schuschny, 2005).

Figura 1.

Característica	CHAID	CHAID Exhaustivo	C&RT	QUEST
Tipo de Partición	Múltiple	Múltiple	Binaria	Binaria
Dependiente Continua	Sí	Sí	Sí	No
Predictoras Continuas	Si (*)	Si (*)	Si	Si
Coste de Mala Clasificación (Crecimiento del Arbol)	No	No	Sí	Sí
Pruebas Estadísticas (Selección del Predictor)	Sí	Sí	No	Sí
Pruebas Estadísticas (Particionar)	Sí	Sí	No	No
Velocidad	Moderada	Moderada	Lento	Moderada/Lento
Utiliza A priori?	No	No	Sí	Sí
Valores Faltantes para los Predictores Usados?	Si, como una categoría	Si, como una categoría	No, Sustitutos usados para partición	No, Sustitutos usados para partición

Fuente: Elaborado por Jorge Del Río L. Consultor Estadístico. (Del Río L., SN)  
Características de los métodos de crecimiento.

Las variables de tipo social y académicas presentan un comportamiento muy heterogéneo, esto afecta la construcción de los nodos e incrementa el grado de impureza.

Algunas de las variables incluidas en el modelo, fueron transformadas o se crearon otras nuevas. De igual forma, se añadieron nuevas variables a partir de fuentes externas: Instituto Nacional de Estadística e Informática (INEI., 2016) y Ministerio de Educación (MINEDU, 2015).

La variable “colegio de procedencia” se cambió por “nivel del colegio de procedencia”, que representa la categoría de pago asignada al postulante en su inscripción, según el colegio de procedencia también se clasifico en forma categoría.

Por último, al tener en la base de datos el 72,4% de alumnos aprobados y solo el 27,6% de alumnos desaprobados a la hora de realizar las simulaciones, se tendrá un pronóstico de acierto en aprobados mayor que de desaprobados y la variable en estudio es el “desempeño académico”.

*Rendimiento o desempeño académico.*

Investigaciones donde se estudió el desempeño (rendimiento) académico desarrolladas por Vélez y Roa González enfocan el desempeño desde una perspectiva operativa. Puede ser medido en forma cualitativa o cuantitativa y, en muchos casos, el resultado es de bajo desempeño y, como consecuencia, el alumno puede perder la materia (Vélez van Meerbeke & Roa González, 2005).

Estudios que desarrollan la intervención de la universidad encontraron la relevancia de las tutorías para el desempeño académico. González y col., investigan el abandono de estudios universitarios y medidas preventivas. Encuentran que el nivel de abandono en el primer año está en 28% (González Afonso, Álvarez Pérez, Pérez, & Bethencourt Benítez, 2007; Unesco. Organización de las Naciones Unidas para la Educación, 2009), a través de la Asociación Internacional de Universidades (IAU), ha constatado esta grave realidad en 180 países del planeta. El estudio encuentra que la asistencia a clase, estar al día en la materia y asistir a las tutorías mejora el desempeño. En la misma línea, García Jiménez, y col. comparan resultados con modelos de regresión lineal y regresión logística. Advierten que la segunda variable que impacta más al desempeño académico es la asistencia a clase y la participación de los estudiantes (García Jiménez, Alvarado Izquierdo, & Jiménez Blanco, 2000).

En el 2014, Beltrán y La Serna, realizan un estudio sobre la base de datos de alumnos universitarios para determinar qué variables no pedagógicas impactan el rendimiento en el primer año de estudios. La investigación plantea como variables predictoras: colegio y/o academia pre universitaria donde estudiaron”, variables socio familiares (nivel académico de los padres), si los alumnos son de Lima o de otro lugar (provincia o extranjeros) y variables psicológicas. Las variables que se usaron para medir el rendimiento fueron: promedio ponderado (acumula solo un año) y el número de créditos aprobados (Beltrán & La Serna Studzinski, 2014) y difiere de nuestra investigación en: la estructura organizacional, la primera ofrece nueve carreras y la segunda, en la base de datos, considera 28 carreras; la primera fue fundada en 1962 y la segunda en 1994. Las variables independientes que se consideran en ambos estudios son distintas. Además, los registros de la primera corresponden a dos semestres académicos y la base de datos de esta investigación mide registros de ocho semestres.

Tejedor y García-Valcárcel afirman que en los estudios en los cuales se investiga el impacto de indicadores pedagógicos, sociales y psicológicos en el desempeño académico con modelos de interacciones de estas variables que sostiene, estos afectan al bajo desempeño académico (González y otros, 1998; Lozano y otros, 2001; De Miguel y otros, 2002; Martín del Buey y Romero, 2003, Tejedor, 2003)”. Se considera, entre otras medidas, la importancia de la intervención por parte de las universidades, las cuales deben ofrecer un buen sistema de tutoría, mayor orientación al alumno para mejorar actitudes, técnicas, auto exigencia y asesoría permanente en orientación vocacional. Para el caso de los docentes, es necesario enseñar por competencias, desarrollar mayor motivación

en sus alumnos y desarrollo de todos los materiales para el curso (Tejedor & García-Valcárcel, 2007).

*Modelos predictivos. Tendencias actuales.*

El desarrollo de modelos predictivos investigados en los últimos cuatro años es muy amplio es por ello se agruparon en tres grupos que sean los más homogéneos posible.

*Modelos estadísticos clásicos probabilísticos* (paramétricos) tienen propuestas donde se consideran variables cuantitativas y modelos estadísticos que requiere de cumplir con una serie de supuestos, que variables sociales difícilmente los cumplirían, la mayoría de las soluciones algorítmicas asumen que todas las variables de probabilidad son discretas, mientras que en la práctica muchas son continuas, nosotros proponemos transformar las variables cuantitativas en categóricas y el desarrollo de un modelo de árbol de decisiones no paramétrico que nos indique una medida de asociación. Con los modelos clásicos no se puede comprender el rendimiento de los modelos cuando la reducción del número de fallas y falsas alarmas es una prioridad **G1**.

*Modelos con herramientas Bayesianas y árbol de decisiones*, las investigaciones de este grupo utilizan un bayesiano para mejorar la selección de variables ordinales en la evaluación de la calificación crediticia. Sin embargo, no se han realizado comparaciones con otros métodos y no se probó el poder predictivo, algunos utilizan una red Bayesiana dinámica para luego generar un árbol de decisión simplificado para calcular y presentar las decisiones óptimas en diferentes escenarios de decisión pero el método usa 'evidencia inconsistente' para modelar la asimetría funcional y estructural, la necesidad de incluir variables cualitativas como el ingreso familiar el pago que realizan del colegio de procedencia entre otras en nuestra investigación se transformaron en categóricas de esta manera reducimos en estas variables que son heterogéneas **G2**.

*Modelos complementados con herramientas de Redes Neuronales, regresión logística, herramientas Bayesianas y árbol de decisiones*. La mayoría de investigaciones de este grupo inician un enfoque bayesiano para seleccionar variables ordinales, luego algunos desarrollan un marco integrado de métodos basados en bosques aleatorios y el promedio bayesiano pero no realizan comparaciones con otros métodos para evaluar el poder predictivo, otras investigaciones desarrollan diferentes modelos, redes neuronales, regresión vectorial, bosque aleatorio, regresión logística, regresión adaptativa multivalente y árboles de clasificación con regresión se validaron usando bondad de ajuste el resultado del rendimiento de los modelos cuando la reducción del número de fallas y falsas alarmas es una prioridad **G3**.

## **Metodología**

Para medir la impureza en la clasificación de los datos se utiliza el índice de “Gini”, con lo que se obtienen divisiones que maximizan la homogeneidad de los nodos con respecto al valor de la variable dependiente.

“Gini” se basa en el cuadrado de las probabilidades de pertenencia de cada categoría de la variable dependiente. El valor mínimo (cero) se alcanza cuando todos los casos de un nodo corresponden a una sola categoría. Esta es la medida por defecto. Como este tipo de partición binaria y dependiente continua, el método de crecimiento adecuado es el CRT.

El primer árbol se construye sobre la base de todos los casos, dejando de lado los que corresponden al primer pliegue de la muestra; el segundo árbol se basa en todos los casos, dejando de lado los del segundo pliegue de la muestra y así sucesivamente.

En cada uno de los árboles, el software SPSS realiza un cálculo de riesgo de clasificación errónea, aplicando el árbol a la sub muestra que se excluyó al generarse este. Para la presente investigación, se estableció que, para particionar un nodo, este debería contener al menos 25 casos.

La conclusión es que este tipo de validación genera un modelo de árbol único y final. La estimación de riesgo mediante validación cruzada para el árbol final se calcula como promedio de los riesgos de todos los árboles.

En segundo lugar, se presenta el análisis de regresión logística con las mismas variables utilizadas en el método anterior: “Desempeño Académico” donde el éxito (bajo desempeño académico o desaprobado, cuando el promedio ponderado sea menor de 13) y el fracaso (aprobado, cuando el promedio ponderado es 13 o más). Asimismo, las variables independientes serán colegio de procedencia, categoría de pago, carrera, tipo de ingreso, modalidad de ingreso, distrito de residencia, entre otras

El modelo genera todas las predicciones, errores, estadísticos de impacto y pruebas de ajuste utilizando los datos a nivel de los casos individuales. No requiere cumplir con los supuestos de normalidad de la manera como se distribuyen las variables. Esto es de suma importancia, ya que todas las variables, con la excepción de la escala de pensión, son categóricas, algunas por su naturaleza y otras porque se transformaron. Una ventaja de este modelo, a pesar de no tener el rigor estadístico de los modelos clásicos de regresión, es que mide la probabilidad de la variable en estudio (CEA, A., 2002, p.128).

Algunas limitaciones, al ser un modelo no estadístico, no medirá las correlaciones entre nuestra variable de estudio “desempeño académico” con cada una de las demás. El modelo mostrará una medida de asociación y también de jerarquía, donde se observa un ranking de mayor a menor impacto.

### **Desarrollo de la investigación**

*Resultados: Comparación entre los dos modelos.*

**Modelo 1 - Modelo de Regresión Logística**

Se realizó la selección del mejor subconjunto de variables utilizando el método Forward criterio de razón de verosimilitud. Para las variables categóricas se usó a la primera categoría como referencia. Resultados sobre el impacto en la ventaja de ser admitido (exponencial de coeficientes).

**Modelo 2 – Modelo de Árboles de Clasificación (algoritmo CRT)**

Se utilizó validación cruzada para evaluar la eficiencia del modelo y un cambio de al menos 0,0001 en el índice de impureza (GINI) para dejar crecer el árbol en la construcción del modelo.

Para comparar los modelos, se consideraron como criterios la exactitud del modelo, la sensibilidad (ambos indicadores a partir de la matriz de confusión) y el área debajo de la curva ROC (AUC).

Para la elección del mejor modelo se utilizaron la matriz de confusión observando el porcentaje correcto de cada modelo y la clasificación de las áreas bajo la curva COR. (tabla 1 y tabla 2)

**Tabla 1. Clasificación con el modelo de regresión Logística.**

Observado	Pronosticado		
	Desaprobado	Aprobado	Porcentaje correcto
Desaprobado	4338	8997	32,5
Aprobado	3374	31 574	90,3
Porcentaje global			74,4

**Tabla 2. Clasificación con el modelo de árbol de decisiones.**

Observado	Pronosticado		
	Desaprobado	Aprobado	Porcentaje correcto
Desaprobado	5891	7444	44,2
Aprobado	4061	30 888	88,4
Porcentaje global	18,5	81,5	76,2
Métodos de crecimiento: CRT Variable dependiente: Bajo desempeño			

Como se observa en la tabla 2 mayor exactitud en comparación con el modelo logístico (76,2% vs. 74,4%). De igual forma, con el árbol de decisiones se aprecia una diferencia importante en la categoría “Desaprobado” mayor sensibilidad (44,2% vs. 32,5%) (tabla 1 y tabla 2).

En relación con el indicador AUC (área debajo de la curva ROC) y a partir de los límites de confianza asintótico al 95% (figura 2), ambos modelos son adecuados y no hay mayores diferencias en el desempeño.

Finalmente, por los resultados obtenidos, se puede concluir que el modelo de árboles de clasificación, en comparación con el modelo logístico, es más adecuado para predecir que un alumno tenga bajo desempeño, así como identificar cuáles son los principales factores asociados.

Considerando el modelo elegido, se presenta la importancia de las variables en relación con la información aporta al modelo (reducción de la entropía basada en el coeficiente de GINI). También

se indica la importancia de las variables independientes para el desempeño académico con el modelo de árbol de decisiones.

Curvas de COR y área debajo de la curva

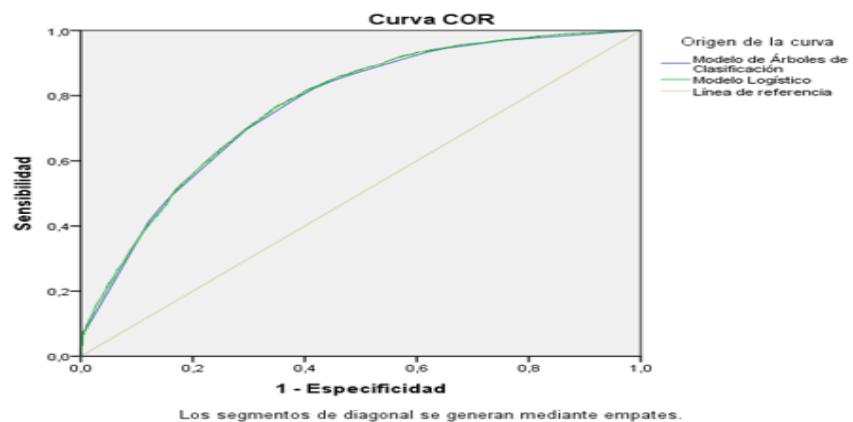


Figura 2.

Tabla 3. Clasificación del porcentaje correcto de las áreas bajo la curva COR.

Variables de resultado de prueba	Área	Error estándar	Significación asintótica	95% de intervalo de confianza asintótico	
				Límite inferior	Límite superior
Modelo de Árboles de Clasificación	0,768	0,002	0,000	0,763	0,772
Modelo Logístico	0,773	0,002	0,000	0,768	0,777
a. Bajo el supuesto no paramétrico b. Hipótesis nula: área verdadera = 0,5					

Fuente: Base de datos. Elaboración propia con el software SPSS ver. 20

La “importancia” de la viable (tabla 4) que indica cómo ayuda cada variable a reducir la entropía (coeficiente de GINI). El SPSS calcula la “importancia normalizada” (figura 3), se diferencia en que hace la comparación en términos relativos, tomando la reducción de la más importante como el 100%.

Se concluye, que la variable “ciclo académico” es la variable que se asocia con mayor fuerza con el “bajo desempeño académico”.

En segundo lugar, se encuentra la variable “modalidad de ingreso”, en este aspecto cabe resaltar que, en las universidades privadas, los departamentos de marketing ofrecen una serie de opciones de ingreso con la finalidad de captar un mayor porcentaje de ingresantes. Esto genera que un postulante, que no cumpla con el perfil adecuado, sea incorporado. En tercer lugar, se encuentra la “escala de pensión”, la cual es diferenciada por la carrera que sigue el alumno; las restantes variables están asociadas con el “desempeño académico” con un menor impacto.

El método de selección de variables utilizado descarta aquellas, que no tienen relación significativa con la variable dependiente; y por lo tanto no aparecen en la tabla 4.



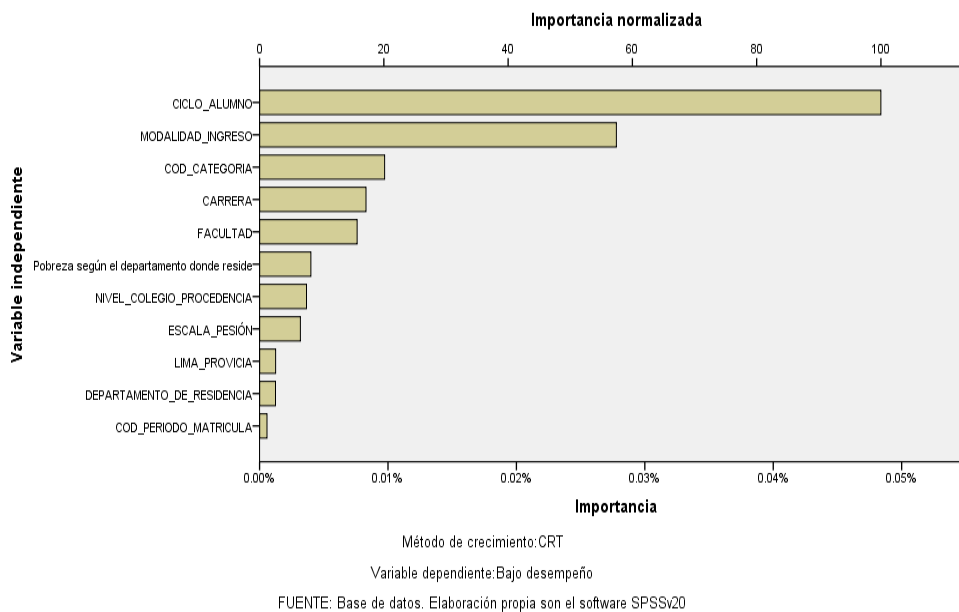
Para poder realizar el contraste de hipótesis, tenemos que resaltar que el modelo de árbol de decisiones no es un modelo estadístico, así que no hay un “p-valor” asociado a cada una de las variables. Por este motivo, podemos aceptar que la variable que aparece en el modelo ya se encuentra asociada y el nivel de aporte se puede medir con la importancia normalizada.

Tabla 4. Importancia de las variables independientes.

Variable independiente	Importancia	Importancia normalizada (%)
ciclo alumno	0,0484	100%
modalidad ingreso	0,0278	57,4%
cod_categoria	0,0097	20,1%
carrera	0,0083	17,1%
facultad	0,0076	15,7%
pobreza según el departamento donde reside	0,0040	8,2%
nivel_colegio_procedencia	0,0037	7,6%
escala_pesión	0,0032	6,6%
lima_provincia	0,0012	2,6%
departamento_de_residencia	0,0012	2,6%
cod_periodo_matricula	0,0006	1,2%

Métodos de crecimiento: CRT  
Variable dependiente: Bajo desempeño  
Fuente: Base de datos. Elaboración propia con el software SPSS ver. 20

Figura 3



Indica la importancia de las variables independientes para el desempeño académico. Además, también podemos confirmar el contraste con las tablas de contingencia (tabla 5).

Tabla 5. Impacto en por ciento del ciclo en el desempeño académico.

Desempeño	Ciclo del Alumno										
	0	1	2	3	4	5	6	7	8	9	10
Desaprobado	57	42	22	18	13	8	5	5	6	4	7

Aprobado	43	58	78	82	87	92	95	95	94	96	93
----------	----	----	----	----	----	----	----	----	----	----	----

*Relación entre el ciclo y el desempeño académico del estudiante.*

Conforme avanza el ciclo, el porcentaje de estudiantes desaprobados tiende a disminuir (tabla 5). Al finalizar el ciclo cero (los cursos de nivelación) el 57% de desaprobados; al término del primer ciclo disminuye a 42%; para el segundo ciclo al 22%.

Sobre este tema, diferentes investigaciones dan cuenta de la adaptación de los estudiantes universitarios al primer año. Provinciaisky, Perry y Hall, encuentran que existe una fuerte correlación entre el bajo desempeño académico y variables que las denominan PAC (control académico percibido) y las relacionas con la autoeficacia, el manejo del estrés y el optimismo, entre otras. Concluye que todas se correlacionan con el desempeño académico. Cuando los alumnos van avanzado a semestres superiores, aprenden a manejarlas de forma que no impacte en su desempeño académico (Stupnisky, Perry, Hall, & Guay, 2012).

En la tabla 6, las variables: rendimiento progresivo, evaluación integral, exoneración de grados, traslado externo medicina y traslados, la variable “desempeño académico” para Desaprobados , se encuentre entre el 28% y 44%.

En nuestro país hasta el año 2016, las universidades podían tener diferentes modalidades de ingreso con el objetivo de la captación de una mayor cantidad de alumnos. A partir del año 2018 la SUNEDU observa el cumplimiento de la “Ley Universitaria”, diversas modalidades de ingreso ya no se pueden ofrecer a los alumnos.

Según la modalidad de ingreso se medían diferentes niveles de conocimientos y competencias de los postulantes con los mismos indicadores, es por lo que también se tiene un grupo muy heterogéneo, en lo que se refiere a los saberes previos que un alumno debe tener para iniciar sus estudios universitarios, se observa que esta variable impacta el bajo desempeño académico de forma significativa.

*Contribución científica*

El modelo seleccionado de árbol de decisiones es un modelo no estadístico y, por lo tanto, no medirá las correlaciones entre nuestra variable de estudio “bajo desempeño académico” con cada una de las demás. No obstante, sí ofrecerá una medida de asociación, y también de jerarquía, para poder establecer un ranking de mayor a menor impacto.

Como modelo predictivo, debe tener una base de datos con las mismas variables y las mismas categorías en cada una de ellas para poder predecir a priori el bajo desempeño académico del alumno con nivel de acierto de casi el 45% de este nuevo grupo de alumnos. De lo anterior, se puede concluir que el modelo de árboles de clasificación que desarrollamos con la transformación

de los datos es más adecuado para poder predecir que un alumno tenga bajo desempeño y explicar cuáles son los principales factores asociados. Además, se observa, en la matriz de confusión, que el modelo tiene una adecuada precisión.

Según el modelo, podemos observar que la variable para el caso de las universidades peruanas constituidas como empresas, el ciclo académico que está cursando es la variable que se asocia con mayor fuerza con el bajo desempeño académico. El modelo con la identificación de las variables que impactan al bajo desempeño académico permite planificar un conjunto de estrategias de gestión administrativa y académica, por parte de la universidad, para minimizar esa deserción y convertirla en retención.

Tabla 6. Como la modalidad de ingreso impacta al desempeño académico

Porcentaje	Modalidad_ingreso															total
	Admisión 30+	Eval integral medicina	Evaluación integral	Exon grado/titulo medicina	Exoneración grado titulo	Formación complementaria	Hijo funcionario interna	PPU	Rendimiento progresivo	Rendimiento progresivo pq	Sel preferente medicina	Selección preferente	Selección preferente	Traslado externo medicina	Traslado	
Desaprobado	40	24	37	38	22	14	19	16	44	32	12	18	27	14	28	28
Aprobado	60	76	63	63	78	86	81	84	56	68	88	82	73	86	72	72

### Comentarios finales

#### *Conclusiones, Análisis descriptivo*

Durante el análisis descriptivo, resultó que, para la variable estado de la matrícula, las categorías retiro temporal, retiro definitivo y sancionado generan el 100% de sus resultados con bajo desempeño académico. Este comportamiento no se repite si el estado de matrícula corresponde a la categoría regular.

Por otro lado, el código de categoría del colegio lo asignan en el área de admisión y el criterio corresponde únicamente al colegio de procedencia del postulante. Esta asignación es temporal para el alumno en su condición de postulante. Al ingresar, se le asigna un código según la carrera en la que se matricule. Por lo tanto, esta variable, al incluirse al modelo, también podría distorsionar los resultados.

Se elaboró tablas de doble entrada donde se apreció a priori las relaciones entre las variables dependientes y las independientes.

También, este desarrollo preliminar permitió reflexionar sobre el futuro de la modalidad de ingreso; ya que, con el nuevo marco regulatorio, las opciones de ingreso se reducirán a su tercera parte. El análisis descriptivo, con todas las diferentes modalidades, permitirá identificar cuáles categorías impactan al bajo desempeño académico que coinciden con el nuevo marco regulatorio.

Se puede observar una asociación por encima del 35%, de alumnos con bajo desempeño académico en las carreras de Odontología, Comunicación e Imagen Empresarial, Administración y Agronegocios, Ingeniería de Telecomunicaciones y Redes, Comunicación y Publicidad, Economía y Negocios Internacionales, Economía y Políticas Públicas y, por último, Hotelería y Administración.

La cantidad de alumnos con un promedio ponderado menor a trece (se considera desaprobado en la universidad privada con fines de lucro Perú) se mantiene con un 25% más o menos uniforme a lo largo de ocho semestres académicos continuos. Este indicador no indica el retiro, ya que solo se está midiendo por semestre en forma independiente, pero ya nos ofrece una clara idea del rendimiento académico, que es una variable que se encuentra muy correlacionada con la deserción de los estudiantes.

Se puede observar un alto nivel de asociación entre el ciclo con respecto al desempeño académico. Inicia el ciclo cero (los cursos de nivelación) con casi un 57% de desaprobados para descender al 42%, 22% y 18%. Conforme el alumno continúa avanzando en sus estudios, el porcentaje va disminuyendo con mucha intensidad.

Es evidente que la modalidad de ingreso fue una variable que impactaba al desempeño académico, no se medían los conocimientos y competencias de los postulantes con los mismos indicadores, es por ello que también se tiene un grupo muy heterogéneo en lo que se refiere a los saberes previos que un alumno debe tener para iniciar sus estudios universitarios.

Sin embargo, las variables asociadas al nivel socio económico (como colegio de procedencia y el distrito donde viven los padres de alumno) son las que tienen una baja asociación con el desempeño académico en universidades privadas con fines de lucro.

El modelo identifica, en segundo lugar, que la modalidad de ingreso impacta al bajo desempeño académico. Esto ya se explicó en este párrafo. Las áreas comerciales y de marketing de las universidades privadas con fines de generan muchas opciones para que el alumno pueda ser admitido sin un filtro adecuado.

*Nuevas líneas de investigación.*

Nuestra investigación deja la oportunidad de nuevas líneas de investigación:

Desarrollo de un modelo que integre la identificación con las estrategias de gestión para disminuir la deserción.

Desarrollar una investigación para que, partiendo del proceso de tutorías, se construya un modelo de tutorías.

### Referencias bibliográficas

- Adejo, O. W., & Connolly, T. (2018). Predicting student academic performance using multi-model heterogeneous ensemble approach. *Journal of Applied Research in Higher Education*, 10(1), 61-75. doi:10.1108/JARHE-09-2017-0113. **G1.**
- Ahamed, A. T. M. S., Mahmood, N. T., & Rahman, R. M. (2017). Prediction of HSC examination performance using socioeconomic, psychological and academic factors. **G2.**
- Aljamaan, H. I., & Elish, M. O. (2009). An empirical study of bagging and boosting ensembles for identifying faulty classes in object-oriented software. **G3.**
- Beltrán, B., & La Serna Studzinski, K. (2014). ¿Qué explica el rendimiento académico en el primer año de estudios universitarios? Un estudio de caso en la Universidad del Pacífico DD/08/09 (pp. 91): Universidad del Pacífico. Perú.
- Berlanga Silvente, V., Rubio Hurtado, M. J., & Vilà Baños, R. (2013). Cómo aplicar árboles de decisión en SPSS. *REIRE, Revista d'Innovació i Recerca en Educació*, 6(1), 65-79. doi:10.1344/reire2013.6.1615
- Challagulla, V. U. B., Bastani, F. B., Yen, I. L., & Paul, R. A. (2005). Empirical assessment of machine learning based software defect prediction techniques. **G3.**
- Chen, Y., Gao, W., Zhu, T., & Ling, C. (2002). Learning prosodic patterns for Mandarin speech synthesis. *Journal of Intelligent Information Systems*, 19(1), 95-109. doi:10.1023/A:1015568521453. **G3.**
- Chuan, Y. Y., Husain, W., & Shahiri, A. M. (2017) An exploratory study on students' performance classification using hybrid of decision tree and Naïve Bayes approaches. Vol. 538 AISC (pp. 142-152): Springer Verlag.
- Dag, A., Topuz, K., Oztekin, A., Bulur, S., & Megahed, F. M. (2016). A probabilistic data-driven framework for scoring the preoperative recipient-donor heart transplant survival. *Decision Support Systems*, 86, 1-12. doi:10.1016/j.dss.2016.02.007. **G3.**
- Davoodi, R., & Moradi, M. H. (2018). Mortality prediction in intensive care units (ICUs) using a deep rule-based fuzzy classifier. *Journal of Biomedical Informatics*, 79, 48-59. doi:10.1016/j.jbi.2018.02.008. **G3.**
- Del Río L., J. (SN). Técnicas de Árboles de Clasificación. [http://www.captac-dr.org/c/document\\_library/get\\_file?folderId=312918&Técnicas de Árboles de Clasificación.pdf](http://www.captac-dr.org/c/document_library/get_file?folderId=312918&Técnicas de Árboles de Clasificación.pdf)

- Devasia, T., Vinushree, T. P., & Hegde, V. (2016). Prediction of students performance using Educational Data Mining. **G2**.
- Educación superior en Iberoamerica: Informe 2011. Centro Interuniversitario de Desarrollo (CINDA) - Universia: Editores: Brunner, José Joaquín; Ferrada Hurtado, Rocío.
- García Jiménez, M. V., Alvarado Izquierdo, J. M., & Jiménez Blanco, A. (2000). La predicción del rendimiento académico: regresión lineal versus regresión logística. *Psicothema*, 12(Suplemento), 248-252.
- González Afonso, M. C., Álvarez Pérez, P. R., Pérez, L. C., & Bethencourt Benítez, J. T. (2007). El abandono de los estudios universitarios: factores determinantes y medidas preventivas. *Revista española de pedagogía*, 71-85.
- González, L. E. (2006). Repitencia y deserción en América Latina. Disponible en internet: <http://www.intercontacto.com/gente>.
- Han, D. (2017). Researches of detection of fraudulent financial statements based on data mining. *Journal of Computational and Theoretical Nanoscience*, 14(1), 32-36. doi:10.1166/jctn.2017.6119. **G3**.
- Hsieh, Y. M. (2016). Construction of financial distress prediction models using stepwise regression and data mining techniques. *ICIC Express Letters, Part B: Applications*, 7(3), 591-597.
- INEI. (2016). Perú: Síntesis Estadística 2015 Instituto Nacional de Estadística e Informática. INEI (Ed.) (pp. 107). Retrieved from [https://www.inei.gob.pe/media/MenuRecursivo/publicaciones\\_digitales/Est/Lib1292/libr](https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1292/libr)
- Khokhar, R. H., & Md Sap, M. N. (2003). Design and development of neural Bayesian approach for unpredictable stock exchange databases. **G3**.
- Kibis, E. Y., Büyüктаhtakin, I. E., & Dag, A. (2017). Data analytics approaches for breast cancer survivability: Comparison of data mining methods. **G3**.
- La Red Martínez, D. L., Karanik, M., Giovannini, M., & Scappini, R. (2016). Towards to a Predictive model of academic performance using data mining in the UTN - FRRe. **G2**.
- Latchoumi, T. P., Jayakumar, L., Ezhilarasi, T. P., Parthiban, L., & Mahalakshmi, K. (2018). Comparison of classification techniques on data mining. *International Journal of Pure and Applied Mathematics*, 118(11 Special Issue), 357-364. doi:10.12732/ijpam.v118i11.43. **G2**.
- Lee, K. J., & Chang, W. (2009). Bayesian belief network for box-office performance: A case study on Korean movies. *Expert Systems with Applications*, 36(1), 280-291. doi:10.1016/j.eswa.2007.09.042. **G3**.

- Lim, C. P., Abeynayake, C., Sato-Ilic, M., & Jain, L. C. (2013). Special issue: Computational intelligence models for image processing and information reasoning. *Journal of Intelligent and Fuzzy Systems*, 24(2), 199-200. doi:10.3233/IFS-2012-0546. **G3**.
- López López, C. P. (2009). *Ingeniería comercial*.
- Martínez Martínez, R., & Heredia Escorza, Y. (2010). Tecnología Educativa en el Salón de Clase: Estudio Retrospectivo de su impacto en el desempeño académico de estudiantes universitarios del área de Informática. *Revista Mexicana de Investigación Educativa*, 15(45), 371-390.
- MINEDU. (2015). *Infraestructura educativa Perú*. Retrieved from <http://datos.minedu.gob.pe/group/infraestructura-educativa>
- Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3), 559-569. doi:10.1016/j.dss.2010.08.006. **G3**.
- Pietruszkiewicz, W., & Dzega, D. (2010) The large scale artificial intelligence applications - An analysis of AI-supported estimation of OS software projects. Vol. 6040 LNAI (pp. 223-232). **G3**.
- Rashid, T. A., & Jabar, A. L. (2016). Improvement on predicting employee behaviour through intelligent techniques. *IET Networks*, 5(5), 136-142. doi:10.1049/iet-net.2015.0106. **G2**.
- Sabino, C. (2014). *El proceso de investigación*. Guatemala: Editorial Episteme.
- Schuschny, A. R. (2005). *Tópicos sobre el modelo de insumo-producto: teoría y aplicaciones* (No. 37): United Nations Publications.
- Shaukat, K., Nawaz, I., Aslam, S., Zaheer, S., & Shaukat, U. (2017). Student's performance in the context of data mining. **G2**.
- Sheng, Y. C., Mustafa, M. B., Alam, S., Hamid, S. H., Sani, A. A., & Gani, A. (2016). Personal CGPA planning system for undergraduates: Towards achieving the first class CGPA. **G2**.
- Shrestha, R. M., Orgun, M. A., & Busch, P. (2016). Offer acceptance prediction of academic placement. *Neural Computing and Applications*, 27(8), 2351-2368. doi:10.1007/s00521-015-2085-7. **G2**.
- Stupnisky, R. H., Perry, R. P., Hall, N. C., & Guay, F. (2012). Examining perceived control level and instability as predictors of first-year college students' academic achievement. [Examen del nivel de control percibido y la inestabilidad como predictores de los logros académicos de los estudiantes universitarios de primer año]. *Contemporary Educational Psychology*, 37(2), 81-90. doi:https://doi.org/10.1016/j.cedpsych.2012.01.001

- Tejedor, F. J., & García-Valcárcel, A. (2007). Causas del bajo rendimiento del estudiante universitario (en opinión de los profesores y alumnos). Propuestas de mejora en el marco del EEES. *Revista de Educación*, 342(1), 443-473.
- Unesco. Organización de las Naciones Unidas para la Educación, I. C. y I. C. (2009). Conferencia Mundial sobre la Educación Superior 2009. La nueva dinámica de la educación superior y la investigación para el cambio social y el desarrollo: UNESCO París.
- Varis, O. (1997). Bayesian decision analysis for environmental and resource management. *Environmental Modelling and Software*, 12(2-3), 177-185. doi:10.1016/S1364-8152(97)00008-X
- Vélez van Meerbeke, A., & Roa González, C. N. (2005). Factores asociados al rendimiento académico en estudiantes de medicina. *Educación Médica*, 8(2).