

Rendimiento académico empleando minería de datos

Academic performance using data mining

QUIÑONES HUATANGARI, Lenin.¹

CARRASCO VEGA, Yajaira L.²

Resumen

El objetivo del trabajo fue pronosticar el rendimiento académico en los estudiantes de la Carrera de Ingeniería de Industrias Alimentarias de la Universidad Nacional de Jaén (UNJ), empleando técnicas de minería de datos. La matriz de datos, fue obtenida a través de una ficha y oficinas de la institución. La metodología empleada, ha sido la CRISP-DM; donde los pronósticos de tres algoritmos de clasificación realizados en el software weka obtuvieron una confiabilidad superior al 83%.

Palabras clave: evaluación de la educación, aplicación informática, previsión

Abstract

The objective of the work was to forecast the academic performance of students in the Food Industry Engineering career at the National University of Jaén (UNJ), using data mining techniques. The data matrix was obtained through a card and offices of the institution. The methodology used was the CRISP-DM; where the forecasts of three classification algorithms made in the weka software obtained a reliability higher than 83%.

Key words: education evaluation, computer application, forecasting

1. Introducción

La preocupación por el desempeño de los estudiantes de las carreras universitarias, surge de los desfavorables indicadores de abandono y rendimiento académico ha llevado a las universidades a investigar sobre las causas que subyacen esta problemática (Dapozo et al., 2008). La educación superior en los tiempos actuales, no responden a los retos que plantea el desarrollo y progreso del país, situación que pone de relieve las deficiencias de nuestro sistema académico. En la enseñanza superior, todavía continúa primando el memorismo e incentivándose la repetición de contenidos; y en lo que respecta a los docentes de educación superior universitaria y de institutos, ellos aún mantienen el uso de una metodología educativa dogmática y represiva. Esta represión conduce a la censura de aquello que es creativo, impide la formación de habilidades creativas e innovadoras, obteniendo como resultado, un bajo rendimiento académico. Eso conlleva a que los egresados de las instituciones no se inserten a la comunidad en forma eficiente y no contribuyan con el desarrollo y progreso tecnológico, en las distintas áreas (Cuadros Luyo et al., 2017).

Probablemente estamos viviendo en el período definitorio de la historia humana; lo que hace que este período sea emocionante y fascinante es la democratización de las diversas herramientas y técnicas de la ciencia de datos, que siguió al impulso de la informática. La Minería de Datos Educativa (MDE) es una de las áreas de aplicación

¹ Docente. Universidad Nacional de Jaén. Carrera de Ingeniería de Industrias Alimentarias. Perú. Email: lenin.quinones@unj.edu.pe

² Docente. Universidad Nacional de Barranca. Carrera de Contabilidad y Finanzas. Perú. Email: ycarrasco@unab.edu.pe

de la minería de datos que ayudan a predecir la deserción, hábitos de aprendizaje y por lo tanto permite proporcionar ayuda oportuna a los estudiantes. Sin embargo, no ha sido introducida en la educación de nivel superior o universitaria (M. Kumar et al., 2017). Los algoritmos de uso común de minería de datos son: Regresión lineal, regresión logística, árbol de decisión, máquina de vectores de soporte (SVM), red bayesiana, kNN, K-medias, bosque al azar, algoritmos de reducción de dimensionalidad y algoritmos de aumento de gradiente.

Los modelos constituyen una herramienta para entender la complejidad que caracteriza los diversos sistemas. Estos permiten realizar análisis de impactos tecnológicos, económicos, ambientales, evaluación de estrategias productivas, etc. Su empleo se enfoca generalmente a comprender mejor los problemas y anticipar la realidad que se investiga (Hernández, 2009). Mediante modelos predictivos de minería de datos es posible identificar características y patrones relacionadas con el comportamiento del rendimiento académico del estudiante, analizando su historial académico junto a los factores socio económicos y otros (Martínez et al., 2015), (Ahmad et al., 2015).

Se mencionan algunos trabajos previos: Realizaron un análisis del rendimiento académico de los estudiantes de nuevo ingreso en la titulación de Ingeniería Técnica en Informática de Sistemas de la Universidad Politécnica de Valencia (UPV) a lo largo de tres cursos, aunque también se trabajó con las titulaciones de Ingeniería Técnica en Informática de Gestión y de Ingeniería Informática. Este análisis relacionó el rendimiento con las características socioeconómicas y académicas de los estudiantes, que se obtienen en el momento de su matrícula, y que se recogen en la base de datos de la universidad. Se definió un indicador del rendimiento para cada estudiante, teniendo en cuenta las calificaciones obtenidas y las convocatorias utilizadas. Para el estudio se utilizó algoritmos de minería de datos, que determinaron el nivel de condicionamiento existe entre dicho rendimiento y características como el nivel de conocimientos de entrada del estudiante, su contexto geográfico y sociocultural, etc. Esto proporciona una herramienta importante para la acción tutorial, que puede apoyarse en las predicciones de los modelos que se obtienen para encauzar sus recomendaciones y encuadrar las expectativas y el esfuerzo necesario para cada estudiante (Alcover et al., 2007). Otra investigación ha demostrado la incidencia de la formación en Matemática sobre el desempeño de los estudiantes, en este contexto esta fortaleza resulta clave para el avance en las carreras que se ofrecen en la Facultad de Ciencias Exactas de la Universidad Nacional del Nordeste (UNNE) de Argentina, dado que todas ellas tienen esta asignatura en el primer año. El trabajo realizó un estudio comparativo de algoritmos de clasificación del campo de la minería de datos utilizando una herramienta de software libre. Se contrastó el rendimiento de las redes neuronales versus la aplicación de estadísticas convencionales. Se obtuvieron buenos resultados que permiten generar información útil para orientar políticas institucionales que contribuyan a mejorar los indicadores preocupantes de fracaso traducidos en deserción y bajo rendimiento de los estudiantes del primer año de carreras universitarias (Dapozo et al., 2008). En (Satyanarayana & Nuckowski, 2016), utilizaron múltiples clasificadores (Árboles de decisión-J48, Naïve Bayes y Random Forest) para mejorar la calidad de los datos de los estudiantes eliminando las instancias ruidosas, y, por lo tanto, mejorar la precisión de la predicción. También identificaron las reglas de asociación que influyen en los resultados del estudiante utilizando una combinación de técnicas basadas en reglas (Apriori, Filtered Associator y Tertius). Proporcionan modelos con mejores precisiones de predicción sobre el rendimiento de los estudiantes, sino que también proporcionan mejores reglas para entender los factores que influyen en los mejores resultados de los estudiantes.

En el marco de los antecedentes, se puede inferir que se emplean múltiples algoritmos de minería de datos para estimar el rendimiento académico en todos los niveles educativos en diversos países. Esta información valiosa, ha sido considerada como referencia en el presente estudio, tales como los algoritmos de minería de datos empleados, las variables consideradas en las bases de datos recolectadas de los estudiantes, etc. En este sentido, la motivación del trabajo de investigación, fue desarrollar modelos basados en minería de datos que nos permitan pronosticar el rendimiento académico de los estudiantes de la carrera profesional de Ingeniería de

Industrias Alimentarias de la UNJ. La misma que será insumo para la toma de decisiones o planes de trabajo que puedan elaborar las autoridades correspondientes, respecto al rendimiento académico.

2. Metodología

2.1. Matriz de datos

Tomando en cuenta la información obtenida de la Universidad Nacional de Jaén y ficha de recolección a 142 estudiantes de la Carrera Profesional de Ingeniería de Industrias Alimentarias que se encontraban cursando entre el segundo y el décimo ciclo hasta noviembre del 2019. Se realizó una base de datos conformada por 142 instancias y 27 variables. Su fin fue generar y validar modelos basados en minería de datos que nos permitieron determinar el rendimiento académico.

Las variables contempladas fueron: nombre del colegio, lugar de colegio, tipo de dependencia del colegio, orientación vocacional, modalidad de ingreso, ciclo de ingreso, actividad económica de la madre, actividad económica del padre, situación laboral del estudiante, posee beca, nivel de educación del padre, nivel de educación de la madre, condición familiar, horas dedicadas para estudio, empleo de TICs, año de nacimiento, género, estado civil, residencia del distrito, residencia provincia, residencia departamento, créditos aprobados, cursos aprobados, cursos desaprobados, total de cursos, repeticiones de matemáticas y rendimiento académico. Han sido determinadas en base al análisis bibliográfico (Ahmad et al., 2015), (Porcel et al., 2009); además tienen relación directa con el alumno y son de características demográficas y académicas, ver Cuadro 1.

Cuadro 1
Descripción, tipo y rango de
veintisiete variables recolectadas.

Variables	Descripción	Tipo	Rango
Nombre_Colegio	Determina el nombre del colegio, donde el estudiante ha terminado sus estudios secundarios.	Nominal	
Lugar_Colegio	Determina el nombre del Distrito en donde se encuentra ubicado el centro educativo.	Nominal	
Dependencia_Colegio	Determina la condición del colegio.	Nominal	Estado Particular Religioso
Orientación_Vocacional	Especifica si recibieron información correspondiente a orientación vocacional.	Nominal	Si No
Modalidad_Ingreso	Especifica la modalidad que empleo el estudiante para ingresar a la UNJ.	Nominal	Examen ordinario Examen extraordinario Centro pre universitario
Ciclo_Ingreso	Determina el ciclo de ingreso del estudiante a la UNJ.	Nominal	• 2012-I • 2016-I • 2012-II • 2016-II • 2013-I • 2017-I • 2013-II • 2017-II • 2014-I • 2018-I • 2014-II • 2018-II • 2015-I • 2019-I • 2015-II • 2019-II
Activ_Econ_Madre	Especifica cual es la principal actividad que la madre realiza para la generación de ingresos.	Nominal	Oficio Profesional Técnico
Activ_Econ_Padre	Especifica cual es la principal actividad que el padre realiza para la generación de ingresos.	Nominal	Oficio Profesional Técnico
Sit_Laboral_Estud	Determina la situación laboral actual del estudiante.	Nominal	Estudia Trabaja y estudia

Variables	Descripción	Tipo	Rango
Posee_Beca	Recibe beneficio por parte del estado por medio de alguna beca.	Nominal	Si No
Niv_Educ_Padre	Nivel educativo del padre.	Ordinal	Sin estudios Primaria incompleta Primaria completa Secundaria incompleta Secundaria completa Superior incompleta Superior completa
Niv_Educ_Madre	Nivel educativo de la madre	Ordinal	Sin estudios Primaria incompleta Primaria completa Secundaria incompleta Secundaria completa Superior incompleta Superior completa
Condición_Familiar	Especifica la condición familiar de los padres del estudiante.	Nominal	Separados Convivientes Casados
Horas_Estudio	Cuántas horas al día emplea el estudiante para labores académicas.	Discreto	A: De 0 a 4 horas B: De 5 a 8 horas C: De 9 a 12 horas D: De 13 a 16 horas.
Empleo_Tics	Determina si emplea las TIC's en su actividad académica.	Nominal	Si No
Año Nacimiento	Especifica el año de nacimiento del estudiante.	Discreto	
Genero	Género del alumno	Nominal	Rango: • M: Masculino • F: Femenino
Estado_Civil	Especifica el estado civil del estudiante.	Nominal	Solo Conviviente
Residencia_Distrito	Especifica el distrito donde proviene el estudiante.	Nominal	
Residencia_Provincia	Especifica la provincia donde proviene el estudiante.	Nominal	Bagua Chota Condorcanqui Cutervo Jaén San Ignacio Utcubamba
Residencia_Departamento	Especifica el departamento donde proviene el estudiante.	Nominal	Cajamarca Amazonas
Creditos_Aprobados	Especifica la cantidad de créditos que el estudiante a aprobado hasta el semestre 2019-I.	Discreto	
Cursos_Aprobados	Especifica la cantidad de cursos que el estudiante a aprobado hasta el semestre 2019-I.	Discreto	
Cursos_Desaprobados	Especifica la cantidad de cursos que el estudiante a desaprobado hasta el semestre 2019-I.	Discreto	
Total_Cursos	Especifica la cantidad del total de cursos que el estudiante a llevado hasta el semestre 2019-I.	Discreto	
Matemáticas_repet	Especifica el número de veces que ha repetido los cursos de matemáticas.	Discreto	
Rendimiento_Académico	Especifica el rendimiento académico de acorde al promedio global.	Ordinal	Satisfactorio (17-20) Esperado (14-16) Proceso (11-13) Inicio (0-10)

2.2. Software Weka

Es el software libre que se ha empleado, debido a que contiene las herramientas necesarias para realizar transformaciones sobre los datos, tareas de clasificación, regresión, clustering, asociación y visualización. Desarrollado por la Universidad de Waikato de Nueva Zelanda, implementado en Java (Azoumana, 2013).

2.3. Metodología para emplear minería de datos

Se empleó la propuesta por la Cross Industry Standard Process for Data Mining, CRISP-DM. Trata de un modelo de proceso de minería de datos que describe los enfoques comunes que utilizan los expertos en minería de datos. Algunas fases permitieron revisar parcial o totalmente las fases anteriores. Estas fueron (Gallardo Corrales, 2017):

Entendimiento del problema: Se realizó una revisión de la literatura referente a la minería de datos empleado a la educación superior universitaria. Se recolectó información de los estudiantes del II al X ciclo de la Carrera Profesional de Ingeniería de Industrias Alimentarias a través de una ficha y se requirió información por parte de la Oficina de Asuntos Académicos. Los estudiantes del primer ciclo no fueron considerados debido a no poseer el promedio académico global.

Comprensión de los datos: Los datos de las veintiocho variables recolectadas fueron estandarizadas, comprendidas. Posteriormente fueron alojados en Excel para su preparación correspondiente.

Preparación de los datos: Se obtuvieron datos sin valores nulos o anómalos, que permitieron obtener resultados confiables, para ello se ha utilizado técnicas estadísticas como la media, mediana y la moda. Se transformó la fuente de datos en Excel al formato ARFF (Attribute Relation File Format) requerido por Weka.

Modelación: Para seleccionar variables independientes que más cercanamente afectan el rendimiento académico, se empleó el atributo evaluador InfoGainAttributeEval de Weka que permitió evaluar el valor de una variable midiendo la ganancia de información con respecto al rendimiento académico. Además, el método de búsqueda weka.attributeSelection.Ranker, permitió clasificar las variables por sus evaluaciones individuales.

Se aplicó las tareas de descubrimiento de clasificación en el software Weka. Se ha seleccionado los algoritmos de clasificación, porque o se obtienen reglas de clasificación del tipo “Si – Entonces” o árboles de decisión y en la revisión previa de la literatura son empleados (Chandra & Nandhini, 2010), (Ramesh et al., 2013), (S. A. Kumar & Vijayalakshmi, 2012). Las reglas del tipo “Si – Entonces” son una manera comprensible de representar el conocimiento. Una regla tiene dos partes, el antecedente y el consecuente. El antecedente de la regla (la parte del “Si”) contiene una combinación de condiciones respecto a los variables de predicción. El consecuente de la regla (la parte del “Entonces”) contiene el valor predicho para el rendimiento académico.

Evaluación del modelo: Se evaluó e interpretó el rendimiento académico para determinar su calidad y consolidar el conocimiento descubierto e incorporarlo en otro sistema para posteriores acciones. En los experimentos se han empleado tres medidas para la evaluación de la calidad de la clasificación: Exactitud, precisión y recuerdo. Las medidas calculadas usan el Cuadro 2, donde muestra la matriz de confusión de la clasificación basados en las ecuaciones (1), (2) y (3).

Cuadro 2
Matriz de confusión de la clasificación.

		Detectado	
		Positivo	Negativo
Actual	Positivo	Positivo Verdadero (<i>PV</i>)	Negativo Falso (<i>NF</i>)
	Negativo	Positivo Falso (<i>PF</i>)	Negativo Verdadero (<i>NV</i>)

La exactitud es la proporción del número total de predicciones donde correctamente ha calculado. La precisión es el radio de los casos correctamente clasificados a el número total de casos mal clasificados y casos correctamente clasificados. El recuerdo es el radio de los casos clasificados correctamente del número total de casos no clasificados y casos clasificados correctamente (Powers, 2011).

$$Exactitud = \frac{PV + NV}{PV + NF + NV + PF} \quad (1)$$

$$Presición = \frac{PV}{PV + PF} \quad (2)$$

$$Recuerdo = \frac{PV}{PV + NF} \quad (3)$$

Además, se pudo incluir la visualización de los patrones extraídos, la remoción de los patrones redundantes o irrelevantes y la traducción de los patrones útiles en términos que sean entendibles para el usuario. Con el fin de evaluar la calidad y precisión de la predicción de las reglas de clasificación obtenidas se utilizó el método de validación cruzada con 10 pliegues (n-fold cross validation).

3. Resultados

El atributo evaluador InfoGainAttributeEval de Weka permitió evaluar el valor de una variable midiendo la ganancia de información con respecto a la variable rendimiento académico. Además, el método de búsqueda weka.attributeSelection.Ranker, permitió clasificar las variables por sus evaluaciones individuales, ver Cuadro 3. Las variables empleadas difieren con el trabajo de (Alcover et al., 2007).

Cuadro 3
Rango e importancia de los veintiséis atributos recolectados.

Rango	Variables	Importancia	Rango	Variables	Importancia
1	NombreColegio	0.60986	14	DependColegio	0.03425
2	Cursosdesaprobados	0.35181	15	ResidProvincia	0.02478
3	CicloIngreso	0.26474	16	ModalidadIngreso	0.02318
4	LugarColegio	0.23439	17	HorasEstudio	0.016
5	Matemáticarepeticiones	0.13402	18	EstadoCivil	0.00799
6	Totalcursos	0.12037	19	CondicionFamiliar	0.00692
7	NivelEducatPadre	0.10474	20	ResidDepart	0.00597
8	NivelEducMadre	0.09488	21	ActivEconomMadre	0.00568
9	AñoNacimiento	0.07865	22	Sexo	0.00558
10	SitLaboralEstud	0.07334	23	OrientacionVocac	0.00528
11	ActivEconPadre	0.04726	24	EmpleoTics	0.00411
12	PoseeBeca	0.04402	25	Credaprobados	0
13	ResidDistrito	0.03587	26	Cursosaprobados	0

En el Cuadro 4, se muestra los algoritmos empleados, además de las instancias correctamente e incorrectamente clasificadas, mediante validación cruzada.

Cuadro 4
Exactitud de los algoritmos

Algoritmo	Instancias correctamente clasificadas	Instancias incorrectamente clasificadas
J48graft	84.507 %	15.493 %
J48	83.8028 %	16.1972 %
PART	83.0986 %	16.9014 %

La matriz de confusión es presentada en el Cuadro 5, 6 y 7, compara las clasificaciones actuales y predichos. Además, las precisiones para las tres clases de las categorías de salida son presentadas. Se realizó un conjunto de experimentos con el objetivo de pronosticar el rendimiento académico de los estudiantes de Ingeniería de Industrias Alimentarias de la UNJ. Esto se pudo conseguir, debido a que tres algoritmos de clasificación (J48graft, J48 y PART) obtuvieron un porcentaje de instancias de buena clasificación mayor que el 83% del rendimiento académico; el mismo que se relacionó con el promedio global de los estudiantes hasta el semestre 2019-I.

Cuadro 5
Matriz de confusión para el algoritmo J48graft

		Actual			Precisión de la clase %
		Proceso	Inicio	Esperado	
Predicción	Proceso	115	2	0	85.2
	Inicio	3	5	0	71.4
	Esperado	17	0	0	0
Recuerdo de la Clase %		98.3	62.5	0	

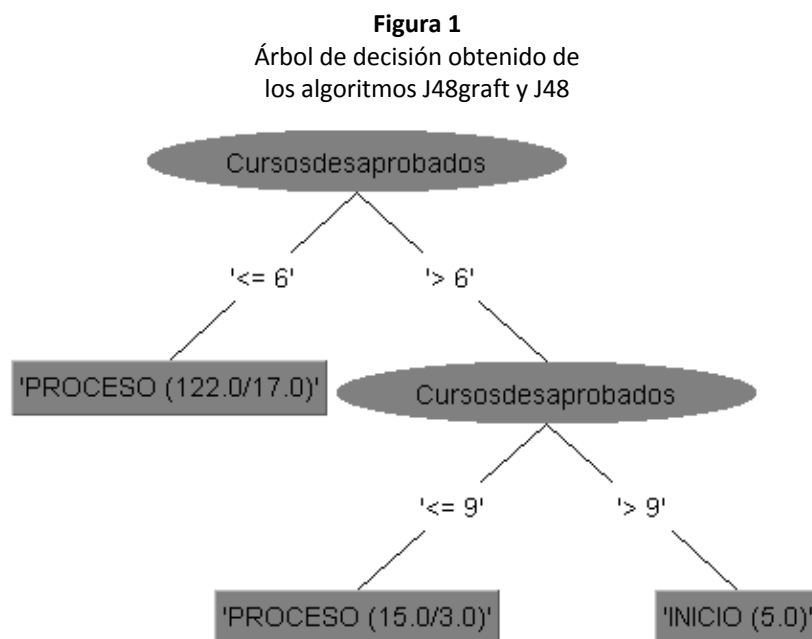
Cuadro 6
Matriz de confusión para el algoritmo J48

		Actual			Precisión de la clase %
		Proceso	Inicio	Esperado	
Predicción	Proceso	114	3	0	85.1
	Inicio	3	5	0	62.5
	Esperado	17	0	0	0
Recuerdo de la clase %		97.4	62.5	0	

Cuadro 7
Matriz de confusión para el algoritmo PART

		Actual			Precisión de la clase %
		Proceso	Inicio	Esperado	
Predicción	Proceso	113	4	0	85.0
	Inicio	3	5	0	55.6
	Esperado	17	0	0	0
Recuerdo de la clase %		96.6	62.5	0	

Para determinar el rendimiento académico, se empleó tres reglas extraídas del uso de los algoritmos dados en el Cuadro 8. Se observa en general que el algoritmo J48 descubre un árbol con tres hojas y de tamaño cinco, ver Figura 1. Con respecto a la primera regla nos dice que, si la cantidad de los cursos desaprobados es menor a seis, entonces el estudiante de la carrera de Industrias Alimentarias de la UNJ tiene un nivel de logro en proceso. La segunda regla nos dice que, si los cursos aprobados son mayores que seis y menores o iguales que nueve, entonces el rendimiento académico está en proceso. Una tercera regla se puede deducir que, si los cursos desaprobados son mayores que nueve, entonces se puede decir que el rendimiento académico del mencionado estudiante es inicial.



Cuadro 8
Reglas obtenidas a través de los algoritmos

Nombre	Descripción de la regla
Regla 1	Si Cursosdesaprobados <= 6, entonces Rendimiento académico = PROCESO.
Regla 2	Si Cursosdesaprobados > 6 y Cursosdesaprobados <= 9, entonces Rendimiento académico= PROCESO.
Regla 3	Si Cursosdesaprobados >9, entonces Rendimiento académico = INICIO.

4. Conclusiones

Se creó una base de datos compuesta de información académica y personal de los estudiantes de la Carrera de Ingeniería de Industrias Alimentarias de la Universidad Nacional de Jaén. A partir de ello, se identificaron siete variables, tales como el nombre del colegio, cantidad de cursos desaprobados, ciclo de ingreso a la universidad, lugar del Colegio, número de repeticiones en Matemáticas, el total de cursos y nivel educativo del padre que más inciden en el rendimiento académico.

Tres algoritmos de minería de datos nos permitieron detectar tres reglas para determinar el rendimiento académico de los estudiantes; estos han sido J48graft, J48 y PART. Los mismos que fueron determinados con el software Weka con un porcentaje de clasificación correcta mayor de ochenta y tres (83%).

Referencias bibliográficas

- Ahmad, F., Ismail, N. H., & Aziz, A. A. (2015). The prediction of students' academic performance using classification data mining techniques. *Applied Mathematical Sciences*, 9(129), 6415–6426.
- Alcover, R., Benlloch, J., Blesa, P., Calduch, M. A., Celma, M., Ferri, C., Hernández-Orallo, J., Iniesta, L., Más, J., & Ramírez-Quintana, M. J. (2007). Análisis del rendimiento académico en los estudios de informática de la Universidad Politécnica de Valencia aplicando técnicas de minería de datos. Teruel, España. Disponible en: <http://bioinfo.uib.es/~joemiro/aenui/procJenui/Jen2007/laalta.pdf>.
- Azoumana, K. (2013). Análisis de la deserción estudiantil en la Universidad Simón Bolívar, facultad Ingeniería de Sistemas, con técnicas de minería de datos. *Pensamiento americano*, 6(10).
- Chandra, E., & Nandhini, K. (2010). Knowledge mining from student data. *European journal of scientific research*, 47(1), 156–163.
- Cuadros Luyo, V. C., Morán Pachas, B. V., Jimenez, T., & Yaquel, A. (2017). Hábitos de estudio, disposición hacia el estudio y rendimiento académico en estudiantes de formación profesional técnica de un Instituto Armado.
- Dapozo, G. N., Porcel, E., López, M. V., & Greiner, C. L. (2008). Técnicas de clasificación aplicadas al estudio del rendimiento de ingresantes universitarios. X Workshop de Investigadores en Ciencias de la Computación.
- Gallardo Corrales, D. E. (2017). Análisis de patrones de deserción estudiantil de la unidad educativa Lenin School aplicando minería de datos [Master's Thesis]. Universidad de las Fuerzas Armadas ESPE. Maestría en Gestión de Sistemas de
- Hernández, N. (2009). Modelos de simulación de cultivos. Características y usos. *Revista Cultivos Tropicales*, 30(1), 73-82.
- Kumar, M., Singh, A. J., & Handa, D. (2017). Literature survey on educational dropout prediction. *International Journal of Education and Management Engineering*, 7(2), 8.
- Kumar, S. A., & Vijayalakshmi, M. N. (2012). Mining of student academic evaluation records in higher education. 2012 International Conference on Recent Advances in Computing and Software Systems, 67–70.
- Martínez, D. L. L. R., Karanik, M., Giovannini, M., & Pinto, N. (2015). Perfiles de rendimiento académico: Un modelo basado en minería de datos. *Campus Virtuales*, 4(1), 12–30.
- Porcel, E., Dapozo, G. N., & López, M. V. (2009). Modelos predictivos y técnicas de minería de datos para la identificación de factores asociados al rendimiento académico de alumnos universitarios. XI Workshop de Investigadores en Ciencias de la Computación. <http://sedici.unlp.edu.ar/handle/10915/19846>
- Powers, D. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, 2(1), 37-63. https://www.academia.edu/3090500/Evaluation_From_Precision_Recall_and_F_Measure_to_ROC_Informedness_Markedness_and_Correlation
- Ramesh, V., Parkavi, P., & Ramar, K. (2013). Predicting student performance: A statistical and data mining approach. *International journal of computer applications*, 63(8).
- Satyanarayana, A., & Nuckowski, M. (2016). Data mining using ensemble classifiers for improved prediction of student academic performance. In ASEE Mid-Atlantic Section Spring 2016 Conference.