



XVI
Congreso Nacional de
Investigación Educativa
CNIE-2021

Predicción del éxito académico en estudiantes de primer año de la carrera en la Facultad de Medicina de la UNAM: comparación entre Redes Neuronales Artificiales y Naïve Bayes

Irma Rocío Zavala Sierra

Coordinación de Universidad Abierta, Innovación Educativa y Educación a Distancia
irma_zavala@cuaieed.unam.mx

Patricia González-Flores

Coordinación de Universidad Abierta, Innovación Educativa y Educación a Distancia
patricia_gonzalez@cuaieed.unam.mx

Melchor Sánchez Mendiola

Coordinación de Universidad Abierta, Innovación Educativa y Educación a Distancia
melchor_sanchez@cuaieed.unam.mx

Área temática 09. Sujetos de la educación.

Línea temática: Trayectorias escolares, profesionales e institucionales individuales y colectivas.

Tipo de ponencia: Reportes parciales o finales de investigación.



Resumen

El rezago y el abandono de los estudios en la carrera de Médico Cirujano constituyen problemáticas persistentes e inciden principalmente al inicio de los estudios. En las últimas décadas se han realizado varias investigaciones para estudiar la regularidad y el egreso, e identificar las causas o los factores relacionados con el avance de sus alumnos con metodologías estadística. Aunque hay estudios que utilizan minería de datos en otras carreras, no se identificaron investigaciones sobre la implementación de estos métodos en educación médica. El objetivo de este trabajo fue predecir la pertenencia al grupo de regularidad o irregularidad de los estudiantes de primer año de la carrera de medicina de la Facultad de Medicina de la UNAM a través de los métodos de redes neuronales artificiales (RNA) y Naïve Bayes e identificar los factores que caracterizan a los estudiantes regulares e irregulares. Asimismo, se buscó reconocer las diferencias entre ambos métodos. Para realizar este estudio, se utilizó información de 7066 alumnos de las cohortes 2011 a 2017. Con ambos métodos, se obtuvieron buenos resultados de predicción, con un ligero mejor desempeño de redes neuronales; sin embargo, Naïve Bayes permitió discriminar los factores de mayor relevancia en la predicción de la irregularidad o regularidad.

Palabras clave: éxito escolar, educación médica, minería de datos, Naïve Bayes, redes neuronales artificiales.

Introducción

El rezago escolar y el abandono de los estudios son problemáticas persistentes en todas las instituciones de educación superior, incluyendo la carrera de Médico Cirujano en la UNAM, y se ha identificado que tienen mayor incidencia al inicio de los estudios. Una investigación de la trayectoria de estudiantes de esta carrera, con datos de 25 generaciones (1985-2010), reveló que solo el 54% de los estudiantes aprobaron todas las materias durante su primer año y del total sólo el 41% egresa en los cinco años que estipula el plan de estudios (Campillo et al., 2019). Así, estudiar los factores relacionados con el rezago e identificar a los estudiantes en riesgo lo más temprano posible resulta importante para implementar intervenciones que promuevan el avance en los estudios.

En las últimas décadas, se han realizado investigaciones para estudiar la trayectoria de los estudiantes del nivel superior. Una primera línea de investigación se ha orientado a identificar los factores relacionados con el avance del alumnado en medicina utilizando métodos estadísticos. Se reporta que los conocimientos previos -en particular sobre matemáticas, física, español- y los atributos psicológicos son las principales variables relacionadas con el desempeño escolar (Vargas et al., 2011; Ponce de León, Ortiz, y Morán, 2003; Gatica-Lara et al., 2010; Urrutia et al., 2014; Martínez-González et al., 2018); algunos factores vinculados con la escuela donde se cursó el bachillerato también han mostrado estar correlacionados con el rendimiento, mientras que no se comprobó una relación con el género ni con estilos de aprendizaje (Nieto, Varela y Fortoul, 2003). Entre las características de la escuela de procedencia, se ha identificado que el subsistema o programa de bachillerato constituye un factor relevante para el desempeño escolar (Ponce de León, Ortiz y Morán 2003; Guevara-Guzman, Galván y Muñoz-Comonfort 2006; Gatica-Lara et al., 2010). En la UNAM, se consideran tres tipos de subsistemas que se diferencian por la naturaleza y estructura del plan de estudios:

1. El subsistema A que imparte una formación cultural, desarrollo integral y capacitación para estudios profesionales.
2. El subsistema B que combina los estudios con una capacitación práctica para preparar al estudiante también para el trabajo.
3. El subsistema C comprende otros programas de estudio de este nivel educativo.

Investigaciones realizadas en otras instituciones educativas en México y el mundo confirman que las calificaciones en estudios previos constituyen uno de los principales indicadores del éxito académico de los alumnos de medicina y se reporta también la incidencia de factores socio-económicos, la violencia intrafamiliar, el nivel de inglés, la carencia de hábitos y estrategias de estudio, la escuela de procedencia y la edad (Valero, Salvador y García, 2010). Este último factor, también incide en la reprobación de asignaturas, junto con atributos del docente; por ejemplo, es interesante notar que entre 73 y 100% de los docentes con mayor índice de reprobación tienen el grado de doctor (Rodallegas et al., 2010).

Otra línea de investigación emplea métodos de minería de datos (en inglés, *Educational Data Mining* o EDM) para generar modelos predictivos del desempeño escolar en diversos programas de educación superior a partir de los datos disponibles en las instituciones educativas (Amaya et al., 2015, Abu Amra y Maghari, 2017). A nivel internacional, existe una gran variedad de modelos aplicados con éxito para este fin en distintas disciplinas, como los árboles de decisiones, la regresión logística, las redes bayesianas, las máquinas de soporte de vectores, las máquinas de aprendizaje extremo, el vecino más cercano-k, el algoritmo de aglomeración k-means, el J48, el zeroR y los árboles aleatorios, entre otros. De acuerdo con los resultados reportados, se han logrado predicciones precisas con los modelos de máquina de soporte de vectores, vecino más cercano-k (k-nearest neighbour), redes neuronales artificiales (RNA), Naïve Bayes y árboles aleatorios (Amaya et al., 2015; Mason et al., 2018; Nida Uzel et al., 2018; Solis et al., 2018). En el contexto mexicano, sólo se identificaron investigaciones en disciplinas relacionadas con las ciencias de la computación (López, González y Sandoval, 2019). Cabe resaltar una investigación que logró predecir el rendimiento académico de los estudiantes de primer año de ingenierías usando el modelo Naïve Bayes con una precisión del 50% y alcanzando en algunas subpoblaciones hasta el 70% (Ibarra y Medina, 2011).

Sin embargo, los estudios realizados en educación médica para explorar y comparar distintos modelos de minería de datos con la finalidad de predecir el desempeño escolar son escasos y se enfocan a utilizar datos recabados de un Learning Management System (LMS) o de recursos digitales para identificar patrones de comportamiento en línea o predecir el desempeño (Saqr, 2018). Así, resulta necesario explorar la aplicación de la minería de datos para estudiar las trayectorias escolares en educación médica. El objetivo de esta investigación es valorar la utilidad de dos modelos en particular, Naïve Bayes y RNA, para predecir el desempeño de los estudiantes e identificar los factores que inciden en el rezago y la regularidad estudiantil, en el caso de la carrera de Médico Cirujano de la Facultad de Medicina, de la UNAM.

Desarrollo

La carrera de Médico Cirujano en la UNAM cuenta con una matrícula numerosa: 9,045 alumnos, 5,820 mujeres y 3,225 hombres (FacMed, 2019). Al ingreso, se recaba información relacionada al contexto familiar y trayectoria académica previa mediante una encuesta y se aplica una evaluación diagnóstica sobre su nivel de conocimientos en ocho materias, incluyendo dominio de inglés y español. Asimismo, durante los estudios, se registran datos sobre el avance escolar de los estudiantes, tales como la calificación obtenida en cada asignatura, el tipo de examen presentado (ordinario o extraordinario) y los créditos obtenidos. A partir de la comparación entre los créditos logrados por un alumno y aquellos establecidos en el plan de estudios, se clasifica a los estudiantes como regulares o irregulares. En esta investigación, se consideró como variable dependiente el avance de los alumnos en créditos al término del primer año y se definieron dos categorías:

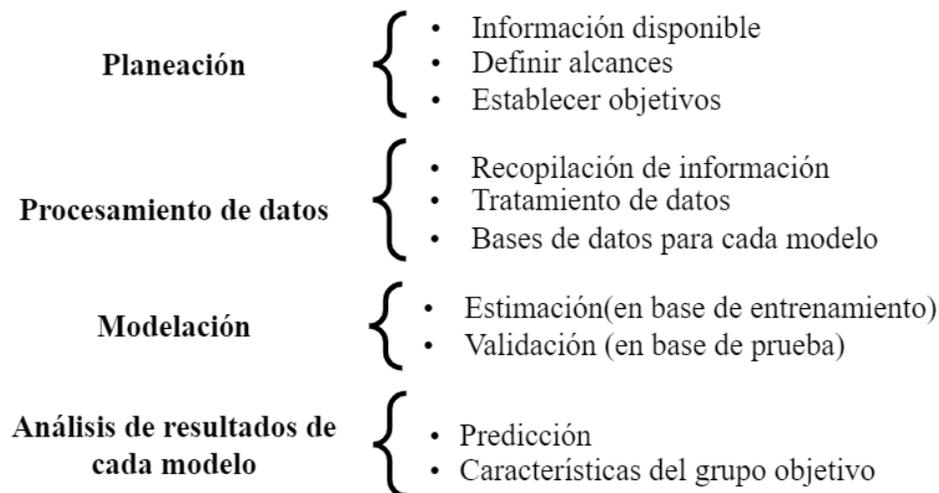
- Regularidad: estudiantes que obtuvieron el 100% de los créditos.
- Irregularidad: estudiantes que completaron entre 0 y 99% de los créditos.

Como variables explicativas, se consideraron atributos relacionados con el nivel de conocimientos al ingreso de la carrera, la trayectoria escolar en el bachillerato y el contexto socio-económico de los estudiantes.

Método

La investigación se realizó en cuatro etapas (Figura 1). En la planeación, se construyó el marco teórico y se establecieron los alcances y objetivos de la investigación. Después se realizó el preprocesamiento de datos con base en las características de cada modelo y a continuación se generaron los modelos para cada método. En la última etapa, se compararon y analizaron los resultados obtenidos con ambos modelos.

Figura 1: Metodología de la investigación



Procesamiento de datos

La base de datos utilizada comprendió información de 7,976 alumnos de las cohortes 2011 a 2017 de la carrera de Médico Cirujano, recabados tanto al ingreso como durante los estudios. Incluyó 47 variables -categóricas, numéricas discretas o numéricas continuas- que fueron clasificadas en nueve grupos: avance en créditos, información personal, contexto familiar, situación económica, trayectoria previa, nivel de conocimientos generales, dominio del idioma español e inglés al ingreso a la carrera, desempeño durante la licenciatura e identificadores del estudiante. Los registros con más del 40% de los campos con valores perdidos fueron retirados de la base de datos (910 registros). El catálogo completo de estas variables puede consultarse en <https://analiticadelaprendizaje.cuaieed.unam.mx/>.

Los dos métodos comparados en este estudio, RNA y Naïve Bayes, requieren procesamiento de datos diferente, por lo cual se prepararon dos bases de datos. En ambos casos, los datos fueron a su vez divididos en datos de entrenamiento (80%) y datos de prueba (20%).

Redes Neuronales Artificiales

Los valores faltantes se sustituyeron empleando una imputación suave, es decir, el valor faltante en las variables numéricas se reemplaza por la media, mientras que en las variables categóricas se recurre a la moda. De las 47 variables totales, sólo 27 presentaron casos con valores faltantes; el porcentaje máximo de valores faltantes dentro de una variable fue menor del 5%.

Naïve Bayes

Los valores faltantes se trataron como una categoría más de la variable. Las variables numéricas continuas asociadas a porcentajes de respuesta se categorizaron en cinco grupos a través de los percentiles poblacionales (Moreno y Stephens , 2015); los valores de cada grupo se calcularon de manera que concentraran aproximadamente el 20% de los casos totales. Las variables numéricas discretas se reagruparon para homogenizar el porcentaje de respuesta en cada una de ellas.

Modelación

Con cada uno de los métodos, RNA y Naïve Bayes, se estimaron dos modelos: uno para predecir la regularidad de los estudiantes y otro, la irregularidad. Después, se buscó definir perfiles de estudiantes regulares e irregulares.

Redes Neuronales Artificiales

El ajuste de los modelos, se realizó empleando las librerías de Python y Tensor Flow de Google. Se realizó una validación cruzada y se verificó la precisión en cada uno de ellos para validar que ésta se mantuviera con distintos subconjuntos de datos.

Además, se realizó un análisis de sensibilidad de las distintas variables para identificar la influencia que cada variable tiene sobre la red neuronal y sus resultados. El análisis consistió en entrenar una red eliminando del conjunto de datos una variable y registrando la precisión en cada caso; posteriormente, se integró de nuevo la variable removida y se quitó otra. Al final, se comparó la precisión obtenida con los múltiples modelos mediante validación cruzada.

Naïve Bayes

En el caso de Naïve Bayes, los modelos se ajustaron con el software R. Primero, se calcularon la probabilidad y el score para cada una de las categorías de respuesta de las variables que representa la relevancia de esta categoría para la pertenencia en el grupo objetivo. A partir del score resultante de la categoría, se estimó el

score para cada estudiante. Además, para analizar la significancia de cada categoría de respuesta, se calcularon los valores de ϵ :

Donde representa la clase, el atributo según la categoría de respuesta y el número de estudiantes con el atributo . Las categorías de respuesta con valores de ϵ mayores a 2 o menores a -2 son consideradas significativas para la predicción (Stephens et al., 2009; Moreno y Stephens, 2015).

Finalmente, se definió el punto de corte con base en los resultados de sensibilidad y especificidad: se eligió aquel en el cual ambos valores fueron lo más cercanos a 1, es decir, se buscó el punto más cercano al (1,0) en la curva de ROC (Fawcett, 2006). Utilizando el punto de corte elegido, se realizó la validación del modelo calculando los valores de precisión, sensibilidad y especificidad.

Resultados

Los resultados de predicción de los modelos fueron parecidos para ambos métodos: mayores o iguales al 70% en sensibilidad, especificidad y precisión en ambos casos (Tabla 1). En la validación cruzada de los modelos de RNA se obtuvo una precisión media de 0.72 y una varianza de 0.01 tanto para la predicción de regularidad como irregularidad.

Tabla 1: Comparación de los resultados de predicción

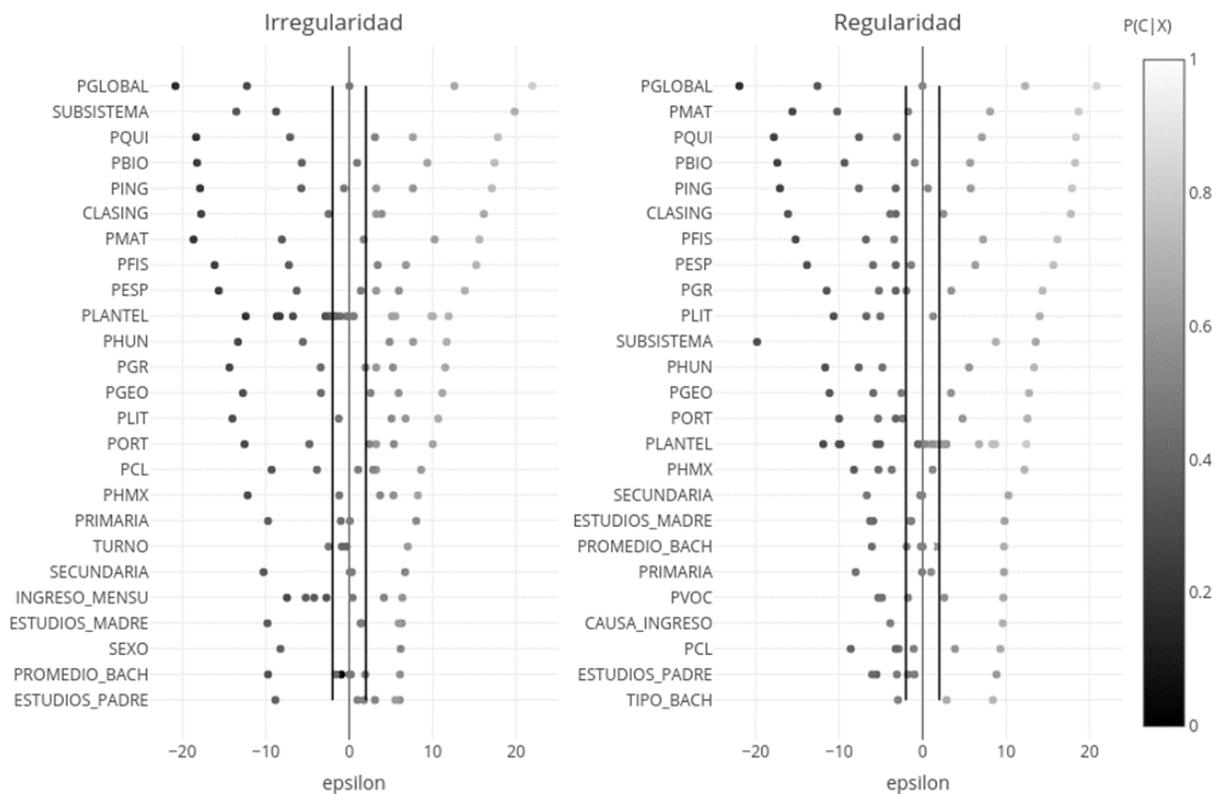
Grupo objetivo	Modelo	Sensibilidad	Especificidad	Precisión
Irregularidad	Redes neuronales	0.72	0.75	0.74
	Naïve Bayes	0.72	0.7	0.71
Regularidad	Redes neuronales	0.75	0.72	0.73
	Naïve Bayes	0.7	0.72	0.71

El análisis de sensibilidad aplicado para el método de RNA reflejó que ninguna de las variables explicativas genera una diferencia en precisión mayor a 0.02. Los valores de precisión media oscilaron entre 0.72 y 0.74 mientras que el valor de precisión del modelo con todas las variables explicativas fue de 0.74 para el modelo de irregularidad y 0.73 para el de regularidad. Dado lo anterior, no fue factible tener una caracterización adecuada de los perfiles de los estudiantes usando redes neuronales artificiales, aunque cabe aclarar que este análisis no consideró la relevancia que pueden tener las variables en combinación debido a que sólo se evaluó su impacto de manera individual.

Los resultados de Naïve Bayes para regularidad e irregularidad se resumen en la Figura 2, que muestra las 25 variables con mayor valor predictivo para cada caso. A la izquierda, se presentan los resultados para la irregularidad y a la derecha, para regularidad. Las variables se ordenan con base en el valor de ϵ de forma descendente; como ϵ es un indicador de la relevancia del atributo para la predicción, aquellas más importantes aparecerán arriba.

Las líneas verticales representan los puntos donde ϵ es igual a 2 y -2, respectivamente. Los valores positivos de ϵ (a la derecha de la línea vertical) señalan las categorías que favorecen pertenecer al grupo objetivo, mientras que los valores negativos indican aquellas que favorecen no pertenecer a esta clase. La intensidad del sombreado de los puntos representa la probabilidad de pertenecer al grupo objetivo; aquellos más oscuros tienen una probabilidad cercana a 0, mientras que los más claros, cercana a 1. Una versión interactiva de estos gráficos puede consultarse en <https://analiticadelaprendizaje.cuaieed.unam.mx/>.

Figura 2: Variables con mayor valor predictivo de acuerdo con ϵ para regularidad e irregularidad



En ambos casos, el atributo con mayor valor predictivo fue el porcentaje global de aciertos en el examen diagnóstico. La mayoría de las variables mostraron un impacto similar o muy parecido en la predicción de la regularidad y la irregularidad. Sin embargo, se identificaron algunas diferencias en el ordenamiento: dos variables relacionadas con la escuela donde los estudiantes completaron el bachillerato tienen mayor valor predictivo para los estudiantes irregulares (subsistema y plantel), mientras que el conocimiento previo en matemáticas tienen mayor incidencia en la regularidad. Por otro lado, hay algunos atributos que sólo aparecieron entre las 25 variables con mayor impacto para uno de los modelos: en el caso de la irregularidad, el sexo y el ingreso familiar; para la regularidad, el porcentaje de aciertos en vocabulario y la causa de ingreso.

Conclusiones

Los resultados obtenidos permiten comprobar que la aplicación de Naïve Bayes y RNA para la predicción del avance regular o irregular de los estudiantes de la carrera Médico Cirujano generaron valores adecuados y consistentes con los hallazgos de otras investigaciones sobre su utilidad para predecir el desempeño escolar en otras disciplinas. A partir de Naïve Bayes, se pudieron identificar los atributos con mayor relevancia para la predicción del avance regular o irregular.

De estos resultados, se desprenden varias reflexiones. En primer lugar, se corrobora la relevancia de los conocimientos previos de los estudiantes como predictor del avance en los estudios de la carrera Médico Cirujano, en particular de química, biología, física y matemáticas, así como de temas vinculados con el dominio del español e inglés, hallazgo consistente investigaciones previas (Ponce de León, Ortiz y Morán, 2003; Gatica-Lara et al., 2010; Urrutia et al., 2014). Se refuerza la necesidad de implementar intervenciones para fortalecer los conocimientos previos de los estudiantes de nuevo ingreso con la finalidad de promover el avance en los estudios. Cabe aclarar que aunque el dominio del inglés también resultó relevante, habría que indagar si su relevancia predictiva se relaciona con el uso de este idioma para los estudios o si refleja una relación indirecta con el nivel socio-económico de la familia.

En segundo lugar, es necesario analizar por qué el tipo de subsistema donde se estudió el bachillerato es buen predictor de la irregularidad y por qué este valor predictivo se reduce para la regularidad. Pareciera que hay factores del Subsistema B que afectan más a los estudiantes irregulares. Una primera hipótesis podría ser que la estructura del plan de estudios está generando una desventaja para los estudiantes de este subsistema B. A diferencia del subsistema A donde los alumnos tienen que cursar las materias del área de conocimientos de la licenciatura que quieren estudiar, en este subsistema los estudiantes eligen las materias que cursarán en el último año de bachillerato entre un grupo de asignaturas. Por ello, los alumnos del subsistema B pueden no cursar Biología, Química o Física avanzada. Asimismo, en el subsistema B, sólo completan un curso de matemáticas avanzadas en el último año de bachillerato y no toman inglés, mientras que en el subsistema A tienen dos asignaturas de matemáticas avanzadas y estudian inglés.

Una segunda hipótesis a explorar se relaciona con el impacto de las expectativas de los docentes sobre el éxito de los alumnos, fenómeno denominado como “principio Pygmalion” o profecía auto-cumplida que ha sido reportado en la literatura. Diversas investigaciones han reportado que los estudiantes del bachillerato B obtienen calificaciones más bajas que el subsistema A en el examen diagnóstico (Sánchez-Mendiola et al, 2020) y el primer año de la licenciatura (Sánchez-Mendiola et al 2017; Ponce de León, Ortiz, Morán 2003; Guevara-Guzman, Galván y Muñoz-Comonfort 2006; Gatica-Lara et al., 2010). En tanto esta información es conocida por la planta académica, los docentes pueden tener expectativas de un menor desempeño de los estudiantes de este subsistema y con ello influir en su menor rendimiento académico.

También, un sesgo en los modelos hacia un tipo de subsistema podría explicar el comportamiento diferencial de la predicción, por lo que sería importante estudiar si existe algún tipo de discriminación en los algoritmos. Por lo tanto, sería importante valorar si existen sesgos relacionados con la variable subsistema así como factores socio-económicos, culturales y sexo.

Esta investigación comprobó que tanto las RNA como Naïve Bayes permiten generar modelos para predecir el avance regular o irregular en los estudiantes de la carrera Médico Cirujano con resultados similares, a partir de información sobre conocimientos previos y factores socio-demográficos de los estudiantes. Las RNA tuvieron resultados ligeramente superiores; por su parte, Naïve Bayes permitió estudiar el valor predictivo de los atributos considerados para cada una de las clases con la finalidad de profundizar el análisis de su incidencia en las distintas trayectorias escolares. De esta manera, ambos métodos pueden resultar complementarios.

Limitaciones

Existen factores como la motivación intrínseca, las preferencias personales, el nivel de integración social y académica y datos de la trayectoria escolar del alumno durante su primer año de licenciatura, que no se capturan en este estudio. En relación con las RNA no se analizó la relevancia para la predicción considerando conjuntos de variables.

Agradecemos las contribuciones del Maestro Manuel Mejía Minjares, el Maestro Diego Monteverde y el Biólogo Roberto Santos Solórzano, quienes han sido parte importante en el desarrollo, ejecución y análisis de esta investigación.

Referencias

- Abu Amra, I. A., y Maghari, A. Y. A. (2017). Students Performance Prediction Using KNN and Naïve Bayesian. In IEEE (Ed.), 8th International Conference on Information Technology (ICIT) (pp. 29–35). 10.1109/ICITECH.2017.8079967
- Amaya, Y., Barrientos, E. y Heredia, D. (2015). Student Dropout Predictive Model Using Data Mining Techniques. *IEEE Latin America Transactions*, 13(9), 3127–3134. <https://doi.org/10.1109/TLA.2015.7350068>
- Campillo Labrandero, M., Martínez González, A., García Minjares, M., Guerrero Mora, L., y Sánchez Mendiola, M. (2019). Desempeño académico y egreso en 25 generaciones de estudiantes de la Facultad de Medicina de la UNAM. *Educación Médica*. <https://doi.org/10.1016/j.edumed.2019.05.003>
- de Boer, H., Timmermans, A. C., y van der Werf, M. P. C. (2018). The effects of teacher expectation interventions on teachers' expectations and student achievement: narrative review and meta-analysis. *Educational Research and Evaluation*, 24(3–5), 180–200. <https://doi.org/10.1080/13803611.2018.1550834>
- FacMed. (2019). *Informe Anual. Facultad de Medicina*. Universidad Nacional Autónoma de México. México.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>

- Gatica-Lara, F., Mendez-Ramirez, I., Sanchez-Mendiola, M., y Martinez-Gonzalez, A. (2010). Variables asociadas al éxito académico en estudiantes de la Licenciatura en Medicina de la UNAM. *Revista de la Facultad de Medicina de la UNAM*, 53(5), 9–18.
- Guevara-Guzmán, R., Galván, A., y Muñoz-Comonfort, M. (2006). El desempeño académico de los estudiantes de la Facultad de Medicina en el primer año de la carrera. *Gaceta Médica de México*, 143(1), 27–32.
- Ibarra García, E. P., y Medina Mora, P. (2011). Model Prediction of Academic Performance for First Year Students. En IEEE (Ed.), *2011 10th Mexican International Conference on Artificial Intelligence* (pp. 169-174). <https://doi.org/10.1109/MICAL.2011.28>
- López Pedraza, F. J., González Macías, M. del C., y Sandoval García Edgar R. (2019). Minería de Datos: Identificando causas de deserción en las Instituciones Públicas de Educación Superior de México. *Revista de Tecnología e Innovación En Educación Superior*, 1(2), 1–12.
- Martínez-González, A., Manzano-Patiño, A. P., García-Minjares, M., Herrera-Penilla, C. J., Buzo-casanova, E. R., y Sánchez-Mendiola, M. (2018). Grado de conocimientos de los estudiantes al ingreso a la licenciatura y su asociación con el desempeño escolar y la eficiencia terminal. Modelo multivariado. *Revista de la Educación Superior*, 47(188). <https://doi.org/10.36857/resu.2018.188.508>
- Mason, C., Twomey, J., Wright, D., y Whitman, L. (2018). Predicting Engineering Student Attrition Risk Using a Probabilistic Neural Network and Comparing Results with a Backpropagation Neural Network and Logistic Regression. *Research in Higher Education*, 59(3), 382–400. <https://doi.org/10.1007/s11162-017-9473-z>
- Moreno, J. G., y Stephens, C. R. (2015). Applying Data Mining Techniques to Identify Success Factors in Students Enrolled in Distance Learning: A Case Study. En A. F. G. Pichardo Lagunas O., Herrera Alcántara O. (Ed.), *Mexican International Conference on Artificial Intelligence (MICAL) 2015: Advances in Artificial Intelligence and Its Applications* (Vol. 9414, pp. 208–219). Springer, Cham. https://doi.org/10.1007/978-3-319-27101-9_15
- Nida Uzel, V., Sevgi Turgut, S., y Ayşe Özel, S. (2018). Prediction of Students' Academic Success Using Data Mining Methods. En IEEE (Ed.), *2018 Innovations in Intelligent Systems and Applications Conference (ASYU 2018)* (pp. 166–170). doi:10.1109/ASYU.2018.8554006
- Nieto Domínguez, D. M., Varela Ruíz, M., y Fortoul, T. I. (2003). Aprendizaje: género, escuela y rendimiento académico. *Revista de la Educación Superior*, 32(125), 53–66.
- Ponce de León, M. E., Ortiz Montalvo, A., y Morán Álvarez, C. (2003). Valor predictivo del desempeño académico del examen diagnóstico, aplicado a los alumnos de primer ingreso a la carrera de médico cirujano de la facultad de medicina de la UNAM. *Revista de la Facultad de Medicina UNAM*, 46(5), 185–189.
- Rodallegas Ramos, E., Torres González, A., Gaona Couto, B. B., Gastelloú Hernández, E., Lezama Morales, R. A., y Valero Orea, S. (2010). Modelo predictivo para la determinación de causas de reprobación mediante Minería de Datos. En Prieto, M. E., Doderó, J. M., y Villegas, D. O. (Eds.), *Recursos digitales para la educación y la cultura*. Volumen KAAMBAL. (pp. 48–55).
- Sagr M. (2018). A literature review of empirical research on learning analytics in medical education. *International Journal of Health Sciences*, 12(2), 80–85.
- Sánchez-Mendiola, M., López-Marypaola, J. M., Buzo-Casanova, E. R., Herrera-Penilla, C. J., García-Minjares, M., y Martínez-González, A. (2017). El desempeño escolar de los estudiantes de la educación media superior y su transición al nivel superior dentro de la Universidad Nacional Autónoma de México. En *Congresos CLABES VII*, Cordoba, Argentina. Recuperado a partir de <https://revistas.utp.ac.pa/index.php/clabes/article/view/1658>

- Sánchez-Mendiola, M. Martínez-González, A. Buzo-Casanova, E. Goytia-Rodríguez, K. Hernández-Flores, M. García-Minjares, M. y Manzano-Patiño, A. (2020). *Exámenes para el diagnóstico de conocimientos. Resultados de los alumnos que ingresan al nivel licenciatura 2020*. Primera edición. Ciudad de México. Coordinación de Desarrollo Educativo e Innovación Curricular. Universidad Nacional Autónoma de México. Disponible en: <https://www.dee.cuaieed.unam.mx/index.php/resultado-de-diagnostico/>
- Solis, M., Moreira, T., Gonzalez, R., Fernandez, T., y Hernandez, M. (2018). Perspectives to Predict Dropout in University Students with Machine Learning. En IEEE (Ed.), *2018 IEEE International Work Conference on Bioinspired Intelligence, IWOB 2018*. <https://doi.org/10.1109/IWOB.2018.8464191>
- Stephens, C. R., Heau, J. G., González, C., Ibarra-Cerdeña, C. N., Sánchez-Cordero, V., y González-Salazar, C. (2009). Using biotic interaction networks for prediction in biodiversity and emerging diseases. *PLOS ONE*, 4(5). <https://doi.org/10.1371/journal.pone.0005725>
- Urrutia Aguilar, M. E., Ortiz Leon, S., Fouilloux Morales, C., Ponce Rosas, E. R., y Guevara Guzman, R. (2014). El rendimiento académico en el primer año de la carrera de médico cirujano: modelo multivariado explicativo. *Gaceta Médica de México*, 150 (Suppl 3), 324–330.
- Valero Orea, S., Salvador Vargas, A., y García Alonso, M. (2010). Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos. En Prieto, M. E., Doderó, J. M., y Villegas, D. O. (Eds.), *Recursos digitales para la educación y la cultura. Volumen KAAMBAL*. (pp. 48–55)
- Vargas, I., Ramírez, C., Cortés, J., Farfán, A., y Heinze, G. (2011). Factores asociados al rendimiento académico en alumnos de la Facultad de Medicina: estudio de seguimiento a un año. *Salud Mental*, 34(4), 301–308.