



Revista de Gestión Empresarial
y Sustentabilidad



<http://rges.umich.mx>

Modelo Basado en Arbol de Decisiones para Determinar los Factores de Deserción de Estudiantes en una Institución de Educación Superior Mexicana

**Gustavo Alfonso Gutiérrez Carreón¹
María del Pilar Meza Bucio²**

¹Universidad Michoacana de San Nicolás de Hidalgo, maria.meza@umich.mx

²Universidad Michoacana de San Nicolás de Hidalgo, gustavo.gutierrez@umich.mx

Modelo Basado en Arbol de Decisiones para Determinar los Factores de Deserción de Estudiantes en una Institución de Educación Superior Mexicana

Resumen

El contexto actual en el que operan las Instituciones de Educación Superior en México es adverso y multifactorial. En este trabajo se analizan los datos obtenidos de una encuesta aplicada a 1.582 estudiantes para determinar los principales factores que influyen en la deserción escolar en una etapa pre-COVID19. Con esta información se desarrolló un análisis del árbol de decisión, detectando las principales rutas que influyen en la deserción escolar. Este estudio puede ser de utilidad tanto para el público como para las instancias involucradas en la toma de decisiones, para tratar de crear un entorno propicio que permita a los estudiantes continuar con su formación universitaria.

***Palabras claves:** analítica del aprendizaje, educación superior, toma de decisiones*

Introducción

La analítica de datos es una técnica utilizada en muchos campos de investigación, uno de ellos es Educación, siendo de gran utilidad para encontrar patrones en grandes conjuntos de datos, y con ello optimizar y predecir resultados que permitan mejorar la gestión y administración de las Instituciones de Educación Superior (IES), ya que es de gran utilidad para la detección y prevención de problemas específicos y toma de decisiones.

Las instituciones de educación superior (IES) se encuentran en un panorama complejo, donde sus recursos son limitados y las situaciones están cambiando. Según un diagnóstico del año 2019 realizado por la Subsecretaría de Educación Superior (SES), la Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES), y la Asociación Mexicana de Órganos de Control y Vigilancia en Instituciones de Educación Superior, AC (AMOCVIES), hay nueve entidades en México cuyas universidades llevan

varios años en crisis económica: Morelos, Oaxaca, Zacatecas, Chiapas, Estado de México, Tabasco, Michoacán, Nayarit y Sinaloa. La magnitud del déficit acumulado de las nueve universidades equivale al 71% de su subsidio público ordinario (23,461 millones de pesos), con un rango que va del 29% al 190%.

En este entorno económico y social desfavorable, existe la necesidad de detectar problemas transversales como el abandono escolar y las múltiples causas que lo originan, a través de diversas técnicas. Uno de ellos podría estar usando análisis de aprendizaje. Elias (2011) menciona que la analítica del aprendizaje se refiere a la recopilación y análisis de datos sobre los alumnos y sus entornos para comprender y mejorar los resultados del aprendizaje.

La Facultad de Contaduría y Ciencias Administrativas (FCCA) de la Universidad Michoacana de San Nicolás de Hidalgo (UMSNH) puede inscribir aproximadamente a 5,000 estudiantes de varios estados del país en las diferentes titulaciones y modalidades que ofrece. Sin embargo, los datos reportados por la UMSNH en los últimos cinco años del número de estudiantes que inician sus estudios disminuyen significativamente con el número de estudiantes que se gradúan. El principal objetivo de este trabajo es intentar encontrar la combinación de factores que genera un mayor riesgo de deserción en los estudiantes.

Trabajo relacionado

Orea, Vargas & Alonso (2005) proponen un modelo predictivo utilizando técnicas de minería de datos a través de una interfaz web que facilita la identificación de estudiantes vulnerables a la deserción escolar en la Universidad Tecnológica de Izúcar de Matamoros (UTIM), en México.

Felizzola, Arias, Pedroza & Pastrana (2018) diseñan un modelo de clasificación para detectar el abandono temprano en la Facultad de Ingeniería de la Universidad La Salle, mediante la aplicación de la metodología CRISP-D (Cross Industry Standard Process for Data Mining). En este trabajo se realizó una revisión de la literatura de 1982 a 2017, en la que se analizan las aplicaciones del machine learning y la minería de datos para abordar el

problema con métodos como árbol de decisión, redes neuronales, máquinas de soporte vectorial (SVM), k vecino más cercano (KNN), regresión logística (LR), entre otros, con los que se generan reglas de predicción en base a un conjunto de indicadores de gestión que pueden ser utilizados en el diseño de políticas educativas para determinar las razones de algunas ineficiencias de las IES.

Madrid Echeverry (2017) realiza un trabajo de tesis de maestría, donde el objetivo de la investigación se basó en el uso de técnicas estadísticas multivariadas: SVM, Análisis Discriminante (DA), KNN y LR para clasificar a estudiantes de pregrado de la Universidad Nacional de Colombia ubicados en dos localidades. de Medellín (con posibilidad o no de deserción), en base a la información disponible sobre las variables definidas e identificadas como determinantes de la deserción estudiantil universitaria. Para este estudio se utilizó la información brindada por los estudiantes que ingresaron a la Universidad Nacional de Colombia desde el primer semestre de 2009 hasta el primer semestre de 2016, su desempeño académico correspondiente en cada período matriculado y la identificación de cuál de ellos perdió, la calidad del estudiante en la universidad por bajo rendimiento y que continuó con sus estudios, lo que permitió tener un porcentaje de datos que se utilizaron para la formación de los modelos y el resto de los datos como validación. Los resultados permitieron identificar la técnica para obtener el modelo con menor porcentaje de error y mayor sensibilidad, y que podría utilizarse para realizar predicciones de deserción en nuevos individuos a partir de la información de las variables seleccionadas.

Existen trabajos relacionados en los que el uso de técnicas de sistemas expertos y minería de datos permiten establecer modelos de predicción, con los que ayudamos al responsable en la toma de estos. Cadavid, Mendoza & Rodríguez (2016) muestran los resultados de una investigación cuyo propósito es evaluar la eficiencia técnica de las IES en Colombia entre los años 2011-2013, mediante la aplicación de técnicas de análisis de envolvente de datos y minería de datos.

Noboa, Ordóñez & Magallanes (2018) proponen un modelo para detectar posibles deserciones en la Educación Superior en una universidad pública, donde se plantean dos

propuestas para la cuantificación de la deserción: La primera, se establece como la proporción de estudiantes que egresan en un tiempo determinado que corresponde a la duración de la carrera; y el segundo es simplemente el número de estudiantes que abandonan la escuela. Para reducir la deserción, estas investigaciones proponen mejorar los mecanismos de detección temprana de potenciales desertores. Para elaborar su investigación utilizaron algunos métodos: regresión logística, k vecinos más cercanos, árboles de decisión que incluyen bosques aleatorios, redes bayesianas, redes neuronales, entre otros.

Método

Para este proyecto se eligieron variables que identificaron a cada alumno de la Facultad de Contaduría y Ciencias Administrativas de la UMSNH, se tomaron como referencia estudios previamente elaborados por investigadores de otras universidades, como es el caso de la Universidad Complutense de Madrid, donde llevó a cabo un estudio para determinar el éxito / fracaso académico, utilizando las técnicas de regresión lineal múltiple y regresión logística (Jiménez, Izquierdo & Blanco, 2000).

En esta parte del trabajo se describirán las variables que se considera que inciden en la tasa de deserción de los estudiantes de las Licenciaturas de la Facultad de Contaduría y Ciencias Administrativas de la UMSNH. Además, se presenta el modelo conceptual que nos permitirá comprender la interacción de las variables para luego desarrollar el Modelo de Análisis de Datos.

Es importante tomar en consideración una rigurosa separación de los diferentes tipos de deserción para su estudio. Los autores (Vásquez, Castaño, Gallón & Gómez, 2003) explican que la deserción de los estudiantes se puede entender desde dos puntos de vista: temporal y espacial.

Como concepto temporal, identifican tres tipos de deserción:

1. Abandono prematuro: cuando un estudiante deja un programa antes de ser aceptado.
2. Abandono temprano: cuando el programa se abandona durante los primeros cuatro semestres.
3. Abandono tardío: entendido como abandono a partir del quinto semestre.

En segundo lugar, como concepto de espacio, se hace referencia al hecho de que un alumno:

1. Cambiar programas dentro de la misma institución.
2. Cambiar las instituciones educativas.
3. Abandono del sistema educativo, donde existe la posibilidad de reingreso en el futuro, ya sea al mismo u otro campus del país.

Para el análisis que se llevará a cabo en este trabajo, la deserción se tomará como una decisión de abandono del programa académico, que puede llegar a ser temporal o definitivo. Con los datos que tenemos del Control Escolar de la UMSNH, es difícil rastrear si los estudiantes que se han matriculado han abandonado temporalmente, por lo tanto, se debe concebir la deserción en general y por lo tanto se aplicará una encuesta en la que se detectarán los motivos de la deserción para analizar la información proporcionada por los estudiantes.

No es la primera vez que se habla de deserción en la UMSNH, es un problema muy perceptible, especialmente en las escuelas o Facultades con una admisión de alumnos baja, es decir, con menos de 200 nuevos alumnos. Rodríguez (2018) analizó la deserción de estudiantes de la Universidad Michoacana de San Nicolás de Hidalgo (UMSNH), y una comparación con la Universidad Nacional Autónoma de México (UNAM), el Instituto Politécnico Nacional (IPN), la Universidad Autónoma Metropolitana (UAM), la Universidad de Veracruz (UV). Los datos que analizaron la matrícula de ingreso frente a la matrícula de graduación. La siguiente figura muestra los resultados.

Como se observa en la Tabla 1, se analizaron 6 ciclos escolares. La columna con encabezado MI (Matrícula de Ingreso) corresponde al registro de inicio o nuevo ingreso y la columna con encabezado ME (Matrícula de Egreso), se refiere al registro de graduación. Los datos etiquetados como nd indican que no hay datos. Con este estudio se aprecia que no es un problema exclusivo de la UMSNH o de las Universidades del Estado, es un problema a nivel nacional.

Tabla 1. Registro de ingreso contra egreso de estudiantes de cinco universidades de México. El año escolar 2010-2011 a 2015-2016

Año escolar	UNAM		IPN		UAM		UV		UMSNH	
	MI	ME	MI	ME	MI	ME	MI	ME	MI	ME
2010-2011	180,763	58,584	95,743	13,478	40,712	4,854	56,582	5,025	37,264	5,637
2011-2012	187,195	58,855	95,743	13,820	41,325	4,448	58,497	5,625	39,646	5,942
2012-2013	190,707	60,748	98,624	13,077	42,242	4,674	58,212	4,767	38,561	5,706
2013-2014	196,565	60,749	100,854	12,915	43,762	5,334	58,995	5,129	31,439	5,561
2014-2015	201,206	63,346	104,125	13,630	44,301	5,063	59,284	4,779	37,139	4,355
2015-2016	204,940	Nd	104,409	12,684	44,712	5,147	59,583	5,643	nd	nd

Fuente: Abandono de estudiantes de pregrado de la UMSNH, análisis y propuesta de solución (Rodríguez, 2018).

Para la elaboración de la tabla 2, se sumó el total de estudiantes nuevos y graduados para determinar el promedio de cada uno de ellos y poder obtener la tasa promedio de graduación de los últimos cuatro ciclos escolares, que corresponde al 51%, es decir, la mitad de los estudiantes que ingresan completan sus estudios universitarios. Otro dato interesante es la tasa de deserción, que es del 49%.

Posteriormente, se diseñó un instrumento de investigación en forma de encuesta. La cual fue desarrollada en la aplicación Google Forms que permite preparar cuestionarios, almacenar los resultados para poder consultarlos y generar algunas gráficas.

Las preguntas que se hicieron fueron: edad, sexo, estado y ciudad de origen, estado civil, ingresos familiares, licenciatura, modalidad, semestre, el estado del estudiante, que pueden

ser estudiantes regulares (que no repiten asignaturas por segunda vez) y estudiante irregular (que repite o repite una o más asignaturas). Estas preguntas se aplicaron para identificar a los entrevistados, sin embargo, para la elaboración del modelo de árbol de decisión no se utilizó esta información. Se determinó un Factor de deserción, a partir de los estudios relacionados, se identifican una serie de variables cualitativas que pueden influir en la deserción del alumno. Se les pregunta sobre aspectos personales que podrían influir en la decisión de abandonar la universidad. Las preguntas fueron: por falta de tiempo, reprobar una o más asignaturas, por problemas personales, porque no te gustó la carrera, por trabajo, por impedimento familiar, problemas de salud, problemas económicos, porque tienes otra opción académica (cambio de titulación o universidad).

Tabla 2. Cifras de Entrada y Salida de la Facultad de Contaduría y Ciencias Administrativas de la UMSNH. Últimos 4 ciclos escolares (2015-2016 a 2018-2019).

Total: graduados	Total: ingreso	Promedio de graduación	Promedio de ingreso	Promedio de graduación	Promedio de deserción
2,832	5,563	708	1,392	51%	49%

Fuente: Elaboración propia.

Resultados

En el año lectivo 2019-2020 en el que se aplicó la encuesta para determinar las causas de deserción en estudiantes de FCCA, se inscribieron 3,290 y la encuesta se aplicó a 1,582 estudiantes. De acuerdo con la determinación del tamaño muestral realizada, se determinó que el tamaño muestral es adecuado.

Para realizar el análisis de datos se utilizó la herramienta R Studio (Gandrud, 2013). R Studio es el entorno de desarrollo integrado principal para R. Está disponible en ediciones comerciales y de código abierto en el escritorio (Windows, Mac y Linux) y desde un navegador web a un servidor Linux que ejecute R Studio Server o R Studio Server Pro.

En la encuesta aplicada a los estudiantes de la FCCA, la pregunta donde se determina el factor de deserción, cita: "Si tuvieras que dejar la Universidad, ¿cuál sería el motivo? (Puedes seleccionar uno o más)". Las opciones se muestran en la siguiente Tabla 3:

Tabla 3. Opciones de preguntas

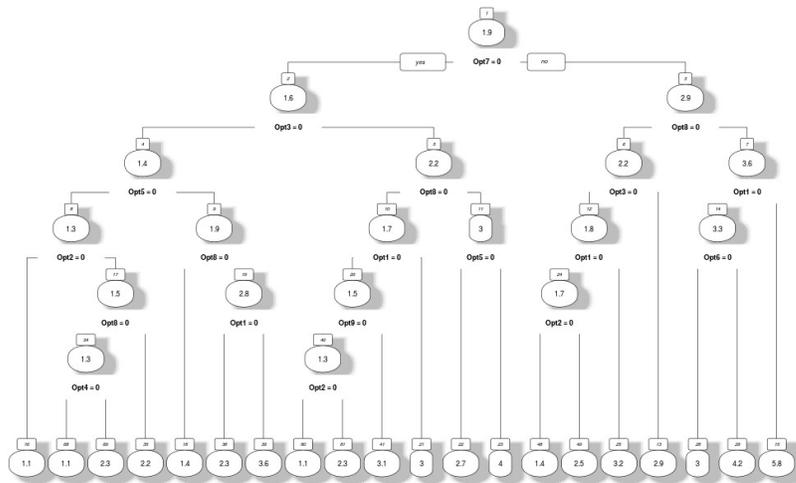
Fuente: Elaboración propia.

No. opción	Pregunta
Opc1	Falta de tiempo
Opc2	No aprobar una o más asignaturas.
Opc3	Por problemas personales
Opc4	Porque no te gustó la carrera
Opc5	Por trabajo
Opc6	Discapacidad familiar
Opc7	Problema de salud
Opc8	Problemas económicos
Opc9	Porque tienes otra opción académica (cambio de universidad o de otro bachillerato)

Fuente: Elaboración propia.

Uno de los principales aportes de este trabajo es el modelo de árbol de decisión a partir de un método predictivo, mostrado en la Figura 1, que a partir de un determinado comportamiento de los datos intenta delimitar el camino por el que se atraviesa para llegar a un determinado punto, además, cada opción tiene un puntaje determinado, las respuestas seleccionadas dependerán de determinar qué puntaje final obtendrá el encuestado. El modelo predice lógicamente si se cumple una determinada condición, la probabilidad de desertar que se obtendría.

Figura 1. Modelo de árbol de decisión predictivo para determinar el factor de deserción en la FCCA de la UMSNH.



Fuente: Elaboración propia.

Se pueden representar diferentes situaciones personales en los alumnos de la FCCA del Modelo de Árbol de Decisión, algunas de ellas podrían ser las siguientes: En el modelo indica que, si el alumno no tiene problemas de salud, opción 7, pero si tiene problemas económicos, opción 8 y falta de tiempo, opción 1, la probabilidad de desertar es la más alta, tendría un puntaje de 5.8. , ahora, si un alumno tiene problemas de salud, opción 7, problemas personales, opción 3, problemas económicos, opción 8, y falta de tiempo, opción 1, su puntaje es 3. Otra situación personal que no afectaría considerablemente la decisión de deserción sería que, si alguien tiene problemas de salud, opción 7, problemas personales, opción 3, y opción 2, que desaprueba una o más asignaturas, la puntuación será de 1,1. Las opciones que puede elegir un alumno son muy diversas, por tanto, las puntuaciones obtenidas en función de la situación personal de cada alumno.

Conclusiones

Este trabajo abordó temas que actualmente son de gran preocupación en diferentes áreas de la investigación administrativa, educativa y tecnológica, para tratar de resolver problemas comunes. Si bien es cierto que la mayoría de las Instituciones de Educación Superior tienen serios problemas, los directores de la Facultad de Contaduría y Ciencias Administrativas de la UMSNH deben tomar decisiones con la ayuda de herramientas de investigación y tecnológicas adecuadas como la minería de datos a través del modelo predictivo de árboles de decisión donde jerarquiza las características más significativas que inciden en el nivel de deserción de los estudiantes.

Aunque los problemas de financiación pueden prevalecer en la UMSNH, se deben evitar las tasas de deserción tardía (Vélez, Gómez, Portilla & Velásquez, 2006) para aumentar la matrícula y disminuir sus tasas de deserción, de lo contrario, habrá consecuencias negativas, que pueden afectar no solo el aspecto de la matrícula de la FCCA sino también en la universidad.

Referencias Bibliográficas

- Cadavid, DV, Mendoza, AM y Rodríguez, EC (2016). Eficiencia en las instituciones de educación superior públicas colombianas: una aplicación del análisis envolvente de datos. *Civilizar: Ciencias Sociales y Humanas*, 16 (30), 105-118.
- Elías, T. (2011). *Analítica de aprendizaje*. Aprendizaje, 1-22.
- Felizzola, HA, Arias, YAJ, Pedroza, FV y Pastrana, AMC (2018). Modelo de predicción para la deserción temprana en la Facultad de Ingeniería de la Universidad de la Salle. *Encuentro Internacional de Educación en Ingeniería*.
- Gandrud, C. (2013). *Investigación reproducible con R y R studio*. Prensa CRC.
- Jiménez, MVG, Izquierdo, JMA y Blanco, AJ (2000). La predicción del rendimiento académico: regresión lineal versus regresión logística. *Psicothema*, 12 (Su2), 248-525.

- Madrid Echeverry, 2017, Propuesta de un modelo estadístico para caracterizar y predecir la deserción estudiantil Universitaria. Escuela de Ingeniería de la Organización.
- Noboa, C., Ordóñez, M. y Magallanes, J. (2018). Aprendizaje estadístico para detectar posibles abandonos en la educación superior: un estudio de caso de una universidad pública. *Learning Analytics para América Latina 2018*, 2231, 12-21.
- Orea, SV, Vargas, AS y Alonso, MG (2005). Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos. *Ene*, 779 (73), 33.
- Rodríguez, MGO (2018). Deserción de estudiantes de licenciatura de la UMSNH Análisis y propuesta de solución. *Economía y Sociedad*, 22 (38), 15-32.
- Vásquez Velásquez, J., Castaño Vélez, EA, Gallón Gómez, SA, & amp; Gómez Portilla, K. (2003). Determinantes de la deserción estudiantil en la Universidad de Antioquia.
- Vélez, EC, Gómez, SG, Portilla, KG y Velásquez, JV (2006). Análisis de los factores asociados a la deserción y graduación estudiantil universitaria. *Lecturas de economía*, (65), 9-36.