

# Propuesta para el uso de árboles de decisión en la predicción de la trayectoria de los estudiantes

Silverio Pérez-Cáceres, Carlos Rodríguez-Flores, Efrén Morales-Mendoza,

Heide G. Cruz-Ramírez y Araceli Carballo-Franco

Facultad de Ingeniería en Electrónica y Comunicaciones

Universidad Veracruzana

Poza Rica, Ver.; México

[sperez, carodriguez, efmorales]@uv.mx, [c.ramirezheide, aarecarballo]@gmail.com

**Abstract**— The decision tree technique was used to make a system for the prediction of success or failure in the school trajectory of the new students in computational technologies. The tree was generated from the personal, family, socioeconomic and academic information of graduates and those who didn't finish the career, after which the system that makes the prediction was designed, with this proposal it is intended that the tutors have reliable information, with a precision of 0.935 for success and 1.000 for failure that allows to identify the students who are at risk and apply the relevant strategies to avoid failure by reducing the school dropout rate.

**Keyword**— *school trajectory, data mining, decision tree, school dropout.*

**Resumen**— La técnica de árboles de decisión fue utilizada para realizar un Sistema para la predicción del Éxito o Fracaso en la trayectoria escolar de los estudiantes de nuevo ingreso a Ingeniería en Tecnologías Computacionales. El árbol fue generado a partir de la información personal, familiar, socioeconómica y académica de egresados y de aquellos que truncaron la carrera, después de lo cual se diseñó el Sistema que realiza la predicción; con esta propuesta se pretende que los tutores tengan información confiable, con una precisión de 0.935 para el Éxito y de 1.000 para el Fracaso, que permita identificar a los estudiantes que se encuentren en riesgo y aplicar las estrategias pertinentes para evitar el fracaso abatiendo el índice de deserción escolar.

**Palabras claves**— *trayectoria escolar, minería de datos, árbol de decisión, deserción escolar.*

## I. INTRODUCCIÓN

Hoy en día, la tecnología está cada vez más inmersa en nuestro mundo globalizado, desde cualquier ámbito o sector que lo veamos, la salud, las industrias, las comunicaciones, la educación, el transporte el entretenimiento, etc.; siendo cada vez más imprescindible el saber cómo usarla, para qué usarla, pero sobretodo que sea utilizada para fines que proporcionan un beneficio apropiado.

En este trabajo se rescata la importancia de utilizar la tecnología como apoyo para predecir el éxito o fracaso de una trayectoria escolar, satisfaciendo la necesidad de información de los Tutores respecto de sus tutorados y considerando que el análisis de los factores que inciden en el éxito o fracaso, pueden propiciar el diseño de estrategias que impacten en la deserción escolar y por ende, reflejarse en la mejora de los índices de eficiencia terminal. Por lo anterior se ha identificado que una predicción de este tipo puede ser de gran apoyo para el asesoramiento que brindan los tutores del nivel de licenciatura, mediante los resultados de un sistema computacional, del cual se presenta su diseño dividido en dos etapas: la aplicación de la técnica de la minería de datos y el desarrollo del sistema Web. En la primera se define la base de la información sobre la cual se realizará la predicción, seleccionando los factores que pueden incidir en una trayectoria de éxito o de fracaso, así como la fuente de información que en este caso corresponde a egresados y estudiantes del Programa Educativo (PE) de Ingeniería en Tecnologías Computacionales (ITC) de la Universidad Veracruzana (UV), a partir de lo cual se aplican las etapas de la minería de datos mediante el software denominado Weka, obteniendo como resultado final un árbol de decisión.

El árbol de decisión se empleó como insumo principal de la segunda etapa, el diseño del sistema Web, para lo cual se utilizó la ingeniería del software, bajo la estructura del modelo cliente-servidor, con desarrollos en lenguajes de programación PHP, HTML y JavaScript así como el servidor XAMPP.

Este sistema proporciona la interfaz para los estudiantes de nuevo ingreso, quienes capturarán su información y el sistema les enviará como resultado la predicción de éxito o de fracaso, identificando los factores principales de esta predicción, lo cual será de gran utilidad para la asesoría que brinde el tutor, mediante acciones pertinentes, oportunas y necesarias para lograr una trayectoria escolar exitosa.

## II. DESARROLLO

### A. *La deserción y la trayectoria escolar*

El éxito y fracaso en la trayectoria escolar son temas de gran importancia en el sistema de enseñanza actual. Específicamente, la deserción estudiantil, el rezago y los bajos índices de eficiencia terminal se encuentran entre los problemas más complejos y frecuentes que enfrentan las Instituciones de Educación Superior del país [1].

Se considera a las trayectorias escolares como el comportamiento académico de un individuo e incluye el desempeño escolar, la aprobación, la reprobación, el promedio logrado, etcétera, a lo largo de los ciclos escolares. El término trayectoria escolar está estrechamente vinculado a la eficiencia terminal, considerado como el indicador más importante en las evaluaciones institucionales; es la relación cuantitativa entre alumnos que ingresan y los que egresan de una determinada cohorte [2].

La trayectoria escolar es entendida como el conjunto de factores y datos que afectan y dan cuenta del comportamiento escolar de los estudiantes durante su estancia en la universidad [3]. Considerando con una trayectoria de éxito a aquellos estudiantes que concluyeron satisfactoriamente la licenciatura, para este caso que nos ocupa, y con una trayectoria de fracaso, aquellos que truncaron su carrera por diversos factores, lo cual se refleja en la disminución de la eficiencia terminal.

Diversas instituciones han realizado investigaciones acerca de los factores, indicadores o variables que intervienen en las trayectorias escolares, como los realizados en la Universidad Veracruzana [4] y en la UAEH [5]; algunos de estos factores son: personales, académicos, físicos, económicos, familiares, sociales, laborales, adicciones, entre otros [6]. Pero son escasos los estudios que exploren la probabilidad de éxito o fracaso en la trayectoria escolar de quienes ingresan a una licenciatura. Actualmente las Instituciones de Educación Superior enfrentan el gran problema de deserción escolar, en relación con este tema se han hecho una gran cantidad de esfuerzos para combatirlo como programas de tutorías, asesorías, cursos, entre otros, sin lograr mejorar este problema [7] [8].

México presenta un índice de deserción escolar del 50 por ciento, uno de los más elevados en América Latina; esta problemática es aún más compleja en los estados con altos niveles de desigualdad respecto a su ingreso, según establece la Organización para la Cooperación y el Desarrollo Económico (OCDE) [9].

El gran problema se centra sobre todo en los jóvenes de 15 años en adelante, los cuales por tener carencias económicas, problemas familiares, de adicciones, entre otras, abandonan la escuela, sin haber alcanzado un nivel de educación mínimo para poder desenvolverse con facilidad en un mundo globalizado en el que el conocimiento es clave para el desarrollo económico de las naciones, además, para que cada persona pueda acceder a un trabajo, un hogar y un estilo de vida digno [10].

En el estudio realizado en la Universidad Veracruzana [4], sólo se detectaron trayectorias estándar, las cuales fueron diseñadas para estudiantes modelo. Tal situación obliga a la institución a realizar investigaciones que permitan disminuir el índice de deserción mediante la predicción oportuna del éxito

o fracaso en la trayectoria escolar de los estudiantes y a su vez llevar un control de su información para su posible estudio.

De acuerdo con lo anterior, la necesidad de identificar y predecir la deserción escolar de los estudiantes en los primeros semestres es indispensable para tomar las acciones pertinentes y poder disminuir este índice, y no menos importante, predecir su deserción en cualquier momento para su correcto seguimiento [7]

### B. Diseño del Sistema Computacional

El diseño del sistema se realizó en dos etapas, en la primera se aplicó una encuesta a los egresados y a estudiantes que truncaron su trayectoria escolar del PE de ITC, obteniéndose el árbol de decisión, el cual fue programado en el Sistema y mediante el cual se realizará la predicción del éxito o fracaso. En la segunda etapa se llevó a cabo el desarrollo técnico del sistema web.

Entre los casos de uso del sistema, el usuario (estudiante) dispone de las siguientes opciones:

1. *Ingresar al sistema.* El sistema cuenta con un proceso de registro e inicio de sesión para cada uno de los usuarios del sistema.
2. *Iniciar pronóstico.* Los usuarios registrados pueden generar el pronóstico de su trayectoria escolar de la siguiente manera:

a. Responder el cuestionario. Después de haber iniciado sesión, el usuario podrá responder el cuestionario que le envía el sistema, siempre y cuando no lo haya respondido antes. En caso de haberlo respondido, sólo podrá generar el resultado de su trayectoria escolar.

b. Guardar cuestionario. Una vez que haya finalizado la opción de Responder el cuestionario, el usuario podrá guardar sus respuestas que se enviarán a la base de datos para ser almacenadas.

c. Obtener resultado y recomendaciones del pronóstico. Después del ingreso de los datos, a través del cuestionario, el estudiante podrá consultar cuál ha sido el resultado de su pronóstico en comparación con las trayectorias escolares de las generaciones anteriores y de esta manera recibir una serie de estrategias y recomendaciones que lo apoyen en su trayectoria escolar.

En la siguiente imagen se muestra el envío de información por parte del estudiante a través del portal web, una vez enviados el servidor recibe los datos y los anexa a la base de datos del sistema que está bajo el control de un administrador.

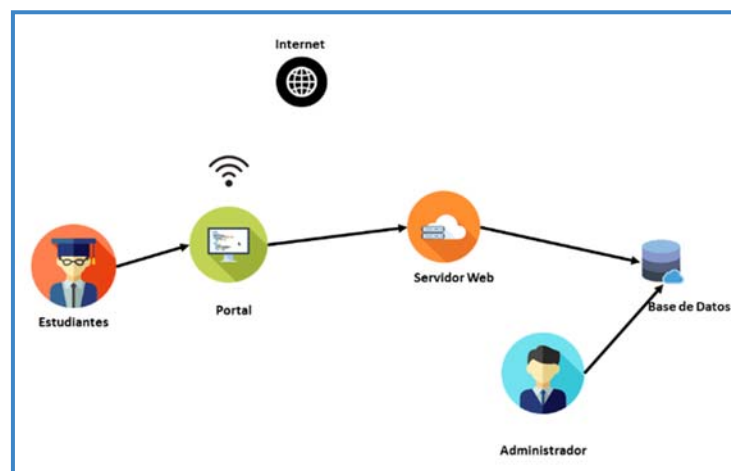


Fig. 1. Diseño conceptual del sistema para el pronóstico del éxito o fracaso de la trayectoria escolar.

C. Obtención de datos aplicando minería de datos

Para la obtención de datos, primeramente, se identificaron los factores que influyen en el éxito y fracaso de la trayectoria escolar de los estudiantes, de acuerdo a investigaciones realizadas por instituciones como la Universidad de Sonora, la Universidad Autónoma de Yucatán, la Universidad Autónoma del Estado de Hidalgo, Universidad Nacional Autónoma de México, la Benemérita Universidad Autónoma de Puebla, la Universidad Autónoma de Aguascalientes, la Universidad Autónoma Metropolitana, y también la Universidad Veracruzana [4]. Después de obtener una serie de factores, se realizó una segunda selección en la cual se encuestó a los docentes de la carrera para determinar qué factores tienen mayor incidencia en el éxito o fracaso de la trayectoria escolar de los estudiantes de ITC.

En base a lo anterior, se pudo identificar que los factores que inciden en el éxito y fracaso de la trayectoria escolar de los estudiantes de ITC son: personales, académicos, socioeconómicos y familiares.

Para proceder con la recolección de datos, se realizó una encuesta a los egresados del PE de ITC y a los estudiantes que truncaron su carrera, donde se plantearon preguntas relacionadas a su información personal, académica, socioeconómica y familiar. Con la información recopilada se creó una base de datos que contiene 37 registros, entre los cuales se encuentran trayectorias de éxito y de fracaso. Posteriormente, se utilizó esta información para poder aplicar la técnica de árboles de decisión con el algoritmo J48 de minería de datos, mediante el software llamado Weka [11], plataforma de software libre para el aprendizaje automático y la minería de datos, a través del cual se generó al árbol de decisión presentado de manera gráfica en la figura 2.

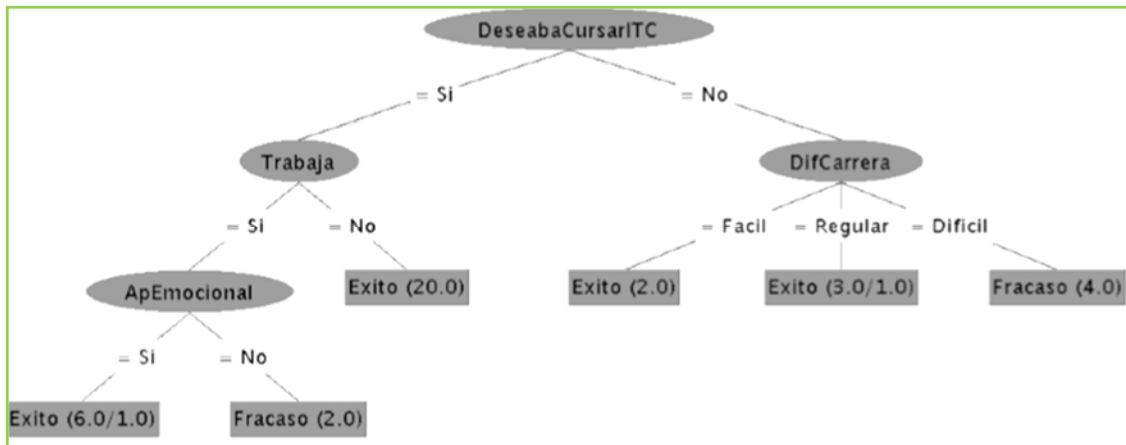


Fig. 2. Árbol de decisión generado en Weka.

Los diferentes trayectos que sigue el árbol indican que el éxito y fracaso de las trayectorias escolares dependen de los factores: *personal* y *económico* del estudiante de acuerdo a su situación laboral; *académico*, reflejado en la percepción de la dificultad para cursar la carrera; y *familiar*, si se le brinda el apoyo emocional para continuar con sus estudios.

Este árbol de decisión se utilizó en el sistema dado que presentó una precisión de 0.935 para el caso de Éxito y 1.000 para el caso de Fracaso, como se muestra en la tabla I, lo cual indica que es un modelo predictivo eficaz para generar un pronóstico adecuado del éxito o fracaso de la trayectoria escolar de los estudiantes de ITC en base a su información.

Tabla I. Datos del árbol de decisión.

Descripción	Valor
Total de Registros	37
Éxito	29
Fracaso	8
Clasificados como correctos	35
% Correctos	94.60
Precisión caso de Éxito	0.935
Precisión caso de Fracaso	1.000
Número de hojas	6
Tamaño del árbol	10

#### D. Desarrollo del sistema web

El sistema computacional fue desarrollado en el editor de texto Sublime Text en los lenguajes de programación PHP, HTML y JavaScript para ser implementado como sistema web desde cualquier navegador con un servidor de modo local en esta primera versión. Para lo cual se ha utilizado XAMPP que es una distribución de Apache que contiene un sistema de gestión de base de datos MySQL para el almacenamiento de la información de los estudiantes y de los resultados que se obtendrán del pronóstico generado por el sistema, además de contar con el servidor web Apache y los intérpretes para el lenguaje script: PHP que, a través de estas herramientas, permitirá la ejecución del sistema web [12]. Para generar el árbol de decisión que se utiliza en el sistema se ha empleado el software llamado Weka, herramienta para experimentación de análisis de datos que permite aplicar, analizar y evaluar las técnicas más relevantes de análisis de datos, principalmente las provenientes del aprendizaje automático, sobre cualquier conjunto de datos del usuario [11].

Una vez obtenida la arquitectura del proyecto, así como el proceso de obtención de datos, lo siguiente fue desarrollar el sistema web. El sistema presenta una encuesta por medio de la cual los estudiantes ingresan la información necesaria que servirá para generar el pronóstico.

El ingreso al sistema se realiza con la opción de inicio de sesión, previo registro de los estudiantes, después de lo cual se presenta el menú principal que permite al estudiante dirigirse a las diferentes secciones con las que cuenta el sistema, tales como:

- *Leer más.* Esta sección muestra información relacionada con el sistema y los vínculos a la página web de la Universidad Veracruzana (UV) y de la página Facebook de la Facultad de Ingeniería en Electrónica y Comunicaciones (FIEC).

- *Iniciar pronóstico.* Este apartado realiza dos funciones importantes: la primera consta de direccionar al estudiante a la encuesta del sistema en caso de que no la haya contestado y en caso contrario, se realizará su segunda función, enviará al estudiante a la sección de resultados y recomendaciones.

- *Resultados y recomendaciones.* Muestra el resultado del pronóstico del éxito o fracaso del estudiante y una serie de recomendaciones de acuerdo con el tipo de pronóstico.

Las recomendaciones son generales como se muestran en la tabla II.

Tabla II. Recomendaciones proporcionadas por el Sistema para pronóstico de éxito o fracaso.

Pronóstico	Recomendación
Éxito	Continuar así, con esa responsabilidad y compromiso que tienes como estudiante, no olvides que es importante acudir a tus sesiones de tutorías. Solicita información a tu Tutor sobre los beneficios que puedes obtener al continuar con una trayectoria de éxito. Además puedes ser candidato a participar en las diferentes convocatorias como: Becas, Estímulos, Programas de Movilidad, etc.
Fracaso	Investigar y repasar un poco más los temas que se te hagan difíciles. Solicitar apoyo a tus compañeros o profesores para resolver las dudas que tengas sobre algún tema. Entregar tareas y trabajos en tiempo y forma. Poner mayor empeño y dedicación a tus estudios. Confiar en ti y en tu aprendizaje para lograr los objetivos.

### III. EVALUACIÓN DEL SISTEMA

La evaluación del sistema se realizará considerando dos vías:

- *Evaluación de recomendaciones y resultados*: Las recomendaciones y resultados serán evaluados por los estudiantes de nuevo ingreso en base a los resultados y recomendaciones que brinda el sistema para generar un pronóstico adecuado del éxito o fracaso en la trayectoria escolar. Dicha variable será medida en los rangos: malo, regular, bueno, excelente.

- *Evaluación técnica del sistema*: El Sistema será evaluado por los estudiantes de nuevo ingreso a través de una encuesta en línea, donde se les preguntará cómo valoran el sistema en cuánto a su diseño y desempeño. La variable será medida en los rangos: malo, regular, bueno y excelente.

### IV. CONCLUSIONES

Con base al proyecto realizado, el Sistema presenta a los estudiantes un pronóstico del éxito o fracaso en su trayectoria escolar a partir de su información personal, académica, socioeconómica y familiar; el Sistema se probó con algunos datos de estudiantes de la generación 2016, esperando obtener resultados con la aplicación a los estudiantes de nuevo ingreso, los cuales sean analizados por los Tutores, de tal forma que se tenga la oportunidad de evaluar la utilidad de la información proporcionada.

La desventaja principal de esta primera versión del Sistema es que el número de registros para generar el árbol de decisión es bajo, y en su mayoría corresponden a casos de éxito (tabla I), dado que no fue posible contactar a aquellos que desertaron de las primeras generaciones, por lo que se ve reflejado en la precisión del árbol de 0.935 para el caso de Éxito, así como en su tamaño.

La precisión del Sistema se podrá ir incrementando mediante la actualización del árbol de decisión, con la información de los estudiantes que vayan truncando su carrera o que la terminen exitosamente, lo cual generará un árbol de decisión que incluya un mayor número de factores, los cuales podrán ser considerados para recomendaciones individualizadas.

Finalmente, si bien el Sistema está basado en la precisión obtenida al generar de árbol de decisión, la exactitud del pronóstico sólo se podrá corroborar al realizar la comparación cuando los estudiantes encuestados hayan concluido su trayectoria escolar.

### RECONOCIMIENTOS

Este proyecto fue realizado con el apoyo de la Facultad de Ingeniería en Electrónica y Comunicaciones Poza Rica, de la Universidad Veracruzana.

## REFERENCIAS

- [1] S. Daniel, D. Alamilla y O. García, “Factores que influyen en la trayectoria escolar de la licenciatura en Derecho y Jurisprudencia,” EDÄHI ICSHU, Hidalgo, 2011.
- [2] G. Octaviano y C. Barrón, “Un estudio sobre la trayectoria escolar de los estudiantes de doctorado en Pedagogía,” *Perfiles Educativos*, México, 2011, pp 94-113.
- [3] J. A. Fernández, A. Peña, F. Vera, “Los estudios de trayectoria escolar. Su aplicación en la educación media Superior,” *Revista de la Facultad de Filosofía y Letras*, México, 2006, N° 6, pp. 24-29.
- [4] O. Juan, R. López y E. Alarcón, “Trayectorias Escolares en Educación Superior,” Xalapa, Ver.: Universidad Veracruzana, 2015.
- [5] P. d. L. María, “Guía para el seguimiento de trayectorias escolares,” Hidalgo, 2003.
- [6] N. Rodríguez, L. Reyes, A. Jiménez, C. Brian, J. Ruiz, “Minería de Datos: Modelo Predictivo de Deserción Escolar,” *IEEE*, 89-93, 2017.
- [7] S. Valero, “Aplicación de Técnicas de Minería de Datos para Predecir Deserción,” Universidad Tecnológica de Izúcar de Matamoros, Tesis 2009.
- [8] ANUIES. “Deserción, rezago y eficiencia terminal en las IES. Propuesta metodológica para su estudio,”. México: Facultad de Ciencias de la Electrónica, 2001.
- [9] OCDE. “México. Políticas prioritarias para fomentar las habilidades y los conocimientos de los mexicanos para la productividad y la innovación” Serie Mejores Políticas, 2015.
- [10] F. S. López, “La deserción escolar de jóvenes en México tiene una caída "brutal" en comparación con países de América Latina:” *Boletín 218*. Senado de la República, 2015.
- [11] M. García, A. Álvarez., “Análisis de datos en Weka: Pruebas de selectividad,”. Madrid, 2007.
- [12] J. M. González. “Desarrollo de sitios web con PHP y MySQL. Departamento de Lenguajes y Sistemas Informáticos”. Universidad de Sevilla, 2017.