



## PERFIL DE INGRESO Y TRAYECTORIA ESCOLAR EN EDUCACIÓN SUPERIOR, SISTEMA PARA LA PREDICCIÓN DE LA PERMANENCIA Y ABANDONO

### **Mtro. Carlos García Trujillo**

*Centro de investigación e innovación en Educación Superior, Universidad Veracruzana*  
cgarciagl@uv.mx

### **Ragueb Chain Revuelta**

*Centro de investigación e innovación en Educación Superior, Universidad Veracruzana*  
ncruz@uv.mx

### **Mtro. Jahir de Jesús Castillo Ochoa**

*Centro de investigación e innovación en Educación Superior, Universidad Veracruzana*  
jahircastilloochoa@gmail.com

**Área temática:** A.9) Sujetos de la educación

**Línea temática:** 2. Trayectorias escolares, profesionales e institucionales individuales y colectivas

**Tipo de ponencia:** Reporte parciales o final de investigación



## Resumen

Al ingresar a la Educación Superior los aspirantes documentan rasgos que dan indicios de su perfil de ingreso, hay pocas descripciones de la diversidad de sus condiciones socioeconómicas y culturales, así como las habilidades académicas y conocimientos obtenidos en su formación previa. Los procesos de ingreso registran en bases de datos la información necesaria para dar cuenta del perfil de ingreso. Aunque no se ha percibido la importancia de esta información, sistematizarla para su consulta es un buen punto de partida pues dar cuenta de los rasgos de la población es crucial para el diseño de políticas institucionales de atención a la población estudiantil. A partir de lo anterior, establecemos las relaciones entre el perfil de ingreso y el desempeño escolar entendido este como la trayectoria que los estudiantes despliegan para avanzar en su plan de estudios, es decir la especial combinación de promedio, aprobación-reprobación, promoción de asignaturas y avance en créditos. A partir de los registros del perfil de ingreso y el seguimiento de la trayectoria escolar de 77 mil estudiantes exploramos mediante técnicas de descubrimiento de conocimiento en bases de datos, usando reyes bayesianas y la técnica de árbol de decisiones, la posibilidad de realizar predicciones sobre la permanencia, el abandono y la trayectoria a partir de los información disponible. En consecuencia, establecer procesos que permitan en los procesos de ingreso de nuevas generaciones calcular la probabilidad de éxito escolar, a través de crear un semáforo de riesgo que identifique estudiantes en riesgo aun antes de su ingreso a la institución.

**Palabras clave:** Trayectoria escolar, Perfil de ingreso, Probabilidad de Abandono, Predicción del desempeño

## Introducción

El incremento de la matrícula es uno de los principales rasgos de las políticas en la educación superior, la mayoría de las propuestas se orientan a la ampliación de la oferta mediante el incremento en el número de espacio disponibles para el primer ingreso a las IES públicas, hoy según el INEGI el 44.9% del grupo de edad 18 a 24 años asisten a la educación superior. Sin embargo, si bien es necesario garantizar las oportunidades de acceso también resulta indispensable incrementar la permanencia y el egreso, no son pocos los estudiantes que abandonan los estudios por muy diversas razones. En el caso particular de la Universidad Veracruzana en los últimos 5 años de cada 100 estudiantes que ingresaron a las aulas 17 las abandonaron en algún momento del primer año de sus estudios los periodos.

Son numerosos los estudios sobre el abandono y la deserción en el sistema educativo y múltiples variables han sido asociadas al fenómeno (Rochin 2021), pero en todo caso se analiza el abandono una vez que ha ocurrido, desde nuestra perspectiva es necesario, en complemento a los avances realizados incluir una mirada preventiva, es decir explorar la posibilidad de identificar con la mayor precisión posible a quienes son potenciales estudiantes en riesgo de abandonar la universidad.

Frente a la abundancia de estudios de las variables asociadas al abandono nos interesa abordar el problema desde una perspectiva diferente más orientada a explorar la posibilidad contribuir a prevenir el fenómeno a partir de identificar la población en riesgo al inicio de su incorporación a las instituciones. En consecuencia, las preguntas centrales consisten en determinar el volumen y distribución del abandono, explorar las variables que caracterizan el perfil de ingreso y cuales se asocian al abandono,

La incógnita para despejar es si a partir de la información disponible y acumulada durante cinco años, la cual por un lado da cuenta del perfil de ingreso y por otro refiere el abandono y la trayectoria escolar de los estudiantes, es posible desarrollar procesos que permitan formular un algoritmo que con relativa certeza derive en la creación de un sistema de alerta sobre la población en riesgo que requiera una mayor atención mediante políticas específicas.

Nuestro objetivo es el desarrollo de procesos de análisis de información para construir conocimiento, pero sobre todo crear un sistema de alerta que identifique desde los procesos de ingreso y la mayor precisión posible la probabilidad de abandono, es decir identificar la población en riesgo como punto de partida para llevar a cabo acciones institucionales para contribuir a reducir el abandono.

## El enfoque

Reconociendo que el éxito escolar es resultado de múltiples factores que se relacionan e interactúan de diversas maneras (Martínez, 1989; Chain, 1995), se encuentra el supuesto, no siempre demostrado, que determinados rasgos socioculturales que caracterizan a los solicitantes de ingreso, se asocian la probabilidad de éxito escolar.

Cuando se pretende identificar a través del perfil de ingreso de los estudiantes admitidos las mayores probabilidades de éxito escolar, resulta conveniente explorar ésta mediante estudios cuidadosos.. La mayoría de los estudios orientados al análisis de la capacidad de predicción de una o varias características (variables) de la población son realizados con base en poblaciones reducidas y aún se realizan mediante enfoques que buscan demostrar estadísticamente el grado de asociación entre una o más variables, promedio en el examen de admisión, promedio de bachillerato, etc.

Son escasos los estudios con base en poblaciones completas, orientados a explorar la capacidad de los perfiles de ingreso para discriminar entre alumnos de alta o baja probabilidad de permanencia y éxito escolar. Un análisis de este tipo requiere de un cuidadoso diseño experimental y el seguimiento de una población de estudiantes a lo largo de su trayectoria universitaria.

Desde nuestra perspectiva, mediante un enfoque de análisis de datos diferente es posible construir aproximaciones razonables en torno al grado de asociación entre los rasgos específicos de la población de primer ingreso y la permanencia en la universidad.

En este sentido, el interés se centra en determinar si es posible calcular la probabilidad de permanencia y éxito escolar a partir de la información derivada de los procesos de ingreso, pero más allá de demostrar los grados de correlación que indican una determinada tendencia y la fuerza de ésta, pretendemos determinar cuál o cuáles variables, o combinación de ellas son más o menos relevantes para determinar la mejor aproximación posible la probabilidad de permanencia y éxito escolar.

La investigación se realiza con la información del conjunto de estudiantes que se inscribieron a la Universidad Veracruzana en los procesos de ingreso 2017,2018, 2019,2020 y 2021, que El grupo se compone de 77,161 alumnos pertenecientes a las más de 150 opciones profesionales carreras por la UV, distribuidas en 6 Areas Académicas en las cinco zonas geográficas .

A partir del análisis de los registros derivados del proceso de ingreso, de cada uno de los estudiantes se recopiló la información sobre el conjunto de variables que mostraron en estudios previos una relativa asociación con la trayectoria escolar (Chain, 1995); en consecuencia destacamos como variables del perfil de ingreso: el promedio de bachillerato, la calificación en el examen de ingreso, el año de egreso del bachillerato, la escolaridad de los padres, la escuela de bachillerato de la cual egreso y el grado de marginación de la población donde se encuentra la escuela donde cursó el bachillerato, además de identificar si cuenta con una computadora en casa, así como una conexión a internet.

## Las variables

1. El promedio de bachillerato se mueve en un rango de 6 a 10, mismos que se categorizan en 4 niveles a partir de su particular distribución:

1	2	3	4
6 a 7.7	7.8 a 8.4	8.5 a 9.0	9.1 – 10.00

2. En cuanto a la calificación en el examen de ingreso, considerando que se trata de los estudiantes que fueron aceptados después de un proceso de selección basado en los resultados del mismo, se categorizo en 4 niveles:

1	2	3	4
[1 – 52]	(53 – 63]	(64– 73.00]	(74 – 100]

3. Por otro lado, se considera que la continuidad en los estudios, es decir ingresar a la universidad inmediatamente después de concluido el bachillerato se asocia a la permanencia y al éxito escolar, en este caso se categoriza como **1** a quienes ingresan a la universidad en un año posterior a su egreso de bachillerato y como **2** a quienes lo hacen el mismo año.

4. Otra variable considerada y que está debidamente documentada es la escolaridad de padre y madre, considerando que refleja en general un mayor o menor contacto con el ámbito escolar, ante la amplia diversidad de niveles de escolaridad se agruparon en 4 categorías

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
Primaria o menos	Secundaria	preparatoria	Licenciatura o más

5. Así mismo, se incluyó como variable la escuela de procedencia, pero la enorme diversidad de escuelas de bachillerato exigió clasificarlas, en este caso se ordenaron mediante un análisis **clúster**, que es una técnica multivariante cuya idea básica es clasificar objetos formando grupos/conglomerados (*clúster*) que sean lo más homogéneos posible, en este caso se agruparon en función de sus resultados en las pruebas nacionales de conocimientos que documentan la proporción de estudiantes de una escuela que se ubican en cada nivel de dominio de las la lengua y las matemáticas. En consecuencia a cada estudiante se le asigno la categoría correspondiente a la escuela de la que procede, en este sentido los estudiantes que tienen asignado un determinado cluster sabemos que provienen de una institución en la cual la mayor proporción de estudiantes se ubica en determinados niveles de habilidades y conocimientos de Matemáticas y Lengua, y que por lo tanto provienen de escuelas diferentes a los asignadas estudiantes.

6. Finalmente se documentó en el registro de cada estudiante indicando si cuenta con equipo de cómputo y conexión a internet en su casa.

7. Por otro lado, para construir una medida de la permanencia se revisaron los cardex electrónicos de cada uno de los estudiantes, en este sentido se documentó la inscripción a los tres semestres posteriores al primer ingreso, en consecuencia categorizamos como 1 al

estudiantes sólo estuvo inscrito en el primer periodo, como 2 cuando estuvo inscrito en el primer periodo y el segundo y 3 si se inscribió a los tres periodos. En consecuencia, cuando, un estudiante no está inscrito en el tercer periodo escolar se clasifica en términos de **1. y 2. No permanece**, en su caso si está inscrito al tercer periodo que corresponde a su generación se le clasifica como **3. Permanece**.

Generación	Área académica	Region	Periodos Inscrito	Permanece
2017	ECONOMICA	XALAPA	1	No
2017	ARTES	XALAPA	2	No
2017	SALUD	CORDOBA	3	SI

### El Modelo de análisis de los datos

El análisis se realizara mediante procesos de Descubrimiento de Conocimiento en Bases de Datos con métodos que sean capaces de encontrar patrones, regularidades y conocimientos implícitamente contenidos en los datos, lo cual nos permita comprender el fenómeno que se estudia, específicamente con base en las redes bayesianas mediante las cuales es posible reconocer fácilmente las interacciones que hay entre las variables ( promedio de bachillerato, escuela de procedencia, escolaridad de padres, calificación en el examen de ingreso) y el fenómeno en estudio (la permanencia). Al mismo tiempo que medir numéricamente las interacciones y realizar inferencias tales como diagnóstico y pronóstico.

A través del análisis con algoritmos que construyen redes bayesianas a partir de datos generan un gran número de hipótesis (redes)y, de acuerdo con cierta heurística seleccionan el mejor modelo. Así mismo, se trata de construir modelos sino también de describir distribuciones de probabilidad de estos, medir el impacto que el valor específico de una o varias variables tienen sobre otras y evaluar que tan adecuado es el modelo construido en términos de clasificación.

Todo lo anterior a través del sistema de software para el análisis de datos basado en redes bayesianas denominado Bansy.

Se trata en resumen de construir sistemas de clasificación y arboles e decisiones que muestren con la mayor certeza las relaciones debidamente jerarquizadas entre las variables antes descritas y a permanencia e los estudiantes en la universidad.

### Aproximación y avances

Hasta ahora el avance consiste en la construcción y consolidación de una bases de datos con más de 77 mil registros que contienen cada uno de ellos los indicadores relativos las variables arriba señaladas nos permiten describir con precisión los rasgos del perfil de ingreso de la población

estudiantil, como califican en el examen de ingreso, cual es su promedio de bachillerato, de que tipo de escuela provienen, si cuentan o no con computadora e internet, datos que por si mismo son relevantes, pero también dar cuenta del volumen y distribución del abandono y la permanencia durante el primer año en la universidad.

En un segundo momento, nos ha permite explorar las interacciones entre los atributos de los estudiantes y su permanencia en la universidad, es decir apuntar a quienes son los que permanecen y apuntar que atributos están o no asociado a ello.

Un primer ejemplo lo documenta el promedio de bachillerato, en la siguiente tabla es notorio que el total de estudiantes que no permanecieron en las aulas el 67% tenia un promedio menor a 8.5 en tanto que de ellos sólo el 15 % corresponde a promedios de 9 o más. En otras palabras, que quienes abandonan se caracterizan mas por promedios más bajos.

Absolutos y porcentaje de los estudiantes que no permanecen según promedio de bachillerato.

Promedio de bachillerato			
<b>[6 - 7.7]</b>	<b>(7.8 - 8.4]</b>	<b>(8.5 - 8.9]</b>	<b>(9.0 - 10.00]</b>
4452	4616	2346	2075
33.0	34.2	17.4	15.4

Un segundo ejemplo lo aporta la calificación en el examen, una vez masa menor puntajes corresponde una mayor proporción de quienes no permanecen en la educación superior, el 57.6 % de quienes abandonan presentan puntajes menores a 64 puntos y solo el 19% corresponden a puntajes superiores a 74 puntos

Absolutos y porcentaje de estudiantes que NO PERMANECEN según calificación en el examen

Puntaje en el examen			
<b>[1 - 52]</b>	<b>(53 - 63]</b>	<b>(64- 73.00]</b>	<b>(74 - 100]</b>
4251	3519	3152	2567
31.5	26.1	23.4	19.0

Ambas aproximaciones sugieren una tendencia la cual indica que a menor promedio y menor calificación en el examen es mayor la proporción de quienes no permanecen, pero analizado desde cada grupo de promedio o calificación en el examen es notorio que la proporción de quienes permanecen es muy significativa y difícilmente se puede discriminar la probabilidad de abandono y mucho menos predecirla.

En la siguiente tabla, se puede percibir que de cada 100 alumnos que ingresaron con un promedio menor a 7.8 el 72.3 % permanece en tanto que el aso de promedios de 9 o mas el 83% permanece.

Absolutos y porcentaje de los estudiantes que PERMANECEN según promedio de bachillerato.

	Promedio de bachillerato			
	<b>[6 - 7.7]</b>	<b>(7.8 - 8.4]</b>	<b>(8.5 - 8.9]</b>	<b>(9.0 - 10.00]</b>
NO	27.7	21.5	19.7	16.1
SI	72.3	78.5	80.3	83.9

Una vez mas se refleja una tendencia per igualmente muestra que si bien indica que constituye un elemento que diferencia la probabilidad de permanencia su contribución a explica el fenómeno es débil e insuficiente para hacer predicciones y afirmaciones contundentes.

### Un comentario final

Se puede seguir ejemplificando con cada una de las variables en esta primera aproximación a la información que contienen los datos, pero de lo que se trata es de explorar, no las relaciones específicas de cada variable con la permanencia, sino en todo caso si existe una combinación de ellas que permite determinar con mayor precisión una mayor probabilidad de no permanencia y en consecuencia realizar clasificaciones y predicciones sobre el futuro de los estudiantes que ingresan a la universidad.

Nuestra hipótesis es que eso es posible mediante la propuesta de análisis comentada antes, pero teniendo presente o incorporando al análisis las carreras a las que asisten los estudiantes, es decir analizando la información mediante redes bayesianas para carreras específicas o grupos de carreras. En el reporte final que da continuidad a estos avances lo sabremos.

### Referencias

- Cruz Ramírez, N. (2001). *Building bayesian networks from data: a constraintbased approach*. Tesis doctoral no publicada. The University of Sheffield, Sheffield, Gran Bretaña.
- Chain, R. (1995). *Estudiantes universitarios: Trayectorias escolares*. México: Universidad Veracruzana- Universidad Autónoma de Aguascalientes.
- Chain, R., Jácome, N. y Rosales, O. (2000). Estudiantes, exámenes y trayectorias. En CENEVAL, *Memoria del Cuarto Foro de Evaluación Educativa* (páginas 29-32). México: Centro Nacional de Evaluación para la Educación Superior.
- Chain, R., Martínez Morales, M. Jácome, N. y Rosales, O. (2001). *Estudiantes, demanda y elección*. México: Universidad Veracruzana.
- Jensen, F. (2001). *Bayesian networks and decision graphs*. Nueva York: Springer-Verlag.

- Martínez Rizo, F. (1989). Diseño de investigación para el estudio de la deserción. Enfoque cuantitativo transversal. En ANUIES, *La trayectoria escolar en la educación superior* (pp. 281-294). México: Asociación Nacional de Universidades e Instituciones de Educación Superior.
- Martínez Rizo, F., Backhoff, E., Castañeda, S., De la Orden, A., Schmelkes, S., Solano-Flores, G., et. al. (2000). *Estándares de calidad para instrumentos de evaluación educativa*. México: Centro Nacional de Evaluación para la Educación Superior.
- Pearl, J. (1998). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Nueva York: Morgan Kauffman.
- Rochin Berumen, F. L. (2021). Deserción escolar en la educación superior en México: revisión de literatura. *RIDE Revista Iberoamericana Para La Investigación Y El Desarrollo Educativo*, 11(22). <https://doi.org/10.23913/ride.v11i22.821>