



Análisis de datos educativos utilizando Modelos Gráficos Probabilísticos

**Analysis of educational data using Probabilistic
Graphical Models**

Centro Sur.
Social Science Journal
2022 – Número continuo
<http://centrosureditorial.com/index.php/revista>
eISSN: 2600-5743
revistacentrosur@gmail.com

Atribución/Reconocimiento-
NoComercial-CompartirIgual 4.0
Licencia Pública Internacional —
CC BY-NC-SA 4.0
<https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.es>

Byron Oviedo Bayas
Amilkar Puris Cáceres
Jorge Gómez-Gómez
Cristian Zambrano-Vega

Resumen

Este artículo presenta una propuesta para implementar un método de conglomerados que mejor involucre los datos educativos (socioeconómicos, rendimiento académico y deserción) en la Facultad de Ingeniería de la Universidad Técnica del Estado de Quevedo. Para esta investigación se ha propuesto el uso de modelos gráficos probabilísticos en el campo de la educación. Para completar el diagnóstico de los estudiantes, y también para predecir su comportamiento, primero se realizó un análisis de modelos de aprendizaje de redes bayesianas, como la optimización de PC, K2 y EM. Debe haber una prueba para cada caso donde la probabilidad se mide en cada modelo usando algoritmos de propagación. Luego, se aplica el logaritmo de probabilidad a cada caso y se suman los resultados en cada modelo para determinar el mejor ajuste para el propuesto. Los resultados de esta investigación ayudarán a crear conciencia sobre los diversos factores que afectan el desempeño de los

estudiantes. Además, esto permitirá a las autoridades institucionales identificar mecanismos para mejorar el índice de retención y el rendimiento académico de los estudiantes, lo que sirve para mejorar los indicadores de calidad de cara a los procesos de evaluación y acreditación institucional y de programas.

Palabras Clave: Redes Bayesianas, K2, PC, EM, Rendimiento Académico

Universidad Técnica Estatal de Quevedo,
<https://orcid.org/0000-0002-5366-5917>,
boviedo@uteq.edu.ec, 120508, Quevedo-Ecuador

Universidad Técnica Estatal de Quevedo,
<https://orcid.org/0000-0002-7288-7451>,
apuris@uteq.edu.ec, 120508, Quevedo-Ecuador

Universidad de Córdoba, <https://orcid.org/0000-0001-8746-9386>,
jeliecergomez@correo.unicordoba.edu.co,
Montería-Colombia

Universidad Técnica Estatal de Quevedo,
<https://orcid.org/0000-0001-8568-8024>,
czambrano@uteq.edu.ec, 120508, Quevedo-Ecuador

Abstract

This paper presents a proposal to implement a cluster method that best engage the educational data (socio-economic, academic achievement and dropouts) at the Engineering Faculty of Quevedo State Technical University. The use of graphical probabilistic models in the field of education has been proposed for this research. To complete the student diagnosis, and to predict their behavior as well, an analysis of such Bayesian networks learning models, as PC, K2, and EM optimization was made first. There should be a test for each case where the probability is measured in every model using propagation algorithms. Then, probability logarithm is applied to each case and the results are added in each model to determine the best fit for the proposed. The results of this research will help raise awareness of the various factors affecting students' performance. Besides, this will allow institutional authorities to identify mechanisms for improving retention index and students' academic achievement, what serves the improvement of quality indicators in face of institutional and program evaluation and accreditation processes.

Key words: Bayesian networks, K2, PC, EM, Academic Performance

Introducción

Las redes bayesianas se conocen en la literatura existente con otros nombres, como redes causales o redes causales probabilísticas, redes de creencia, sistemas probabilísticos, sistemas expertos bayesianos, o también como diagramas de influencia. Las redes bayesianas son métodos estadísticos que representan la incertidumbre a través de las relaciones de independencia condicional que se establecen entre ellas (Edwards, 1998). Este tipo de redes codifica la incertidumbre asociada a cada variable por medio de probabilidades (Fernández, 2004) afirman que una red bayesiana es un conjunto de variables, una estructura gráfica conectada a estas variables y un conjunto de distribuciones de probabilidad.

Una red bayesiana es un grafo acíclico dirigido en el que cada nodo representa una variable y cada arco una dependencia probabilística en la cual se especifica la probabilidad condicional de cada variable dados sus padres; la variable a la que apunta el arco es dependiente (causa-efecto) de la que está en el origen de éste. La topología o estructura de la red nos da información sobre las dependencias probabilísticas entre las variables, pero también sobre las independencias condicionales de una variable (o conjunto de variables) dada otra u otras variables; dichas independencias simplifican la representación del conocimiento (menos parámetros) y el razonamiento (propagación de las probabilidades).

El obtener una red bayesiana a partir de datos es un proceso de aprendizaje que se divide en dos etapas: el aprendizaje estructural y el aprendizaje paramétrico (Pearl, 2011). La primera de ellas consiste en obtener la estructura de la red bayesiana, es decir, las relaciones de dependencia e independencia entre las variables involucradas. La segunda etapa tiene como finalidad obtener las probabilidades a priori y condicionales requeridas a partir de una estructura dada.

En el presente trabajo hace referencia a la investigación del uso de modelos probabilísticos gráficos en el campo de la enseñanza para la realización del diagnóstico de estudiantes y poder determinar el problema de deserción estudiantil en las universidades, el mismo que ha sido ya estudiado por algunos investigadores. Magaña Echeverría lo analiza haciendo uso de clúster agrupando a individuos u objetos en conglomerados de acuerdo con sus semejanzas, maximizando la homogeneidad de los objetos dentro de los conglomerados a la vez que maximiza la heterogeneidad entre agregados (Martha Alicia Magaña, 2006). Otro caso de estudio para predecir la probabilidad de que un estudiante abandone la institución educativa se han realizado utilizando técnicas de minería de datos para lograr el objetivo; entre ellos tenemos a (Ramón García Martínez, 2010), quienes realizaron un trabajo basado en el uso del conocimiento, en reglas de descubrimiento y en el enfoque TDIDT (Top Down Induction of Decision Trees) sobre la base de datos de la gestión académica del consorcio SIU de Argentina (que reúne 33 universidades de Argentina), lo cual permite un interesante análisis para encontrar las reglas de conducta que contienen variables de ausencia. En este trabajo nos centraremos en aplicaciones de los modelos gráficos probabilistas en la educación superior tal como lo menciona (Oviedo, 2018), (Oviedo B. M., 2016) que indica que los resultados obtenidos en su investigación permitirán una toma de decisiones inmediata para solucionar los problemas

detectados y de esta manera cumplir con la misión institucional de formar profesionales con visión científica y humanista capaces de desarrollar investigaciones, crear tecnologías, mantener y difundir nuestros saberes y culturas ancestrales, para la construcción de soluciones a los problemas de la región y el país (Oviedo B. Z.-V.-G., 2018)

Este trabajo se desarrolla con una base de conocimientos de 773 estudiantes matriculados en el periodo 2012-2013 en la Facultad de Ciencias de la Ingeniería de la Universidad Técnica Estatal de Quevedo.

Metodología

Si tenemos un conjunto de variables denotadas $X = \{X_1, X_2, \dots, X_n\}$ donde cada variable X_i toma valores dentro de un conjunto finito Ω_{x_i} ; entonces, usamos x_i para expresar uno de los valores de X_i , $x_i \in \Omega_{x_i}$. Ahora si tenemos un conjunto de índices denotado como I , simbolizamos x_i , $i \in I$. Notaremos el conjunto de todos los índices $N = \{1, 2, \dots, n\}$ y Ω_i representa a los elementos de Ω_{x_i} y se les llama configuraciones de X_i que serán representadas con x o x_i .

Definición de Red Bayesiana: es un grafo acíclico dirigido que representa un conjunto de variables aleatorias y sus dependencias condicionales (véase Figura 1), donde X_i es un suceso aleatorio representado por cada nodo (conjunto de variables), y la topología del gráfico muestra las relaciones de independencia entre variables de acuerdo con el criterio de separación (Peter Cheeseman, 1993), (Pearl, Probabilistic reasoning in intelligent systems: networks of plausible inference, 1988). Cada nodo X_i tiene una distribución de probabilidad condicional $p_i(X_i | \Pi(X_i))$ para esa variable, dadas sus padres. Por lo tanto, una red bayesiana determina una distribución única probabilidad conjunta:

$$P(X=x) = \prod p_i(x_i | \Pi(X_i)), \forall x \in \Omega_x \quad (1)$$

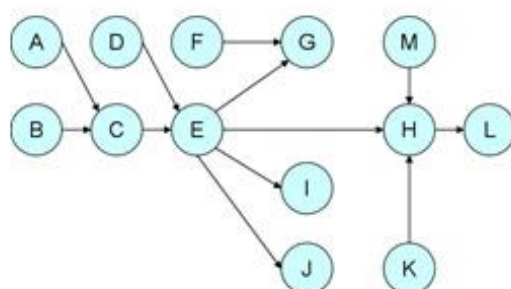


Figura 1. Ejemplo de red bayesiana

El aprendizaje automático se divide en aprendizaje supervisado y aprendizaje no supervisado. En el aprendizaje supervisado la atención se centra en el vaticinio exacto, mientras que en el aprendizaje no supervisado el objetivo es encontrar descripciones compactas de los datos. En ambos aprendizajes el interés es en los métodos que generalizan bien a los datos nuevos. En este sentido, se diferencia entre los datos que se utilizan para entrenar un modelo y los datos que se utiliza para probar el funcionamiento del modelo en formación.

Definición de Aprendizaje Supervisado: Dado un conjunto de datos $D = \{(x_n, c_n), n = 1, \dots, N\}$; donde (x_n) es el conjunto de observaciones donde tenemos una variable especial C que es la clase, hay que aprender la relación entre la entrada x y la salida y , de manera que cuando se tenga una nueva entrada x^* la salida predicha c^* sea exacta.

La pareja (x^*, c^*) no está en D , pero se asume que se generan por el mismo proceso desconocido que generó a D . para especificar explícitamente con precisión se define una función de pérdida $L(c^{\text{pred}}, c^{\text{true}})$ o, a la inversa, una función utilidad $U = -L$. aquí la utilidad suele ser 0 o 1, $L(c^{\text{pred}}, c^{\text{true}}) = 1$ si $(c^{\text{pred}} \neq c^{\text{true}})$, $L(c^{\text{pred}}, c^{\text{true}}) = 0$ si $(c^{\text{pred}} = c^{\text{true}})$.

En este tipo de aprendizaje lo importante es la distribución condicional $p(y | x, D)$. La salida también es llamada una "etiqueta", sobre todo cuando se habla de clasificación. El término "clasificación" denota el hecho de que la etiqueta es discreta, es decir, que consiste en un número finito de valores (Murphy, 2012).

Clasificadores bayesianos: cualquier red bayesiana puede ser usada para realizar una clasificación solo con distinguir la variable de interés del problema como la variable clase para luego aplicarle algún algoritmo de propagación con las nuevas evidencias y una regla de decisión para que los resultados sea un valor para la clase. Para determinar el mejor clasificador o el que más se ajuste a nuestro estudio se determinará el que mayor tasa de aciertos de clasificación obtiene, permitiéndonos una aproximación a la mejor representación bayesiana de la base de datos.

El modelo gráfico probabilístico para clasificación supervisada más usado y sencillo es Naive Bayes que se basa en dos supuestos: Primero, que cada atributo es condicionalmente independiente de los atributos dada la clase y segundo, que todos los atributos tienen influencia sobre la clase. NB ha demostrado ser comparable, en términos de precisión de clasificación en muchos dominios de muchos algoritmos más complejos, como las redes neuronales y árboles de decisión (Pazzani, 1998). El algoritmo utiliza los datos de entrenamiento para estimar todos los valores de la probabilidad requeridos.

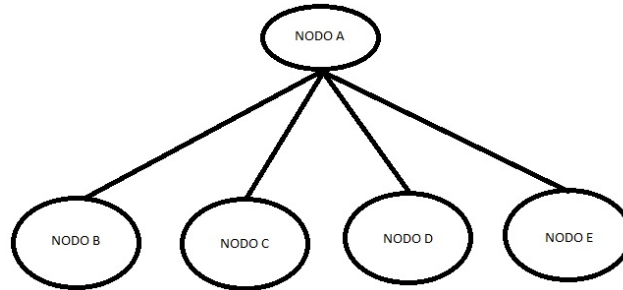


Figura 2. Clasificador Naive Bayes

Definición de Aprendizaje No Supervisado: dado un conjunto de datos $D = \{x_n, n = 1, \dots, N\}$ el objetivo es encontrar una descripción compacta de los datos. El mismo que se utiliza para cuantificar la exactitud de la descripción. En este tipo de aprendizaje no hay una variable de predicción especial, desde el punto de vista probabilístico, el interés es en el modelamiento de la distribución $p(x)$. La probabilidad del modelo para generar los datos es una medida popular de la exactitud de la descripción (John Robert Anderson, 1986).

El aprendizaje no supervisado, son conjuntos de observaciones que no tienen clases asociadas y la finalidad es detectar regularidades de los datos de cualquier tipo. A diferencia del supervisado este aprendizaje no trata los objetos de entrada como un conjunto de variables aleatorias. Son muy usados para la compresión de datos y agrupación.

¿Podemos agrupar los ejemplos en base a sus características?

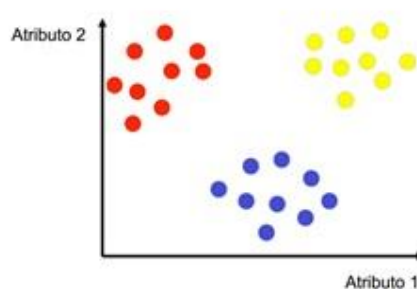


Figura 3. Agrupamiento con aprendizaje no supervisado

Algunas técnicas de aprendizaje no supervisado son clustering (agrupan objetos en regiones donde la similitud mutua es elevada), visualización (permite observar el espacio de instancias en un espacio de menor dimensión), reducción de la dimensionalidad (Los datos de entrada son agrupados en subespacios de una dimensión más baja que la inicial), extracción de características (construyen nuevos atributos a partir de los muchos atributos originales).

En este trabajo se usará la técnica de clustering descubrir la estructura en los datos de entrada buscando agrupamientos entre las variables y grupos de casos de forma que cada grupo sea homogéneo y distinto de los demás. Hay varios métodos de clustering entre los más conocidos tenemos: jerárquicos (los datos se agrupan de manera arborescente), no jerárquicos (genera particiones a un solo nivel), paramétricos (asume que las densidades condicionales de los grupos tienen cierta forma paramétrica conocida y se reduce a estimar los parámetros), no paramétricos (no asumen nada sobre el modo en el que se agrupan los objetos).

En muchas ocasiones es posible que no se hayan determinado algunas variables que eran importantes para la explicación del fenómeno bajo estudio, y es notado que existe una influencia en los datos que no es percibida fácilmente, entonces es posible postular la existencia de alguna o algunas variables ocultas como responsables de explicar esta producción anormal de los datos (Cruz-Ramírez, 2011).

Como deseamos aprender la red que mejor se adapte a los datos se han analizado varios algoritmos de aprendizaje bayesiano.

Algoritmo PC: Es uno de los algoritmos utilizados para aprendizaje bayesiano. Este Algoritmo se basa en pruebas de independencias entre variables $I(X_i, X_j | A)$, donde A es un subconjunto de variables, donde se asume que se tienen los datos suficientes y que las pruebas estadísticas no tienen errores. El algoritmo PC empieza con un grafo completo no dirigido para, posteriormente, ir reduciéndolo. Primero elimina las aristas que unen dos nodos que verifican una independencia condicional de orden cero, después las de orden uno, y así sucesivamente. El conjunto de nodos candidatos para formar el conjunto separador (el conjunto al que se acondiciona) es el de los nodos adyacentes a alguno de los nodos que se pretende separar. En este trabajo usamos la implementación hecha de PC en el programa Elvira, con un nivel de significancia de 0.05. PC realiza su cálculo basado en independencia y no en optimización del score.

Los algoritmos basados en funciones de score+search pretenden tener un grafo que mejor modelice a los datos de entrada, de acuerdo con un criterio específico. Cada uno de estos algoritmos utilizan una función de evaluación junto con un método que mide la bondad de cada estructura explorada en el conjunto total de estructuras. Durante el proceso de exploración, la función de evaluación es aplicada para evaluar el ajuste de cada estructura candidata a los datos. Cada uno de estos algoritmos se caracteriza por la función de evaluación y el método de búsqueda utilizados.

Cada uno de los algoritmos como función de evaluación utilizan una métrica no especializada en clasificación usada comúnmente en el aprendizaje de redes bayesianas como BIC (et, 1978), BDEu (David Heckerman, 2007), K2 (Herskovits, 1992). Por eficiencia la métrica usada debe ser descomponible, así sólo tendremos la parte que el operador modifica y no el grafo entero.

La métrica K2 para una red G y una base de datos D es la siguiente: (2)

$$f(G:D) = \log P(G) + \sum_{i=1}^n \left[\sum_{k=1}^{s_i} \log \left(\frac{\Gamma(n_{ik})}{\Gamma(N_{ik} + n_{ik})} \right) + \sum_{j=i}^{r_i} \log \left(\frac{\Gamma(N_{ijk} + n_{ijk})}{\Gamma(n_{ijk})} \right) \right]$$

Dónde: N_{ijk} es la frecuencia de las configuraciones encontradas en la base de datos D de las variables x_i ; n es el número de variables, tomando su j -ésimo valor y sus padres en G tomando su k -ésima configuración; s_i es el número de configuraciones posibles del conjunto de padres r_i es el número de valores que puede tomar la variable x_i ;

$$N_{ik} = \sum_{j=1}^{r_i} N_{ijk} \quad (3)$$

Γ es la función Gamma

La métrica BIC se define de la siguiente manera:

$$f(G : D) = \sum_{i=1}^n \sum_{j=1}^{r_i} \sum_{k=1}^{s_i} N_{ijk} \log \left(\frac{N_{ijk}}{N_{ik}} \right) - \frac{1}{2} C(G) \log(N) \quad (4)$$

Dónde: N es el número de registros de la base de datos; $C(G)$ es una medida de complejidad de la red G, definida como:

$$C(G) = \sum_{i=0}^n (r_i - 1) s_i \quad (5)$$

La métrica BDeu depende de un solo parámetro, el tamaño muestral a y se define de la siguiente manera:

$$gBDeu(G:D) = \log(p(D)) \sum_{i=1}^n \left[\log \left(\frac{\Gamma \frac{\alpha}{q_i}}{\Gamma(N_{ij} + \frac{\alpha}{q_i})} \right) + \sum_{k=1}^{r_i} \log \left(\frac{\Gamma(N_{ijk} + \frac{\alpha}{q_i})}{\Gamma \frac{\alpha}{q_i}} \right) \right] \quad (6) +$$

Algoritmo K2: Este algoritmo está basado en la búsqueda y optimización de una métrica bayesiana. K2 realiza una búsqueda voraz y muy eficaz para encontrar una red de buena calidad en un tiempo aceptable. K2 Es un algoritmo heurístico codicioso basado en la optimización de una medida. Esa medida se usa para explorar, mediante un algoritmo de

ascensión de colinas, el espacio de búsqueda formado por todas las redes que contienen las variables de la base de datos. Este algoritmo supone un orden entre las variables.

Para definir la red con este algoritmo debemos seguir tres pasos importantes:

- Se tiene un grafo inicial sin arcos
- Se elige un arco para añadir al grafo
 - Calculamos el score de la nueva red con un arco nuevo en cada paso
 - Se escoge el arco que da mayor score
 - Si el arco nuevo aumenta la probabilidad de la nueva red, se añade y se va al paso 2, caso contrario esa es la red y finalizará
 - Se supone un orden de las variables.

Algoritmo EM (Expectation Maximization) (Arthur P Dempster, 1977): es un método para encontrar el estimador de máxima verosimilitud de los parámetros de una distribución de probabilidad, este algoritmo es muy útil cuando parte de la información está oculta. Para encontrar estos parámetros óptimos se deben realizar dos pasos: primero (Expectation) calcular la esperanza de la verosimilitud con respecto a la información conocida y unos parámetros propuestos, luego (Maximization) maximizar con respecto a los parámetros; estos dos pasos se repiten hasta alcanzar la convergencia (Araujo, 2006)

Resultados

En esta sección se ha preparado un estudio experimental para realizar el análisis de los datos utilizando las técnicas de RB presentadas en el software Elvira. A continuación, se presentan las fases que se llevaron a cabo para el análisis

A. *Análisis de datos, identificación de atributos e imputación.*

Tabla I. Descripción del archivo de datos

Variable	Descripción
A	Carrera
B	Curso
D	Tiene discapacidad
E	Costo educación
F	Vive domicilio diferente de la familia
G	Tipo vivienda familia
H	Propietario de vivienda
I	Servicios tv cable
J	Servicios tarjeta crédito
K	Servicios acceso internet
L	Servicios básicos
M	Servicios transporte privado
N	Servicios celular plan
O	Servicios vehículo propio
P	Servicios viene vehículo propio
Q	Trabaja actualmente

La Tabla I identifica las variables con las que se trabajó, aquí podemos encontrar variables socioeconómicas representadas con las letras desde la D hasta la Q y se puede visualizar en la tabla I, variables que describen o identifican a los estudios que se las representó con las letras A y B que pueden ser visualizadas en la tabla III y por último variables de los

resultados académicos representadas con las letras R y S.

TABLA II: Carreras de la Facultad de Ciencias de la Ingeniería

Código	Carrera
FI024	Ingeniería en Sistemas
FI025	Ingeniería en Diseño Gráfico
FI026	Ingeniería Mecánica
FI027	Ingeniería Industrial
FI028	Ingeniería Telemática
FI029	Ingeniería Eléctrica
FI030	Ingeniería Agroindustria
FI031	Ingeniería Seguridad Industrial

Para el caso de las variables estado de aprobación y deserción estudiantil se codificaron con valores 1 y 0 para identificar los casos de estudio.

B. Discretización de variables

Elvira proporciona una gran cantidad de métodos de discretización entre los más destacados se encuentran: Discretización por igual frecuencia, Discretización por igual ancho, Suma de cuadrados diferentes, Contraste Mono técnico sin supervisión, Discretización K-medias. Los métodos de selección de variables continuas buscan obtener un nuevo conjunto de variables discretas a partir del conjunto original continuo para la clasificación, persiguiendo como objetivo simplificación del problema, mayor velocidad de aprendizaje, mayor interpretabilidad y un aumento del porcentaje de acierto.

En el caso de este trabajo el proceso de discretización se llevó a cabo de manera manual a través de criterio de expertos. La Tabla III muestra los valores discretizados de cada una de las variables.

C. Resultados

Para empezar a realizar las diferentes pruebas y experimentos se utilizó la técnica de Clúster cuyo objetivo es obtener clasificaciones (clusterings), de los casos que tienen características comunes; es decir, que sean similares, siendo los distintos grupos entre ellos tan distintos como sea posible.



TABLA III: Variables Discretizadas

Var	Atributo	Discretización
D	Tiene discapacidad	SI= 1; NO=0
E	Costo educación	X<200=0; 200>X<800=1; X>800=2
F	Vive domicilio diferente de la familia	SI= 1; NO=0
G	Tipo vivienda FAMILIA	MEDIA AGUA=0; CASA/VILLA=1; DEPARTAMENTO=2
G	Tipo vivienda FAMILIA	CUARTO DE INQUILINO=3; OTRA=4; RANCHO=5
H	Propietario VIVIENDA FAMILIA	PADRE Y MADRE=0; PADRE=1; MADRE=2; OTRO PARIENTE=3; OTRO=4
I	Servicios tv cable	SI= 1; NO=0
J	Servicios tarjeta crédito	SI= 1; NO=0
K	Servicios acceso internet	SI= 1; NO=0
L	Servicios básicos	SI= 1; NO=0
M	Servicios transporte privado	SI= 1; NO=0
N	Servicios celular plan	SI= 1; NO=0
O	Servicios vehículo propio	SI= 1; NO=0
P	Servicios viene vehículo propio	SI= 1; NO=0
Q	Trabaja actualmente	SI= 1; NO=0
R	Aprueba	SI= 1; NO=0
S	Deserta	SI= 1; NO=0

Como queremos que nuestra red aprenda nos ubicamos en aprendizaje automático, seleccionamos clasificador con corrección de Laplace y procesamos teniendo en consideración las variables mostradas en la Tabla IV

TABLA IV: variables utilizadas en el estudio

Variable	Descripción
A	Carrera
B	Curso
D	Tiene discapacidad
E	Costo educación
F	Vive domicilio diferente de la familia
G	Tipo vivienda familia
H	Propietario de vivienda
I	Servicios tv cable
J	Servicios tarjeta crédito
K	Servicios acceso internet
L	Servicios servicios básicos
M	Servicios transporte privado
N	Servicios celular plan
O	Servicios vehículo propio
P	Servicios viene vehículo propio
Q	Trabaja actualmente
R	Aprueba
S	Deserta

En la figura 4 se puede visualizar la red bayesiana que se ha generado a partir de un clasificador Naive Bayes que a través de una predicción supervisada nos construye este modelo que nos permite predecir la probabilidad de posibles resultados.

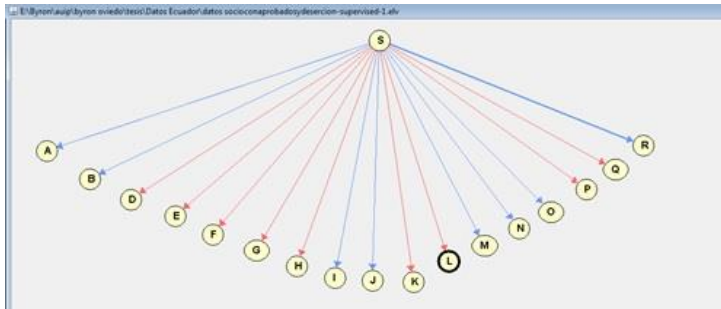


Figura. 4 clasificador Naives Bayes para los datos de estudio

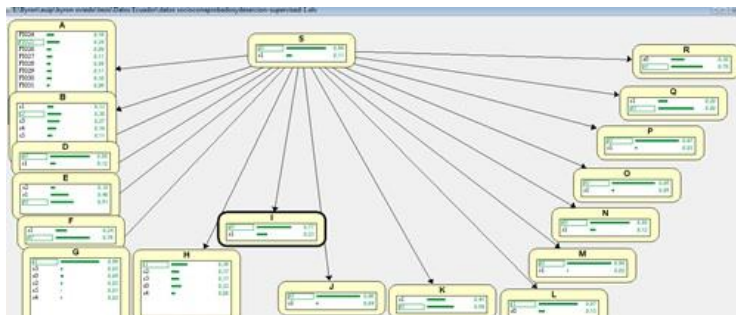


Figura. 5 red Bayesiana a partir de los datos donde se marca el nodo de influencia de la red (modo inferencia)

En la Figura 5 podemos visualizar la misma red, pero en el modo de inferencia donde se muestra la probabilidad de cada valor mediante una barra de longitud proporcional a la probabilidad y mediante un número. Como aún no hemos introducido ningún hallazgo, aparecen las probabilidades a priori.

Seguidamente se presenta un estudio para encontrar grupos de variables estrechamente relacionadas utilizando los algoritmos y parámetros mostrados en la Tabla V.

TABLA V: Métodos utilizados

Método	Clasificación	Clas-mét	Clus-pad	Estimación
K2	FACTORIZACIÓN	BDE	5	LAPLACE
PC	FACTORIZACIÓN	BIC	5	LAPLACE
EM	NO SUPERVISADA	CBN	2	LAPLACE

Como se puede visualizar en la Figura 6, El software Elvira ha entregado tres nodos padres que son Carrera, tipo de vivienda de la familia y si tiene discapacidad el estudiante o no, cada uno de ellos relacionados con otras variables, se puede notar que la variable Aprobar (R) se relaciona con la variable Deserción (S) y que tanto Carrera (A) y Curso (B) de manera indirecta influyen

sobre la deserción (S). Por otro lado se encuentran las variables socio económicas muy relacionadas entre ellas, pero sin influencia en la aprobación o deserción estudiantil.

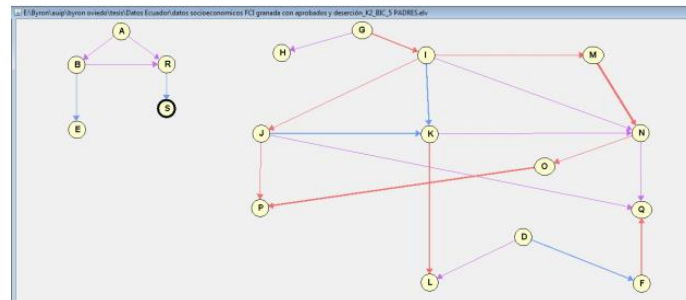


Figura. 6 red Bayesiana probando K2, máximo 5 padres, BIC, estimación bayesiana

De acuerdo con lo presentado por la figura 7 al ya trabajar con los 773 casos (estudiantes legalmente matriculados durante el periodo lectivo 2012-2013) se puede determinar la fuerte influencia del que si se cuenta con vehículo propio asistirá a la Universidad en ese medio de transporte, de igual manera se visualiza que si cuenta con transporte privado es más que seguro que tiene un plan de servicio celular al igual que tv cable.

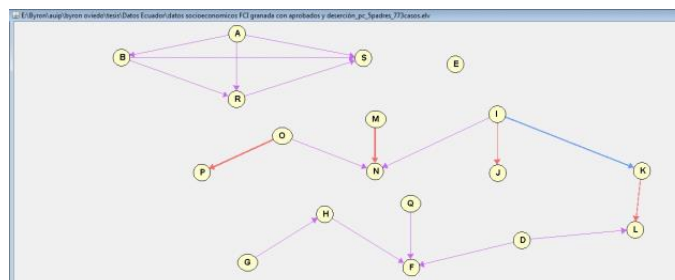


Figura. 7 red Bayesiana probando PC, máximo 5 padres.

Podemos identificar como se van agrupando las variables, por un lado, la carrera se relaciona con el curso, aprobación y deserción de una manera retirada a los otros grupos. La variable costo se encuentra aislada de todos los grupos.

Se puede fácilmente ir identificando que el tipo de relaciones no difieren mucho con el algoritmo anteriormente analizado (muchos se mantienen); pero por otro lado se empiezan a generar grupos muchos de ellos aislados entre las otras variables que no mantienen la relación similar al obtenido con K2.

Al trabajar con dos clústeres nos permite obtener un nodo clase del cual depende las variables, pudiéndose evidenciar la relación más fuerte con servicio de tv cable y servicio de internet tal como está representado en la figura 8.

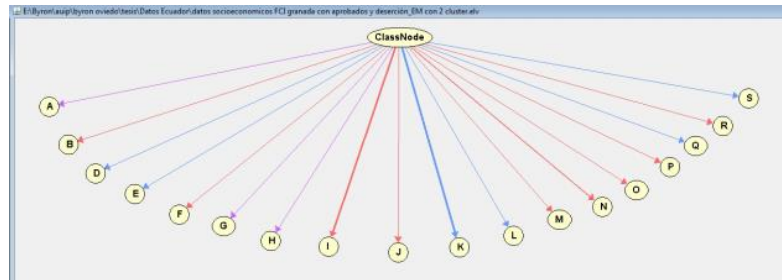


Figura. 8 red Bayesiana probando EM con 2 clúster (modo edición).

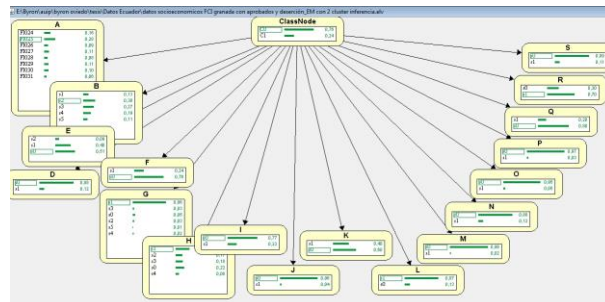


Figura. 9. Red Bayesiana probando EM con 2 clúster (modo inferencia).

De acuerdo con esta figura 9 se puede determinar el porcentaje de los estudiantes que desertan de sus estudios universitarios, de igual manera que la carrera que cuenta con mayor índice de deserción es la de Diseño Gráfico con un 28 por ciento, en lo referente al año en el que más desertan los estudiantes es segundo año con un 30 por ciento seguido muy de cerca de tercer año con un 27 por ciento. Los estudiantes en un 30 por ciento no logran aprobar las unidades de aprendizaje que están cursando, pero la mayoría de ellos se vuelven a matricular en la carrera por segunda ocasión.

El problema de deserción estudiantil en las carreras de la Facultad de Ciencias de la Ingeniería de la Universidad Técnica Estatal de Quevedo durante el periodo lectivo 2012-2013 tiene relación directa con el factor socio-económico, donde se puede determinar que el 24 por ciento de los estudiantes viven en un domicilio diferente al de la familia, donde el 86 por ciento de estas viviendas son de tipo villa, y el propietario de la vivienda en un 33 por ciento es el padre, hay que indicar también que el 77 por ciento de los estudiantes no tienen servicio de tv cable, que solo el 4 por ciento cuenta con tarjeta de crédito, que el 40 por ciento de los estudiantes pueden conectarse a internet en sus hogares; por otro lado solo el 2 por ciento de los estudiantes cuentan con transporte privado, 12 por ciento cuenta con plan celular, el 5 por ciento cuenta con vehículo propio, pero solo el 3 por ciento llega a la Universidad en vehículo propio; también se debe manifestar que el 20 por ciento de estos estudiantes trabajan, teniendo como consecuencia que el 30 por ciento de los estudiantes reprueban, de los cuales solo el 11 por ciento desertan, tomando segunda matrícula en la unidad de aprendizaje que reprobó. Por otro lado, estos datos son el resultado de la convergencia de estas variables de bajos ingresos

familiares, el desempleo y la incompatibilidad entre trabajo y estudio; como la causante principal.

En este trabajo se realiza un estudio de este problema aplicando el algoritmo k2, PC y EM, que son algoritmos que nos ayudan a agrupar variables como modelos para encontrar grupos de variables fuertemente relacionadas. Los resultados encontrados se pueden resumir de la siguiente manera:

Después de ejecutar el algoritmo k2 hemos podido determinar que K2 nos arroja grupos de variables agrupadas que definieron una posible causa de la deserción estudiantil, tomando como referencia el primero de estos grupos podemos observar que las variables Carrera (A) y Curso (B) influyen fuertemente sobre la variable Aprobar (R), así como estos también influyen sobre la deserción (S), hablando inclusive sobre los datos, notamos que hay un 11 por ciento de probabilidad que un estudiante deserte durante el transcurso de la carrera debido a la mala elección de la misma, también se puede apreciar que el 30 por ciento de estos estudiantes desertan en el segundo año de la carrera.

- a) El algoritmo PC con 5 padres y 733 casos, nos muestra un grupo similar de variables, pero, con diferentes conjuntos de relaciones, este algoritmo nos permite observar que la variable A influye altamente en las variables B, S y R, tomando como referencia que el mayor número de deserciones proviene de una sola carrera y que este se da en el segundo año de esta.
- b) *b)* Con respecto al caso de EM con dos clústeres podemos determinar una relación entre una variable clase y las demás variables, tomando en consideración la variable C como la variable clase y las demás como nodos hijos.

Luego de aplicar los algoritmos citados se puede notar que estos tuvieron en común el agrupamiento de las variables A, B, R y S en un solo conjunto de variables; así como también, la gran influencia que tiene la variable A hacia las demás variables del conjunto, siendo una posible causal directo de la deserción estudiantil.

Conclusiones

La permanencia o deserción de un estudiante en la universidad y la culminación con éxito de sus estudios están influenciadas por diferentes factores: individuales, académicos, socioeconómicos e institucionales. En este sentido el fenómeno de la deserción seguirá así cambien las Instituciones de Educación Superior. El investigar este problema permite establecer soluciones que controlen de manera parcial los índices de deserción y se logre retener a los estudiantes.

En este trabajo se ha podido evidenciar que un factor de impacto para la deserción estudiantil es el año en el que está cursando y los diferentes factores socio económicos que de manera directa influyen en el desempeño académico del estudiante

Se sugiere revisar los diferentes procesos de selección y admisión de los aspirantes a ingresar a la Universidad Técnica Estatal de Quevedo para poder detectar de entrada los posibles desertores. Además, se sugiere implementar políticas de becas y apoyo económico a estudiantes que lo necesiten y de esta manera disminuir la deserción cuando se sabe de antemano que los problemas económicos ya no son una barrera de ingreso por la gratuidad de la educación.

Referencias

- Araujo, B. S. (2006). *Aprendizaje automático: conceptos básicos y avanzados. Aspectos prácticos utilizando el software Weka.* .
- Arthur P Dempster, N. M. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1-38.
- Barros-Bastidas, C., & Turpo, O. (2020). La formación en investigación y su incidencia en la producción científica del profesorado de educación de una universidad pública de Ecuador. *Publicaciones*, 50(2), 167–185. doi:10.30827/publicaciones.v50i2.13952
- Barros, C., & Turpo, O. (2017). La formación en el desarrollo del docente investigador: una revisión sistemática. *Revista Espacios*, 38(45).
- Cruz-Ramírez. (2011). *Times Series Discretization Using Evolutionary Programming.*
- David Heckerman, C. K. (2007). Leveraging information across hlaalleles / supertypes improves epitope prediction . *Journal of Computational Biology*, 736-746.
- Edwards, W. (1998). Hailfinder: tools for and experiences with bayesian normative modeling. *American Psychologist*, 416.
- et, G. S. (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.
- Fernández, E. (2004). *Análisis de clasificadores bayesianos - Trabajo Final de Especialidad en Ingeniería de Sistemas Expertos.* Buenos Aires: Escuela de Postgrado. Instituto Tecnológico de Buenos Aires.
- Herskovits, G. F. (1992). A bayesian method for the induction of probabilistic networks from data. *Machine learning*, 309-347.
- John Robert Anderson, R. S. (1986). *Machine learning: An artificial intelligence approach.* Morgan Kaufmann.
- Martha Alicia Magaña, O. A. (2006). Análisis de la evolución de los resultados obtenidos por los profesores en las evaluaciones esdedep y las realizadas por los estudiantes. *Revista de la Educación Superior*, 29-48.
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective.* MIT Press.
- Oviedo, B. M. (2016). A hierarchical clustering method: Applications to educational data. *Intelligent Data Analysis*, 933-951.
- Oviedo, B. P. (2018). Algoritmos meta heurísticos para el aprendizaje de redes bayesianas. *Revista Lasallista de Investigación*, 353-366.
- Oviedo, B. Z.-V.-G. (2018). RED BAYESIANA JERARQUICA PARA ANALIZAR DATOS EDUCATIVOS. *Revista Investigación Operacional.*
- Pazzani, G. I. (1998). Adjusted probability naive bayesian induction. *In Proceedings of the Eleventh Australian Joint Conference on Artificial Intelligence* (págs. 285–295). Springer-Verlag.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan Kaufmann.
- Pearl, J. (2011). *Bayesian networks.* Department of Statistics, UCLA.
- Peter Cheeseman, J. K. (1993). *Autoclass: a bayesian classification system In*

Readings in know-ledge acquisition and learning. Morgan Kaufmann Publishers Inc.

Ramón García Martínez, H. K. (2010). Identificación de causales de abandono de estudios universitarios. *TE & ET*.

von Feigenblatt, Otto Federico (2007). *Japan and Human Security: 21st Century ODA Policy Apologetics and Discursive Co-optation* (2nd ed.). Delray Beach: Academic Research International.

von Feigenblatt, Otto Federico (2009a). Anomie, Racial Wage, and Critical Aesthetics: Understanding the Negative Externalities of Japanese and Thai Social Practices. *Journal of Asia Pacific Studies*, 1(1), 69-75.