

RED BAYESIANA JERÁRQUICA PARA ANALIZAR DATOS EDUCATIVOS

Byron Oviedo*, Cristian Zambrano-Vega*, Jorge Gómez-Gómez**

*Universidad Técnica Estatal de Quevedo, Ecuador.

**Universidad de Córdoba, Colombia.

ABSTRACT

This document presents a proposal to implement a cluster method that best matches the educational data. The use of probabilistic graphic models in the field of education has been considered for this research. But the problem with these general learning procedures comes from the presence of a high number of variables that measure different aspects of the same concept. In this case, we have that all variables have some degree of dependence between them, without a true causal structure. Therefore, a new procedure is presented that makes a hierarchical grouping of the data while learning a joint probability distribution. We try to make a test for each case where the probability is measured in each model using propagation algorithms. Then, the probability logarithm is applied to each case and the results are added in each model to determine the best fit for the proposed one. The method is applied to the analysis of a data set of educational data: evaluation of students of the Quevedo State Technical University (Ecuador)

KEYWORDS: Bayesian Networks, K2, PC, EM, Hierarchical Clustering, Academic Performance, Student Evaluation.

MSC: 62F15;62C10.

RESUMEN

Este documento presenta una propuesta para implementar un método de clster que mejor acople los datos educativos. El uso de modelos gráficos probabilísticos en el campo de la educacin ha sido considerado para esta investigacin. Pero el problema con estos procedimientos generales de aprendizaje proviene de la presencia de un alto nmero de variables que miden diferentes aspectos del mismo concepto. En este caso, tenemos que todas las variables tienen algñ grado de dependencia entre ellas, sin una verdadera estructura causal. Por lo tanto, se presenta un nuevo procedimiento que hace una agrupacin jerrquica de los datos mientras se aprende una distribucin de probabilidad conjunta. Se busca hacer una prueba para cada caso donde la probabilidad se mide en cada modelo usando algoritmos de propagacin. Luego, el logaritmo de probabilidad se aplica a cada caso y los resultados se agregan en cada modelo para determinar el mejor ajuste para el propuesto. El mtodo se aplica al anlisis de un conjunto de datos educativos: evaluacin de estudiantes de la Universidad Tecnica Estatal de Quevedo (Ecuador).

PALABRAS CLAVE: Redes Bayesianas, K2, PC, EM, Clustering Jerrquico, Desempeo Acadmico, Evaluacin Estudiantil.

*boviedo@uteq.edu.ec

1. INTRODUCCIÓN

En este trabajo, en primer lugar se propone un nuevo clasificador bayesiano simple (SBND) que aprende forma rápida una frontera de Markov de la variable clase y una estructura de red que relaciona las variables de la clase y su frontera de Markov. Este modelo se compara con otros clasificadores bayesianos. A continuación, se considera el uso de modelos gráficos probabilísticos 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: [5] lo analiza haciendo uso de clúster, agrupando a individuos u objetos en conglomerados de acuerdo a sus semejanzas, maximizando la homogeneidad de los objetos dentro de los conglomerados a la vez que maximiza la heterogeneidad entre agregados. Otro caso de estudio para predecir la probabilidad de que un estudiante abandone la institución educativa se ha realizado utilizando técnicas de minería de datos; entre ellos tenemos a [3], quienes realizaron un trabajo basado en el uso del conocimiento, en reglas de asociación 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 comportamiento.

[4] usa un método de predicción de deserción en los cursos de e-learning, basado en tres técnicas populares de aprendizaje automático: redes neuronales feedforward, máquinas de soporte vectorial y métodos de ARTMAP difuso simplificado.

[1] comparan distintos modelos para predecir las tasas de abandono durante el primer semestre de los estudios de grado en la Universidad de Eindhoven. Utilizan árboles de clasificación, naive Bayes, regresión logística y bosques de árboles, obteniendo unas tasas de acierto entre el 75 y el 80 por ciento. En este artículo empezaremos con una sección 2. en la que introducimos un nuevo clasificador que llamaremos *clasificador bayesiano simple*, que será una red bayesiana genérica, pero aprendida con una técnica voraz.

En la sección 3. se realiza un análisis de deserción de los estudiantes legalmente matriculados en la Facultad de Ciencias de la Ingeniería de la Universidad Técnica Estatal de Quevedo en el periodo 2012-2013 en base a los datos socio económicos, [6]. En la subsección (3.1.) se hace uso de la herramienta Weka para realizar la clasificación. Primero empezamos usando clasificadores bayesianos (3.2.). En este caso se utiliza naive Bayes, BayesNet con K2 y un solo padre, BayesNet con K2 y un máximo de 5 padres, BayesNet con TAN, BayesNet con Hill Climber y un solo padre, BayesNet con Hill Climber y un máximo de 5 padres. Luego en la subsección (3.3.), hacemos uso de los clasificadores de árboles. Aquí se comparan los resultados obtenidos al hacer uso de árboles J48 y Random Forest. A continuación se analizan los datos con clasificadores que hacen uso de reglas de clasificación en la subsección (3.4.).

2. UN CLASIFICADOR BAYESIANO SIMPLE

Supongamos un problema de clasificación supervisada con clase C y atributos $\mathbf{X} = (X_1, \dots, X_n)$ y un conjunto de observaciones completas para la clase y los atributos \mathcal{D} . El objetivo es aprender una red bayesiana para un conjunto de variables $\mathbf{X}' \cup \{C\}$, donde $\mathbf{X}' \subseteq \mathbf{X}$. El método tendrá así un sistema

implícito de selección de variables, descartando aquellas variables en $\mathbf{X} \setminus \mathbf{X}'$. Todas las variables seleccionadas estarán relacionadas con la variable clase C .

La idea es simple: se comienza introduciendo a C como nodo raíz en la red bayesiana \mathcal{B} resultado. Se mantiene un conjunto de nodos \mathbf{X}' de los atributos ya introducidos en la red (inicialmente vacía).

Se pueden usar distintas métricas $Score(X_i, \mathbf{A} | \mathcal{D})$ que miden la idoneidad de \mathbf{A} como conjunto de padres de X_i (esta métrica puede ser BDEu, BIC, K2 o Akaike).

Se supone que tenemos un procedimiento $PADRES(X_i, CANDIDATOS, \Pi_i)$ que calcula el mejor conjunto de padres Π_i de X_i con la métrica seleccionada y devuelve el valor de esa métrica óptima. La implementación actual de esta función es un algoritmo voraz que comienza con un conjunto Π_i vacío y va añadiendo o quitando de Π_i la variable que produce un mayor incremento de la métrica, hasta que ya no hay mejoras posibles.

En estas condiciones se calcula para cada variable $X_i \in \mathbf{X} \setminus \mathbf{X}'$ el valor:

$$Infor(X_i, C) = PADRES(X_i, \mathbf{X}' \cup \{C\}, \Pi_i) - PADRES(X_i, \mathbf{X}', \Pi_i)$$

Es decir $Infor(X_i, C)$ calcula la diferencia entre las mejores métricas de X_i con conjunto de padres elegido entre \mathbf{X}' incluyendo C y sin incluir C entre los candidatos. Intuitivamente, es una medida de la dependencia condicional de X_i y C dadas las variables ya incluidas \mathbf{X}' . Este valor teóricamente siempre es mayor o igual a cero, pero podría ser negativo dado que el mejor conjunto de padres se calcula de forma aproximada.

Una vez calculado este valor para cada variable $X_i \in \mathbf{X} \setminus \mathbf{X}'$, se elige la variable $X_{max} = \arg \max_{X_i \in \mathbf{X} \setminus \mathbf{X}'} Infor(X_i, C)$. Esta sería la variable que más informa sobre la clase C condicionado a las variables ya introducidas. Si $Infor(X_{max}, C) > 0$, entonces esta variable aporta información adicional sobre C y se inserta en la red y en \mathbf{X}' . Su conjunto de padres es Π_i calculado con $PADRES(X_{max}, \mathbf{X}' \cup \{C\}, \Pi_i)$. En teoría, siempre $C \in \Pi_i$, ya que en otro caso $Infor(X_{max}, C) = 0$, aunque debido a la naturaleza voraz del procedimiento podría ser que $C \notin \Pi_i$, aunque esta posibilidad sería rara.

El proceso continúa de forma iterativa tratando de añadir cada vez una nueva variable y termina cuando en un paso $Infor(X_{max}, C) \leq 0$.

Las principales características de este clasificador son:

- Aprende una red bayesiana arbitraria con un subconjunto de las variables iniciales que influyen directamente en esta variable. En este sentido se puede considerar como un algoritmo que calcula una frontera de Markov, ya que trata de conseguir un conjunto de variables tales que dadas estas variables, el resto de las variables son independientes.
- La variable clase es siempre un nodo raíz. y hay enlaces desde este nodo al resto de los atributos (excepto en raras ocasiones debido a la naturaleza aproximada del cálculo de los padres). En este sentido se parece a otros clasificadores bayesianos en los que siempre existe un enlace de la clase a cada uno de los atributos.
- El orden de los atributos se basa en elegir de forma voraz primero los atributos que más informan sobre la clase, dados los atributos ya seleccionados. De esta forma se introducen primero los atributos más relevantes. No se busca en el espacio de los órdenes de los atributos para obtener

la red con mejor métrica, sino que se concentra sólo en obtener la máxima información para la clase. En este sentido puede haber alguna pérdida en la calidad de la red, pero el algoritmo gana en rapidez.

3. ANÁLISIS DE DESERCIÓN EN LA FCI-UTEQ

En esta sección se desarrollan experimentos 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, de los cuales se han obtenido sus datos socio-económicos y académicos para poder ser usados en clasificación con la herramienta Weka [2]. Los diferentes valores que asumen cada una de esas variables, se muestran en las tablas 1, 2, 3,4.

Valor	Descripción
FI024	Ingeniería en Sistemas
FI025	Ingeniería en Diseño Gráfico
FI026	Ingeniería Mecánica
FI027	Ingeniería Industrial
FI028	Ingeniería en Telemática
FI029	Ingeniería Eléctrica
FI030	Ingeniería Agroindustrial
FI031	Ingeniería en Seguridad Industrial y Salud Ocupacional

Tabla 1: Valores y descripción de la variable carrera

Valor	Descripción
1	Primero
2	Segundo
3	Tercero
4	Cuarto
5	Quinto

Tabla 2: Valores y descripción de la variable cursos

Variable	D	F	I	J	K	L
Descripción	{1, 0}	{1, 0}	{1, 0}	S{1, 0}	{1, 0}	{1, 0}

Variable	M	N	O	P	Q	R	S
Descripción	{1, 0}	{1, 0}	{1, 0}	{1, 0}	{1, 0}	{1, 0}	{1, 0}

Tabla 3: Variables y descripción de valores, SI=1; NO=0

Variable	Descripción
E	$X < 200 = 0; 200 > X < 800 = 1; X > 800 = 2$
G	MEDIA AGUA=0;CASA/VILLA=1;DEPARTAMENTO=2; CUARTO DE INQUILINATO=3;OTRA=4;RANCHO=5
H	PADRE Y MADRE=0;PADRE=1;MADRE=2;OTRO PARIENTE=3;OTRO=4

Tabla 4: Variables y consideraciones a discretizar

3.1. Clasificación usando Weka

Weka es una colección de herramientas de visualización y algoritmos de inferencia y clasificación con librerías en JAVA que permite la extracción de conocimiento desde bases de datos. Permite trabajar con distintas herramientas de minería de datos como reglas de asociación, agrupación, clasificación y regresión. Podemos visualizar en las figuras 1, los resultados obtenidos por cada uno de los atributos en referencia a la clase *S* deserta. De forma visual, ya se puede observar que no hay variables que de forma individual den información significativa sobre la deserción.

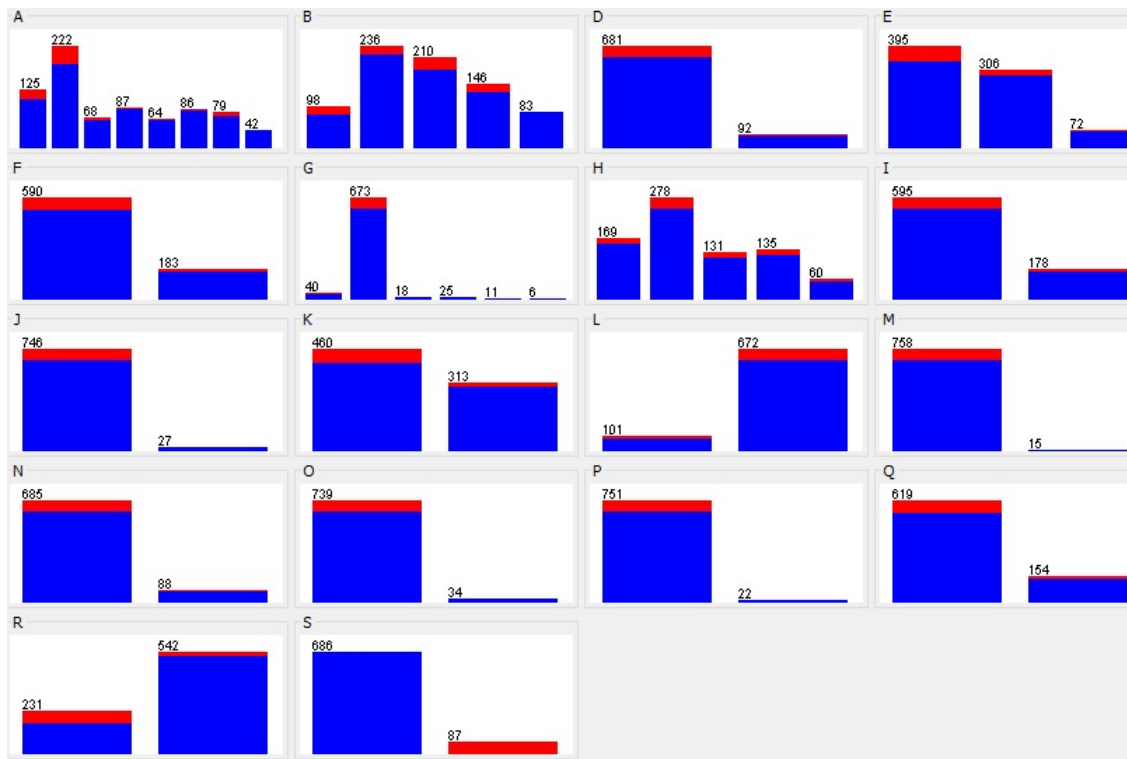


Figura 1: Resultados obtenidos por cada uno de los atributos en referencia a la clase

De acuerdo a la figura 1 se ha construido las tablas 5, 6 y 7, en el que podemos encontrar el análisis descriptivo de las variables donde se puede visualizar el porcentaje de cada uno de sus valores.

Variable	Descripción	Cantidad	Porcentaje
Q	no	619	80,08
	si	154	19.92
R	no	231	29.89
	si	542	70.11
S	no	686	88.75
	si	87	11.25

Tabla 5: Variables y análisis descriptivo

Variable	Descripción	Cantidad	Porcentaje
A	FCI024	125	16.17
	FCI025	222	28.72
	FCI026	68	8.80
	FCI027	87	11.25
	FCI028	64	8.28
	FCI029	86	11.13
	FCI030	79	10.22
	FCI031	42	5.43
B	1	98	12.68
	2	236	30.53
	3	210	27.17
	4	146	18.89
	5	83	10.74
D	no	681	88.10
	si	92	11.90

Tabla 6: Variables y análisis descriptivo

Variable	Descripción	Cantidad	Porcentaje
E	$X < 200$	395	51.10
	$200 > X < 800$	306	39.59
	$X > 800$	72	9.31
F	no	590	76.33
	si	183	23.67
G	media agua	40	5.17
	casa villa	673	87.06
	departamento	18	2.33
	cuarto inquilinato	25	3.23
	otro	11	1.42
	rancho	6	0.78
H	padre y madre	169	21.87
	padre	278	35.97
	madre	131	16.95
	otro pariente	135	17.46
	otro	60	7.76
I	no	595	76.97
	si	178	23.03
J	no	746	96.51
	si	27	3.49
K	no	460	59.51
	si	313	40.49
L	no	101	13.07
	si	672	86.93
M	no	758	98.06
	si	15	1.94
N	no	685	88.62
	si	88	11.38
O	no	739	95.60
	si	34	4.40
P	no	751	97.15
	si	22	2.75

Tabla 7: Variables y análisis descriptivo

3.2. Usando clasificadores bayesianos

Se obtuvieron resultados usando como clasificadores Naive Bayes y BayesNet con diferentes alternativas como K2, TAN, Hill Climber. con un padre y también con un máximo de 5 padres.

Los algoritmos voraces se utilizan generalmente para resolver problemas de optimización (obtener máximo o mínimo) van tomando decisiones en función de la información que está disponible en cada momento, una vez tomada la decisión, ésta no vuelve a replantearse en el futuro. Normalmente son muy rápidos y son bastante sencillos de programar, en cada paso eligen una solución local óptima, pero en la parte negativa no siempre conducen al óptimo.

Para obtener estos valores en la herramienta Weka hemos clasificado usando una validación cruzada de 10. Como se puede observar en la tabla 8 se ha trabajado con 773 casos, de los cuáles BayesNet con K2 y máximo 5 padres es el que mejor ha clasificado correctamente (91.0737 por ciento) y adicionalmente nos indica la tasa de verdaderos negativos (TN) y la tasa de verdaderos positivos (TP). Podemos observar que hay un 32.20 por ciento de sensibilidad. Este es el porcentaje de estudiantes que se han clasificado correctamente entre aquellos que desertan. éstos son los que deberían de recibir alguna atención y sobre los que habría que aplicar acciones especiales para disminuir este índice. Aunque no es una tasa muy alta, es importante sealar que es un problema difícil de predecir y por este procedimiento se detectan principalmente la tercera parte de los estudiantes que desertan. Además, el costo en términos de falsos positivos es muy bajo.

Clasificador	Clasificados correctamente	Tasa TN	Tasa TP
NaiveBayes	88.0983	0.965	0.218
BayesNet con K2-1 padre	87.9690	0.964	0.218
BayesNet con K2-5 padres	91.0737	0.985	0.322
BayesNet con TAN	89.9094	0.974	0.318
BayesNet con Hill Climber-1	88.6158	0.974	0.195
BayesNet con Hill Climber-5	89.6507	0.975	0.276

Tabla 8: Resultados obtenidos con los diferentes clasificadores

La tasa de verdaderos negativos o especificidad corresponde a la probabilidad de que un estudiante que esté bien en su proceso académico tenga un resultado negativo en la prueba. En este caso solo se llega a detectar como falsos positivos un 1.5 por ciento ($1 - 0.985$).

En la figura 2 podemos visualizar que todas las variables están relacionadas directamente con la clase deserción (S). La variable curso (B) depende también de la carrera (A) e influyen sobre el resultado académico (R). Por otro lado, se puede también considerar que la variable contar con servicio de tv cable (I) influye directamente sobre (J, K, M), servicio de plan celular (N), y éstas sobre servicio de vehículo propio (O) y trabajar actualmente (Q).

En la figura 3 podemos visualizar la dependencia de las variables con la clase S ; así como también que el curso (B) y la variable trabaja actualmente (Q) van a estar dependiendo directamente de la variable carrera (A). El TAN considera sólo las relaciones más relevantes. Entre ellas la dependencia de la carrera para saber si se aprueba o no el curso (R), de igual manera que si el estudiante trabaja (Q) vivirá en un domicilio diferente al de la familia (F).

En la figura 4 podemos visualizar la red obtenida con con BayesNet y un solo padre. Hay influencia de 5 variables carrera (A), curso (B), costo de educación (E), servicio de internet (K) y aprueba (R)

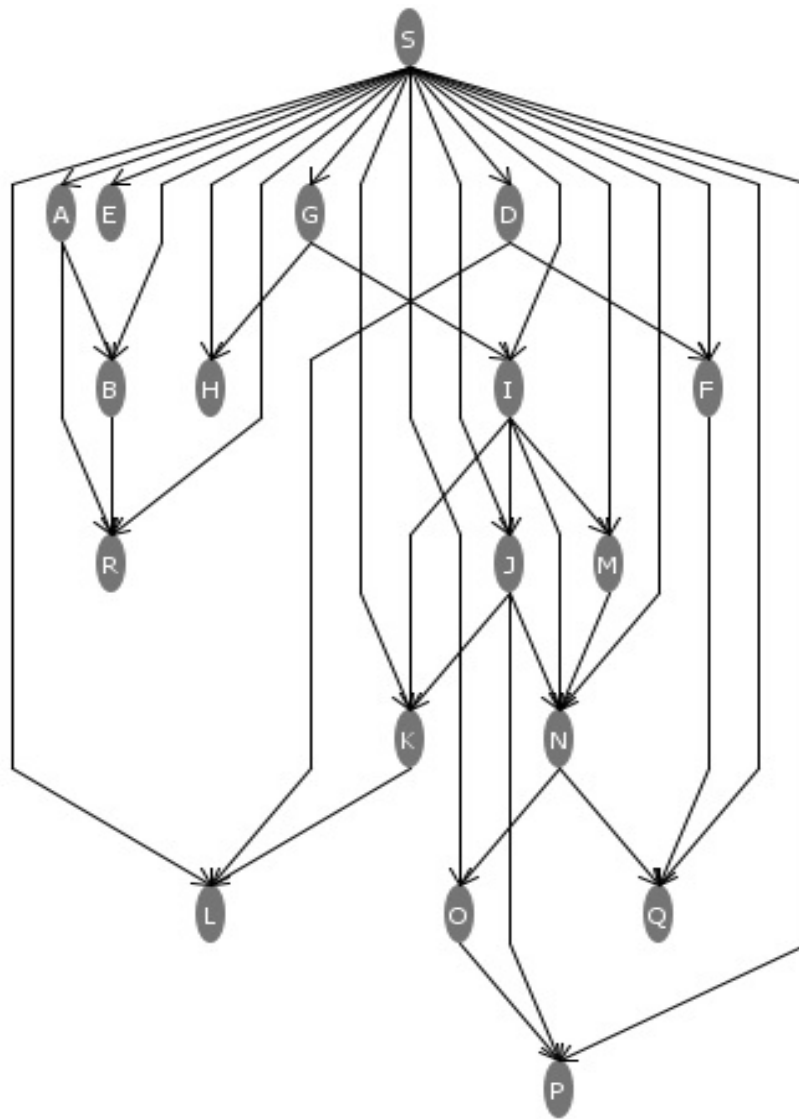


Figura 2: Red obtenida con clasificador BayesNet con K2 y un máximo de 5 padres

de manera directa con la clase deserta (S); así como también la de la variable estudiante trabaja actualmente (Q) con carrera (A). También se demuestra una dependencia fuerte de la variable tener servicio de internet (K) con las variables servicios básicos (L) y con servicio de tv cable (I).

En la figura 5 podemos visualizar la dependencia de las variables carrera (A), curso (B), costo de educación (E), servicio de internet (K) y aprueba (R) con la clase deserta (S); se sigue manteniendo la dependencia tanto de servicios básicos (L) como servicio de plan celular (N) de servicio de acceso a internet (K) y está de servicio de tarjeta de crédito (J).

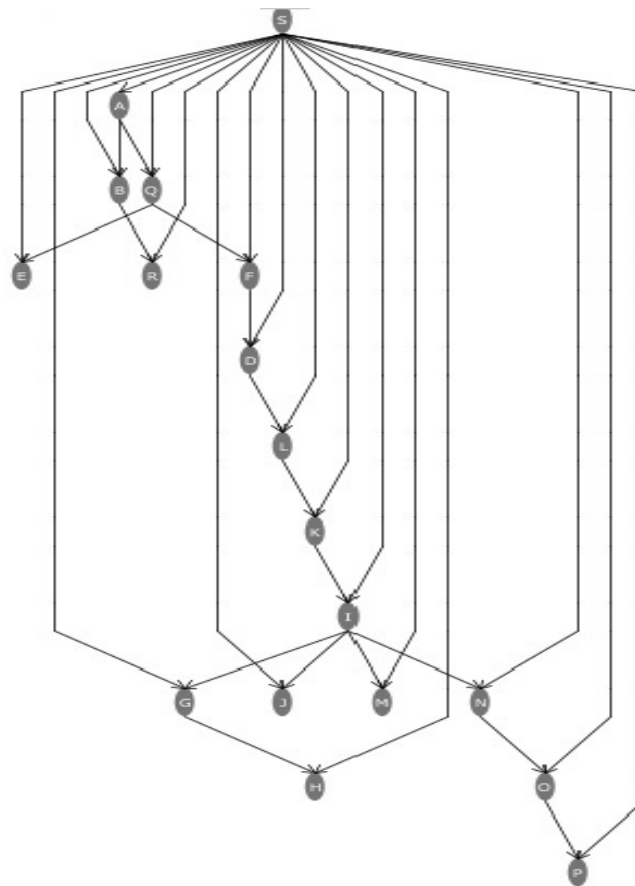


Figura 3: Red obtenida con clasificador BayesNet con TAN

3.3. Usando clasificadores de árboles

Los resultados que se obtuvieron usando como clasificadores de árboles J48 y Random Forest pueden ser visualizados en la tabla 9.

Clasificador	Clasificados correctamente	Tasa TN	Tasa TP
J48	88.2277	0.993	0.011
RandomForest	89.6507	0.978	0.253

Tabla 9: Resultados obtenidos con clasificadores de árboles

Como se puede observar trabajando con un clasificador de árbol J48 los casos clasificados correctamente equivalen al 88.2277 por ciento. Además se indica el porcentaje de sensibilidad y especificidad. Estos valores no mejoran al del clasificador BayesNet con K2 y 5 padres. De igual manera se puede observar que trabajando con un clasificador de árbol Random Forest, los casos clasificados correctamente mejoran en referencia a J48. Se debe indicar que es un bosque aleatorio de 100 árboles de los

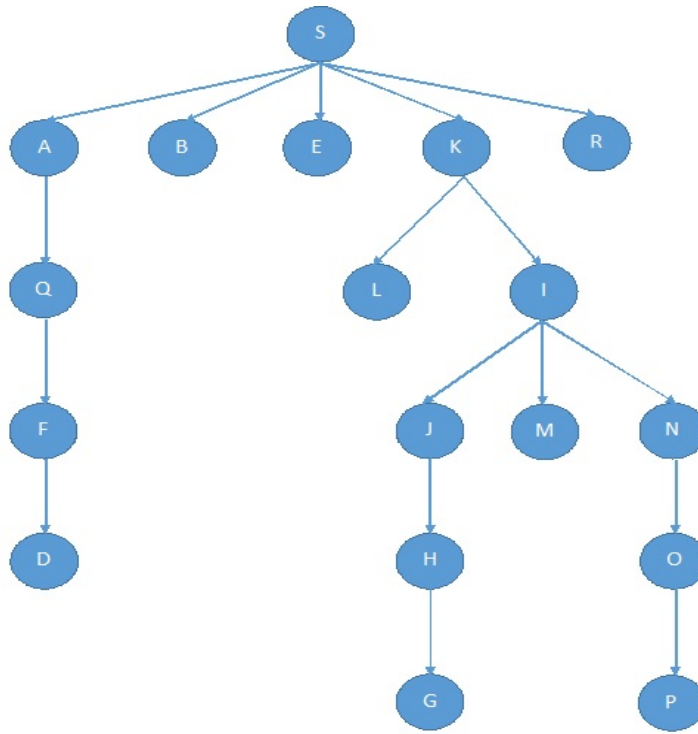


Figura 4: Red obtenida con clasificador BayesNet con HILL CLIMBER y un solo padre

cuáles cada uno está construido con 5 características.

3.4. Usando reglas de clasificación

Los resultados usando como reglas de clasificación ZeroR y tablas de decisiones se pueden ver en la tabla 10.

Clasificador	Clasificados correctamente	Tasa TN	Tasa TP
ZeroR	88.7451	1	0
Tabla de decisiones	89.0039	0.981	0.172

Tabla 10: Resultados obtenidos con diferentes reglas de clasificación

Como se puede observar trabajando con reglas de clasificación ZeroR que corresponde a clasificadores triviales que siempre responden a la clase más frecuente ya sea negativa o positiva, los casos clasificados correctamente equivalen al 88.7451 por ciento. En el caso del trabajo con tablas de decisiones, los casos clasificados correctamente mejoran en referencia a ZeroR.

Se realizará una comparación de los resultados obtenidos con los diferentes algoritmos del estado del arte con la base de datos de variables socio-económicas de los estudiantes de la Facultad de Ciencias de la Ingeniería de la Universidad Técnica Estatal de Quevedo tal como se indica en la tabla 11.

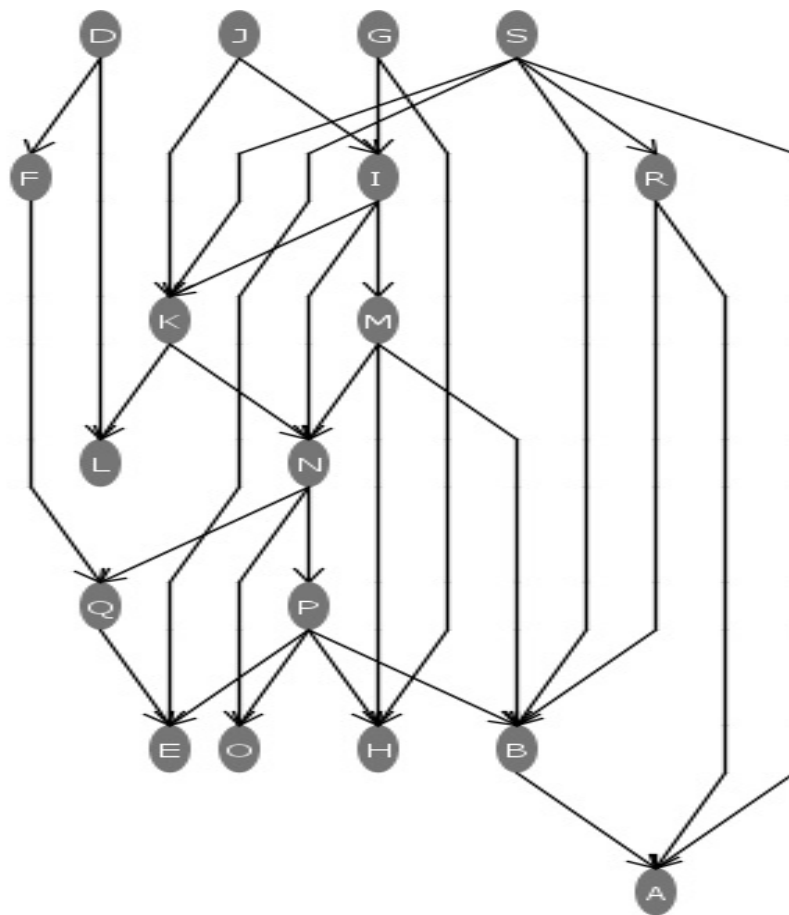


Figura 5: Red obtenida con clasificador BayesNet con HILL CLIMBER y un máximo de 5 hijos

Datos	SBND BDE	SBND BIC	SBND Ak	SBND K2	BAN BDEu	BAN BIC
socioeconomico	88.232	88.745	88.357	89.910	87.581	87.711
Datos	BAN K2	RPDag BDEu	RPDag BIC	RPDag K2	TAN	NaiveBayes
socioeconomico	87.723	87.584	87.972	89.657	88.743	87.456

Tabla 11: Resultados con base de datos de estudiantes de la UTEQ

De igual manera se obtiene el ranking de los algoritmos en la tabla 12. Estos valores servirán también para las comparaciones con los datos de la UCI en los cursos de matemáticas y portugués.

Algoritmos	Puntaje
SBND1	6.00
SBND2	3.33
SBND3	8.00
SBND4	5.33
BAN BDe	7.00
BAN BIC	8.33
BAN K2	7.66
RPDag BDe	5.5
RPDag BIC	5.5
RPDag K2	8.33
TAN	6.0
NaiveBayes	7.0

Tabla 12: Puntaje promedio de los algoritmos

Como se puede observar el algoritmo que entrega mejores resultados es SBND con K2 con mucha diferencia de los otros con los que se han comparado y que los de peor resultado son con métrica BDEu y BAN con la métrica BIC.

En virtud de que valor que se obtiene con el test de Friedman es mayor que 0.05, la hipótesis nula no es rechazada y se determina que no hay diferencias significativas entre las distribuciones y por lo tanto no es necesario seguir realizando pruebas. Estos resultados se originan en virtud de que se han comparado pocas bases de datos.

4. CONCLUSIONES

En este trabajo hemos introducido un clasificador bayesiano conocido como SBND que se basa en obtener de forma rápida una frontera de Markov que es fácil de aprender y muy competitivo. SBND fue aplicado al análisis de datos en problemas de educación. Este clasificador es rápido de aprender y muy competitivo en relación a los otros clasificadores del estado del arte. Se realizaron varios experimentos tanto con base de datos socio-económicos de estudiantes legalmente matriculados en la Facultad de Ciencias de la Ingeniería de la Universidad Técnica Estatal de Quevedo en el periodo

2012-2013.

Se pudo determinar en esta investigación que la métrica BIC no da buenos resultados con pruebas no paramétricas, mientras que Akaike entrega un excelente resultado en referencia a la media, pero mal en el orden en términos de clasificación.

El comportamiento del clasificador en determinados ejemplos es dependiente de la métrica que se use. BIC no da buen resultado, Akaike da buenos resultados en referencia a la media y K2 da buenos resultados en tests no paramétricos.

Se puede indicar que el problema de deserción estudiantil que se analiza es complejo y difícil y que ha sido necesario utilizar métodos que usan una combinación de factores (clasificadores bayesianos) para poder obtener algunas mejoras sobre el clasificador trivial que determina que ningún estudiante deserta.

Aunque la tasa de éxito no es muy alta se puede determinar que usando un clasificador bayesiano (bayesNet con K2 y un máximo de 5 padres) se puede detectar el 32.20 por ciento de los estudiantes que van a desertar y poder aplicar metodologías que ayuden a trabajar con este individuo de manera tal que se evite ese paso.

El coste medido como el porcentaje de alumnos que se consideran potenciales desertores entre los no desertores es muy bajo y equivale al 1.5 por ciento lo que no permite determinar con exactitud los desertores.

Como trabajo futuro, pensamos que se debe incluir los costes de las clasificaciones incorrectas en el problema, ya que no es lo mismo un falso positivo que un falso negativo. Si consideramos que el coste de un falso negativo es mejor que el de un falso positivo se podría detectar más alumnos que abandonarían, aunque se aumentaría el número de estudiantes que se consideran en peligro de desertar.

RECEIVED: NOVEMBER, 2019.

REVISED: JANUARY, 2020.

REFERENCIAS

- [1] DEKKER, G., PECHENIZKIY, M., and VLEESHOUWERS, J. (2009): Predicting students drop out: A case study In **Educational Data Mining 2009**.
- [2] GARNER, S. R. et al. (1995): Weka: The waikato environment for knowledge analysis In **Proceedings of the New Zealand computer science research students conference**, pages 57–64.
- [3] KUNA, H., GARCÍA-MARTÍNEZ, R., and VILLATORO, F. (2010): Identificación de causales de abandono de estudios universitarios. uso de procesos de explotación de información **Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología**, 5:39–44.
- [4] LYKOURENTZOU, I., IOANNIS, G., VASSILIS, N., GEORGE, M., and LOUMOS, V. (2009): Dropout prediction in e-learning courses through the combination of machine learning techniques **Computers & Education**, 53(3):950–965.

- [5] MAGAÑA, M., MONTESINOS, O., and HERNÁNDEZ, C. (2006): Análisis de la evolución de los resultados obtenidos por los profesores en las evaluaciones ESDEPED y las realizadas por los estudiantes **Revista de la Educación Superior**, 35(140):29–48.
- [6] OVIEDO, B., PURIS, A., VILLACÍS, A., DELGADO, D., and MORENO, A. (2015): Análisis de datos educativos utilizando redes bayesianas **Latin American and Caribbean Conference for Engineering and Technology LACCEI 2015**.