El clasificador Naïve Bayes en la extracción de conocimiento de bases de datos

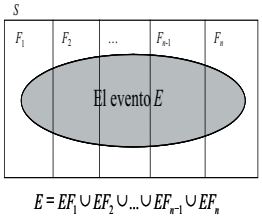
Samuel D. Pacheco Leal, Luis Gerardo Díaz Ortiz,
Rodolfo García Flores
Posgrado en Ingeniería de Sistemas, FIME-UANL
samuel@yalma.fime.uanl.mx
rodolfo@yalma.fime.uanl.mx

RESUMEN

El análisis de la gran cantidad de datos generados día a día en las actividades productivas y que se almacenan en grandes bases de datos electrónicas resulta cada vez más complicado. La minería de datos es un área de la Inteligencia Artificial que se ha desarrollado para facilitar el análisis de estos registros y se define como la búsqueda automática o semiautomática de patrones no triviales en bases de datos. En este artículo se hace una breve introducción a sus técnicas y aplicaciones a problemas reales y se muestra en detalle el funcionamiento de uno de estos algoritmos, conocido como el método Naïve Bayes. Se demuestra su utilidad práctica a través de un caso de estudio relacionado al diagnóstico de uso de lentes de contacto.

PALABRAS CLAVE:

Minería de datos, Naïve Bayes, reconocimiento de patrones, aprendizaje de máquina.



ABSTRACT

Day after day, the analysis of the huge amount of generated data by productive activities and stored in electronic databases becomes more complicated. Data mining is an emerging field of Artificial Intelligence which purpose is to facilitate the analyses of such data repositories, and it is defined as the automated or semi-automated search of non-trivial patterns in databases. This paper briefly introduces its main techniques and makes an overview of data mining applications to real-life problems. To illustrate the practical value of data mining, a detailed explanation of one of its methods, known as the Naïve Bayes classifier, is given. The argument is enhanced through the presentation of a case study related to contact lens diagnosis.

KEYWORDS:

Data Mining, Naïve Bayes, pattern recognition, machine learning.

INTRODUCCIÓN

Hoy en día existe tal cantidad de datos generados en todas las actividades productivas almacenados en medios electrónicos, que su análisis manual resulta

imposible. Es muy fácil que información relevante quede escondida entre montañas de datos, cuando la velocidad con la que se mueve la sociedad hoy en día requiere que esta información sea extraída de forma rápida, sin descuidar su confiabilidad.

Recientemente se han desarrollado diferentes algoritmos con el fin de cubrir esta necesidad de información, los cuales forman parte de la denominada minería de datos. La minería de datos es una etapa del proceso de descubrimiento en bases de datos que utiliza diversas herramientas de análisis para descubrir patrones y relaciones entre los datos que puedan ser usados para realizar predicciones válidas.[1] El fin de la minería de datos es encontrar, de una forma u otra, patrones útiles y significativos para quien generó los datos originales. Los patrones encontrados pueden ser utilizados para diversos fines, como por ejemplo comprender mejor una situación, hacer predicciones ante casos nuevos, servir como una herramienta para la toma de decisiones o simplemente para adquirir conocimiento. No hay que perder de vista que la minería de datos es solamente un instrumento que se aplica a datos ya existentes y que por ello no genera información por sí sola.

La minería de datos ha encontrado numerosas aplicaciones entre las cuales se pueden mencionar:

- *Banca.* Diversas instituciones bancarias han usado modelos basados en la minería de datos para evaluar y aprobar créditos.[2]
- *Pronósticos del clima.* Se han analizado registros históricos de fenómenos atmosféricos para pronosticar eventos climáticos. Por ejemplo, Liong y Sivapragasam[3] utilizaron la minería de datos para pronosticar inundaciones en la región de Dhaka, Bangladesh, logrando muy buenos resultados.
- *Medicina.* No todos los medicamentos surten el mismo efecto ni actúan con la misma intensidad en todos los pacientes; algunas personas pueden presentar ciertas reacciones negativas a determinado tipo de medicamento o grupo de medicamentos. La investigación de reportes acerca de estas reacciones adversas a medicamentos ha sido estudiada usando la minería de datos.[4]
- *Bibliotecas.* El interés en el análisis del comportamiento de los usuarios de bibliotecas se ha incrementado rápidamente a partir del desarrollo de las bibliotecas digitales e Internet. En este contexto, se reporta en[5] que se analizaron las consultas realizadas a bibliotecas digitales y el registro de estas consultas fue almacenado. Posteriormente y a partir de este registro se construyeron comunidades de usuarios con intereses similares utilizando la minería de datos, con el fin de que estas comunidades puedan mejorar su acceso a la información.
- *Seguridad Nacional (EUA).* El gobierno de los EU maneja un proyecto llamado "conciencia total de la información", o TIA por sus siglas en inglés (Total Information Awareness) en conjunto con la Agencia de Proyectos de Investigación de Defensa Avanzados, DARPA (Defense Advanced Research Projects Agency). Este proyecto busca recolectar información de transacciones financieras individuales, registro de viajes, registros médicos y otras actividades con el fin de prevenir el terrorismo. La minería de datos es utilizada para analizar esta información desde el año 2003.[6]

La gran cantidad de aplicaciones que ha encontrado la minería de datos ha requerido el rápido desarrollo de una variedad de técnicas de análisis. A continuación se mencionan las principales de estas técnicas junto con algunos algoritmos representativos.

- *Agrupamiento.* El propósito de estas técnicas es agrupar un conjunto de elementos relacionando aquellos que sean semejantes y al mismo tiempo que sean suficientemente diferentes de otros grupos de elementos formados. A este tipo de algoritmo se le conoce como no dirigido, pues no se conoce con antelación el grupo específico al que pertenece una instancia, sino que de acuerdo a los datos, los grupos se van formando, según sus semejanzas y diferencias. Dentro de las aplicaciones del agrupamiento se encuentran: reducción de datos, generación de hipótesis, prueba de hipótesis, y predicción basada en grupos.[7] Como ejemplo de esta técnica, Strehl y Ghosh[8] aplicaron un algoritmo de agrupamiento para disminuir la dimensión de una base de datos a matrices de 2 dimensiones, lo cual es de gran ayuda al momento de visualizar los resultados.
- Análisis de series de tiempo. El pronóstico de series de tiempo pronostica valores aún no conocidos, utilizando resultados conocidos para guiar sus predicciones. Como ejemplo, el análisis de series de tiempo fue utilizado, junto con otros algoritmos, para identificar fenómenos atmosféricos que pudieran surgir sobre las islas de Japón como tifones y frentes fríos, en imágenes de satélites.[9]
- Asociación. El objetivo de la asociación es encontrar aquellos artículos (sucesos) que tienden a aparecer juntos en algún evento dado. El campo donde más se ha desarrollado este tipo de algoritmo es el de los supermercados. Este problema, mejor conocido como el problema del análisis del carrito de supermercado, consiste en encontrar aquellos artículos que los consumidores adquieren juntos, con el fin de diseñar mejores estrategias de venta. Por ejemplo, una estrategia de venta puede consistir en ubicar los productos asociados en estantes cercanos para facilitar a los consumidores su adquisición.[10] Las transacciones en el supermercado proporcionan los datos y debido a la enorme cantidad que se genera diariamente es necesaria la automatización del análisis mediante la minería de datos.
- *Predicción.* Existen dos tipos de algoritmos utilizados para realizar predicciones:

  * *Regresión.* El objetivo de este tipo de análisis es determinar, de acuerdo a un resultado dado, el valor de los parámetros que produjeron ese resultado. Por ejemplo, se ha reportado el uso de métodos de regresión para asegurar la calidad de los sistemas de cómputo mediante el análisis de los errores de ejecución.[11]
  * *Clasificación.* La clasificación trata de encontrar las características que identifican a un grupo para ser clasificado dentro de cierta clase. Este conocimiento puede ser utilizado para entender el comportamiento del sistema que generó los datos y de esta forma predecir la clase a que pertenecerá una nueva instancia. Entre los algoritmos de clasificación se encuentran:
    - ♦ *Análisis discriminante.* La forma en la que opera este algoritmo, es determinando la localización óptima de una línea que actúa como frontera entre los diferentes casos. El algoritmo trata de ubicar la línea de tal manera que el margen de separación entre casos de diferente clase sea máximo. Este método tiene la ventaja de ser muy fácil de visualizar, sin embargo no siempre se puede hacer este tipo de discriminaciones. Por ejemplo, R. Kholi et al[12] reporta que se utilizó un análisis discriminante para presentar evidencia estadística de características que discriminan entre estudios de rentabilidad de tecnología de información que presentan efectos positivos y los que no los presentan.
    - ♦ *k-vecinos más cercanos.* Conociendo ciertos individuos similares, el algoritmo forma un grupo de k individuos, de acuerdo a sus características. Cuando aparece un nuevo individuo, éste se puede clasificar en cierto grupo de acuerdo a su semejanza con los k individuos pertenecientes a ese grupo. Por ejemplo, en la referencia[13] se reporta que la producción de madera de piceas en Noruega requiere una evaluación muy precisa de la calidad interna de la madera de los árboles. La calidad interna se predijo de acuerdo a ciertas variables externas que son fácilmente medibles o que se pueden conocer sin necesidad de cortar el árbol, como por ejemplo su edad, diámetro medio,

área en la base, altura y volumen. La relación de estas variables externas con las variables que reflejan la calidad de la madera como la densidad, cantidad de núcleo y cantidad de nudos que presenta cada árbol fue determinada por los investigadores gracias a la experiencia de los aserraderos a lo largo de los años. El método de k-vecinos más cercanos fue utilizado para predecir la calidad de estos árboles utilizando una amplia base de datos que fue obtenida en una extensa investigación llevada a cabo por el Instituto Finlandés de Investigación de Bosques.

♦ *Redes neuronales.* Este tipo de algoritmos intenta emular el funcionamiento de los cerebros de los seres vivos mediante capas de "neuronas", que son funciones matemáticas con un comportamiento determinado. Existe una capa de entrada seguida de una o varias capas intermedias, para finalizar en una capa de salida. Por ejemplo, en[14] se reporta que, en búsqueda de la disminución de gases producidos por los automóviles, se utilizó una red neuronal para ajustar los parámetros de operación en un tipo de motor.

♦ *Árboles de decisión.* Estos algoritmos "aprenden" reglas a partir de datos, tratando de obtener la descripción más sintética (i.e., de menor tamaño) que represente de forma más cercana los datos originales. Cuando se presenta un nuevo caso, simplemente se siguen las reglas extraídas por el algoritmo. Por ejemplo, en el tratamiento de cáncer, se ha reportado el uso de los árboles de decisión para mostrar los resultados posibles a partir de los síntomas e historias clínicas de los pacientes.[15]

♦ *Vectores soporte.* Estos algoritmos están relacionados con los de análisis discriminante. Se han utilizado técnicas de vectores soporte para el pronóstico de inundaciones.[16]

♦ *Naïve Bayes.* Es un método basado en la teoría de la probabilidad, usa frecuencias para calcular probabilidades condicionales para calcular predicciones sobre nuevos casos. Naïve Bayes es una técnica tanto predictiva como descriptiva. A pesar de ser simple, ha sido desarrollada con éxito, produciendo buenos resultados en sus aplicaciones. Este método será descrito en detalle en la siguiente sección y posteriormente utilizado para nuestro caso de estudio.

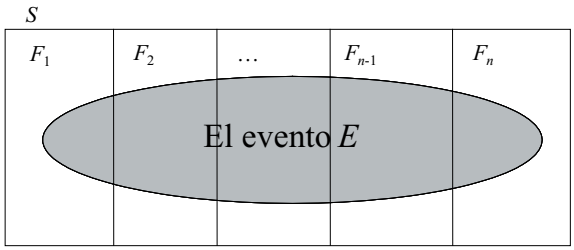


Fig. 1. El evento E sucede junto con alguno de los eventos mutuamente excluyentes F_j

EL CLASIFICADOR NAÏVE BAYES

Sean E y F eventos. Podemos expresar a E como

$$E = EF \cup EF^c \quad (1)$$

es decir, para que ocurra un evento E, deben suceder E y F, o bien debe suceder E y no suceder F.

Thomas Bayes

Debido a que EF y EF^c son mutuamente excluyentes, tenemos que

$$P(E) = P(EF) + P(EF^c)$$

$$= P(E|F)P(F) + P(E|F^c)P(F^c)$$

$$= P(E|F)P(F) + P(E|F^c)(1 - P(F)) \quad (2)$$

La ecuación (2) establece que la probabilidad del evento E es una ponderación de la probabilidad condicional de E dado que F ha ocurrido y la probabilidad condicional del evento E dado que F no ha ocurrido. Cada probabilidad condicional proporciona tanta ponderación como el evento condicionado tiende a ocurrir.

La ecuación (2) puede generalizarse de la siguiente manera: supongamos que los eventos $F_1, F_2, \ldots F_n$ son mutuamente excluyentes tal que $\bigcup_{i=1}^{n} F_i = S$, donde S es el espacio muestral. En otras palabras, exactamente uno de los eventos ocurrirá (figura 1).

Podemos escribir lo anterior como

$$E = \bigcup_{i=1}^{n} EF_i$$

De la definición de probabilidad condicional tenemos que

$$P(EF_i) = P(E|F_i)P(F_i) \quad (3)$$

Además, usando el hecho de que los eventos EF_i, $i = 1, \ldots, n$ son mutuamente excluyentes, obtenemos que

$$P(E) = \sum_{i=1}^{n} P(EF_i)$$

$$= \sum_{i=1}^{n} P(E|F_i)P(F_i) \quad (4)$$

Así, la ecuación (4) muestra como, para eventos dados $F_1, F_2, \ldots F_n$ de los cuales uno y solamente uno puede ocurrir, se puede calcular $P(E)$ condicionando a que ocurra F_1. Esto es, se establece que $P(E)$ es igual al promedio de las ponderaciones de $P(E|F_i)$ y cada término es ponderado por la probabilidad del evento en el cual es condicionado.

Supóngase ahora que E ha ocurrido y que se quiere determinar la probabilidad de que el evento F_j haya ocurrido. Por la ecuación (4) tenemos que

$$P(F_j|E) = \frac{P(EF_j)}{P(E)}$$

$$= \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)} \quad (5)$$

La ecuación (5) es conocida como la fórmula de Bayes.[17] Así, podemos considerar a E como evidencia de F_j, y calcular la probabilidad de que F_j ocurra dada la evidencia, $P(E|F_i)$. Supóngase ahora que se tiene evidencia de múltiples fuentes. De la ecuación (3):

$$P(F_j|E_1E_2\ldots E_m) = \frac{P(E_1E_2\ldots E_m|F_j)P(F_j)}{P(E_1E_2\ldots E_m)} \quad (6)$$

lo que dificulta el cálculo, pues el término $P(E_1E_2\ldots E_m|F_j)$ no es sencillo de obtener. Para resolver el problema, se supone que los E_i son independientes dado F_j, lo que nos permite escribir:

$$P(F_j|E_1E_2\ldots E_m) = \frac{P(E_1|F_j)P(E_2|F_j)\ldots P(E_m|F_j)P(F_j)}{P(E_1E_2\ldots E_m)} \quad (7)$$

la cual es la ecuación que se utilizará para la obtención de resultados.

La suposición que da origen al adjetivo Naïve (ingenuo) es la independencia entre las variables, lo cual no es siempre cierto. Sin embargo, el método ha sido exitoso en su aplicación debido a que la información relevante está contenida en las magnitudes relativas entre las cantidades y no tanto en los valores de las probabilidades en sí.

PRESENTACIÓN DEL PROBLEMA

Un campo de aplicación muy fértil para la minería de datos ha sido tradicionalmente el de la salud. En esta sección presentaremos una problemática que se encuentra dentro del área de la oftalmología. Entre los padecimientos de los ojos que puede sufrir un individuo dentro de una población, se encuentra la deficiencia visual, esto es, padecer hipermetropía, miopía, o astigmatismo, entre otras enfermedades. Una forma de manejar estas deficiencias es mediante el uso de lentes de contacto, de los cuales existen dos tipos de acabado: los lentes suaves y los lentes duros. Por otro lado, existen individuos con deficiencias visuales cuyo organismo no es capaz de aceptar el uso de ningún tipo de lente. Esto significa que aun sufriendo algún padecimiento, es posible ser diagnosticado como un paciente no apto para el uso de lentes de contacto. Para el diagnóstico de lentes de contacto se toman como base atributos relevantes como la edad, padecimiento, astigmatismo y lagrimeo. De acuerdo con estos datos se puede predecir el uso o no de lentes y de qué tipo.

Para demostrar la aplicación del clasificador Naïve Bayes se utilizará una base de datos que contiene 24 entradas que representan las características de los pacientes (las instancias) y que se muestra en la tabla I. A cada instancia corresponde el diagnóstico de no usar lentes de contacto o de usar lentes duros o suaves. Este diagnóstico es la clase que se quiere predecir a partir de los valores de los demás atributos (las características de los pacientes), que son edad, padecimiento, si se sufre astigmatismo y lagrimeo. Estos atributos se muestran en ese orden en la tabla I.

Los valores que puede tomar cada atributo se explican a continuación. Para la edad existen 3 posibles valores: joven, pre-presbiópico y presbiópico; como padecimiento se tomarán: miope o hipermétrope. El astigmatismo puede presentarse o no, por lo tanto este atributo tomará el valor de sí o no; y el lagrimeo se presentará normal o reducido.

Se presenta a continuación la tabla de instancias para el problema abordado. Se muestran las características de los pacientes y en la última columna el resultado obtenido por el clasificador, el

Tabla I. Base de datos a analizar

	Edad	Padecimiento	Astigmatismo	Lagrimeo	Tipo de lente	Pct.
1	joven	hipermétrope	si	reducido	ninguno	89.06%
2	joven	hipermétrope	si	normal	duro	45.53%
3	joven	hipermétrope	no	reducido	ninguno	83.28%
4	joven	hipermétrope	no	normal	suave	67.17%
5	joven	miope	si	reducido	ninguno	80.71%
6	joven	miope	si	normal	duro	66.28%
7	joven	miope	no	reducido	ninguno	83.35%
8	joven	miope	no	normal	suave	57.85%
9	pre-presbiópico	hipermétrope	si	reducido	ninguno	92.97%
10	pre-presbiópico	hipermétrope	si	normal	ninguno	51.02%
11	pre-presbiópico	hipermétrope	no	reducido	ninguno	86.11%
12	pre-presbiópico	hipermétrope	no	normal	suave	65.25%
13	pre-presbiópico	miope	si	reducido	ninguno	87.80%
14	pre-presbiópico	miope	si	normal	duro	53.24%
15	pre-presbiópico	miope	no	reducido	ninguno	86.73%
16	pre-presbiópico	miope	no	normal	suave	57.77%
17	presbiópico	hipermétrope	si	reducido	ninguno	94.48%
18	presbiópico	hipermétrope	si	normal	ninguno	57.74%
19	presbiópico	hipermétrope	no	reducido	ninguno	91.28%
20	presbiópico	hipermétrope	no	normal	suave	52.21%
21	presbiópico	miope	si	reducido	ninguno	89.78%
22	presbiópico	miope	si	normal	duro	51.76%
23	presbiópico	miope	no	reducido	ninguno	91.31%
24	presbiópico	miope	no	normal	ninguno	43.21%

Tabla II. Tabla de conteos

Conteo

Edad

	ninguno	suave	duro
joven	4	2	2
pre-presbiópico	5	2	1
presbiópico	6	1	1
		Total	24

Padecimiento

	ninguno	suave	duro
hipermétrope	8	3	1
miope	7	2	3
		Total	24

Astigmatismo

	ninguno	suave	duro
si	8	0	4
no	7	5	0
		Total	24

Lagrimeo

	ninguno	suave	duro
normal	4	5	4
reducido	11	0	0
		Total	24

Tipo de lente

ninguno	suave	duro
15	**5**	**4**

cual se describe más adelante.

Por ejemplo, la instancia 7 muestra que a un individuo joven que padece de miopía, no sufre astigmatismo y presenta lagrimeo reducido, se le diagnostica como no apto para uso de lentes de contacto.

Organizar los datos de manera práctica agiliza mucho el análisis. En este caso se manipularon los datos en Excel, en el cual se programaron las formulas presentadas en la sección anterior para obtener los resultados que se muestran más adelante.

La forma de organizar los datos es a través de un conteo en las instancias, realizado de la siguiente forma: para cada atributo, se cuenta el número de veces que aparece determinado valor junto con una clase en particular. Ponemos como ejemplo el caso de la edad. Se cuenta el número de veces que se presenta cada valor del atributo con cada valor de la clase. Así, joven y lente suave aparecen juntos un total de 2 veces, joven y lente duro un total de 2, joven y no apto para lente un total de 4, para un total de 8 veces que se presenta el atributo de joven. El total de conteos para cada clase por atributo se presenta en la tabla II.

Antes de determinar las probabilidades condicionales de los casos posibles, consideremos los siguientes ejemplos:

1. Se tiene un total de 4 pacientes jóvenes a quienes se les diagnosticó como no aptos para el uso de lentes, según la información recabada en la tabla II. El total de los pacientes diagnosticados de esta misma forma es 15 (esto es, 4 jóvenes más 5 pre-presbiópicos más 6 presbiópicos, como puede apreciarse en el recuadro "Edad" de la misma tabla). Así, asignaríamos la probabilidad de 4/15 a no usar lentes siendo joven. Hasta este punto no se ha encontrado ningún problema.
2. Ahora consideremos el caso de padecer astigmatismo y usar lentes suaves. La probabilidad asignada entonces es de 0/5, lo cual eliminaría la posibilidad de que se presente este caso si utilizáramos el mismo procedimiento que en el ejemplo anterior, pero en circunstancias reales no se puede asegurar *a priori* que nunca se presentará al consultorio una persona así. Este problema surge al asignar probabilidades como se describió en el ejemplo anterior.

Para eliminar el problema referido en el ejemplo 2, se utilizó el *estimador de Laplace*, el cual considera que cada valor del atributo v_i, es equiprobable respecto a todos los posibles valores de ese atributo y que existe una constante μ tal que $\mu = |v|$ ($|v|$ es el número posible de distintos valores que puede tomar dicho atributo). Considerando esto, la probabilidad para cada combinación está dada por

$$P[c_j | v_i] = \frac{(casos \quad favorables + \mu p_i)}{(casos \quad totales + \mu)}, \quad (8)$$

Tabla III. Tabla de probabilidades

Probabilidades

Edad

	ninguno	suave	duro
joven	5/18	3/8	3/7
pre-presbiópico	1/3	3/8	2/7
presbiópico	7/18	1/4	2/7

Padecimiento

	ninguno	suave	duro
hipermétrope	9/17	4/7	1/3
miope	8/17	3/7	2/3

Astigmatismo

	ninguno	suave	duro
si	9/17	1/7	5/6
no	8/17	6/7	1/6

Lagrimeo

	ninguno	suave	duro
normal	5/17	6/7	5/6
reducido	12/17	1/7	1/6

Tipo de lente

ninguno	suave	duro
5/8	**5/24**	**1/6**

$$i \in \{Atributos\}$$
$$j \in \{Clases\}$$

donde p_i, es la probabilidad de v_i, para todos los valores de c_j y v_i.

Recalculando la probabilidad de los ejemplos anteriores, tenemos:

1. El parámetro $|v|$, los posibles valores del atributo "Edad", es 3. Se considera a priori que los tres valores de este atributo son igualmente probables, de forma que la probabilidad de "ser joven" es 1/3. La probabilidad condicional de no usar lentes siendo joven utilizando el estimador de Laplace (ecuación 8) es (4 + 3*1/3) / (15 + 3) = 5/18.
2. Al reevaluar la probabilidad de padecer astigmatismo y usar lentes suaves mediante el estimador de Laplace, tenemos que el número de casos favorables = 0; μ = 2, {si, no}; *pi*= 1/2, por ser equiprobable; mientras que los casos totales son 5, así obtenemos la probabilidad de 1/7.

Aplicando el estimador para cada una de las combinaciones, se eliminan todos los ceros que se presentan y obtenemos los resultados que se muestran en la tabla III.

Se puede ver que cada combinación de atributos tiene una probabilidad asignada diferente de 0, lo cual permite continuar con el análisis usando la ecuación (7). Los resultados se muestran en la última columna de la tabla I, dentro de la columna de porcentajes (Pct.).

Aunque la tabla I ya contiene todas las posibles combinaciones de atributos (excluyendo la clase) y el diagnóstico podría leerse directamente de ella, se supondrá para ilustrar la aplicación práctica del método que se presenta un nuevo paciente con las características que se muestran en la tabla IV.

Para pronosticar el tipo de lente recomendado para un paciente con estas características, se multiplican las probabilidades de ser joven y usar lente suave (3/8), ser hipermétrope y usar lente suave (4/7), no padecer astigmatismo y usar lente suave (6/7), tener lagrimeo normal y usar lente suave (6/7), y la probabilidad a priori de utilizar lente suave (5/24). Esto da como resultado (2/61). Se realiza el mismo cálculo respecto a los diagnósticos "no apto" y "lente duro", usando los mismos atributos (joven, hipermétrope, no padecer astigmatismo y lagrimeo normal). Una vez que se tienen los 3 resultados

Tabla IV. Paciente nuevo.

Edad	Padecimiento	Astigmatismo	Lagrimeo	Tipo de lente	Pct.
joven	hipermétrope	no	normal	suave	¿?

(5/393, 2/61 y 2/605) se normalizan dividiendo cada resultado entre las suma de los tres. Para el caso de lente suave, [(2/61) / (5/393 + 2/61 + 2/605)] da como resultado 67.17%. Este resultado muestra que la probabilidad de ser diagnosticado con lentes de contacto suaves siendo joven, hipermétrope, no padecer astigmatismo y presentar lagrimeo normal es de 67.17%, según la información contenida en la base de datos original.

Los resultados finales obtenidos mediante el clasificador Naïve Bayes mostrados en la tabla I, última columna, son mayores que 33.33% (43.21% en el peor de los casos). Esto significa que el tipo de lente asignado tiene preferencia respecto a los otros 2 tipos de lentes no asignados. Por ejemplo, para un paciente con las características de la instancia 7, se favorece la decisión de no prescribir lentes de contacto, respecto a prescribir lentes duros o suaves. Esto tiene sentido, pues las clases (esto es, el tipo de lente de contacto) de las instancias ya presentes en la base de datos original tienden a prevalecer sobre instancias nuevas. En otras palabras, el método tiende a favorecer instancias semejantes a las que ya existen. Sin embargo, la utilidad real del método está en que permite asignar probabilidades cuando aparecen instancias totalmente nuevas o desconocidas a partir de la información existente.

COMENTARIOS FINALES

En este artículo se presentó un panorama de las técnicas y aplicaciones de la minería de datos y se mostró en detalle el funcionamiento de una de sus técnicas: el clasificador Naïve Bayes, a través de un caso de estudio de diagnóstico de lentes de contacto. Con este método se obtuvieron las probabilidades de diagnóstico para cada tipo de lente, dependiendo de la combinación de los atributos de cada instancia. El método Naïve Bayes tiene varias ventajas, como el hacer predicciones a partir de datos parciales y el ser rápido. Entre sus principales desventajas está el no ser apto para el manejo de variables aleatorias continuas.

El caso de estudio presentado solo toma en cuenta para la clasificación 4 atributos, mostrando un desempeño satisfactorio. En problemas reales el número de atributos a considerar puede ser mayor, sin implicar esto la degradación en el desempeño del clasificador Naïve Bayes.

La minería de datos tiene múltiples aplicaciones, varias de las cuales se mencionaron en el desarrollo de este artículo. El campo de aplicación de la minería de datos es muy extenso y el buen uso de las técnicas ya existentes puede llevar a un ahorro sustancial de recursos para la obtención de conocimiento a partir de bases de datos.

REFERENCIAS

1. Introduction to Data Mining and Knowledge Discovery, Tercera Edición, Two Crows Corporation, 1999.
2. K. Chye, T. Chin, G. Peng. Credit scoring using data mining techniques. Singapore Management Review. Tomo 26, No. 2: 25-47, 2004.
3. T. Chin. Data mining. American Medical News. Tomo 46 No. 46:19-20, 2003.
4. W. DuMochel, R. O'Neill, A. Szarfman, T. Louis. Bayesian data mining in large frequency tables, with an aplication to the FDA spontaneous repoting system. The American Statistician. Tomo 53, No. 3: 177-202, 1999.
5. C. Papatheodorou, S. Kapidakis, M. Sfakakis, A. Vassiliou. Mining user communities in digital libraries. Information Technology and Libraries. Tomo 22, No. 24:152-157, 2003.
6. J. Bagner, A. Evansburg, V. Watson, J. Welch. Senators seek on DoD mining of personal data. Intellectual Property & Technology Law Journal. Tomo 15, No. 5: 19-20, 2003.
7. M. Halkidi, Y. Batistakis, M. Vazirgiannis. On Clustering Validation Techniques. Journal of Intelligent Information System. Tomo 17, No. 2-3: 107-145, 2001.
8. A. Strehl, J Ghosh. Relationship-Based Clustering and Visualization for High-Dimensional Data Mining. INFORMS Journal of computing. Tomo 15, No. 2: 208-230, 2003.
9. R. Honda, S. Wang, T. Kikuchi, O. Konishi. Mining of Moving Objects from Time-Series Images and its Application to Satellite Weather Imagery. Journal of Intelligent Information System. Tomo 19 No. 1: 79-93, 2002.

10. B. Barber, H. Hamilton. Parametric Algorithms for Mining Share Frequent Itemsets. Journal of Intelligent Information Systems. Tomo 16, No. 3: 277-293, 2irm-Level Empirical Research. Information System Research. Tomo 14, No. 2: 127-145, 2003.

13. J. Malinen, M. Maltamo. E. Verkasalo. Predicting the internal quality and value of Norway spruce trees by using two non-parametric nearest neighbor methods. Forest Products Journal. Tomo 53, No. 4:85-94, 2003.

14. N. Canter. Development of a lean, green automobile. Tribology & Lubrication Technology. Tomo 60, No. 7: 15-16, 2004.

15. G. Elwyn, A. Edwards, M. Eccles, D. Rovner. Decision analysis in patient care. The Lancet. Tomo 358, No. 9281: 571-574, 2001.

16. S. Liong, Ch. Sivapragasam. Flood stage forecasting with support vector machines. Journal of the American Water Resources Association. Tomo 38, No. 1: 173-186, 2002.

17. Ross S. Introduction to Probability Models, 7ª. Edición, Harcourt Academic Press, 2000.