


Death Row Information in Texas Department of Criminal Justice

Reglas de Asociación Aplicada a Crimenes

Karen Jaqueline Reyes Flores
Lic.Tecnologías para la Información en Ciencias
UNAM, ENES CAMPUS MORELIA
karenreyes346@gmail.com

Alejandra Guadalupe Esquivel Guillén
Lic.Tecnologías para la Información en Ciencias
UNAM, ENES CAMPUS MORELIA
alejandraeg9899@gmail.com




Texas Department of Criminal Justice

"...to provide public safety, promote positive change in offender behavior, reintegrate offenders into society, and assist victims of crime."

Death Row Information

Offender Information



Name	Cardenas, Ruben Ramirez
TDCJ Number	999275
Date of Birth	04/07/1970
Date Received	07/29/1998
Age (when Received)	28
Education Level (Highest Grade Completed)	11
Date of Offense	02/22/1997
Age (at the time of Offense)	26
County	Hidalgo
Race	Hispanic
Gender	Male
Hair Color	Black
Height	5' 6"
Weight	204
Eye Color	Brown
Native County	Guanajuato
Native State	Mexico

Figura 1: Ficha Informativa del departamento de justicia de Texas.

ABSTRACT

En el estado de Texas el registro de armas de fuego no es requerido, aunque para la portabilidad de una, por ley se requiere una identificación emitida por el estado, esto se aplica a todos los compradores. La falta del registro de armas de fuego en el estado dificulta informar con exactitud cuántas personas poseen una. Existen 4,900 comerciantes de armas en Texas registrados [?].

En pleno siglo XXI existen prejuicios respecto a los delincuentes, juzgando a estos por la raza, el nivel académico logrado y los lugares de origen.

El objetivo en el presente trabajo es realizar un análisis en el cual se tratará de encontrar cierto patrón que nos lleve a determinar que no es necesario ser de tal o cual raza para cometer un delito, mostrando que los prejuicios no están del todo bien. La idea principal a trabajar es llevar a cabo una tarea de aprendizaje denominada Reglas de Asociación, esto con la finalidad de saber cómo es que

con los datos que se obtuvieron sobre una base de datos criminal del estado de Texas están relacionados.

Por ejemplo: Deseábamos saber si el grado de escolaridad, la raza y el lugar de origen influyen en los crímenes.

KEYWORDS

Algoritmo A priori, Minería de Datos, Reglas de Asociación, Discretización

ACM Reference format:

Karen Jaqueline Reyes Flores and Alejandra Guadalupe Esquivel Guillén. 2017. Death Row Information in Texas Department of Criminal Justice Reglas de Asociación Aplicada a Crimenes. In *Proceedings of UNAM, CAMPUS MORELIA, ENES*, 4 pages.
<https://doi.org/10.1145/nnnnnnnn.nnnnnnn>

1. INTRODUCCIÓN

En minería de datos y aprendizaje automático, las reglas de asociación se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos. El problema de minería de reglas de asociación se define como:

Sea $I = i_1, i_2, \dots, i_n$ un conjunto de n atributos binarios llamados ítems. Sea $D = t_1, t_2, \dots, t_m$ un conjunto de transacciones almacenadas en un conjunto de datos. Cada transacción en D tiene un ID

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
UNAM, ENES, CAMPUS MORELIA

© 2017 Copyright held by the owner/author(s).
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
<https://doi.org/10.1145/nnnnnnnn.nnnnnnn>

(identificador) único y contiene un subconjunto de items de I . Una regla se define como una implicación de la forma: $X \Rightarrow Y$ Donde:

$$X, Y \subseteq I$$

$$X \cap Y = \emptyset$$

Los conjuntos de items X y Y se denominan respectivamente "antecedente"(o parte izquierda) y "consecuente"(o parte derecha) de la regla. La ventaja de los algoritmos de reglas de asociación sobre los algoritmos de árboles de decisión es que las asociaciones pueden existir entre cualquiera de los atributos. Un algoritmo de árbol de decisión generará reglas con una única conclusión, mientras que los algoritmos de asociación tratan de buscar muchas reglas, cada una de las cuales puede tener una conclusión diferente. El algoritmo Apriori extrae un conjunto de reglas de los datos y destaca aquellas reglas con un mayor contenido de información.

Precisamente es lo que deseamos encontrar en nuestro análisis, un conjunto de reglas que nos generen información respecto a los criminales y nos den patrones comunes dentro de los asesinos registrados en la base de datos.

Seria muy interesante poderlo aplicar para saber las tendencias que existen en los crímenes de asesinatos en Texas y así tomar medidas de precaución tal vez, en tutelares de menores o en las mismas escuelas.

Death Row Information

Executed Offenders

Execution	Link	Link	Last Name	First Name	TDCJ Number	Age	Date	Race	County
545	Offender Information	Last Statement	Cardenas	Ruben	999275	47	11/09/2017	Hispanic	Hidalgo
544	Offender Information	Last Statement	Pruett	Robert	999411	38	10/12/2017	White	Bee
543	Offender Information	Last Statement	Preyor	Tachin	999494	46	7/27/2017	Black	Bexar
542	Offender Information	Last Statement	Bighy	James	997	61	3/14/2017	White	Tarrant
541	Offender Information	Last Statement	Ruiz	Rolando	999145	44	3/07/2017	Hispanic	Bexar
540	Offender Information	Last Statement	Edwards	Terry	999463	43	1/26/2017	Black	Dallas
539	Offender Information	Last Statement	Wilkins	Christopher	999533	48	01/11/2017	White	Tarrant
538	Offender Information	Last Statement	Fuller	Barney	999401	58	10/05/2016	White	Houston
537	Offender Information	Last Statement	Vasquez	Pablo	999297	38	04/06/2016	Hispanic	Hidalgo
536	Offender Information	Last Statement	Ward	Adam	999525	33	03/22/2016	White	Hunt
535	Offender Information	Last Statement	Westbrook	Coy	999281	58	03/09/2016	White	Harris

Figura 2: Listado de ejecuciones (13/oct/2017)

2. DATASET

Del Departamento de Justicia de Texas [?] se obtuvo un listado de las ejecuciones que se realizaron desde el 7 de Diciembre de 1982 hasta 9 de Noviembre del 2017 en el cual se incluyen las atributos descritos a continuación:

El hipervínculo Offender Information en el cuadro 1 vincula la información que se muestra en la figura 2, de la cual se decidió considerar únicamente los atributos mostrados en el cuadro 2 debido a que estos son relevantes para nuestro estudio.

3. DESARROLLO

Se desarrolló una herramienta que descargara los vínculos de la tabla de Information Offender los cuales venían con formato html y se convirtieron a csv, de ahí se eliminaron las columnas innecesarias y con la ayuda de open refine se editaron y eliminaron las palabras

Atributo	Significado	Tipo de Dato
Execution	Número de ejecuciones	numérico
Offender Information	Información detallada	hipervínculo
Last Statement	Últimas palabras	hipervínculo
Last Name	Apellido	texto
First Name	Nombre	texto
TDCJ Number	Número de expediente	numérico
Age	Edad ejecución	numérico
Date	Fecha de ejecución	numérico
Race	Raza	texto
Country	Lugar donde se cometió el crimen	texto

Cuadro 1: Listado de atributos del Departamento de Justicia

Name Cardenas, Ruben Ramirez
 TDCJ Number 999275
 Date of Birth 04/07/1970
 Date Received 07/29/1998
 Age (when Received) 28
 Education Level (Highest Grade Completed) 11
 Date of Offense 02/22/1997
 Age (at the time of Offense) 26
 County Hidalgo
 Race Hispanic
 Gender Male
 Hair Color Black
 Height 5' 6"
 Weight 204
 Eye Color Brown
 Native County Guajalajara
 Native State Mexico

Figura 3: Atributo: Offender Information

Atributo
Education Level(Highest Grade Completed)
Gender
Native Country
Native State

Cuadro 2: Atributos seleccionados

no necesarias, además se convirtieron a atributos numéricos. El proceso anterior generó 170 registros, además se encontraron 381 con formato jpg los cuales fueron procesados manualmente. Aquellos registros cuya información no quedó completa fueron marcados como NA, finalmente sin considerar los NA se obtuvieron un total de 492 registros viables de análisis.

El DataSet contaba con 492 registros, 2 atributos continuos (Edad y Escolaridad) y 4 categóricos (Género, Condado Nativo, Raza, Condado donde se cometió el delito). Para aplicar el algoritmo Apriori de esta investigación (se describe a continuación) se requería que el Nivel de Escolaridad fuera una variable categórica tomando los valores de Primaria en el rango de [3.. 6], Secundaria en el rango de [7.. 9], Prepa en el rango de [10.. 12] y Licenciatura en el rango de [13.. 20] también la Edad se requiere que sea una variable categórica tomando valores de Veinte en el rango de [20.. 30], Treinta en el rango de [31.. 40], Cuarenta en el rango de [41.. 50], Cincuenta en el rango de [51.. 60], Sesenta en el rango de [61.. 70].

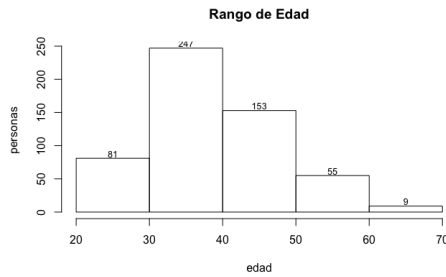


Figura 4: Histograma que muestra el rango de edades.

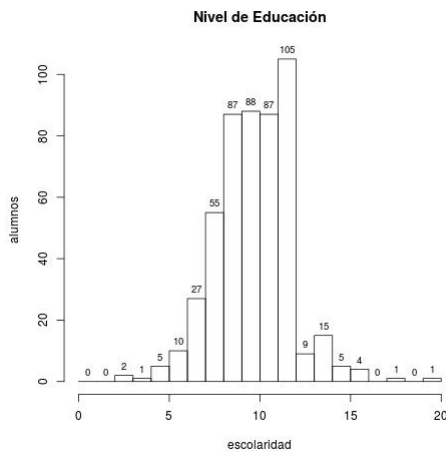


Figura 5: Histograma que muestra el rango del nivel de escolaridad.

3.1. A priori

El algoritmo a priori [?] es un algoritmo utilizado en minería de datos, sobre bases de datos transaccionales, que permite encontrar de forma eficiente "conjuntos de ítems frecuentes", los cuales sirven de base para generar reglas de asociación.

El algoritmo a priori recibe una matriz binaria y regresa los patrones frecuentes.

educationLevel	nativeCounty	Age	Race	County
prepa	Guanajuato	cincuenta	Hispanic	Hidalgo
secundaria	Harris	cuarenta	White	Bee
prepa	Bexar	cuarenta	Black	Bexar
prepa	Dallas	cuarenta	Black	Dallas
prepa	Harris County	cincuenta	White	Tarrant

Cuadro 3: Dataset resultante de eliminación de atributos continuos por discretos.

4. RESULTADOS

El resultado del análisis fue congruente con lo que observamos en la figura 4 y la figura 5 en donde si ordenamos por nivel de confianza y lift, lo que arroja la regla de asociación nos dicen que lo más frecuente es que el criminal tenga 30 años al momento del crimen y su grado de escolaridad sea secundaria que corresponde a un rango de entre [7.. 9 años], la segunda regla nos dice que tiene la escolaridad antes mencionada y es un hombre blanco, la tercera es que tenga la misma edad antes mencionada y sea de Harris, Texas.

antecedants	consequents	support	confidence	lift
(treinta, secundaria)	(Male)	0.168699	1.000000	1.00613
(White, secundaria)	(Male)	0.144309	1.000000	1.00613
(treinta, Harris)	(Male)	0.130081	1.000000	1.00613
(White, cuarenta)	(Male)	0.140244	1.000000	1.00613
(treinta, White)	(Male)	0.182927	1.000000	1.00613
(secundaria, Black)	(Male)	0.117886	1.000000	1.00613
(Harris, Black)	(Male)	0.142276	1.000000	1.00613
(cuarenta, prepa)	(Male)	0.156504	1.000000	1.00613
(treinta, prepa)	(Male)	0.247967	0.991803	0.99788
(White, prepa)	(Male)	0.225610	0.990991	0.99707
(treinta, Black)	(Male)	0.186992	0.989130	0.99519
(Harris, prepa)	(Male)	0.146341	0.986111	0.99216
(Black, prepa)	(Male)	0.247967	0.983607	0.98964
(treinta, Black, prepa)	(Male)	0.119919	0.983051	0.98908

Cuadro 4: Resultado de Reglas de Asociación ordenados de forma Descendente por Lift y Confianza

antecedants	consequents	support	confidence	lift
(secundaria)	(Male)	0.337398	1.000000	1.006135
(treinta)	(Male)	0.453252	0.995516	1.001623
(White)	(Male)	0.441057	0.995392	1.001498
(Black)	(Male)	0.384146	0.989418	0.995488
(prepa)	(Male)	0.558943	0.989091	0.995159

Cuadro 5: Atributos frecuentes e independientes junto con su nivel de soporte

4.1. Evaluación de Resultados

Para determinar que una regla es significativa se requiere de varias medidas como **el soporte, la confianza y el lift**:

El **soporte** de un conjunto de ítems se define como la proporción de transacciones en la base de datos que contiene dicho conjunto de ítems:

$$\text{Soporte}(X) = \frac{|X|}{|D|} \quad (1)$$

La **confianza** de una regla puede interpretarse como un estimador de $P(Y|X)P(Y|X)$, la probabilidad de encontrar la parte derecha de una regla condicionada a que se encuentre también la parte izquierda.:

$$\text{conf}(X \Rightarrow Y) = \frac{\text{sop}(X \cup Y)}{\text{sop}(X)} = \frac{|X \cup Y|}{|X|} \quad (2)$$

El **lift** expresa cuál es la proporción del soporte observado de un conjunto de productos respecto del soporte teórico de ese conjunto dado el supuesto de independencia.

$$\text{lift}(A \Rightarrow B) = \frac{P(B | A)}{P(B)} = \frac{P(A \cap B)}{P(A)P(B)} \quad (3)$$

Un valor de lift = 1 indica que ese conjunto aparece una cantidad de veces acorde a lo esperado bajo condiciones de independencia. Nuestros resultados fueron con a lo más lift de 1, nivel de confianza mayor o igual a 0.5 y soporte mayor a 0.15:

Estos datos nos dicen que nuestras reglas mostraron asociaciones de atributos independientes, en el cual aparecían en un 50 % juntas como regla, todo lo mencionado se muestra en el cuadro 4.

5. CONCLUSIONES

A contrario de lo que se piensa comúnmente en los estados unidos, las reglas de asociación nos demostraron que es más común que los criminales que fueron asesinados mediante la pena de muerte, son personas de raza blanca con un 44 % en nivel de soporte en contradicción de la raza negra con un 38 % de soporte. Y respecto al grado de escolaridad, los criminales con mayor nivel de soporte fueron los que estudiaron la preparatoria (9-12 años) con un nivel de 55 % y los que estudiaron la secundaria fue de 33 %.

Creemos que sería de gran ayuda este análisis de patrones o reglas frecuentes para localizar ciertos focos rojos en los tutelares para menores de Texas e intentar prevenir más penas de muerte en este estado, y reducir la cantidad de crímenes que en estos últimos años ha ido en incremento en dicha zona.



ESCUELA
NACIONAL
DE ESTUDIOS
SUPERIORES

UNIDAD MORELIA

