

DETERMINACIÓN DE PERFILES DELICTIVOS EN EL ESTADO DE JALISCO UTILIZANDO WEKA A TRAVÉS DE MINERÍA DE DATOS

M. Amador García¹
M. C. Baltazar Martínez²
M. C. Rodríguez Camacho³
C. Ruiz Perales⁴

RESUMEN

El uso de la minería de datos en diversos aspectos es muy útil para identificar patrones así como para ayudar en la toma de decisiones, y la perfilación de delitos no es la excepción, en este trabajo se presenta el análisis realizado en la Dirección de Prevención Social en la Fiscalía General del Estado de Jalisco, en donde uno de los puntos centrales es minimizar los delitos, dentro de los cuales se encuentran: robo a vehículo, robo a casa habitación, robo a negocio y robo a transeúnte. El proceso seguido fue determinar primero qué tipos de delitos eran propicios para iniciar el análisis, lo cual dio pauta para continuar con exploración de los datos así como con la limpieza de los mismos, lo que permitió seguir con el análisis por medio de la herramienta WEKA, haciendo uso de los algoritmos Random Forest, K Means y A Priori.

La información obtenida fue precisa y contundente, permitió identificar que el rango de edades de victimarios está entre 27 y 40 años; se definió que Guadalajara y Zapopan son las ciudades más propensas a tener un mayor número de delitos; los hombres son más proclives a ser atacados. Lo anterior ayudó a generar estrategias para reducir este tipo de delitos. El análisis realizado no es extenso a todos los tipos de delitos por lo que se espera complementar estos resultados con los delitos faltantes, de esta manera tener un perfil más completo para definir estrategias que ayuden a minimizar los ataques con el objetivo de mantener la seguridad en la sociedad.

ABSTRACT

The use of data mining in several respects is very useful to identify patterns and to assist in decision-making, and the profiling of crime is no exception, in this work the analysis in the Social Prevention presents in the Attorney General of the State of Jalisco, where one of the central points is to minimize the crimes, among which are: car theft, burglary, robbery and burglary business passerby. The first proceedings was to determine what types of crimes were favorable to start the analysis, which gave guidelines for further exploration of the data as well as the cleaning of the same, allowing continue the analysis through WEKA tool, using algorithms Random Forest, K Means and A Priori.

The information obtained was precise and forceful, he identified the perpetrators age range is between 27 and 40 years; defined Guadalajara and Zapopan cities are more likely to have a greater number of crimes; men are more likely to be attacked. This helped generate strategies to reduce this type of crime. The analysis is extended to all types of offenses for which it is expected to complement these results with the missing crimes thus have a more complete profile to define strategies to minimize attacks in order to maintain security in society.

Referencias del autor(es)

¹ Docente Informática - Sistemas. Instituto Tecnológico de Rioverde, mony_951@hotmail.com

² Alumno. Instituto Tecnológico de Rioverde, moni_ca_13@hotmail.com

³ Alumno. Instituto Tecnológico de Rioverde, kassie_chan@hotmail.com

⁴ Alumno. Instituto Tecnológico de Rioverde, cristianperales24@hotmail.com

INTRODUCCIÓN

La seguridad en la sociedad es un tema muy importante a nivel nacional y el estado de Jalisco en la Dirección General de Prevención Social se trabaja de manera ardua para lograr reducir los crímenes y aumentar la seguridad en el estado. El punto en este sentido es que se trabaja en base a estadísticas que emite el Departamento de Estadística, las cuales son muy generales omitiendo información útil que ayude al cumplimiento del objetivo de la Dirección. Debido a lo anterior se planteó la Minería de Datos como una nueva opción para el análisis de la información que permita extender el conocimiento a nivel detallado, entendiendo que la minería de datos permite descubrir información relevante que a simple vista no se percibe.

Para determinar de forma precisa los perfiles de víctimas, victimarios, edades y ciudades se hizo uso de la herramienta WEKA, así como de los algoritmos Random Forest, Simple K Means y A Priori. Lo cual permitió definir estrategias que contribuyan a reducir los índices delictivos en el estado de Jalisco.

El presente paper se encuentra desglosado de la siguiente manera: se inicia con un resumen del análisis realizado e indicando los resultados obtenidos; se muestra la metodología seguida en donde se identifica los tipos de delitos a analizar con los algoritmos mencionados anteriormente para determinar los rangos de edades de los victimarios así como de las víctimas, ciudades más propensas en cuanto a delitos se refiere; la anterior información se ve reflejada en la sección de resultados, en la cual se muestra que las ciudades con mayor probabilidad de comisión de delitos es Guadalajara y Zapopan. Se finaliza con las conclusiones en donde se manifiesta el trabajo futuro y lo que se espera de esta etapa de trabajo.

METODOLOGÍA

En este trabajo, como primer punto fue identificar el uso de la minería de datos, la cual, como se menciona en Slimani 2014, es el resultado de la integración de diversas técnicas de análisis de datos, mediante las cuales busca comportamientos parecidos en los datos, los describe y los orienta hacia la obtención de algún beneficio. Además, contribuye para la extracción de información importante (Ankita B.R, Amit H.R. 2015), y útil, con el fin de mejorar la toma de decisiones.

El lugar donde se llevó a cabo el análisis fue dentro de la Dirección General de Prevención Social la cual pertenece a la Fiscalía General de Jalisco, para identificar las zonas más vulnerables en el Estado de Jalisco respecto a los diferentes tipos de delitos criminales y determinar el programa adecuado que se deberá impartir para con ello minimizar el número de delitos ocurridos en dichas zonas; de igual manera, descubrir información destacada, características como edad y sexo de víctima y victimario, en base a esto saber a qué población aplicar los programas y actividades preventivas.

El modelo de proceso de Minería de Datos utilizado fue el Cross Industry Estándar Process for Data Mining (CRISP-DM), que define una "estructura para proyectos de minería de datos y suministra la orientación para su ejecución" (Gallego and Careces, 2015). CRISP-DM consiste en un conjunto de tareas, las cuales se muestran en la figura 1.

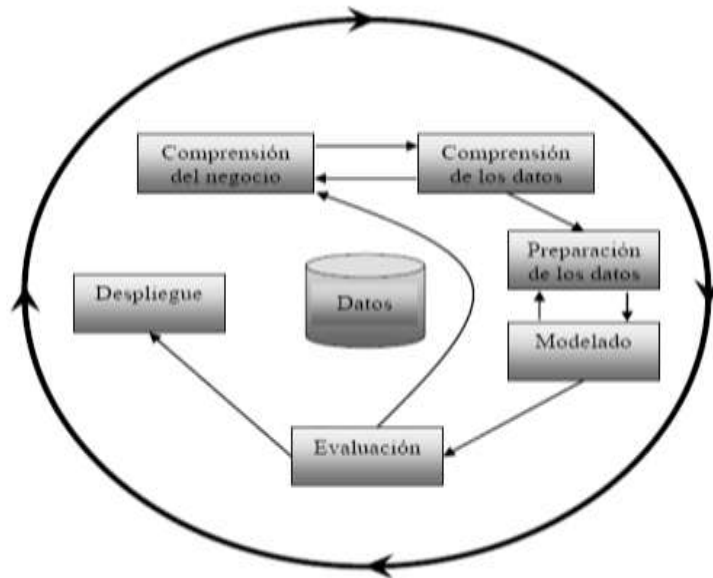


Figura 1 Etapas del Modelo CRISP-DM. Extraído de (Vanrell, 2011, página 22)

Se trabajó con 323 registros los cuales fueron proporcionados por la Dirección de Comunicación Social perteneciente a la Fiscalía General de Jalisco, los registros fueron recolectados de extractos de denuncias, ya que por protocolos de seguridad no se permite tener acceso a toda la información, se valoraron los meses de Diciembre 2014 y Enero 2015 de los siguientes delitos criminológicos:

1. Robo a vehículo.
2. Robo a casa habitación.
3. Robo a negocio.
4. Robo a transeúnte.

Una vez identificados los tipos de delitos a analizar se continuo con la exploración de datos, lo cual resulto en la selección de los siguientes atributos para cada uno de los delitos: delito, calle, número, colonia, hora, sexo de victimario, edad de victimario, arma utilizada, número de atacantes, objeto robado, monto, sexo de víctima y edad de víctima. Otro aspecto que se llevo a cabo fue la limpieza de los datos. De acuerdo a esto se llevo a cabo el análisis, experimentando con los algoritmos:

1. Random Forest, algoritmo de clasificación orientado a organizar grandes cantidades de datos con exactitud, es una combinación de árboles predictores (Sunil Kumar, G., Sirisha C.V.K, Kanaka Durga.R, Devi A., 2012) en la que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos. Mejora la precisión en la clasificación mediante

la incorporación de aleatoriedad en la construcción de cada clasificador individual.

2. Simple K Means, algoritmo de cuantificación vectorial que permite dividir varias observaciones en conglomerados en los que cada observación pertenece al grupo más cercano con la media, sirve como un prototipo de clúster (Kumar, V., Chauhan, H., Panwar, Dheeraj. 2013).
3. A Priori, ayuda a asociar, formar las relaciones y obtener los atributos ocultos dentro de los registros. Se utiliza para encontrar reglas de asociación en los registros, se basa en el conocimiento previo en conjuntos previamente clasificados con los algoritmos anteriores, esto con el objeto de reducir la búsqueda y aumentar la eficiencia en la misma (Karimuddin, M., M.Prudhvi Ravi Raja Reddy, 2014).

Lo anterior en la herramienta de WEKA. Para ello, primero se identificó, con Random Forest, las zonas más vulnerables, en cuanto a delitos en mención se refiere; en seguida con el algoritmo Simple K Means se generaron los clúster identificando los atributos, mencionados anteriormente, para confirmar lo obtenido.

RESULTADOS

El algoritmo de árbol de decisiones (Random Forest), permitió trabajar con los atributos, la variable estuvo en función del delito; el objetivo fue conocer las zonas con más alto impacto.

Los resultados obtenidos del análisis de los diferentes atributos, debido a que son varios se van a presentar solamente algunos, uno de ellos se puede ver en la figura 2, el cual se refiere a un resumen de los atributos seleccionados, con las ciudades con mayor impacto de delitos, utilizando el algoritmo Simple K Means, en donde se observa que resultaron dos clúster, y las ciudades identificadas fueron Guadalajara y Zapopan, con los delitos de robo a transeúnte y robo a negocio respectivamente.

Cluster centroids:			
Attribute	Full Data (323)	Cluster#	
		0 (240)	1 (83)
Delito	Robo Transeunte	Robo Transeunte	Robo Negocio
Lugar	Guadalajara	Guadalajara	Zapopan
Colonia	Oblatos	Oblatos	El Mante
Hora	23:00:00	23:00:00	13:00:00
Sexo_Victimario	Masculino	Masculino	Masculino
Edad_Victimario	'All'	'All'	'All'
Arma_Utilizada	Arma de fuego	Arma de fuego	Arma de fuego
Numero Atacantes	'(-inf-1.5]'	'(-inf-1.5]'	'(-inf-1.5]'
Sexo_victima	Masculino	Masculino	Masculino
Edad_Victima	'(24.5-inf)'	'(24.5-inf)'	'(24.5-inf)'
Clustered Instances:			
0	240 (74%)		
1	83 (26%)		

Figura 2 Generación de Cluster por el Algoritmo Simple K Means

Al utilizar el algoritmo A Priori se identificó información interesante, la cual se muestra en la figura 3, que contienen las características más sobresalientes en cuanto a la comisión de delitos se refiere.

- 1. Sexo_Victimario=Masculino 299==>Edad_Victimario='A11' 299 <conf:(1) lift:(1) lev:(0) [0] conv:(0)
- 2. Edad_Victima='(24.5-inf)' 255==> Edad_Victimario='A11' 255 <conf:(1) lift:(1) lev:(0) [0] conv:(0)
- 3. Numero_Atacantes='(-inf-1.5)' 253==> Edad_Victimario='A11' 253 <conf:(1) lift:(1) lev:(0) [0] conv:(0)
- 4. Sexo_Victimario=Masculino Edad_Victima='(24.5-inf)' 239==>Edad_Victimario='A11' 239 <conf:(1) lift:(1) lev:(0) [0] conv:(0)
- 5. Sexo_Victimario=Masculino Numero_Atacantes='(-inf-1.5)' 234==> Edad_Victimario='A11' 234 <conf:(1) lift:(1) lev:(0) [0] conv:(0)
- 6. Edad_Victima='(24.5-inf)' 255==> Sexo_Victimario=Masculino 239 <conf:(0.94) lift:(1.01) lev:(0.01) [2] conv:(1.11)
- 7. Edad_Victimario='A11' Edad_Victima='(24.5-inf)' 255==> Sexo_Victimario=Masculino 239 <conf:(0.94) lift:(1.01) lev:(0.01) [2] conv:(1.11)
- 8. Edad_Victima='(24.5-inf)' 255==> Sexo_Victimario=Masculino Edad_Victimario='A11' 239 <conf:(0.94) lift:(1.01) lev:(0.01) [2] conv:(1.11)
- 9. Edad_Victimario='A11' 323==> Sexo_Victimario=Masculino 299 <conf:(0.93) lift:(1) lev:(0) [0] conv:(0.96)
- 10. Numero_Atacantes='(-inf-1.5)' 253==> Sexo_Victimario=Masculino 234 <conf:(0.92) lift:(1) lev:(0) [0] conv:(0.94)

Figura 3. Características principales de Algoritmo Apriori

Como se aprecia en la figura 3, la mayoría de los que cometen delitos son del sexo masculino con rango de edades entre 18 a 50 años con una confiabilidad del 100%. También se observa que la edad de las víctimas que son atacadas en su mayoría están entre 24 a 73 años y que los delitos son cometidos por hombres victimarios entre 18 a 50 años con una confiabilidad del 100%. A lo anterior se le agrega que los delitos son cometidos por menos de 2 personas y cuando ataca solamente un victimario en su mayoría es hombre con una confiabilidad del 0.94%.

Los

Delito	Perfil Victimario	Perfil Víctima	Horas	Lugar
--------	-------------------	----------------	-------	-------

resultados por delito, los cuales ayudarán a mejorar estrategias de la Dirección de Prevención Social perteneciente a la Fiscalía General de Jalisco, se muestran en la tabla 1.

Los

Robo a Vehículo	+27 a 30 Masculino -43 a 46 Femenino	+33 a 40 Masculino -59 a 66 Femenino	12:00 - 19:00	+Guadalajara -Tlaquepaque
Robo a Casa Habitación	+27 a 30 Masculino -12 a 34 Femenino	+27 a 33 Masculino Femenino	13:00 - 17:30	+Zapopan -Tonalá
Robo a Negocio	+27 a 40 Masculino -40 a 43 Femenino	+27 a 33 Masculino -59 a 66 Femenino	10:00 - 12:00	+Guadalajara -Puerto Vallarta
Robo a Transeúnte	+27 a 30 Masculino -21 a 24 Femenino	+20 a 27 Masculino -66 a 73 Femenino	20:00 - 23:00	+Guadalajara -Chapala

Tabla 1 Evaluación de resultados

algoritmos utilizados realizados arrojaron un porcentaje de confianza del 99.0712%.

CONCLUSIONES

Se propuso una nueva manera de obtener la información en tiempo real y así poder mejorar la tomar decisiones, observando que la aplicación de la minería de datos tuvo éxito dentro de la Dirección General de Prevención Social, brindando resultados destinados en la generación de estrategias más exactas para la prevención de los delitos ocurridos en el estado de Jalisco, reestructuración de los programas de prevención en base a características como edad y sexo de víctimas y victimarios.

Como se puede apreciar, la minería de datos que es una buena opción de análisis de datos para diversos campos, por otro lado puede resultar conveniente para compartir información entre diferentes áreas.

Las recomendaciones sugeridas son las siguientes: expandir el análisis de la información criminal de un año o más de las denuncias del estado de Jalisco, con el fin de reducir la delincuencia contando con información precisa y de tiempo real así como la generación de estrategias para la resolución de problemas comúnmente relacionados con la falta de información asociativa.

Otro aspecto interesante a destacar, es que el uso de una herramienta como lo es la minería de datos no sólo sirve para ser aplicada al departamento ya que se podría aplicar en los distintos departamentos pertenecientes a la Fiscalía General del Estado de Jalisco, como lo son el departamento de portación de armas, jurisdicción, entre otros. También es importante

mencionar que como trabajo futuro se puede trabajar en el desarrollo de un software, el cual almacene todos los resultados obtenidos a través de la herramienta Weka y con ello poder generar las estrategias efectivas.

BIBLIOGRAFÍA

Slimani, T., 2014, *RST Aproach for efficient CARs Mining*. Bonfring International Journal of Data Mining, Vol. 4, No. 4, November 2014. Pags. 34-40. DOI: 10.9756/BIJDM.10365

Ankita B.R, Amit H.R. 2015. *A survey on mining uncertain frequent item set effectively using pattern growth approach*. International Journal of Engineering Sciences & Research. Pags. 42-47.

Gallego Gallego, M., Hernández Cáceres J., 2015, *Identificación de factores que permitan potencializar el éxito de proyectos de desarrollo de software*. Red de Revistas Científicas de América Latina, el Caribe, España y Portugal. Scientia et Technica. Vol. 20, No. 1, Marzo de 2015. Pags. 70-80.

Sunil Kumar, G., Sirisha C.V.K, Kanaka Durga.R, Devi A., 2012, *Robust Preprocessing and Random Forests Technique for Network Probe Anomaly Detection*. International Journal of Soft Computing and Engineering (IJSCE). Volume-1, Issue-6, January 2012. Pags. 391-395.

Kumar, V., Chauhan, H., Panwar, Dheeraj. 2013. *K-Means Clustering Approach to Analyze NSL-KDD Intrusion Detection Dataset*. International Journal of Soft Computing and Engineering (IJSCE). Volume-3, Issue-4, September 2013. Pags. 1-4

Karimuddin, M., M.Prudhvi Ravi Raja Reddy. 2014. *Apriori Algorithm for Vertical Association Rule Mining*. International Journal of Engineering Sciences & Research Technology. Pags. 659-665

Vanrell, J. A. 2011. *Un modelo de procesos para proyectos de explotación de información*. Argentina. Recuperado el 07 de 06 de 2015, de UNLa:
<http://sistemas.unla.edu.ar/sistemas/gisi/tesis/vanrell-tesisdemagister.pdf>