



Ciencias Jurídicas

Artículo Científico

Detección automática del nivel de crimen basado en el análisis de puntos calientes en la ciudad de Guayaquil

Automatic crime level detection based on the analysis of hot spots in the city of Guayaquil

Detecção automática de nível de criminalidade com base na análise de pontos quentes na cidade de Guayaquil

Juan C. García-Plúa ^I
juan.garciapl@ug.edu.ec

Wellington R. Villota-Oyarvide ^{II}
wellington.villotao@ug.edu.ec

Julio G. Litardo-Unda ^{III}
julio.litardou@ug.edu.ec

Recibido: 30 de enero de 2017 * **Corregido:** 2 de febrero de 2017 * **Aceptado:** 1 mayo de 2017

- ^{I.} Cursando PhD en Ciencias de la Computación Aplicada; Magister en Sistemas de Información Gerencial; Magister en Administración de Empresas; Diplomado Superior en Administración de Empresas; Licenciado en Sistemas de Información; Escuela Superior Politécnica del Litoral, Ecuador
- ^{II.} Magister en Diseño y Evaluación de Modelos Educativos; Diploma Superior en Diseño y Gestión Curricular; Diploma Superior en Modelos Educativos; Licenciado en Sistemas de Información; Universidad de Guayaquil, Ecuador.
- ^{III.} Magister en Sistemas de Información Gerencial; Ingeniero en Sistemas Computacionales; Universidad de Guayaquil, Ecuador.

Resumen.

La detección temprana de los lugares del delito es importante para que las ciudades puedan tomar decisiones preventivas que permitan aumentar la percepción de la seguridad pública. En este contexto, tomando datos históricos de crímenes ocurridos en la ciudad de Guayaquil se realizó exitosamente un proceso de clasificación que mide a un determinado punto geográfico en cuatro niveles de delito: extremo, alto, moderado y bajo. Proporcionando sólo la dirección de un lugar en la ciudad, esta es fácilmente convertida en coordenadas polares expresadas como latitud y longitud para predecir automáticamente el nivel de delito de ese lugar con un 93 por ciento de precisión.

Palabras Clave: Crimen; clasificación; redes neuronales; regresión.

Abstract.

Early detection of crime scenes is important for cities to take preventive decisions to increase the perception of public safety. In this context, taking historical data of crimes in the city of Guayaquil it was successfully carried out a classification process that measures a certain geographical point in four levels of crime: extreme, high, moderate and low. Providing only the address of a place in town, this is easily converted into polar coordinates is expressed as latitude and longitude to automatically predict the level of crime that place with 93 percent accuracy.

Keywords: Crime; classification; neural networks; regression.

Resumo.

A detecção precoce de cenas de crime é importante para que as cidades tomem decisões preventivas para aumentar a percepção de segurança pública. Neste contexto, levando dados históricos de crimes na cidade de Guayaquil foi realizado com sucesso um processo de classificação que mede um determinado ponto geográfico em quatro níveis de criminalidade: extrema, alta, moderada e baixa. Fornecer apenas o endereço de um lugar na cidade, isto é facilmente convertido em coordenadas polares é expressa como latitude e longitude para prever automaticamente o nível de crime que lugar com precisão de 93 por cento.

Palavras chave: Crime; classificação; redes neurais; regressão.

Introducción.

El rápido progreso de las ciudades de todo el mundo está estimulando las transformaciones económicas y sociales. Este fenómeno mejora la vida de las personas, pero también trae problemas como el aumento de la delincuencia y las presiones a los presupuestos de la ciudad (Cesario et al. 2016). A nivel mundial, se espera que la población urbana crezca 2,86 billones en 2000 a 4,98 mil millones en 2030 (Tayebi et al. 2014); como resultado más del 66 por ciento de la población mundial se prevé que sea urbana (United Nations 2014). Del mismo modo en Ecuador la proporción de población urbana está aumentando desde 64 por ciento en 2014 hasta el 75 por ciento en 2050.

Las actividades criminales están distribuidas aleatoriamente sobre un espacio geográfico, sin embargo, tienden a concentrarse en ciertos lugares por razones como escasez de vigilancia, o una relación entre la víctima y el delincuente (Chainey et al. 2008; Mohamad No et al. 2015). Un punto caliente de crimen es un área con un número de incidentes de delito por arriba del promedio. Esto denota que las áreas debajo del promedio se consideran puntos fríos (Zhang et al. 2015). Además, se puede advertir que hay puntos más calientes que otros, por el grado de desviación desde el promedio.

Importancia del problema

Los problemas de seguridad pública aumentan a medida que crecen las ciudades, convirtiéndose en un problema social que afecta a la calidad de vida de la población (Bogomolov et al. 2014). Esta no es la excepción para la ciudad de Guayaquil, que es uno de los puertos más activos en América del Sur (Taylor et al. 2012), también en los últimos años se ha convertido en un importante destino para el turismo, sin embargo, hay muchas dificultades, como la necesidad de mejorar la percepción de seguridad.

Investigaciones criminológicas sugieren la importancia de identificar los lugares donde ocurren los crímenes, para el control efectivo y proactivo de la seguridad pública (Zhang & Goncalves 2016; Azeez & Aravindhar 2015). Un lugar es una calle, esquina, dirección, edificio o segmento de la calle (Eck & Weisburd 1995; Davies & Bishop 2013). La detección de esos lugares ayudará a las autoridades y los ciudadanos a sensibilizar y prevenir a la población (Kadar & Cvijikj 2014).

El objetivo de esta investigación es diseñar un método de clasificación para predecir de forma automática si un lugar específico en la ciudad de Guayaquil será un centro de delincuencia o no. El paso inicial de este estudio consiste en identificar un conjunto de datos geospaciales sobre crímenes pasados en la ciudad de Guayaquil, para visualizar las áreas de puntos calientes en la ciudad y finalmente entrenar dos clasificadores para encontrar la mejor precisión al momento de predecir si un lugar es un punto caliente.

Materiales y métodos.

La metodología aplicada en esta investigación se divide en las siguientes fases:

- 1) Preprocesamiento de datos
- 2) Exploración de datos
- 3) Diseño del Modelo
- 4) Clasificación
- 5) Evaluación del modelo

El área de estudio es de aproximadamente 344,5 km² corresponden a la ciudad de Guayaquil, con una población de 3,5 millones de personas en el área metropolitana. El conjunto de datos crimen fue proporcionado por la Facultad de Ciencias Naturales y Matemáticas en la

Detección automática del nivel de crimen basado en el análisis de puntos calientes en la ciudad de Guayaquil

Universidad ESPO. Hay 11.165 delitos denunciados por las víctimas en 2014 año, sin embargo, hay algunos registros que han desaparecidos los errores de datos o error tipográfico.

indice	delito	fechaev	horaev	calle
1	1HOMICIDIO	1/1/2014	16:00	LOMAS DE LA FLORIDA
2	1HOMICIDIO	29/12/2013	18:00	SOPENA
3	1HOMICIDIO	2/1/2014	23:00	AV. LAS EXCLUSAS Y DOMINGO COMIN
4	1HOMICIDIO	5/12/2013	0:55	CENTRO DE PRIVACION DE LIBERTAD ZONAL 8
5	1HOMICIDIO	13/1/2014	12:50	CALLE RUMICHACA

Figura N° 1.- Datos originales

Los registros de crímenes donde no hay ubicación fueron ignorados. También se normalizaron las direcciones, por ejemplo: "CDLA", "CIUD", se convirtió a "CIUDADELA". Se utilizó un programa escrito en Python que a través de la API de geolocalización de Google (Google Inc. 2017) se obtiene la latitud / longitud a partir de una dirección. Después de este proceso se obtiene un mejor conjunto de datos con 7.277 registros de criminalidad, incluyendo la latitud y la longitud necesaria para realizar el análisis espacial y clasificación:

	calle	latitudo	longitudo
0	LOMAS DE LA FLORIDA	-2.135384	-79.962232
1	SOPENA	-2.244973	-79.899679
2	AV. LAS EXCLUSAS Y DOMINGO COMIN	-2.250411	-79.890124
3	CENTRO DE PRIVACION DE LIBERTAD ZONAL 8	-2.047480	-79.945152
4	CALLE RUMICHACA	-2.195975	-79.886863

Figura N° 2.- Datos depurados con latitud y longitud

La exploración de los datos se refleja en un mapa de calor que permite ver la distribución de datos sobre la ciudad. Esto revela que las coordenadas X, Y que fueron calculadas en el proceso de localización anterior son correctas. El color rojo en la figura 3 muestra los lugares que concentra un alto nivel de delincuencia. Se estableció un umbral inicial basado en la distancia de Euclides con un

Detección automática del nivel de crimen basado en el análisis de puntos calientes en la ciudad de Guayaquil

valor por defecto que garantiza que cada punto tenga un mínimo de vecindad (Tollenaar & van der Heijden 2013).

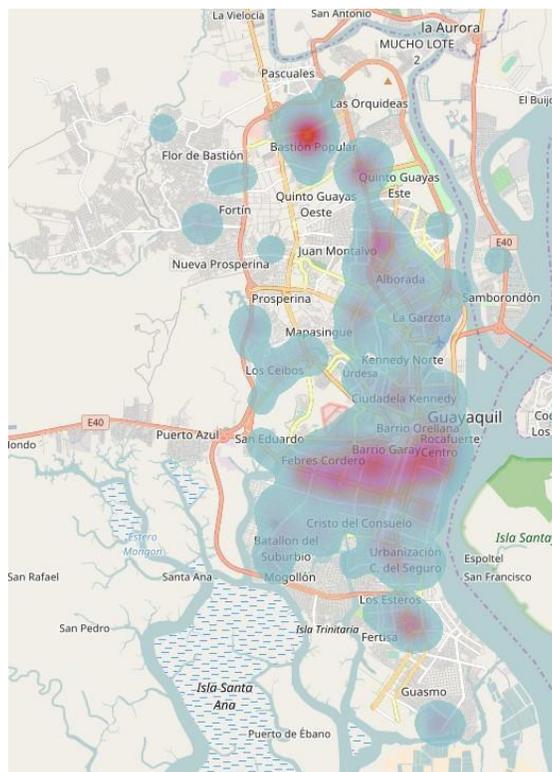


Figura N° 3.- Mapa de calor de distribución de crimen en la ciudad de Guayaquil

Los registros de delitos deben agruparse en clusters que corresponden a un nivel extremo, alto, moderado y bajo de actividad criminal. Los umbrales que definen estas etiquetas se basan en la frecuencia de la actividad de la delincuencia en un clúster específico. Estos umbrales se definen tomando como referencia la frecuencia de delitos ocurridos cada cluster.

Nivel	Rango
Extremo	<15
Alto	≥ 15 AND <25
Moderado	≥ 25 AND <40
Bajo	≥ 40

Tabla N° 1.- Umbrales que definen a un punto caliente

En muchas investigaciones se ha usado dos técnicas de cluster para la agrupación espacial que son K-Means (Imas Sukaesih Sitanggang; Tsamrul Fuad; Annisa n.d.; Chandra et al. 2008) y el agrupamiento espacial basado en densidad de aplicaciones con ruido (DBScan) (Vural & Gök 2016; Nisa et al. 2014).

La primera técnica de agrupamiento aplicada fue k-Means convencional pero los resultados no representaron la visualización inicial en la figura 3, esto se debe a que la latitud y longitud son coordenadas polares en lugar de coordenadas cartesianas (Joshi 2016). En este contexto, DBScan se probó con muchas combinaciones de radio (tamaño del hotspot) y la frecuencia mínima de incidentes. Se obtuvo un buen resultado con un radio de 200 metros y un mínimo de 5 eventos delictivos en el interior, sin embargo, una cantidad considerable de puntos no se agruparon.

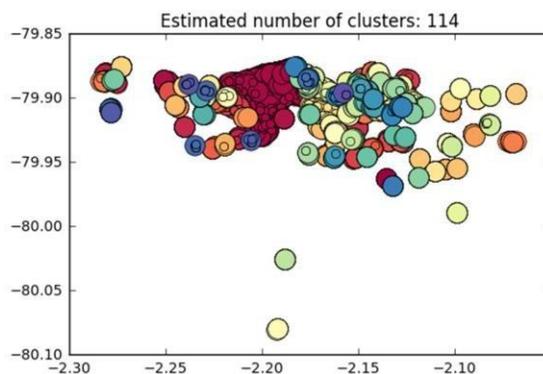


Figura N° 4.- DBScan con un radio de 200 metros y un mínimo de 5 incidentes

Lo que falta es un mecanismo para calcular la distancia entre dos puntos basado en latitud y longitud (Alam et al. 2016). Este mecanismo es la fórmula de Haversine que determina la distancia del círculo en una esfera (Fernando MaldonadoAlberto OchoaJulio ArreolaDaniel AzpeitiaAriel De la TorreDiego CanalesSaúl González 2013), la cual se usó para calcular las distancias de los datos de delito. Después de calcular las distancias de Haversine se aplicó k-means y DBScan obteniendo resultados similares y todos los puntos agrupados en su respectivo cluster.

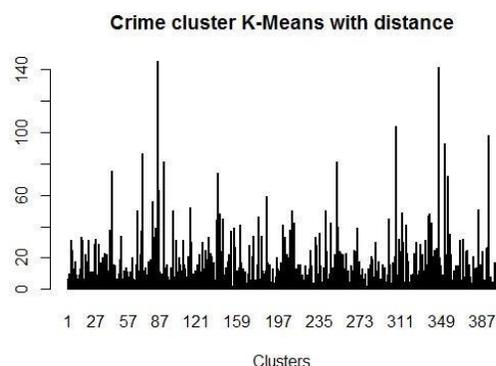


Figura N° 5.- Cluster de delitos con k-Means

Para predecir si un lugar específico es un punto caliente se evaluarán dos algoritmos de clasificación: 1) Máquina de Vector de Soporte (SVM) (Kianmehr & Alhadj 2008) y 2) Redes Neuronales Artificiales (NNE) (Weihong, Lei y Yebin 2016). El proceso de clasificación tiene dos

Detección automática del nivel de crimen basado en el análisis de puntos calientes en la ciudad de Guayaquil

pasos. Primero selección de parámetros y segunda evaluación de precisión que se explicará en la sección de implementación.

El conjunto de datos se tomaron muestras de 1.000 registros. Para obtener la mejor precisión es necesario evaluar cada modelo con diferentes parámetros usando una validación cruzada con un tamaño de dataset de entrenamiento de 60 por ciento de tamaño. El 40 por ciento restante se utiliza para las pruebas más adelante en la sección de evaluación.

El valor de predicción Y es el hotspot multi-clase (Extremo, Alto, Moderado y Bajo). Los valores de entrada X son latitud y longitud.

El método de clasificación SVM se obtuvo mejores resultados con un nucleo radial gaussiano sobre un nucleo lineal. Los parámetros óptimos de coste (C) y sigma se encontraron con una validación cruzada 10 veces repetido proceso. Los valores finales son $C = 32$ y $\sigma = 7,11$.

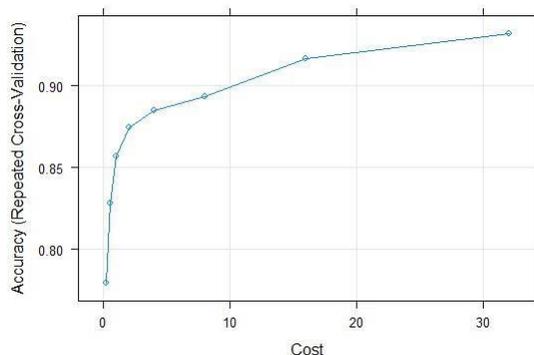


Figura N° 6.- Validación cruzada del método SVM

The neural network is also a nonlinear classification method. The input layer is composed by latitude and longitude, with a 10 hidden neurons. The regularization parameter was set to 0.01.

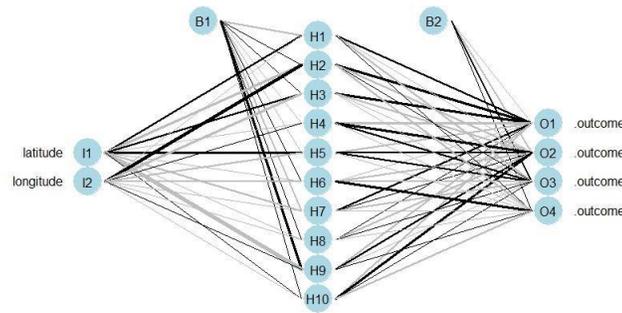


Figura N° 7.- Diseño de la red neuronal

Discusión.

El modelo se evalúa en términos de exactitud haciendo un proceso de re-muestreo que compara SVM y precisión de red neuronal de lado a lado. Podemos ver que SVM tiene una mayor precisión sobre la red neuronal.

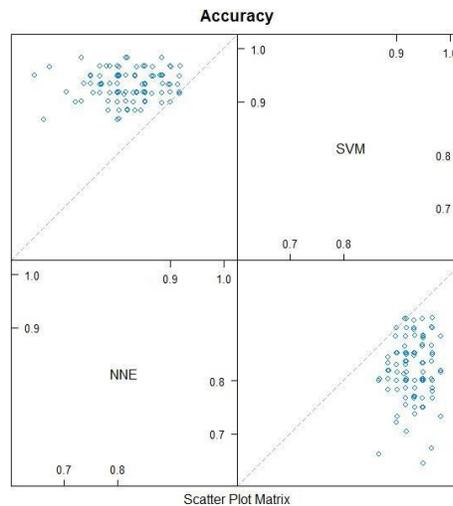


Figura N° 8.- Validación cruzada de SMV y Redes Neuronales

Utilizamos el 40 por ciento restante de los datos para predecir los resultados. Podemos confirmar los resultados de la validación cruzada con la predicción de los datos de la prueba. El SVM tiene mejores resultados como podemos ver en la siguiente matriz:

Detección automática del nivel de crimen basado en el análisis de puntos calientes en la ciudad de Guayaquil

Actual // Predicted	01	02	03	04	
EXTREME 01	99	1			01
MODERATE 02		86	9	5	02
HIGH 03		1	96	3	03
LOW 04		4	9	88	04
	01	02	03	04	

Figura N° 9.- Matriz comparativa de resultados de pruebas

Los valores en verde representan la coincidencia entre lo real versus las predicciones, que se traducen en un 93 por ciento de precisión con el método SVM. Con estos resultados este algoritmo se puede implementar como parte de una aplicación móvil en donde un ciudadano puede ser advertido sobre el nivel de crimen de su ubicación en tiempo real. Inclusive se puede complementar la aplicación de Google Maps obteniendo una trayectoria indicando el nivel de crimen en cada punto.

Conclusión.

El cálculo de la distancia para coordenadas polares es clave en el proceso de agrupación de los incidentes de delitos. El nivel de predicción del 93 por ciento permite la detección temprana de puntos calientes en diferentes niveles. Simplemente de requiere de una dirección en lenguaje natural que puede ser utilizada por un sistema para predecir el nivel de crimen de ese lugar. Es importante obtener datos recientes sobre los incidentes de crimen para alimentar la aplicación y mantener el nivel de precisión en la predicción.

Bibliografía.

- Abbasi, A., Chen, H. & Salem, A., 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. *ACM Transactions on Information Systems ...*, 26(3), pp.1–34. Available at: <http://dl.acm.org/citation.cfm?id=1361685> <http://portal.acm.org/citation.cfm?doid=1361684.1361685>.
- Azeez, J. & Aravindhar, D.J., 2015. Hybrid approach to crime prediction using deep learning. *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp.1701–1710. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7275858>.
- Bogomolov, A. et al., 2014. Once Upon a Crime. *Proceedings of the 16th International Conference on Multimodal Interaction - ICMI '14*, pp.427–434. Available at: <http://dx.doi.org/10.1145/2663204.2663254> <http://dl.acm.org/citation.cfm?doid=2663204.2663254>.
- Cesario, E., Catlett, C. & Talia, D., 2016. Forecasting Crimes Using Autoregressive Models. *Proceedings - 2016 IEEE 14th International Conference on Dependable, Autonomic and Secure Computing, DASC 2016, 2016 IEEE 14th International Conference on Pervasive Intelligence and Computing, PICom 2016, 2016 IEEE 2nd International Conference on Big Data*.
- Chainey, S., Tompson, L. & Uhlig, S., 2008. The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime. *Security Journal*, 21, pp.4–28. Available at: <http://discovery.ucl.ac.uk/112873/>.
- Chandra, B., Gupta, M. & Gupta, M.P., 2008. A multivariate time series clustering approach for crime trends prediction. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, pp.892–896.
- Davies, T.P. & Bishop, S.R., 2013. Modelling patterns of burglary on street networks. *Crime Science*, 2(1), p.10. Available at: <http://www.crimesciencejournal.com/content/2/1/10>.
- Eck, J.E. & Weisburd, D.L., 1995. Crime Places in Crime Theory. *Crime Prevention Studies*, 4, pp.1–33. Available at: http://www.popcenter.org/library/CrimePrevention/Volume_04/01-EckWeisburd.pdf.
- Fernando MaldonadoAlberto OchoaJulio ArreolaDaniel AzpeitiaAriel De la TorreDiego CanalesSaúl González, 2013. Intelligent Application to Reduce Transit Accidents in a City Using Cultural Algorithms. *Advances in Intelligent Systems and Computing*.
- Google Inc., 2017. Google Geolocation API. Available at: <https://developers.google.com/maps/documentation/geolocation/intro> [Accessed January 1, 2017].
- Imas Sukaesih Sitanggang; Tsamrul Fuad; Annisa, K-means clustering visualization of web-based OLAP operations for hotspot data. *2010 International Symposium on Information Technology*.
- Joshi, M.R., 2016. Analysis of Change in Coordinate System on Clustering. *Current Trends in Advanced Computing (ICCTAC)*.
- Kadar, C. & Cvijikj, I.P., 2014. CityWatch : The Personalized Crime Prevention Assistant. *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia*, pp.260–261.

- Kianmehr, K. & Alhadj, R., 2008. Effectiveness of Support Vector Machine for Crime Hot-Spots Prediction. *Applied Artificial Intelligence*, 22(5), pp.433–458. Available at: <http://www.tandfonline.com/doi/abs/10.1080/08839510802028405>.
- Mohamad No, N.M. et al., 2015. A Review on a Classification Framework for Supporting Decision Making in Crime Prevention. *Journal of Artificial Intelligence*, 8(1), pp.17–34. Available at: <http://www.scialert.net/abstract/?doi=jai.2015.17.34>.
- Nisa, K.K., Andrianto, H.A. & Mardhiyyah, R., 2014. Hotspot clustering using DBSCAN algorithm and shiny web framework. *Proceedings - ICACISIS 2014: 2014 International Conference on Advanced Computer Science and Information Systems*, pp.129–132.
- Pang, B. & Lee, L., 2008. Opinion Mining and Sentiment Analysis. *Foundations & Trends in Information Retrieval*, 2(1–2), pp.1–135.
- Tayebi, M.A. et al., 2014. CRIMETRACER: Activity space based crime location prediction. *ASONAM 2014 - Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, (Asonam), pp.472–480.
- Taylor, P.J. et al., 2012. *Global Urban Analysis: A Survey of Cities in Globalization*, Taylor & Francis. Available at: <https://books.google.com.ec/books?id=fmns3RI1xnwC>.
- Tollenaar, N. & van der Heijden, P.G.M., 2013. Which method predicts recidivism best?: A comparison of statistical, machine learning and data mining predictive models. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 176(2), pp.565–584.
- United Nations, 2014. *World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352)*, Available at: <http://esa.un.org/unpd/wup/Highlights/WUP2014-Highlights.pdf>.
- Vural, M.S. & Gök, M., 2016. Criminal prediction using Naive Bayes theory. *Neural Computing and Applications*. Available at: <http://link.springer.com/10.1007/s00521-016-2205-z>.
- Yuan, G.-X. et al., 2010. A Comparison of Optimization Methods and Software for Large-scale L1-regularized Linear Classification. *The Journal of Machine Learning Research*, 11, pp.3183–3234.
- Zhang, N. et al., 2015. Leave me alone: App-level protection against runtime information gathering on android. *Proceedings - IEEE Symposium on Security and Privacy*, 2015–July, pp.915–930.
- Zhang, Q. & Goncalves, B., 2016. Security Analysis and Exploitation of Arduino devices in the Internet of Things. *Acm*, p.6.
- Zhong-Yuan Zhang, 2012. Nonnegative Matrix Factorization: Models, Algorithms and Applications. In *Data Mining: Foundations and Intelligent Paradigms*.