Conceptos de *big data*, minería de datos, aprendizaje automático y aprendizaje profundo en los datos sobre delitos*

Big Data, Data Mining, Machine Learning, and Deep Learning Concepts in Crime Data

*Emre Cihan Ateş***
*Erkan Bostanci****
*Mehmet Serdar Güzel*****

Resumen

Junto con el rápido cambio de las tecnologías de la información y el uso generalizado de Internet a lo largo del tiempo, han surgido baterías de datos con una amplia diversidad y volúmenes bastante grandes. El uso de la minería de datos aumenta día a día debido al gran

---

* Este artículo ha sido publicado originalmente en inglés: Emre Cihan Ateş, Erkan Bostancı y Mehmet Serdar Güzel. "Big Data, Data Mining, Machine Learning, and Deep Learning Concepts in Crime Data", *Ceza Hukuku ve Kriminoloji Dergisi-Journal of Penal Law and Criminology*, vol. 8, n.° 2, 2020, p. 293 y ss., disponible en [https://iupress.istanbul.edu.tr/en/journal/jplc/article/big-data-data-mining-machine-learning-and-deep-learning-concepts-in-crime-data]. Su publicación en español ha sido autorizada por el *Journal of Penal Law and Criminology*.

** Gendarmerie Captain, Lecturer, Department of Security Science, Gendarmerie and Coast Guard Academy, Ankara, Turkey; *e-mail* [emre_cihan_ates@hotmail.com], ORCID [https://orcid.org/0000-0001-9550-4532].

*** Ph.D., Assistant Professor, Department of Computer Science, Ankara University, Ankara, Turkey; *e-mail* [ebostanci@ankara.edu.tr], ORCID [https://orcid.org/0000-0001-8547-7569].

**** Ph.D., Assistant Professor, Department of Computer Science, Ankara University, Ankara, Turkey; *e-mail* [mguzel@ankara.edu.tr], ORCID [https://orcid.org/0000-0002-3408-0083].

papel que desempeña en la adquisición de información al realizar los análisis necesarios, en especial, dentro de una gran cantidad de datos. La obtención de información precisa es un factor clave que afecta a los procesos de toma de decisiones. Los datos sobre delitos se incluyen entre las áreas de aplicación de la minería de datos, siendo una de las baterías de datos que crece con rapidez cada día que pasa. Los eventos delictivos constituyen un comportamiento no deseado en todas las sociedades. Por esta razón, es importante extraer información significativa de los datos sobre delitos. Este artículo tiene como objetivo proporcionar una descripción general del uso de la minería de datos y el aprendizaje automático en los datos sobre delitos y dar una nueva perspectiva sobre los procesos de toma de decisiones, al presentar ejemplos del uso de la minería de datos para un delito. Para este propósito, se presentan algunos ejemplos de minería de datos y aprendizaje automático en áreas de delincuencia y seguridad que brindan un marco conceptual en el tema de *big data*, minería de datos, aprendizaje automático y aprendizaje profundo junto con tipos de tareas, procesos y métodos.

Palabras clave: Crimen; Crimen; Big data; Minería de datos; Seguridad; Vigilancia policial.

Abstract

Along with the rapid change of information technologies and the widespread use of the internet over time, data stacks with ample diversity and quite large volumes have emerged. The use of data mining is increasing day by day due to the huge part it plays in the acquisition of information by making necessary analyses, especially within a large amount of data. Obtaining accurate information is a key factor affecting decision-making processes. Crime data is included among the application areas of data mining, being one of the data stacks which is rapidly growing with each passing day. Crime events constitute unwanted behavior in every society. For this reason, it is important to extract meaningful information from crime data. This article aims to provide an overview of the use of data mining and machine learning in crime data and to give a new perspective on the decision-making processes by presenting examples of the use of data mining for a crime. For this purpose, some examples of data mining and machine learning in crime and security areas are presented by giving a conceptual framework in the subject of big data, data mining, machine learning, and deep learning along with task types, processes, and methods.

Keywords: Crime; Big data; Data mining; Security; Policing.

Fecha de presentación: 20 de enero de 2021. Revisión: 15 de febrero de 2021. Fecha de aceptación: 3 de marzo de 2021.

I. Introducción

Para examinar el concepto de minería de datos (*data mining*), primero se deben examinar con cuidado los conceptos de datos, conocimiento, información y sabiduría. En este contexto, el término "datos"

se expresa en forma de definiciones como conocimiento oculto pero aún no interpretado ni analizado[1], observaciones y símbolos simples, hechos crudos/desorganizados[2] y un texto que no responde a ciertas preguntas[3].

El término "información" se define al utilizar expresiones como el conocimiento valioso en la mente humana, información contextual que incluye experiencia, valores, predicciones, una colección de datos organizados de manera consistente[4] y un texto que responde a preguntas de por qué y que se expresa como un resultado derivado de los datos y filtrado a través de la mente. La palabra "conocimiento" se refiere al mensaje que cambia el nivel de entendimiento y el propósito de una persona, la percepción de su receptor, un texto que responde a preguntas como quién, cuándo, qué o dónde, y datos que tienen sentido. La sabiduría, por otro lado, implica una previsión del futuro que aprovecha el conocimiento existente. La figura 1 ilustra estos conceptos en forma de pirámide.

1 Chaim Zins. "Conceptual approaches for defining data, information, and knowledge", en *Journal of the American Society for Information Science and Technology*, vol. 58, n.° 4, 2007, pp. 479 a 493, disponible en [https://www.researchgate.net/publication/220432993_Conceptual_approaches_for_defining_data_information_and_knowledge].

2 Adel Ahmed. "'From Data to Wisdom' Using Machine Learning Capabilities in Accounting and Finance Professionals", en *Talent Development & Excellence*, n.° 12, 2020.

3 Lixu Shao, Yucong Duan, Xiaobing Sun, Honghao Gao, Donghai Zhu y Welkal Miao. "Answering Who/When, What, How, Why through Constructing Data Graph, Information Graph, Knowledge Graph and Wisdom Graph", en *Seke*, July 2017, pp. 1 a 6, disponioble en [https://ksiresearch.org/seke/seke17paper/seke17paper_79.pdf].

4 Ahmed. "'From Data to Wisdom' Using Machine Learning Capabilities in Accounting and Finance Professionals", cit.; Paul Cooper. "Data, information, knowledge and wisdom", en *Anaesthesia & Intensive Care Medicine*, vol. 18, n.° 1, 2017, pp. 55 y 56.

Figura 1. Datos, información, conocimiento y sabiduría

Fuente: Traducido a partir de Emre Cihan Ateş, Gazy Erkan Bostanci y Mehmet Serdar Güzel. "Big data, data mining, Machine Learnig, and deep learning concepts in crime data", en *Ceza Hukuku ve Kriminoloji Dergisi-Journal of Penal Law and Crimino*logy, vol. 8, n.° 2, 2020, p. 294, disponible en [https://dergipark.org.tr/en/download/article-file/1354184].

En la pirámide de la figura 1 se muestra[5]:

• El concepto de datos expresa un concepto bastante general que aún no se ha analizado;

• El término información se refiere al procesamiento de datos a través de la mente;

• El concepto de conocimiento expresa una forma de datos que se ha vuelto significativa.

El concepto de sabiduría proviene de una combinación de los pasos de datos, información y conocimiento. Junto con el concepto de sabi-

5 Jonathan Hey. "The data, information, knowledge, wisdom chain: The metaphorical link", Intergovernmental Oceanographic Commission, 26, 2004, pp. 1 a 18, disponible en [https://www.docsity.com/en/docs/the-data-information-knowledge-wisdom-chain/9588737/]; Cooper. "Data, information, knowledge and wisdom", cit.

duría, el de *big data* surge en un principio sin procesar, pero se vuelve significativo al procesarse y, por lo tanto, ayuda a nuestras actividades de toma de decisiones[6].

El término *big data* se utiliza para describirla en todos los niveles, ya sea en bruto o procesado. Los datos son como metales preciosos que esperan ser extraídos de montones de piedras. Vivimos en un mundo en el que se recopilan muchos datos todos los días. En paralelo al desarrollo de la tecnología, también es un hecho conocido que los tamaños y tipos de datos han aumentado a medida que los conceptos de computadoras e Internet han entrado en nuestras vidas. Un problema particular que trae consigo este aumento del volumen de datos es el de la complejidad, por lo que analizarlos en forma correcta y significativa cobra cada día más importancia.

Como resultado del hecho que el proceso de obtención de datos significativos mediante la extracción de aquellos no cualificados de los brutos restringía a los investigadores y los métodos estadísticos actuales eran insuficientes para procesar grandes datos, ha surgido el concepto de minería de datos (*data mining*). En otras palabras, la minería de datos es un enfoque sistemático para analizar e identificar diferentes patrones y relaciones dentro de los grandes datos[7]. Una de las cuestiones importantes del método es el análisis de los datos existentes en lugar de la recopilación de una gran cantidad de ellos. La minería de datos tiene como objetivo desarrollar modelos de predicción sobre eventos que se prevé que ocurran en el futuro y respaldar los procesos de toma de decisiones aprovechando la muestra pasada en general[8]. El concepto de minería de datos también se conoce en la literatura como "extracción de conocimiento", "descubrimiento de conocimiento", "arqueología de datos", "análisis de datos/patrones", "minería de conocimiento" y "recolección de información"[9].

6 David J.Pauleen, David Rooney y Ali Intezari. "Big data, little wisdom: Trouble brewing? Ethical implications for the information systems discipline", *Social Epistemology*, vol. 31, n.° 4, 2017, pp. 400 a 416.

7 David M. Blei y Padhraic Smyth. "Science and data science", *Proceedings of the National Academy of Sciences*, vol. 114, n.° 33, August 2017, pp. 8.689 a 8.692, disponible en [https://www.pnas.org/doi/epdf/10.1073/pnas.1702076114].

8 John D. Kelleher y Brendan Tierney. *Data science*, Massachusetts, mit Press, 2018, disponible en [https://mrce.in/ebooks/Data%20Science.pdf].

9 Akhill Rajendra Khare y Pallavi Shrivasta. "Data mining for the internet of things", en Ambati V. Krishna Prasad (ed.). *Exploring the Convergence of Big Data and the Internet*

Las tecnologías de la información e Internet han alterado de modo significativo el campo del crimen y la seguridad[10]. Las estadísticas actuales muestran que el crimen aumenta en todo el mundo. Por esta razón, los conceptos de crimen y criminalidad han aumentado de manera constante en paralelo con un aumento en la cantidad de *big data*. Teniendo en cuenta que el crimen es un comportamiento no deseado en las sociedades, los datos y el conocimiento obtenido de ellos se han vuelto importantes para determinar las estrategias de prevención del crimen. Determinar objetivos y estrategias futuras solo depende de datos actualizados, precisos y confiables.

El objetivo de este artículo de revisión es brindar una perspectiva sobre los procesos de toma de decisiones con el uso de la minería de datos (*data mining*) en aquellos sobre delitos al brindar una descripción general de esta y analizar sus métodos. Después de la introducción, este estudio define los conceptos de *big data*, minería de datos (*data mining*), aprendizaje automático (*machine learning*) y aprendizaje profundo (*deep learning*) respectivamente. En la sección "métodos y modelos de minería de datos" se definen los conceptos, se explica el proceso de esta y se analizan los modelos de análisis de datos. La sección "uso de *big data*, minería de datos, aprendizaje automático y aprendizaje profundo en áreas criminales y de seguridad" se resumen varios casos que se dan en el campo de la seguridad, que tratan la lucha contra el crimen. En la sección de discusión, se examina el uso de las teorías de minería de datos, aprendizaje automático y aprendizaje profundo en la lucha contra el crimen. En la sección "conclusión" se afirma que los conceptos de *big data* y minería de datos se utilizan de manera activa en la lucha contra el crimen y se enfatiza su importancia futura.

of Things, pp. 181 a 191, Hershey, PA, IGI Global, 2018; Alİ Serhan Koyuncugİl y Nermİn Özgülbaş. "Veri madenciliği: Tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları", en *International Journal of Informatics Technologies*, vol. 2, n.° 2, May 2009, pp. 21 a 32, disponible en [https://dergipark.org.tr/tr/download/article-file/75259].

10 Naci Akdemir y Christopher James Lawless. "Exploring the human factor in cyber-enabled and cyber-dependent crime victimisation: A lifestyle routine activities approach", en *Internet Research*, vol. 30, n.° 6, June 2020, pp. 1665 a 1687, disponible en [https://www.semanticscholar.org/reader/f2d52e0caa51f6adb42acb32ace0f305b44058c9].

II. Antecedentes

A. Big Data

Dado que el concepto de *big data* es relativamente nuevo, primero debe examinarse en términos de definiciones.

– El concepto de *big data* es interdisciplinario, abstracto y de rápido crecimiento que ofrece un potencial de aumento en todas las áreas sociales y que se almacena a una escala sin precedentes[11].

– Los datos de gran volumen llegan en su mayoría en forma muy rápida de diversas fuentes, como Internet y las redes informáticas.

– Se define como un conjunto de datos al que es difícil acceder a través de tecnologías de la información estándar y que implica el uso de métodos de captura, procesamiento, recopilación, visualización y análisis en grandes conjuntos de datos[12].

Para investigar más a fondo el concepto de *big data*, será útil en este punto examinar los principios de la literatura que lo definen y que por lo general se denominan "las cinco v" (volumen, velocidad, variedad, valor y veracidad)[13].

11 Min Chen, Shiwen Mao y Yunhao Liu. "Big data: A survey", *Mobile Networks and Applications*, vol. 19, n.° 2, April 2014, pp. 171 a 209.

12 Siva Vaidhyanathan y Chris Bulock. "Knowledge and dignity in the era of 'big data'", en *The Serials Librarian*, vol. 66, n.os 1 a 4, 2014, pp. 49 a 64, disponible en [https://www.tandfonline.com/doi/epdf/10.1080/0361526X.2014.879805].

13 Noraini Abdullah, Saiful Adli Ismail, Siti Sophiayati y Suriani Mohd Sam. "Data quality in big data: A review", en *International Journal of Advances in Soft Computing & Its Applications*, vol. 7, n.° 3, November 2015, pp. 16 a 27, disponible en [https://www.i-csrs.org/Volumes/ijasca/IJASCA-SI-070302_Pg16-27_Data-Quality-in-Big-Data-A-Review.pdf]; Emre Cihan Ateş, Gazy Erkan Bostanci y Mehmet Serdar Güzel. "Security Evaluation of Industry 4.0: Understanding Industry 4.0 on the Basis of Crime, Big Data, Internet Of Thing (IoT) and Cyber Physical Systems", en *Güvenlik Bilimleri Dergisi*, (International Security Congress Special Issue), n.° 9, February 2020, pp. 29 a 50, disponible en [https://www.researchgate.net/publication/339642922_Security_Evaluation_of_Industry_40_Understanding_Industry_40_on_the_Basis_of_Crime_Big_Data_Internet_of_Thing_IoT_and_Cyber_Physical_Systems]; Samuel Fosso Wamba, Shahriar Akter, Andrew James Edwards, Geoffrey Chopin y Denis Gnanzou. "How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study", en *International Journal of Production Economics*, vol. 165, July 2015, pp. 234 a 246, disponible en [https://www.researchgate.net/publication/270276259_How_'big_data'_can_make_big_impact_Findings_from_a_systematic_review_and_a_longitudinal_case_study]; Martin White. "Digital workplaces: Vision and reality", en *Business Information Review*, vol. 29, n.° 4, 2012, pp. 205 a 214, disponible en [https://es.scribd.com/document/886559974/Digital-Workplaces-Vision-and-Reality].

– *Volumen*: Este término se utiliza para describir una cantidad de datos en constante aumento. Junto con el aumento aquellos, también se ha comenzado a formar una masa de conocimiento y se hace más difícil alcanzar uno calificado con tal aumento de volumen.

– *Variedad*: La mayoría de los datos producidos se da en entornos no estructurales como computadoras, redes sociales, teléfonos móviles y tabletas. Las fuentes de datos son relativamente nuevas y se agregan nuevas fuentes día a día. Esto aumentará el número de fuentes de datos, los que estarán en diferentes formatos e incluirán diferentes variables después de su recopilación, esto hará que el trabajo del analista sea más difícil.

– *Velocidad*: Esto expresa la tasa de crecimiento de los datos y cuando es alta, es un signo de que los datos son *big data*.

– *Veracidad*: Esta es una variable importante que garantiza que los datos sean seguros durante su flujo. Es importante protegerlos con medidas de seguridad de alto nivel y no alterarlos mediante intervenciones no autorizadas.

– *Valor*: Esto se refiere a la evaluación de *big data* y a la ayuda que proporciona a las unidades de apoyo para la toma de decisiones para obtener información significativa al transformar los datos brutos en información y conocimiento.

B. Minería de datos (data mining)

El concepto de minería de datos se define en muchas fuentes, algunas de las cuales se detallan a continuación:

– Descubrir conocimiento potencialmente útil sobre datos antes desconocidos de una manera no confidencial[14];

– Permite obtener conocimiento transformable en comprensión, entendimiento y acción mediante la combinación de estadísticas con conceptos, herramientas y algoritmos, aprendizaje automático (*machine learning*) y el análisis de conjuntos de datos muy grandes[15].

14 Konda Srinivas, B. Kavihta Rani y Aliseri Govrdhan. "Applications of data mining techniques in healthcare and prediction of heart attacks", en *International Journal on Computer Science and Engineering*, vol. 2, n.° 2, 2010, pp. 250 a 255, disponible en [https://www.researchgate.net/publication/49617135_Applications_of_Data_Mining_Techniques_in_Healthcare_and_Prediction_of_Heart_Attacks].

15 Graham J. Williams "Rattle: a data mining GUI for R", en *The R Journal*, vol. 1, n.° 2, 2009, pp. 45 a 55, disponible en [https://journal.r-project.org/archive/2009-2/RJournal_2009-2_Williams.pdf].

– La minería de datos es un paso en todas las fases de descubrimiento de conocimiento que puede describirse como un proceso de recuperación de conocimiento o extracción de minería de una gran cantidad de datos y es una forma de descubrimiento de conocimiento que es en particular necesaria para resolver problemas en un área determinada[16].

A partir de estas definiciones, es posible realizar la siguiente definición teniendo en cuenta la relación: La minería de datos (*data mining*) es un proceso que tiene como objetivo recuperar conocimiento haciendo uso de *big data* del pasado y utiliza métodos como la estadística, el aprendizaje automático y la inteligencia artificial, así como utilizar ese conocimiento en las actividades de apoyo a la toma de decisiones relacionadas con el futuro[17].

*C. Aprendizaje automático (*machine learning*)*

La introducción en nuestras vidas del concepto de aprendizaje automático comenzó con la pregunta "¿Pueden pensar las máquinas?" planteada por primera vez por Alan Turing[18-19]. El concepto de aprendizaje automático es la capacidad de una computadora de aprender por sí misma, al utilizar datos y experiencias existentes. En el aprendizaje automático, existen diferentes métodos de aprendiza-

16 Sunita Beniwal y Jitender Arora. "Classification and feature selection techniques in data mining", en *International Journal of Engineering Research & Technology*, vol. 1, n.° 6, August 2012, pp. 1 a 6, disponible en [https://www.researchgate.net/publication/263662705_Classification_and_feature_selection_techniques_in_data_mining]; David L. Olson y Georg Lauhoff. "Descriptive data mining", en *Descriptive Data Mining*, pp. 129 a 130, Singapore, Springer, 2019.

17 Colin Campbell y Yiming Ying. "Learning with support vector machines", *Synthesis Lectures on Artificial Intelligence and Machine Learning*, n.° 10, 1 a 83, San Rafael, CA, Morgan & Claypool, 2011, disponible en [https://digi-lib.stekom.ac.id/assets/dokumen/ebook/feb_B8yCNdjgUActi5qdyUbB6LyV7949c7RegDGg8jiFkgjSqd8C_HvSSZE_1681201865.pdf]; David J. Hand y Niall M. Adams. "Data Mining", en *Wiley StatsRef: Statistics Reference Online*, Hoboken, NJ, John Wiley & Sons, 2014, pp. 1 a 7.

18 Londres, 23 de junio de 1912-Wimsloe, Reino Unido, 7 de junio de 1954.

19 Allan Mathison Turing. "Computing machinery and intelligence", en *Mind*, vol. lix, n.° 236, October 1950, pp. 433 a 460, disponible en [https://courses.cs.umbc.edu/471/papers/turing.pdf].

je según el etiquetado de los datos de entrenamiento[20]. En el aprendizaje supervisado, los datos disponibles se conocen como datos de entrenamiento, que son la entrada básica que se le da a la máquina para aprender el modelo. La computadora intenta aprender los datos de entrenamiento con la ayuda de varios algoritmos de aprendizaje automático. La determinación del éxito del aprendizaje de la información se evalúa con datos de prueba. La diferencia más importante entre los datos de prueba y los datos de entrenamiento es si los datos son predichos o clasificados por la máquina. El aprendizaje no supervisado es el aprendizaje automático en el que no hay datos de entrenamiento y todo el aprendizaje se realiza de acuerdo con las similitudes y diferencias entre los datos. El aprendizaje semi supervisado, por otro lado, es el realizado a través de una estructura mixta, en la que hay datos de entrenamiento y datos no etiquetados.

El concepto de aprendizaje automático se utiliza tanto para el análisis de datos como para la minería de datos y, además, es el tipo de análisis de inteligencia artificial que define y reconoce patrones al utilizar inteligencia artificial[21].

– Se pueden realizar predicciones mediante el aprendizaje de datos a través del aprendizaje automático, que es una subrama de la inteligencia artificial.

– En el aprendizaje automático, un principio básico es que el sistema creado realiza el proceso de aprendizaje por sí mismo.

– La diferencia más importante entre las computadoras y los humanos es que estos aprenden de los eventos que vivieron en el pasado, a lo que llaman experiencia, mientras que el aprendizaje automático tiene como objetivo llegar a este punto a través de un mecanismo de decisión.

– El aprendizaje se proporciona por medio de datos de entrenamiento que mejoran la automatización en el acceso al conocimiento, lo que permite que se lleven a cabo técnicas automatizadas eficientes con poco poder humano[22].

20 Charu C. Aggarwal. *Machine learning for text*, Cham, Springer, 2018.

21 Lawrence McClendon y Natarajan Meghanathan. "Using machine learning algorithms to analyze crime data", en *Machine Learning and Applications: An International Journal*, vol. 2, n.° 1, March 2015, pp. 1 a 12, disponible en [https://www.researchgate.net/publication/275220711_Using_Machine_Learning_Algorithms_to_Analyze_Crime_Data].

22 Joyce Jackson. "Data mining; a conceptual overview", en *Communications of the Association*

– En esencia, el aprendizaje automático (*machine learning*) es un método de aprendizaje de un sistema informático a través del muestreo, existen muchos algoritmos de aprendizaje automático diferentes en varios problemas o tipos de datos.

Cuando se examinan los conceptos de minería de datos y aprendizaje automático a un nivel básico, el uso de inteligencia artificial lleva a la conclusión de que los sistemas de datos obtienen conocimiento de estos sin procesar, sin intervención[23]. Además, la minería de datos proporciona una recuperación de datos significativa de las baterías de datos. También permite una minimización del trabajo humano y una optimización de los resultados a través del concepto de aprendizaje automático. Por esta razón, es posible decir que la minería de datos y el aprendizaje automático se pueden utilizar en forma activa en el proceso de recuperación de conocimiento en todos los campos de la ciencia, en especial, en las naturales.

D. Aprendizaje profundo (deep learning)

El aprendizaje profundo es el aprendizaje automático, uno de los temas más populares en la actualidad, es un campo de estudio que incluye redes neuronales artificiales con una o más capas ocultas y algoritmos de aprendizaje automático similares. El aprendizaje profundo consiste en las estructuras que pueden aprender, estructuras complejas con redes neuronales de múltiples capas y que, a través de capas ocultas, pueden aprender con mayor facilidad y éxito modelos que los algoritmos clásicos de aprendizaje automático no pueden aprender con facilidad[24].

Las capas ocultas son parte de una estructura en la que se realizan funciones matemáticas y cálculos para obtener el resultado deseado. El número de sus capas puede variar. Aunque el aprendizaje

for Information Systems, vol. 8, 2002, pp. 267 a 296, disponible en [https://www.academia.edu/5442707/DATA_MINING_A_CONCEPTUAL_OVERVIEW#:~:text=Download,-PDF].

23 Kelleher y Tierney. *Data science*, cit.

24 Ayon Dey. "Machine learning algorithms: A review", en *International Journal of Computer Science and Information Technologies*, vol. 7, n.° 3, 2016, pp. 1.174 a 1.179, disponible en [https://ijcsit.com/docs/Volume%207/vol7issue3/ijcsit2016070332.pdf].

profundo es un enfoque nuevo, se ha utilizado de manera activa en muchas áreas como la semántica, el aprendizaje por transferencia, el procesamiento del lenguaje natural, la visualización y la investigación criminal. A medida que aumenta el número de capas, también aumentará el *hardware* necesario para el proceso de aprendizaje (sobre todo, el requisito de la unidad de procesamiento gráfico –GPU–) y el tiempo (debido a la gran cantidad de parámetros). Además, aunque se sabe que el aprendizaje profundo requiere muchos más datos que los algoritmos clásicos durante el entrenamiento, esto no es un problema hoy en día debido a la abundancia de datos de que disponemos. Sin embargo, si la cantidad de datos a analizar es pequeña, los algoritmos clásicos pueden producir resultados más exitosos en muchos sentidos.

Si bien los algoritmos clásicos de aprendizaje automático utilizan métodos basados en análisis estadístico para reconocer el patrón, existe un modelo similar a las neuronas del cerebro humano en el aprendizaje profundo. Por esta razón, la optimización de hiper parámetros (*tuning*) que minimiza la función de pérdida, se puede realizar de muchas formas diferentes. Con el aprendizaje profundo, se obtiene un buen rendimiento de clasificación para datos de texto, audio y visuales sin intervención humana, en especial debido a la capacidad de extraer características.

III. Métodos y modelos de minería de datos (*data mining*)

La minería de datos es muy popular hoy y es un hecho conocido que cada día se añade un nuevo método o algoritmo a la literatura, que considera tanto el rápido cambio en la estructura de la información como el aumento de las necesidades de información de las personas. Esto es una indicación que los métodos y algoritmos de minería de datos se mueven de manera dinámica en lugar de permanecer estáticos. Esto se debe a que los métodos utilizados para obtener el conocimiento correcto también pueden cambiar de acuerdo con el estado de los datos, dado que las estructuras de las fuentes de datos cambian. La minería de datos se centra en tres secciones principales. La clasificación realizada de acuerdo con la capacidad de editar los datos se muestra a continuación[25].

– Estructurados;
– No estructurados;
– Semiestructurados.

Los datos estructurados que definen un grupo de datos especificado en forma explícita juegan un papel importante en su análisis en todo momento debido a la conveniencia de su clasificación[26]. La mayoría de las verificaciones biométricas utilizadas en las definiciones en el campo de la seguridad se basan en datos estructurados. Los datos no estructurados no tienen un modelo predefinido, no se pueden clasificar y muestran variabilidad. El 90% de todos los datos se consideran no estructurados, lo que revela el hecho de que se trata de un conjunto mayor de datos que es difícil de analizar[27]. Esto es cierto en el caso de los sitios de redes sociales, que se utilizan cada día con más frecuencia y que como es evidente, desempeñan un papel importante en los datos no estructurados. Los datos no estructurados mejoran la capacidad de obtener una mayor cantidad de información de los conjuntos de datos. Sin embargo, la precisión de aquellos obtenidos puede no ser tan grande como la de los estructurados. Los datos semiestructurados ayudan a la clasificación y definición parcial de la información. Entre los ejemplos más comunes de datos semiestructurados se encuentran los correos electrónicos, los lenguajes de creación de documentos y la utilización de bases de datos web y NoSQL[28].

25 Anurag Agrahari y Dharmaji Rao. "A review paper on Big Data: Technologies, tools and trends", en *International Research Journal of Engineering and Technology*, vol. 4, n.° 10, October 2017, pp. 640 a 649, disponible en [https://www.irjet.net/archives/V4/i10/IRJET-V4I10112.pdf]; Zhiqiang Ge, Zhihuan Song, Steven X. Ding y Biao Huang. "Data mining and analytics in the process industry: The role of machine learning", en *IEEE Access*, vol. 5, 2017, pp. 20.590 a 20.616, disponible en [https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8051033].

26 Ravi Kumar y Bharti Nagpal. "Analysis and prediction of crime patterns using big data", *International Journal of Information Technology*, vol. 11, n.° 4, December 2019, pp. 799 a 805.

27 Vesal Tirgari. "Information technology policies and procedures against unstructured data: A phenomenological study of information technology professionals", en *Journal of Management Information and Decision Sciences*, vol. 15, n.° 2, 2012, pp. 87 a 106, disponible en [https://www.abacademies.org/articles/aimsjvol15no22012.pdf].

28 "NoSQL, también conocido como 'no soólo SQL' o 'no SQL', es un enfoque utilizado en el diseño de bases de datos que permite el almacenamiento y consulta de datos fuera de las estructuras tradicionales que se encuentran en las bases de datos relacionales". IBM.

Figura 2. Métodos de analítica de datos

Fuente: Traducido a partir de Ateş, Bostanci y Güzel. "Big data, data mining, Machine Learnig, and deep learning concepts in crime data", cit., p. 301.

La minería de datos, que se examina en tres grupos como estructurada, no estructurada y semiestructurada, se evalúa en cuatro estructuras de modelo bajo los títulos generales de predictiva, descriptiva, diagnóstica y prescriptiva (mostradas en la figura 2)[29].

– *Métodos descriptivos*: Estos incluyen aquellos que determinan los vínculos y las relaciones entre los datos que respaldarán la toma de decisiones en la batería de datos[30];

"¿Qué es una base de datos nosql?", disponible en [https://www.ibm.com/es-es/topics/nosql-databases].

29 Mohammad Mesgarpour e Ian Dickinson. "Enhancing the value of commercial vehicle telematics data through analytics and optimisation techniques", en *Archives of Transport System Telematics*, vol. 7, n.° 3, September 2014, pp. 27 a 30, disponible en [https://bibliotekanauki.pl/articles/393920.pdf].

30 Frederic E. Bock, Roland C. Aydin, Christian J. Cyron, Norbert Huber, Surya R. Kalidindi y Benjamin Klusemann. "A review of the application of machine learning and data mining approaches in continuum materials mechanics", en *Frontiers in Materials*, vol. 6, May 2019, article 110, pp. 1 a 23, disponible en [https://www.frontiersin.org/journals/materials/articles/10.3389/fmats.2019.00110/full].

– *Métodos de diagnóstico*: Estos son aquellos que dan razones y brindan conocimiento sobre preguntas de "por qué";

– *Métodos predictivos*: Estos incluyen aquellos utilizados para estimar la variable dependiente al utilizar las variables independientes en la base de datos. Se usan para hacer una predicción de eventos futuros con el uso de resultados pasados conocidos;

– *Métodos prescriptivos*: Estos se utilizan para hacer una predicción de eventos futuros al usar resultados pasados conocidos.

Si afirmamos que la mayoría de los datos del mundo no están estructurados, podemos decir con facilidad que los métodos más utilizados son los descriptivos y los de diagnóstico. Los métodos descriptivos, diagnósticos, predictivos y prescriptivos se incluyen en una sub jerarquía de métodos estructurados, no estructurados y semiestructurados. Los métodos descriptivos, diagnósticos, predictivos y prescriptivos se muestran en el orden más alto en la jerarquía de métodos en muchas fuentes, ya que clasifican los métodos de minería de datos.

Los modelos de minería de datos se examinaron bajo los siguientes cuatro títulos principales según sus funciones:

– *Clasificación*: En la clasificación, la categorización de los datos suele estar a la vanguardia, y así es como se deciden las orientaciones de los datos[31]. Por ejemplo, un modelo de clasificación sobre datos de delincuencia puede indicar si la ciudad es segura o peligrosa en función de la intensidad de los eventos delictivos en la misma. En especial en el aprendizaje supervisado dentro del ámbito de la clasificación, es posible estimar la clasificación de problemas agregados en forma reciente al hacer un descubrimiento de patrones a partir de los datos. Muchas operaciones, como el reconocimiento de voz, el desvío de llamadas y la clasificación de texto, se pueden llevar a cabo con estos algoritmos, sobre todo los algoritmos genéticos, cuya clasificación está definida por reglas, es decir, supervisada. Algunos de los algorit-

31 Anirban Mukhopadhyay, Ujjwal Maulik, Sanghamitra Bandyopadhyay y Carlos Artemio Coello Coello. "A survey of multiobjective evolutionary algorithms for data mining: Part I", en *ieee Transactions on Evolutionary Computation*, vol. 18, n.° 1, 2013, pp. 4 a 19; Lleszek Rutkowski, Maciej Jaworski y Piotr Duda. *Stream data mining: Algorithms and their probabilistic properties*, Cham (Switzerland), Springer, 2020.

mos más utilizados, son por ejemplo, el Naive Bayes[32], regresión logística, algoritmos genéticos, máquinas de vectores de soporte, k-NN (algoritmo de los k vecinos más cercanos), razonamiento basado en la memoria y también se utiliza con frecuencia la lógica difusa.

– *Regresión*: Es el método de análisis que puede configurar un modelo al predecir las orientaciones de los datos. A diferencia de la clasificación, se lleva a cabo como una predicción que es esencial en la regresión[33]. Por ejemplo, el modelo de regresión se puede utilizar para predecir posibles datos futuros sobre delitos al analizar los tipos, horarios y frecuencias actuales de delitos en la ciudad.

– *Agrupamiento*: Es un proceso de descomposición en grupos llamados *clusters* en función de un cierto criterio de proximidad[34]. Se espera que los datos de un mismo clúster sean similares entre sí y que la similitud sea mucho menor en diferentes clústeres en general después de que se haya realizado el proceso de agrupamiento[35]. El agrupamiento, en otras palabras, la agrupación, se ha vuelto aún más importante, en especial con el concepto de aprendizaje automático. Se utiliza de manera activa en muchas áreas, como el procesamiento de voz e imágenes, el reconocimiento de voz, las llamadas telefónicas frecuentes, la mensajería y el uso de datos, la clasificación de clientes y, sobre todo, para el análisis de redes sociales. Las redes neuronales artificiales de Korhonen, Canopy, Mean Shift, K-means, Fuzzy C-averages, Latent Dirichlet Allocation, k-medoids y MinHash son ejemplos de algoritmos de agrupamiento que se utilizan con este fin[36].

32 Algoritmo de aprendizaje automático para tareas de clasificación basado en el teorema de Bayes donde todas las variables predictoras son independientes entre sí.

33 Mamta Mittal, Lalit Mohan Goyal, Duraisamy Jude Hemanth y Jasleen Kaur Sethi. "Clustering approaches for high-dimensional databases: A review", en *Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery*, vol. 9, n.° 3, 2019, pp. 1 a 14.

34 Pavel Berkhin. "A survey of clustering data mining techniques", en *Grouping multidimensional data*, Berlin-Heidelberg, Springer, 2006, pp. 25 a 71; Manoj Kr. Gupta y Pravin Chandra, P. "A comparative study of clustering algorithms", en *2019 6th International Conference on Computing for Sustainable Global Development (indiacom)*, New Delhi, ieee Xplore, 13-15 March, 2019, pp. 801 a 805.

35 Rutkowski, Jaworski y Duda. *Stream data mining: Algorithms and their probabilistic properties*, cit.

36 Mukhopadhyay, Maulik, Bandyopadhyay y Coello Coello. "A survey of multiobjective evolutionary algorithms for data mining: Part I", cit.; Ramin Ghorbani y Rouzbeh Ghousi. "Predictive data mining approaches in medical diagnosis: A review of some diseases prediction", en *International Journal of Data and Network Science*, vol. 3, 2019, pp. 47 a 70, disponible en [https://www.growingscience.com/ijds/Vol3/ijdns_2019_12.pdf].

– *Reglas de asociación*: Se utilizan para apoyar estudios futuros al determinar los comportamientos asociados y hacer uso de los datos obtenidos del pasado[37]. Es en particular la revelación de actitudes encontradas en crímenes organizados y cometidos con frecuencia lo que permitirá realizar la detección con mucha más rapidez.

IV. Uso de *big data*, minería de datos, aprendizaje automático y aprendizaje profundo en áreas criminales y de seguridad

Los conceptos de minería de datos (*data mining*) y aprendizaje automático (*machine learning*) se han vuelto más populares de manera reciente, ya que son más fáciles de usar a través de *software* incluido. Esta popularidad también ha aumentado las áreas de uso de la minería de datos. A medida que crece la batería de datos a examinar, la importancia de conceptos como la minería de datos, la inteligencia artificial y el aprendizaje automático se ha demostrado una vez más. Mientras la población de la humanidad crezca y existan reglas en el mundo, los datos sobre delitos seguirán en aumento, al igual que otras baterías de datos.

El concepto de delito se refiere a la práctica de actividades que están prohibidas de acuerdo con las normas y desviaciones en las sociedades. Aunque varía de una sociedad a otra, el concepto de delito por lo general se define como actuar en contra de las leyes existentes. Es posible clasificar los tipos y clases de delitos en categorías –desde delitos sexuales hasta delitos de orden público, desde la trata de personas hasta delitos relacionados con las drogas, desde la delincuencia juvenil hasta los crímenes de guerra–, estos superarán los límites de la imaginación de los seres humanos. En este caso, la lucha contra el crimen cobra cada vez mayor importancia para las sociedades en aras de la continuidad del orden social existente.

Las aplicaciones de cumplimiento de la ley en la lucha contra el crimen se examinan bajo dos títulos, a saber, preventivas y reac-

37 Wai Ting Eric Ngai, Li Xiu y Dorothy C. K. Chau. "Application of data mining techniques in customer relationship management: A literature review and classification", *Expert Systems with Applications*, vol. 36, n.° 2, March 2009, pp. 2.592 a 2.602, disponible en [https://didawiki.di.unipi.it/lib/exe/fetch.php/dm/crm_dm-survey.pdf]; Mittal, Goyal, Hemanth y Sethi. "Clustering approaches for high-dimensional databases: A review", cit.

tivas (en otras palabras, judiciales)[38]. Las investigaciones reactivas de delitos tienen como objetivo identificar elementos como los autores desconocidos o la víctima del delito después de que este se haya cometido. Estas investigaciones se llevan a cabo con el apoyo de la investigación de la escena del crimen. Las investigaciones preventivas de delitos implican la sensibilización de las víctimas de delitos disuadiendo a los delincuentes con operaciones anticrimen eficaces antes de que se cometa el delito. Además, el concepto de análisis del delito es más importante en las medidas preventivas, esto se debe a que es un concepto en el que se detectan los delitos existentes y las tendencias delictivas y también se consideran las posibles medidas a adoptar.

El uso del crimen en la minería de datos a través del aprendizaje automático, la identificación, clasificación y agrupación como patrones también es uno de los temas importantes en varias disciplinas científicas y de ingeniería como biología, química, física, psicología, medicina y información para examinar la evidencia, etc. Esto se debe a que las investigaciones criminales son multidisciplinarias[39].

En distintas ramas de la ciencia que utilizan elementos criminales, algunos ejemplos de uso son los siguientes:

– Examen de cartas manuscritas y firmas pasadas con ejemplos comparativos tomados de certificados legales y documentación;

– Procedimientos de comparación de muestras biológicas existentes como una uña, sangre y hueso con bases de datos existentes dentro del ámbito de la biología forense;

– En escenas donde ocurrieron accidentes graves, reestructuración del área con el objetivo de encontrar razones de esos accidentes dentro del ámbito de la gestión del tráfico y aplicaciones;

– Identificación de una sustancia dentro del ámbito de la química forense, en especial, sus derivados relacionados con el crimen, como las drogas,

38 Curtis Clarke. "Proactive policing: Standing on the shoulders of community-based policing", en *Police Practice and Research*, vol. 7, n.° 1, March 2006, pp. 3 a 17.

39 Hossein Hassani, Xu Huang, Emmanuel Sirimal Silva y Mansi Ghodsi. "A review of data mining applications in crime", en *Statistical Analysis and Data Mining, The ASA Data Science Journal*, vol. 9, n.° 3, April 2016, disponible en [https://www.researchgate.net/publication/301579904_A_Review_of_Data_Mining_Applications_in_Crime].

– Examen de las muestras sujetas a comparación con grabaciones de voz y video pasadas dentro del ámbito de los exámenes forenses de audio y video[40];

– Dentro del ámbito de la contabilidad forense, transferencias de dinero inusuales o transferencias de dinero que sean compatibles con la cantidad sujeta al delito;

– Investigación de la conexión de transacciones realizadas en una computadora, internet, redes sociales, teléfono móvil, etc., con delitos cometidos dentro del ámbito de la informática forense[41];

– En las actividades de prevención del delito que se llevan a cabo en el ámbito de la criminología, como el análisis del delito, la elaboración de perfiles de víctimas y sospechosos, las regiones donde la delincuencia es intensa, etc[42].

Como se puede ver en los ejemplos de áreas de uso anteriores, los conceptos de *big data* y minería de datos son parte de un área multidisciplinaria. Su uso aumenta cada día que pasa, aunque no al nivel deseado, por este motivo, se ofrecen algunos ejemplos de áreas de uso de la minería de datos y el aprendizaje automático bajo los siguientes títulos principales.

A. Corrupción y fraude en cuentas bancarias

El concepto de corrupción, falsificación y fraude es uno de los delitos más difíciles de prevenir por parte de las fuerzas del orden. Es muy difícil identificar el delito en su fase inicial, ya que cada tipo de falsificación emergente tiende a engañar a la gente con nuevas técnicas. Además, los conceptos de corrupción y fraude se vuelven cada vez más complejos con los avances tecnológicos[43]. Junto con el sufrimien-

40 Darren Quick y Kim-Kwang Raymond Choo. "Big forensic data reduction: Digital forensic images and electronic evidence", *Cluster Computing*, vol. 19, n.° 2, 2016, pp. 723 a 740.

41 Idem.

42 Janet Chan y Lyria Bennett Moses. "Making sense of big data for security", en *The British Journal of Criminology*, vol. 57, n.° 2, March 2017, pp. 299 a 319, disponible en [https://uploads-ssl.webflow.com/5cd23e823ab9b1f01f815a54/5cff2fe1af68c502e4474e4b_Making%20Sense%20of%20Big%20Data%20for%20Security.pdf].

43 Duen Horng Chau, Shashank Pandit y Christos Faloutsos. "Detecting fraudulent personalities in networks of online auctioneers", en *European Conference on Principles of Data Mining and Knowledge Discovery*, pp. 103 a 114, Berlin-Heidelberg, Springer, 2006, disponible en [https://link.springer.com/content/pdf/10.1007/11871637_14.pdf].

to injusto, la confianza en el Estado, y en especial en las unidades de seguridad afiliadas al Estado, se ve comprometida en la sociedad y la gente se ve afectada en forma negativa.

Con el desarrollo de la tecnología, es un hecho conocido hoy en día que el dinero tangible se ha convertido en dinero virtual debido a la facilidad de uso de los bancos e incluso se han creado extractos de dinero en cantidades reales mediante transacciones virtuales. En el caso de los datos de transferencias de dinero, en la mayoría de los países del mundo los problemas se presentan solo cuando se trata de grandes cantidades; los bancos comparten la información necesaria con las autoridades judiciales pertinentes, por lo que, de hecho, se utiliza la minería de datos[44]. Sin embargo, notificar a las autoridades judiciales las transferencias de dinero sospechosas, que se repiten con frecuencia, incluso en pequeñas cantidades, facilitará la detección de transacciones ilegales desde el principio[45].

Dado que las transacciones bancarias que realizamos a diario ocurren con bastante frecuencia y en diferentes cantidades, si bien no es fácil identificar los elementos que pueden ser delictivos en las bases de *big data*, es posible determinar los valores que pueden considerarse sospechosos mediante análisis algorítmicos apropiados. Por esta razón, con trabajos de aprendizaje automático realizados en el ámbito de la minería de datos, se pueden detectar transferencias de dinero regulares a una cuenta desde múltiples cuentas o a múltiples cuentas fantasma desde una cuenta y examinar estas transacciones como operaciones sospechosas permite una detección más rápida del delito. Si se detecta un delito en operaciones sospechosas, se involucrarán unidades de aplicación de la ley forense, si no se detecta ningún delito, se llevará a cabo una acción de aplicación de la ley preventiva, lo cual es importante en la prevención del delito. Además, la comparación de los datos estadísticos de los clientes en el banco con

44 James Osabuohien Odia y Osahemi Thaddeus Akpata. "Role of Data Science and Data Analytics in Forensic Accounting and Fraud Detection", en Bhushan Patil y Manisha Vohra (eds.). *Handbook of Research on Engineering, Business, and Healthcare Applications of Data Science and Analytics*, 2020, Hershey, PA, igi Global, 2020, pp. 203 a 227.

45 Sunday C. Agu, Ifeyinwa Ajah y Walter E. Ibe. "Impact of Human Character and Information System on Corruption Risk in Nigeria", en *International Journal of Scientific Research and Engineering Development*, vol. 2, n.° 4, July-August 2019, pp. 481 a 485, disponible en [https://www.madonnauniversity.edu.ng/wp-content/uploads/2019/09/IJSRED-V2I4P50-Sunday-Agu.pdf].

los patrones de comportamiento habituales de estos ayudará a identificar cualquier nueva actividad sospechosa. En este contexto, existen modelos exitosos de aprendizaje automático y aprendizaje profundo en la literatura, en especial en tarjetas de crédito y cuentas bancarias[46].

B. Uso de los datos básicos y del c onocimiento sobre teléfonos móviles

Con el uso de teléfonos móviles en todo el mundo, que alcanza el 68% de la población, sobre todo en los países desarrollados, es obvio que la proporción es bastante alta[47]. Podemos decir que los teléfonos móviles son parte de nuestras vidas, ya que los datos estadísticos también revelan que son comunes. Al tener en cuenta que el tiempo medio de uso del teléfono móvil en Europa es de 251 minutos al mes, se puede decir que existe una alta tasa de uso del teléfono móvil[48].

El uso del teléfono móvil proporciona mucha información, como con quién y con qué frecuencia nos comunicamos en la vida diaria y en qué coordenada geográfica el teléfono recibe la estación base de la empresa gsm (Sistema Global para Comunicaciones Móviles) pertinente. Además, es posible realizar llamadas frecuentes durante el día y cada una de ellas crea baterías de datos en diferentes intervalos de tiempo. Estas baterías se denominan por su sigla en inglés hts

46 Aderemi O. Adewumi y Andronicus A. Akinyelu. "A survey of machine-learning and nature-inspired based credit card fraud detection techniques", en *International Journal of System Assurance Engineering and Management*, vol. 8, n.° 2, November 2017, pp. 937 a 953, disponible en [https://link.springer.com/article/10.1007/s13198-016-0551-y]; Johan Perols. "Financial statement fraud detection: An analysis of statistical and machine learning algorithms", *Auditing: A Journal of Practice & Theory*, vol. 30, n.° 2, May 2011, pp. 19 a 50, disponible en [https://www.researchgate.net/publication/256045322_Financial_Statement_Fraud_Detection_An_Analysis_of_Statistical_and_Machine_Learning_Algorithms]; Abhimanyu Roy, Jingyi Sun, Robert Mahoney, Loreto Alonzi, Stephen Adams y Peter Beling. "Deep learning detecting fraud in credit card transactions", en *2018 Systems and Information Engineering Design Symposium*, Oxford, ieee, 2018, pp. 129 a 134.

47 Simon Kemp. "Digital in 2018: World's Internet users pass the 4 billion mark", en London, *We are social*, 30 January 2018, disponible en [https://wearesocial.com/uk/blog/2018/01/global-digital-report-2018/].

48 Ajans Press. "Türkiye cep telefonuyla konuşmada Avrupa birincisi", en *cnnturk.com*, 29 de noviembre de 2017, disponible en [https://www.cnnturk.com/teknoloji/turkiye-cep-telefonuyla-konusmada-avrupa-birincisi-705837].

(Búsqueda de Tráfico Histórico) (en algunas fuentes se las denomina "Registro de Datos de Llamadas" –CDR–)[49] y proporcionan el conocimiento sobre las llamadas realizadas por las personas en sus teléfonos, incluida información como la persona que llama, la persona llamada, la hora de la llamada, su duración, ubicación y estaciones base recibidas[50]. De esta forma, a partir de estos datos sin sentido es posible determinar la posición aproximada de la persona en el momento del crimen y establecer la relación entre el sospechoso y la escena del crimen. El programa del paquete I2 (i dos) desarrollado por IBM para este análisis es muy utilizado para fines de análisis[51].

Cuando se examinan las investigaciones criminales, los sospechosos suelen ser personas que rodean a la víctima y que tienen relación con ella[52]. Por lo general, existe una relación entre el sospechoso y la víctima, ya que un crimen no se comete sin una razón. Las enfermedades mentales como la psicopatía son una excepción a esta relación. Uno de los hallazgos que revelan esta relación es el conocimiento de entrevistas pasadas y el tráfico de la última llamada antes del evento, en especial cuando la víctima resulta herida, a menudo puede arrojar luz sobre el evento. En este contexto, existen modelos exitosos en la literatura, en lo referente al uso de teléfonos móviles[53].

49 JOHN STEENBRUGGEN, EMMANOUIL TRANOS y PETER NIJKAMP. "Data from mobile phone operators: A tool for smarter cities?", *Telecommunications Policy*, vol. 39, n.os 3 y 4, May 2015, pp. 335 a 346.

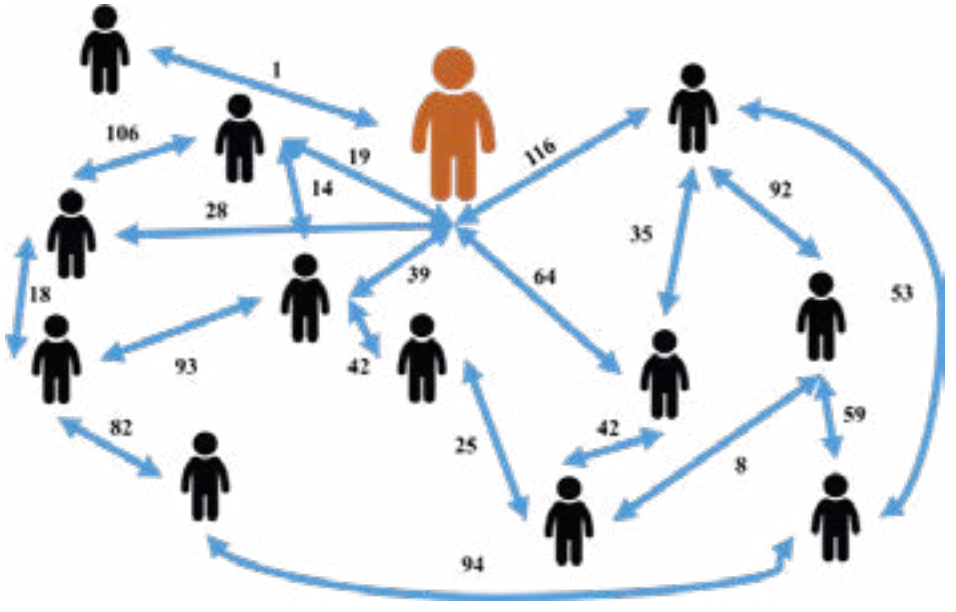
50 DANAH BOYD y KATE CRAWFORD. "Critical questions for big data: Provocations for a cultural, technological and scholarly phenomenon", en *Information, Communication & Society*, vol. 15, n.° 5, January 2012, pp. 662 a 679, disponible en [https://www.researchgate.net/publication/281748849_Critical_questions_for_big_data_Provocations_for_a_cultural_technological_and_scholarly_phenomenon].

51 XIN LI, BING LIU y PHILIP S. YU. "Discovering overlapping communities of named entities", en *Knowledge Discovery in Databases: PKDD 2006*, pp. 593 a 600, Berlin-Heidelberg, Springer, 2006, disponible en [https://link.springer.com/chapter/10.1007/11871637_60]; CHRISTOPHER TASSONE, BEN MARTINI y KIM-KWANG RAYMOND CHOO. "Forensic visualization: Survey and future research directions", en KIM-KWANG RAYMOND CHOO y ALI DEHGHANTANHA (eds.). *Contemporary Digital Forensic Investigations of Cloud and Mobile Applications*, Cambridge, MA, Syngress, 2017, pp. pp. 163 a 184.

52 NICK TILLEY y AIDEN SIDEBOTTOM. *Handbook of crime prevention and community safety*, New York, Routledge, 2017.

53 LI HE, ANTONIO PÁEZ, JIANMIN JIAO, PING AN, CHUNTIAN LU, WEN MAO y DONGPIN LONG. "Ambient Population and Larceny-Theft: A Spatial Analysis Using Mobile Phone Data", *ISPRS International Journal of Geo-Information*, vol. 9, n.° 6, 2020, disponible en [https://www.mdpi.com/2220-9964/9/6/342/pdf?version=1591598730]; MARTIN W. TRAUNMUELLER, GIOVANNI QUATTRONE y LICIA CAPRA. "Mining mobile phone data to investigate urban crime theories at scale", en *International Conference on Social Informatics*, pp. 396 a 411, Cham (Switzerland), Springer, 2014.

FIGURA 3. EJEMPLO DE REGISTRO DE HTS



Como se indica en la figura 3, el análisis de HTS es muy importante en lo que respecta al desarrollo de la dimensión de relación, sobre todo en las organizaciones del crimen organizado. Por ejemplo, si bien uno puede afirmar que se le llama por error en una llamada única de diez segundos por el número X dentro de un período de dos meses, será incoherente afirmar que también se le llama por error durante un total de 590 horas en 47 llamadas realizadas por el número Y.

C. Hallazgos como huellas dactilares y ADN en escenas del crimen

Algunos hallazgos detectados en la escena del crimen son bastante importantes para la identificación particular, tanto las huellas dactilares como las muestras de ADN tienen una importancia separada debido al hecho que proporcionan detección para no dejar lugar a sospechas[54]. Las huellas dactilares comienzan a formarse en los hu-

54 ERKAN BOSTANCI. "3D reconstruction of crime scenes and design considerations for an interactive investigation tool", en *International Journal of Information Security Science*, vol. 4, n.° 2, 2015, pp. 50 a 58, disponible en [https://dergipark.org.tr/en/download/

manos cuando todavía son fetos. Estas se utilizan para identificar a las personas porque son únicas. En otras palabras, no hay dos personas que tengan exactamente las mismas huellas dactilares, por lo que se utilizan con fines de identificación.

Si bien las líneas papilares principales son las mismas, las lesiones dejan marcas en la punta del dedo, pero la estructura característica principal es siempre la misma[55]. Todos los datos de huellas dactilares en cuestión constituyen la batería de datos. En los casos forenses, las identidades de los sospechosos se detectan al hacer uso de los hallazgos de huellas dactilares tomadas en la escena del crimen, siempre que sea posible.

Figura 4. Visualización de la pantalla de interrogación del Sistema Automático de Identificación de Huellas Dactilares –afis–

La Figura 4 muestra una comparación de huellas dactilares en la base de datos del Sistema Automático de Identificación de Huellas Dactilares –afis–[56]. El sistema mencionado antes escanea la nueva huella dactilar cargada en su memoria y la transfiere a su propia base

article-file/147970]; David B.Wilson, David McClure y David Weisburd. "Does forensic dna help to solve crime? The benefit of sophisticated answers to naive questions", en *Journal of Contemporary Criminal Justice*, vol. 26, n.° 4, 2010, pp. 458 a 469.

55 Monowar Hussain Bhuyan, Sarat Saharia y Dhruba Kr Bhattacharyya. "An effective method for fingerprint classification", en *International Arab Journal of e-Technology*, vol. 1, n.° 3, January 2010, pp. 89 a 97, disponible en [https://arxiv.org/pdf/1211.4658].

56 Commission. *Kriminalistik*, Ankara, Gendarmerie and Coast Guard Academy, 2017.

de datos, aquí presenta a los expertos las muestras de huellas dactilares más similares del conjunto de datos mediante algoritmos de aprendizaje automático[57]. En este contexto, existen modelos exitosos de aprendizaje automático y aprendizaje profundo en la literatura, en especial, en el reconocimiento y clasificación de huellas dactilares[58].

Al comparar el adn con la base de datos, como en el ejemplo de las huellas dactilares, si hay datos similares en el conjunto de datos que contiene muestras de adn tomadas antes, se envían a los expertos para su revisión[59]. La única desventaja del adn es que las muestras son idénticas en los gemelos monocigóticos. Existen modelos de aprendizaje automático y aprendizaje profundo exitosos en la literatura, sobre todo en adn[60].

D. Elaboración de normas a partir de estadísticas delictivas y estimación de tendencias delictivas futuras

El análisis de la información estadística sobre delincuencia es una práctica que implica revelar tendencias delictivas mediante la identificación de delitos existentes y también tomar las precauciones nece-

57 Khin Nandar Win, Kenli Li, Jianguo Chen, Philippe Fournier Viger y Keqin Li. "Fingerprint classification and identification algorithms for criminal investigation: A survey", en *Future Generation Computer Systems*, vol. 110, September 2020, pp. 758 a 771.

58 Diaa M.Uliyan, Somayeh Sadeghi y Hamid A. Jalab. "Anti-spoofing method for fingerprint recognition using patch based deep learning machine", en *Engineering Science and Technology, an International Journal*, vol. 23, n.° 2, 2020, pp. 264 a 273, disponible en [https://www.sciencedirect.com/journal/engineering-science-and-technology-an-international-journal/vol/23/issue/2]; M. Arif Wani, Farooq Ahmad Bhat, Saduf Afzal y Asif Iqbal Khan. "Application of Supervised Deep Learning in Fingerprint Recognition", en *Advances in Deep Learning*, Singapore Springer, 2020, pp. 111 a 132, disponible en [https://www.researchgate.net/publication/331790153_Supervised_Deep_Learning_in_Fingerprint_Recognition]; Bhavesh Pandya, Georgina Cosma, Ali A. Alani, Aboozar Taherkhani, Vinayak Bharadi y Thomas Martin McGinnity. "Fingerprint classification using a deep convolutional neural network", en *2018 4th International Conference on Information Management*, Oxford, ieee, 2018, pp. 86 a 91.

59 Hongyan Xu. "Big data challenges in genomics", en Arni S. R. Srinivasa Rao y Calyampudi Radhakrishna Rao (eds.). *Handbook of Statistics*, vol. 43, "Principles and Methods for Data Science", Amsterdam, Elsevier, 2020, pp. 337 a 348.

60 Mohammed Aledhari, Marianne Di Pierro, Mohamed Hefeida y Fahad Saeed. "A deep learning-based data minimization algorithm for fast and secure transfer of big genomic datasets", en *ieee Transactions on Big Data*, vol. 7, n.° 2, 2018; Pui Yin Lau y Wing Kam Fung. "Evaluation of marker selection methods and statistical models for chronological age prediction based on dna methylation", en *Legal Medicine*, vol. 47, November 2020.

sarias contra dichos delitos[61], porque el delito es un comportamiento aprendido en general, también "el pasado es el presagio del futuro" para la predicción de delitos futuros[62]. La detección de conductas delictivas en la lucha contra el delito es el tema más importante de la criminología. La conducta delictiva suele surgir como resultado de diferentes motivaciones delictivas. Los sentidos son los puntos de contacto que conectan a los individuos con el mundo social, muchos comportamientos psicológicos –como el comportamiento egoísta y la violencia– surgen junto con el mundo social. La identificación y categorización de estos comportamientos implica una comprensión del estado emocional del propio sospechoso y de la sociedad que lo incluye[63]. Por esta razón, las estadísticas sobre delincuencia son importantes y se han desarrollado modelos que incluyen una gama de áreas dinámicas micro o macro que predicen el comportamiento delictivo con modelos de este tipo hipotéticos y experimentales para comprender el delito[64]. Sin embargo, el principal problema de los datos sobre delincuencia es que se presentan en forma de una pila de datos compleja y grande antes del análisis. Con los métodos de análisis estadístico-ordinarios, no siempre es fácil obtener conocimiento en este caos, incluso con la adición de variables dependientes e independientes que están en una posición específica e interactúan con otras. Por esta razón, la recuperación de conocimiento significativo de dicha pila

61 Cheng Lei. "Legal Control over Big Data Criminal Investigation", en *Social Sciences in China*, vol. 40, n.° 3, 2019, pp. 189 a 204.

62 Hongjian Wang, Daniel Kifer, Corina Graif y Zhenhui Li. "Crime rate inference with big data", en *Proceedings of the 22nd acm sigkdd International Conference on Knowledge Discovery and Data Mining*, New York, Association for Computing Machinery, 2016, pp. 635 a 644, disponible en [https://dl.acm.org/doi/pdf/10.1145/2939672.2939736].

63 Umair Saeed, Muhammad Sarim, Amna Usmani, Aniqa Mukhtar, Abdul Basit Shaikh y Sheikh Kashif Raffat. "Application of Machine learning Algorithms in Crime Classification and Classification Rule Mining", en Research Journal of Applied Sciences, vol. 4, n.° 3, March 2015, pp. 106 a 114, disponible en [https://www.researchgate.net/publication/344099883_Application_of_Machine_learning_Algorithms_in_Crime_Classification_and_Classification_Rule_Mining].

64 Thom Snaphaan y Wim Hardyns. "Environmental criminology in the big data era", en *European Journal of Criminology*, vol. 18, n.° 5, October 2019, disponible en [https://www.researchgate.net/publication/336352488_Environmental_criminology_in_the_big_data_era]; Elena Bulgakova, Vladimir Bulgakov, Igor Trushchenkov, Dmitry Vasilev y Evgeny Kravets. "Big data in investigating and preventing crimes", en Alla G. Kravets (ed.). *Big Data-driven World: Legislation Issues and Control Technologies*, pp. 61 a 69, Cham (Switzerland), Springer, 2019.

de datos es posible mediante el análisis de datos. El uso de técnicas de minería de datos brinda la capacidad de tomar medidas proactivas contra las actividades delictivas y los posibles riesgos de seguridad[65].

Muchos estudios en criminología y sociología, con independencia del tamaño del análisis específico que definen, proporcionan una cantidad significativa de conocimiento sobre la densidad criminal a niveles como la ubicación geográfica micro y macro, la estructura criminal, etc., con el análisis de los datos sobre delincuencia en cuestión:

– Se dispone de conocimiento en muchos temas, como los tipos de delitos cometidos de manera principal y durante qué época se cometen más delitos;

– Dónde se ubican los puntos críticos de delincuencia, Perfiles de sospechosos;

– Perfiles de víctimas;

– Si existe una relación o correlación entre el tipo de delito y otras variables;

– Estimación de las proporciones y frecuencia de las tasas de delincuencia en el futuro.

Los datos sobre delincuencia son un volumen de datos relativamente grande. Por ejemplo, en 2019, las Fiscalías de la República de Turquía (población: 83.154.997) recibieron 9.342.676 expedientes (transferidos desde el año anterior, el número total de expedientes recibidos durante el año)[66]. Cuando se examina esta cifra, se utilizan bastante

65 Mingchen Feng, Jiangbin Zheng, Jinchang Ren, Amir Hussain, Xiuxiu Li, Yue Xi y Qiaoyuan Liu. (2019). "Big data analytics and mining for effective visualization and trends forecasting of crime data", en *ieee Access*, vol. 7, 2019, pp. 106 a 123, disponible en [https://www.researchgate.net/publication/334617796_Big_Data_Analytics_and_Mining_for_Effective_Visualization_and_Trends_Forecasting_of_Crime_Data].

66 Turkish Ministry of Justice. *Judicial Statistics 2019*, Ankara, Turkish Statistical Institute –tüik–, disponible en [https://data.tuik.gov.tr/Kategori/GetKategori?p=adalet-ve-secim-110&dil=2]; Turkish Statistical Institute –tüik–. "The Results of Address Based Population Registration System, 2019", disponible en [https://data.tuik.gov.tr/bulten/index?p=adrese-dayali-nufus-kayit-sistemi-sonuclari-2019-33705#:~:text=-

las técnicas de minería de datos para el análisis de los correspondientes a delincuencia, basadas en la agrupación en clústeres.

Este tipo de clasificación de sospechas razonables solo es posible con el tipo de actividad de predicción y elaboración de normas que está disponible a través de la minería de datos. Dicha actividad proporcionará aquellos que expliquen el delito que se revelará. En cada delito, la determinación del perfil del delito con los modos de acción objetivos que se presentarán a partir de los datos será el concepto más importante en la lucha contra el delito[67]. Estudios sobre aprendizaje automático y aprendizaje profundo exitosos están disponibles en la literatura, sobre todo en el modelo de predicción del delito[68].

E. Análisis de los datos sujetos al delito en Internet y las redes sociales

En la era de la información, los conceptos de tiempo y espacio se han minimizado con la aparición y el desarrollo de Internet, se han producido cambios importantes en muchos campos, en especial en el campo de la comunicación. Con estos cambios, las actividades de ocio y las actitudes de las personas cambiaron, se descubrieron nuevos métodos de comunicación y muchas empresas, sobre todo de periodismo, educación, banca y servicios comerciales, comenzaron a operar en el entorno virtual. La mayoría de los trabajos recientes muestran que el uso de Internet se ha convertido en una rutina diaria y que las personas pasan mucho tiempo en la red para realizar muchas actividades diarias. Al aumentar la cantidad de dispositivos tecnológicos que poseen, las personas han llegado al punto en que pueden acceder a Internet en cualquier momento, con dispositivos como computadoras portátiles, tabletas y teléfonos móviles, ya sea en el espacio privado o en el espacio público.

Internet y las redes sociales han transformado la forma en que las personas se comunican y socializan, por lo tanto, los entornos virtuales se han convertido en plataformas populares para la comunicación. En un mundo en el que las personas utilizan cada vez más Internet y las redes sociales, la cantidad de datos que se utilizan aumenta día a día[69].

Con la aparición de Internet y las redes sociales como plataforma social, se ha observado un aumento de los delitos cometidos a través de Internet y las redes sociales en los últimos años, en particular, delitos como la pornografía infantil, el acoso cibernético, el acoso (laboral o sexual), los insultos, el fraude en Internet y el terrorismo cibernético[70]. En ocasiones, las redes sociales pueden utilizarse para organizar acciones humanas contra el Gobierno actual, como se vio en la Primavera Árabe y las protestas de Gezi (Turquía). En estos eventos, la organización de las protestas se logró a través de Twitter y los Gobiernos prohibieron Twitter. Se afirma que un sistema desarrollado para clasificar las cuentas de usuario como simpatizantes o no simpatizantes logró una precisión del 90% al analizar las cuentas de Twitter relacionadas con las protestas de Gezi[71].

Las publicaciones realizadas a través de Internet y las redes sociales también tienen un efecto en el esclarecimiento de los delitos. En un estudio realizado en Estados Unidos, más del 80% de las unidades de aplicación de la ley utilizan de manera activa las redes sociales[72].

Türkiye%20nüfusu%2083%20milyon%20154%20bin%20997%20kişi%20oldu&text=Diğer%20bir%20ifadeyle%20toplam%20nüfusun,531%20bin%20180%20kişi%20oldu.], Ankara, TÜIK, 2020.

67 JIN SOUNG YOO. "Crime data warehousing and crime pattern discovery", en *Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems 2019*, pp. 1 a 6, December 2019.

68 RICHARD BERK. "An impact assessment of machine learning risk forecasts on parole board decisions and recidivism", en *Journal of Experimental Criminology*, vol. 13, n.° 2, June 2017, pp. 193 a 216; MAMTA MITTAL, LALIT MOHAN GOYAL, DURAISAMY JUDE HEMANTH y JASLEEN KAUR SETHI. "Monitoring the impact of economic crisis on crime in India using machine learning", en *Computational Economics*, vol. 53, n.° 4, 2018, pp. 1.467 a 1.485, disponible en [https://www.researchgate.net/publication/325000522_Monitoring_the_Impact_of_Economic_Crisis_on_Crime_in_India_Using_Machine_Learning]; ANDREW P. WHEELER y WOUTER STEENBEEK. "Mapping the risk terrain for crime using machine learning", en *Journal of Quantitative Criminology*, vol. 37, n.° 2, June 2021, pp. 445 a 480.

69 AGRAHARI y RAO. "A review paper on Big Data: Technologies, tools and trends", cit.; ATEŞ, BOSTANCI y GÜZEL. "Security Evaluation of Industry 4.0: Understanding Industry 4.0 on the Basis of Crime, Big Data, Internet Of Thing (IoT) and Cyber Physical Systems", cit.

70 ROLAND HEICKERÖ. "Cyber terrorism: Electronic jihad", en *Strategic Analysis*, vol. 38, n.° 4, 2014, pp. 554 a 565.

71 Üyesi URAZ YAVANOĞLU, MEDINE COLAK, BUSRA CAGLAR, SEMRA CAKIR, OZLEM MILLETSEVER y Şeref Sağiroğlu. "Intelligent approach for identifying political views over social networks", en *2013 12th International Conference on Machine Learning and Applications*, vol. 2, pp. 281 a 287, Oxford, IEEE, 2013.

72 A. J. GUENTHER. "Role of Social Media in Law Enforcement Significant and Growing", 2012, disponible en [https://patch.com/district-columbia/georgetown/an--role-of-social-media-in-law-enforcement-significaea1c7d91fd].

Aunque no existe un estudio estadístico claro relacionado con este tema, Internet –en particular las redes sociales–, se utilizan en forma activa para todo tipo de delitos, así como para investigaciones de seguridad en la mayoría de los países del mundo. Por ejemplo, en un caso forense que se considera un suicidio, la normalidad de las publicaciones de la persona levantará la sospecha de que puede tratarse de un asesinato que se ha hecho parecer un suicidio, mientras que en los delitos de terrorismo, las publicaciones de una persona que elogia a una organización terrorista y comparte pensamientos e imágenes del líder de la organización también levantarán sospechas de que esta persona puede estar relacionada con la organización. Aparte de estos, varias aplicaciones y programas en Internet, en especial los escaneos de motores de búsqueda basados en contactos y las aplicaciones para compartir contactos (como getcontact y truecaller) pueden respaldar los datos sobre el crimen.

Además de los datos de Internet y de las redes sociales (normalmente llamados "servicios de redes sociales" –SNS–) que se utilizan a nivel personal, también se usan en tareas preventivas e informativas en el marco general[73]. Entre el 60 y el 80% de la información de inteligencia del mundo se obtiene de fuentes abiertas, por lo que el análisis de los datos de fuentes abiertas es cada vez más importante[74]. Desde los ataques terroristas del 11 de septiembre de 2001, las preocupaciones sobre la seguridad nacional han aumentado bastante. Varias unidades de inteligencia recopilan y analizan conocimientos sobre diversos temas, más las actividades de los terroristas[75]. Si se detecta un delito, se puede volver a contactar a las unidades relacionadas para que inicien procedimientos judiciales. En particular, a través de herramientas de redes sociales que permiten buscar palabras clave, también se puede examinar el intercambio general de temas de actualidad relevantes[76]. Además, existen modelos exitosos de aprendizaje automático y aprendizaje profundo en la literatura, en estudios de Internet y redes sociales[77].

73 Humaira Arshad, Aman Jantan y Oludare Esther Omolara. "Evidence collection and forensics on social networks: Research challenges and directions", en *Digital Investigation*, vol. 28, March 2019, pp. 126 a 138.

74 David J. Power. "'Big Brother' can watch us", en *Journal of Decision Systems*, vol. 25, n.° 51, (sup. 1), 2016, pp. 578 a 588, disponible en [https://www.researchgate.net/publication/304031923_Big_Brother_can_watch_us]; Chen, Mao y Liu. "Big data: A survey", cit.

75 Power. "'Big Brother' can watch us", cit.

76 Lori Bowen Ayre y Jim Craner. "Open data: What it is and why you should care", en *Public Library Quarterly*, vol. 36, n.° 2, April 2017, pp. 173 a 184.

F. Uso de propiedades biométricas en el área de seguridad

La autenticación biométrica se utiliza para medir las propiedades físicas y conductuales de las personas. Se utiliza en diversas áreas de seguridad, en particular, en la validación de identidad, al permitir que dichos valores medibles se distingan mediante sistemas de automatización[78]. La propiedad más importante del concepto de biometría es que solo se utilizan por sí mismos sin utilizar ningún objeto o dato en las identificaciones.

Los sistemas biométricos se estudian en dos grupos: físicos (pasivos) y conductuales (activos)[79]. En la biometría física, se utilizan la voz, el rostro, la geometría de la mano, la huella dactilar, el iris y la retina, mientras que características como el estilo de escritura, la firma, los patrones de marcha y los movimientos de los labios durante el habla se utilizan en la biometría conductual. En esencia, la biometría física se basa en propiedades físicas fijas de un individuo que le

77 Rupa Ch, Thippa Reddy Gadekallu, Mustafa Haider Abidi y Abdulrahman Al-Ahmari. "Computational System to Classify Cyber Crime Offenses Using Machine Learning", *Sustainability*, vol. 12, n.° 10, 2020, pp. 1 a 16, disponible en [https://www.mdpi.com/2071-1050/12/10/4087]; Matthew L.Williams, Pete Burnap, Amir Javed, Han Liu y Sefa Ozalp. "Hate in the machine: Anti-Black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime", en *The British Journal of Criminology*, vol. 60, n.° 1, January 2020, pp. 93 a 117, disponible en [https://academic.oup.com/bjc/article/60/1/93/5537169]; Alina Ristea, Mohammad Al Boni, Bernd Resch, Matthew Gerber y Michael Leitner. "Spatial crime distribution and prediction for sporting events using social media", *International Journal of Geographical Information Science*, 2020, vol. 34, n.° 9, pp. 1 a 32, disponible en [https://www.tandfonline.com/doi/full/10.1080/13658816.2020.1719495#d1e339]; Amgat Muneer y Suliman Mohamed Fati. "A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Twitter", *Future Internet*, vol. 12, n.°11, 2020, 187, disponible en [https://www.mdpi.com/1999-5903/12/11/187/pdf?version=1604400347].

78 Lauren Stewart. "Big Data Discrimination: Maintaining Protection of Individual Privacy without Disincentivizing Businesses' Use of Biometric Data to Enhance Security", *Boston College Law Review*, vol. 60, n.° 1, 2019, 349 a 386, disponible en [https://core.ac.uk/reader/200222570].

79 Shradha Tiwari, J. N. Chourasia y Vijai S. Chourasia. "A review of advancements in biometric systems", *International Journal of Innovative Research in Advanced Engineering*, vol. 2, n.° 1, pp. 187 a 204, January 2015, disponible en [https://www.ijirae.com/volumes/Vol2/iss1/30.JAEC10091.pdf].

permiten distinguirse de otras personas[80]. La biometría del comportamiento se basa en aquellos que llevan a cabo personas distintas entre sí en función de un fin concreto y en un momento determinado. La fiabilidad de estos datos biométricos es muy alta en términos de seguridad porque no son datos transferibles como ocurre con otros métodos de validación.

Los ámbitos de la vida cotidiana en los que se pueden utilizar los sistemas biométricos son por ejemplo, en los pasaportes o sistemas de imagen, en el control de fronteras, en la videovigilancia, la identificación de delincuentes, el control de acceso, el inicio de sesión en ordenadores, la verificación de usuarios en teléfonos inteligentes, el escaneo de multitudes, el comercio electrónico, la banca electrónica, la informática forense y los documentos de identidad.

La verificación biométrica es un método muy fiable porque utiliza la minería de datos basada en el aprendizaje supervisado de patrones. Además, el diseño y la simulación de dichos sistemas también se simplifican mucho al utilizar sistemas nerviosos artificiales y técnicas de procesamiento de señales. Al tener en cuenta sus aspectos positivos de mayor validación de la identidad y seguridad, se espera que la frecuencia de uso de la biometría aumente en un futuro próximo. En la literatura se encuentran modelos exitosos de aprendizaje automático y aprendizaje profundo, sobre todo, en el uso de la biometría[81].

80 Ivan Martinovic, Kasper Rasmussen, Marc Roeschlin y Gene Tsudik. "Authentication using pulse-response biometrics", en *Communications of the ACM*, vol. 60, n.° 2, 2017, pp. 108 a 115, disponible en [https://dl.acm.org/doi/pdf/10.1145/3023359].

81 Soša Adamović, Vladislav Mišković, Nermanja Maček, Milan Milosavljević, Marko Šarac, Muzafer Saračević y Milan Gnjatović. "An efficient novel approach for iris recognition based on stylometric features and machine learning techniques", en *Future Generation Computer Systems*, vol. 107, June 2020, pp. 144 a 157; Shefali Arora, M. P. S. Bhatia y Harshita Kukreja. "A Multimodal Biometric System for Secure User Identification Based on Deep Learning", en *Advances in Intelligent Systems and Computing Proceedings of Fifth International Congress on Information and Communication Technology*, pp. 95 a 103, Singapore, Springer, 2020; Rohit Kumar Pandey, Yingbo Zhou, Bhargava Urala Kota y Venu Govindaraju. "Learning Representations for Cryptographic Hash Based Face Template Protection", en Bir Bhanu y Ajay Kumar (eds.). *Deep Learning for Biometrics*, Cham (Switzerland), Springer, 2017, pp. 259 a 285; Kalaivani Sundararajan y Damon L. Woodard. "Deep Learning for Biometrics: A Survey", ACM Computing Surveys, vol. 51, n.° 3, 2018.

V. Discusión

En las operaciones que se llevan a cabo sobre datos delictivos, la prioridad es que los agentes de la ley actúen de forma proactiva y prevengan el delito antes de que se cometa. Si se previene el delito, no habrá víctimas. Sin embargo, no es fácil intervenir en el momento en que se producen las actividades delictivas. Además, como no se ha cometido ningún delito, el culpable solo recibe un pequeño castigo por el intento de acto. Se debe evitar la victimización, incluso si los sospechosos recibirán un castigo menor. En este contexto, en los últimos años se han llevado a cabo actividades de prevención del delito de gran éxito, como en plataformas digitales, al utilizar algoritmos de aprendizaje profundo que actúan sobre la base del aprendizaje automático. La repetición de conductas delictivas cometidas por las mismas personas en los mismos lugares es un tema bien conocido en la teoría de los puntos calientes y las actividades rutinarias[82]. Como resultado del análisis de datos históricos delictivos con técnicas de minería de datos, se pueden detectar estas repeticiones y se pueden establecer políticas de prevención del delito exitosas.

Una vez que se ha cometido un delito, intervienen los agentes de la ley. En esta etapa, la cuestión más importante es que cada contacto dejará un rastro, como se afirma en el principio de intercambio de Edmond Locard[83]. Los rastros en cuestión son testigos silenciosos del crimen. La escena del crimen puede ser un espacio físico o un espacio virtual. A partir de la escena del crimen, se intenta establecer una relación entre el sospechoso, la víctima, el incidente y la escena[84].

82 Guangwen Song, Wim Bernasco, Lin Liu, Luzi Xiao, Suhong Zhou y Weiwei Liao. "Crime feeds on legal activities: Daily mobility flows help to explain thieves' target location choices", en *Journal of Quantitative Criminology*, vol. 35, n.° 4, December 2019, pp. 831 a 854, disponible en [https://link.springer.com/content/pdf/10.1007/s10940-019-09406-z.pdf]; Shuyu Yao, Ming Wei, Lingyu Yan, Chunzhi Wang, Xinhua Dong, Fangrui Liu y Ying Xiong. "Prediction of Crime Hotspots based on Spatial Factors of Random Forest", en *The 15th International Conference on Computer Science & Education*, pp. 811 a 815, Oxford, ieee, 2020.

83 Ewelina Mistek, Marisia Fikiet, Shelby R. Khandasammy y Igor K. Lednev. "Toward locard's exchange principle: recent developments in forensic trace evidence analysis", en *Analytical chemistry*, vol. 91, n.° 1, 2018, pp. 637 a 654, disponible en [https://pubs.acs.org/doi/10.1021/acs.analchem.8b04704].

84 Jeroen Bode. "Every Contact Leaves a Trace: A Literary Reality of Locard's Exchange Principle", *Outside the Box: A Multi-Lingual Forum*, vol. 9, n.° 1, June 2019. pp. 18 a 22, disponible en [https://otbforum.net/otb_forum_9-1/Bode_2019_otb_forum_9-1.pdf].

Debido al desarrollo de la tecnología en los últimos años, el procesamiento de *big data* y los resultados rápidos basados en el aprendizaje automático han hecho una gran contribución en la resolución de casos criminales, mucho mayor que la lograda por el poder humano[85]. En las operaciones realizadas sobre datos criminales, hacer inferencias significativas constituye evidencia, que es la base fundamental de las autoridades judiciales en el proceso judicial. Las decisiones tomadas sobre la base de la evidencia no dejarán dudas en la mente de nadie y aumentarán la confianza de la sociedad en las autoridades judiciales.

VI. Conclusión

El crimen es uno de los problemas a los que se ha enfrentado la sociedad a lo largo de la historia de la humanidad, por esta razón, muchas áreas de la ciencia han propuesto diversos métodos y tácticas para combatir los delitos existentes. El concepto de crimen se ha transformado a lo largo de la historia de la humanidad, junto con cada nuevo método de lucha contra el crimen, pero en su esencia, se ha mantenido como un fenómeno que causa daño a la otra parte. Por esta razón, cada sociedad lleva a cabo operaciones que ayudan a la determinación del perfil criminal responsable de los actos delictivos, la detección de las causas del delito, la prevención de estos, la lucha contra ellos, la exposición a delitos particulares y los lugares donde se cometen.

Con el desarrollo de la tecnología, tenemos una gran cantidad de datos que no podemos utilizar en muchas materias, experimentamos dificultades en la clasificación de esta cantidad cada vez mayor de datos. Los delictivos relacionados con la criminología y la criminalidad son los principales, por esta razón, el proceso de minería de datos es muy importante, porque lo que importa es la recuperación de conocimiento significativo. Además de la minería de datos, junto con el desarrollo de conceptos como la inteligencia artificial, el aprendizaje automático y los avances tecnológicos, las fuerzas de seguridad han aumentado su trabajo en esta área. Los datos digitales obtenidos de estas áreas se combinan con los datos obtenidos del mundo físico, lo

85 Coleen McCue. *Data mining and predictive analysis: Intelligence gathering and crime analysis*, Oxford, Butterworth- Heinemann, 2014.

que permite a las autoridades judiciales tomar algunas determinaciones para desarrollar una hipótesis sujeta a delito.

El objetivo de este artículo fue proporcionar una descripción de las aplicaciones de la minería de datos en las áreas de criminología y criminalística. Los ejemplos dados en la sección de aplicaciones son solo algunas de las principales áreas de aplicación y no es posible limitar el tema del delito a estos ejemplos. Los autores esperan y predicen que estos métodos se utilizarán bastante tanto en un sentido preventivo como reactivo en la lucha contra el delito.

Bibliografía

Abdullah, Noraini; Saiful Adli Ismail, Siti Sophiayati y Suriani Mohd Sam. "Data quality in big data: A review", en *International Journal of Advances in Soft Computing & Its Applications*, vol. 7, n.° 3, November 2015, pp. 16 a 27, disponible en [https://www.i-csrs.org/Volumes/ijasca/IJASCA-SI-070302_Pg16-27_Data-Quality-in-Big-Data-A-Review.pdf].

Adamović, Soša; Vladislav Miškovic, Nermanja Maček, Milan Milosavljević, Marko Šarac, Muzafer Saračević y Milan Gnjatović. "An efficient novel approach for iris recognition based on stylometric features and machine learning techniques", en *Future Generation Computer Systems*, vol. 107, June 2020, pp. 144 a 157.

Adewumi, Aderemi O. y Andronicus A. Akinyelu. "A survey of machine-learning and nature-inspired based credit card fraud detection techniques", en *International Journal of System Assurance Engineering and Management*, vol. 8, n.° 2, November 2017, pp. 937 a 953, disponible en [https://link.springer.com/article/10.1007/s13198-016-0551-y].

Aggarwal, Charu C. *Machine learning for text*, Cham, Springer, 2018.

Agrahari, Anurag y Dharmaji Rao. "A review paper on Big Data: Technologies, tools and trends", en *International Research Journal of Engineering and Technology*, vol. 4, n.° 10, October 2017, pp. 640 a 649, disponible en [https://www.irjet.net/archives/V4/i10/IRJET-V4I10112.pdf].

Agu, Sunday C.; Ifeyinwa Ajah y Walter E. Ibe. "Impact of Human Character and Information System on Corruption Risk in Nigeria", en *International Journal of Scientific Research and Engineering Development*, vol. 2, n.° 4, July-August 2019, pp. 481 a 485, disponible en [https://www.madonnauniversity.edu.ng/wp-content/uploads/2019/09/IJSRED-V2I4P50-Sunday-Agu.pdf].

Ahmed, Adel. "'From Data to Wisdom' Using Machine Learning Capabilities in Accounting and Finance Professionals", en *Talent Development & Excellence*, n.° 12, 2020.

Ajans Press. "Türkiye cep telefonuyla konuşmada Avrupa birincisi", en *cnnturk.com*, 29 de noviembre de 2017, disponible en [https://www.cnnturk.com/teknoloji/turkiye-cep-telefonuyla-konusmada-avrupa-birincisi-705837].

Akdemir, Naci y Christopher James Lawless. "Exploring the human factor in cyber-enabled and cyber-dependent crime victimisation: A lifestyle routine activities approach", en *Internet Research*, vol. 30, n.° 6, June 2020, pp. 1665 a 1687, disponible en [https://www.semanticscholar.org/reader/f2d52e0caa51f6adb42acb32ace0f305b44058c9].

Aledhari, Mohammed; Marianne Di Pierro, Mohamed Hefeida y Fahad Saeed. "A deep learning-based data minimization algorithm for fast and secure transfer of big genomic datasets", en *ieee Transactions on Big Data*, vol. 7, n.° 2, 2018.

Arora, Shefali; M. P. S. Bhatia y Harshita Kukreja. "A Multimodal Biometric System for Secure User Identification Based on Deep Learning", en *Advances in Intelligent Systems and Computing Proceedings of Fifth International Congress on Information and Communication Technology*, pp. 95 a 103, Singapore, Springer, 2020.

Arshad, Humaira; Aman Jantan y Oludare Esther Omolara. "Evidence collection and forensics on social networks: Research challenges and directions", en *Digital Investigation*, vol. 28, March 2019, pp. 126 a 138.

Ateş, Emre Cihan; Gazy Erkan Bostanci y Mehmet Serdar Güzel. "Big data, data mining, Machine Learnig, and deep learning concepts in crime data", en *Ceza Hukuku ve Kriminoloji Dergisi-Journal of Penal Law and Crimin*ology, vol. 8, n.° 2, 2020, pp. 293 a 319, disponible en [https://dergipark.org.tr/en/download/article-file/1354184].

Ateş, Emre Cihan; Gazy Erkan Bostanci y Mehmet Serdar Güzel. "Security Evaluation of Industry 4.0: Understanding Industry 4.0 on the Basis of Crime, Big Data, Internet Of Thing (IoT) and Cyber Physical Systems", en *Güvenlik Bilimleri Dergisi*, (International Security Congress Special Issue), n.° 9, February 2020, pp. 29 a 50, disponible en [https://www.researchgate.net/publication/339642922_Security_Evaluation_of_Industry_40_Understanding_Industry_40_on_the_Basis_of_Crime_Big_Data_Internet_of_Thing_IoT_and_Cyber_Physical_Systems].

Ayre, Lori Bowen y Jim Craner. "Open data: What it is and why you should care", en *Public Library Quarterly*, vol. 36, n.° 2, April 2017, pp. 173 a 184.

Beniwal, Sunita y Jitender Arora. "Classification and feature selection techniques in data mining", en *International Journal of Engineering Research & Technology*, vol. 1, n.° 6, August 2012, pp. 1 a 6, disponible en [https://www.researchgate.net/publication/263662705_Classification_and_feature_selection_techniques_in_data_mining].

Berk, Richard. "An impact assessment of machine learning risk forecasts on parole board decisions and recidivism", en *Journal of Experimental Criminology*, vol. 13, n.° 2, June 2017, pp. 193 a 216.

Berkhin, Pavel. "A survey of clustering data mining techniques", en *Grouping multidimensional data*, Berlin-Heidelberg, Springer, 2006, pp. 25 a 71.

Bhanu, Bir y Ajay Kumar (eds.). *Deep Learning for Biometrics*, Cham (Switzerland), Springer, 2017.

Bhuyan, Monowar Hussain; Sarat Saharia y Dhruba Kr Bhattacharyya. "An effective method for fingerprint classification", en *International Arab Journal of e-Technology*, vol. 1, n.° 3, January 2010, pp. 89 a 97, disponible en [https://arxiv.org/pdf/1211.4658].

Blei, David M. y Padhraic Smyth. "Science and data science", *Proceedings of the National Academy of Sciences*, vol. 114, n.° 33, August 2017, pp. 8.689 a 8.692, disponible en [https://www.pnas.org/doi/epdf/10.1073/pnas.1702076114].

Bock, Frederic E.; Roland C. Aydin, Christian J. Cyron, Norbert Huber, Surya R. Kalidindi y Benjamin Klusemann. "A review of the application of machine learning and data mining approaches in continuum materials mechanics", en *Frontiers in Materials*, vol. 6, May 2019, article 110, pp. 1 a 23, disponible en [https://www.frontiersin.org/journals/materials/articles/10.3389/fmats.2019.00110/full].

Bode, Jeroen. "Every Contact Leaves a Trace: A Literary Reality of Locard's Exchange Principle", *Outside the Box: A Multi-Lingual Forum*, vol. 9, n.° 1, June 2019. pp. 18 a 22, disponible en [https://otbforum.net/otb_forum_9-1/Bode_2019_otb_forum_9-1.pdf].

Bostanci, Erkan. "3D reconstruction of crime scenes and design considerations for an interactive investigation tool", en *International Journal of Information Security Science*, vol. 4, n.° 2, 2015, pp. 50 a 58, disponible en [https://dergipark.org.tr/en/download/article-file/147970].

Boyd, Danah y Kate Crawford. "Critical questions for big data: Provocations for a cultural, technological and scholarly phenomenon", en *Information, Communication & Society*, vol. 15, n.° 5, January 2012, pp. 662 a 679, disponible en [https://www.researchgate.net/publication/281748849_Critical_questions_for_big_data_Provocations_for_a_cultural_technological_and_scholarly_phenomenon].

Bulgakova, Elena; Vladimir Bulgakov, Igor Trushchenkov, Dmitry Vasilev y Evgeny Kravets. "Big data in investigating and preventing crimes", en Alla G. Kravets (ed.). *Big Data-driven World: Legislation Issues and Control Technologies*, pp. 61 a 69, Cham (Switzerland), Springer, 2019.

Campbell, Colin y Yiming Ying. "Learning with support vector machines", *Synthesis Lectures on Artificial Intelligence and Machine Learning*, n.° 10, 1 a 83, San Rafael, CA, Morgan & Claypool, 2011, disponible en [https://digi-lib.stekom.ac.id/assets/dokumen/ebook/feb_B8yCNdjgUActi5qdyUbB6LyV7949c7RegDGg8jiFkgjSqd8C_HvSSZE_1681201865.pdf].

Ch, Rupa; Thippa Reddy Gadekallu, Mustafa Haider Abidi y Abdulrahman Al-Ahmari. "Computational System to Classify Cyber Crime Offenses Using Machine Learning", *Sustainability*, vol. 12, n.° 10, 2020, pp. 1 a 16, disponible en [https://www.mdpi.com/2071-1050/12/10/4087].

Chan, Janet y Lyria Bennett Moses. "Making sense of big data for security", en *The British Journal of Criminology*, vol. 57, n.° 2, March 2017, pp. 299 a 319, disponible en [https://uploads-ssl.webflow.com/5cd23e823ab9b1f01f815a54/5cff2fe1af68c502e4474e4b_Making%20Sense%20of%20Big%20Data%20for%20Security.pdf].

CHAU, DUEN HORNG; SHASHANK PANDIT y CHRISTOS FALOUTSOS. "Detecting fraudulent personalities in networks of online auctioneers", en *European Conference on Principles of Data Mining and Knowledge Discovery*, pp. 103 a 114, Berlin-Heidelberg, Springer, 2006, disponible en [https://link.springer.com/content/pdf/10.1007/11871637_14.pdf].

CHEN, MIN; SHIWEN MAO y YUNHAO LIU. "Big data: A survey", *Mobile Networks and Applications*, vol. 19, n.° 2, April 2014, pp. 171 a 209.

CHOO, KIM-KWANG RAYMOND y ALI DEHGHANTANHA (eds.). *Contemporary Digital Forensic Investigations of Cloud and Mobile Applications*, Cambridge, MA, Syngress, 2017.

CLARKE, CURTIS. "Proactive policing: Standing on the shoulders of community-based policing", en *Police Practice and Research*, vol. 7, n.° 1, March 2006, pp. 3 a 17.

COMMISSION. *Kriminalistik*, Ankara, Gendarmerie and Coast Guard Academy, 2017.

COOPER, PAUL. "Data, information, knowledge and wisdom", en *Anaesthesia & Intensive Care Medicine*, vol. 18, n.° 1, 2017, pp. 55 y 56.

DEY, AYON. "Machine learning algorithms: A review", en *International Journal of Computer Science and Information Technologies*, vol. 7, n.° 3, 2016, pp. 1.174 a 1.179, disponible en [https://ijcsit.com/docs/Volume%207/vol7issue3/ijcsit2016070332.pdf].

FENG, MINGCHEN; JIANGBIN ZHENG, JINCHANG REN, AMIR HUSSAIN, XIUXIU LI, YUE XI y QIAOYUAN LIU. (2019). "Big data analytics and mining for effective visualization and trends forecasting of crime data", en *IEEE Access*, vol. 7, 2019, pp. 106 a 123, disponible en [https://www.researchgate.net/publication/334617796_Big_Data_Analytics_and_Mining_for_Effective_Visualization_and_Trends_Forecasting_of_Crime_Data].

GE, ZHIQIANG; ZHIHUAN SONG, STEVEN X. DING y BIAO HUANG. "Data mining and analytics in the process industry: The role of machine learning", en *IEEE Access*, vol. 5, 2017, pp. 20.590 a 20.616, disponible en [https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8051033].

GHORBANI, RAMIN y ROUZBEH GHOUSI. "Predictive data mining approaches in medical diagnosis: A review of some diseases prediction", en *International Journal of Data and Network Science*, vol. 3, 2019, pp. 47 a 70, disponible en [https://www.growingscience.com/ijds/Vol3/ijdns_2019_12.pdf].

GUENTHER, A. J. "Role of Social Media in Law Enforcement Significant and Growing", 2012, disponible en [https://patch.com/district-columbia/georgetown/an--role-of-social-media-in-law-enforcement-significaea1c7d91fd].

GUPTA, MANOJ KR. y PRAVIN CHANDRA, P. "A comparative study of clustering algorithms", en *2019 6th International Conference on Computing for Sustainable Global Development (indiacom)*, New Delhi, IEEE Xplore, 13-15 March, 2019, pp. 801 a 805.

HAND, DAVID J. y NIALL M. ADAMS. "Data Mining", en *Wiley StatsRef: Statistics Reference Online*, Hoboken, NJ, John Wiley & Sons, 2014, pp. 1 a 7.

HASSANI, HOSSEIN; XU HUANG, EMMANUEL SIRIMAL SILVA y MANSI GHODSI. "A review of data mining applications in crime", en *Statistical Analysis and Data Mining, The ASA Data Science Journal*, vol. 9, n.° 3, April 2016, disponible en [https://www.researchgate.net/publication/301579904_A_Review_of_Data_Mining_Applications_in_Crime].

HE, LI; ANTONIO PÁEZ, JIANMIN JIAO, PING AN, CHUNTIAN LU, WEN MAO y DONGPIN LONG. "Ambient Population and Larceny-Theft: A Spatial Analysis Using Mobile Phone Data", *ISPRS International Journal of Geo-Information*, vol. 9, n.° 6, 2020, disponible en [https://www.mdpi.com/2220-9964/9/6/342/pdf?version=1591598730].

HEICKERÖ, ROLAND. "Cyber terrorism: Electronic jihad", en *Strategic Analysis*, vol. 38, n.° 4, 2014, pp. 554 a 565.

HEY, JONATHAN. "The data, information, knowledge, wisdom chain: The metaphorical link", Intergovernmental Oceanographic Commission, 26, 2004, pp. 1 a 18, disponible en [https://www.docsity.com/en/docs/the-data-information-knowledge-wisdom-chain/9588737/].

IBM. "¿Qué es una base de datos NOSQL?", disponible en [https://www.ibm.com/es-es/topics/nosql-databases].

JACKSON, JOYCE. "Data mining; a conceptual overview", en *Communications of the Association for Information Systems*, vol. 8, 2002, pp. 267 a 296, disponible en [https://www.academia.edu/5442707/DATA_MINING_A_CONCEPTUAL_OVERVIEW#:~:text=Download,-PDF].

KELLEHER, JOHN D. y BRENDAN TIERNEY. *Data science*, Massachusetts, MIT Press, 2018, disponible en [https://mrce.in/ebooks/Data%20Science.pdf].

KEMP, SIMON. "Digital in 2018: World's Internet users pass the 4 billion mark", en London, *We are social*, 30 January 2018, disponible en [https://wearesocial.com/uk/blog/2018/01/global-digital-report-2018/].

KHARE, AKHILL RAJENDRA y PALLAVI SHRIVASTA. "Data mining for the internet of things", en AMBATI V. KRISHNA PRASAD (ed.). *Exploring the Convergence of Big Data and the Internet of Things*, pp. 181 a 191, Hershey, PA, IGI Global, 2018.

KOYUNCUGİL, ALİ SERHAN y NERMİN ÖZGÜLBAŞ. "Veri madenciliği: Tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları", en *International Journal of Informatics Technologies*, vol. 2, n.° 2, May 2009, pp. 21 a 32, disponible en [https://dergipark.org.tr/tr/download/article-file/75259].

KRAVETS, ALLA G. (ed.). *Big Data-driven World: Legislation Issues and Control Technologies*, Cham, Springer, 2019.

KRISHNA PRASAD, AMBATI V. (ed.). *Exploring the Convergence of Big Data and the Internet of Things*, pp. 181 a 191, Hershey, PA, IGI Global, 2018.

KUMAR, RAVI y BHARTI NAGPAL. "Analysis and prediction of crime patterns using big data", *International Journal of Information Technology*, vol. 11, n.° 4, December 2019, pp. 799 a 805.

LAU, PUI YIN y WING KAM FUNG. "Evaluation of marker selection methods and statistical models for chronological age prediction based on DNA methylation", en *Legal Medicine*, vol. 47, November 2020.

LEI, CHENG. "Legal Control over Big Data Criminal Investigation", en *Social Sciences in China*, vol. 40, n.° 3, 2019, pp. 189 a 204.

LI, XIN; BING LIU y PHILIP S. YU. "Discovering overlapping communities of named entities", en *Knowledge Discovery in Databases: PKDD 2006*, pp. 593 a 600, Berlin-Heidelberg, Springer, 2006, disponible en [https://link.springer.com/chapter/10.1007/11871637_60].

MARTINOVIC, IVAN; KASPER RASMUSSEN, MARC ROESCHLIN y GENE TSUDIK. "Authentication using pulse-response biometrics", en *Communications of the ACM*, vol. 60, n.° 2, 2017, pp. 108 a 115, disponible en [https://dl.acm.org/doi/pdf/10.1145/3023359].

MCCLENDON, LAWRENCE y NATARAJAN MEGHANATHAN. "Using machine learning algorithms to analyze crime data", en *Machine Learning and Applications: An International Journal*, vol. 2, n.° 1, March 2015, pp. 1 a 12, disponible en [https://www.researchgate.net/publication/275220711_Using_Machine_Learning_Algorithms_to_Analyze_Crime_Data].

MCCUE, COLEEN. *Data mining and predictive analysis: Intelligence gathering and crime analysis*, Oxford, Butterworth- Heinemann, 2014.

MESGARPOUR, MOHAMMAD e IAN DICKINSON. "Enhancing the value of commercial vehicle telematics data through analytics and optimisation techniques", en *Archives of Transport System Telematics*, vol. 7, n.° 3, September 2014, pp. 27 a 30, disponible en [https://bibliotekanauki.pl/articles/393920.pdf].

MISTEK, EWELINA; MARISIA FIKIET, SHELBY R. KHANDASAMMY y IGOR K. LEDNEV. "Toward locard's exchange principle: recent developments in forensic trace evidence analysis", en *Analytical chemistry*, vol. 91, n.° 1, 2018, pp. 637 a 654, disponible en [https://pubs.acs.org/doi/10.1021/acs.analchem.8b04704].

MITTAL, MAMTA; LALIT MOHAN GOYAL, DURAISAMY JUDE HEMANTH y JASLEEN KAUR SETHI. "Clustering approaches for high-dimensional databases: A review", en *Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery*, vol. 9, n.° 3, 2019, pp. 1 a 14.

MITTAL, MAMTA; LALIT MOHAN GOYAL, DURAISAMY JUDE HEMANTH y JASLEEN KAUR SETHI. "Monitoring the impact of economic crisis on crime in India using machine learning", en *Computational Economics*, vol. 53, n.° 4, 2018, pp. 1.467 a 1.485, disponible en [https://www.researchgate.net/publication/325000522_Monitoring_the_Impact_of_Economic_Crisis_on_Crime_in_India_Using_Machine_Learning].

MUKHOPADHYAY, ANIRBAN; UJJWAL MAULIK, SANGHAMITRA BANDYOPADHYAY y CARLOS ARTEMIO COELLO COELLO. "A survey of multiobjective evolutionary algorithms for data mining: Part I", en *IEEE Transactions on Evolutionary Computation*, vol. 18, n.° 1, 2013, pp. 4 a 19.

MUNEER, AMGAT y SULIMAN MOHAMED FATI. "A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Twitter", *Future Internet*, vol. 12, n.°11, 2020, 187, disponible en [https://www.mdpi.com/1999-5903/12/11/187/pdf?version=1604400347].

NGAI, WAI TING ERIC; LI XIU y DOROTHY C. K. CHAU. "Application of data mining techniques in customer relationship management: A literature review and classification", *Expert Systems with Applications*, vol. 36, n.° 2, March 2009, pp. 2.592 a 2.602, disponible en [https://didawiki.di.unipi.it/lib/exe/fetch.php/dm/crm_dm-survey.pdf].

ODIA, JAMES OSABUOHIEN y OSAHEMI THADDEUS AKPATA. "Role of Data Science and Data Analytics in Forensic Accounting and Fraud Detection", en BHUSHAN PATIL y MANISHA VOHRA (eds.). *Handbook of Research on Engineering, Business, and Healthcare Applications of Data Science and Analytics*, 2020, Hershey, PA, IGI Global, 2020, pp. 203 a 227.

OLSON, DAVID L. y GEORG LAUHOFF. *Descriptive Data Mining*, Singapore, Springer, 2019.

PANDEY, ROHIT KUMAR; YINGBO ZHOU, BHARGAVA URALA KOTA y VENU GOVINDARAJU. "Learning Representations for Cryptographic Hash Based Face Template Protection", en BIR BHANU y AJAY KUMAR (eds.). *Deep Learning for Biometrics*, Cham (Switzerland), Springer, 2017, pp. 259 a 285.

PANDYA, BHAVESH; GEORGINA COSMA, ALI A. ALANI, ABOOZAR TAHERKHANI, VINAYAK BHARADI y THOMAS MARTIN MCGINNITY. "Fingerprint classification using a deep convolutional neural network", en *2018 4th International Conference on Information Management*, Oxford, IEEE, 2018, pp. 86 a 91.

PATIL, BHUSHAN y MANISHA VOHRA (eds.). *Handbook of Research on Engineering, Business, and Healthcare Applications of Data Science and Analytics*, 2020, Hershey, PA, IGI Global, 2020.

PAULEEN, DAVID J.; DAVID ROONEY y ALI INTEZARI. "Big data, little wisdom: Trouble brewing? Ethical implications for the information systems discipline", *Social Epistemology*, vol. 31, n.° 4, 2017, pp. 400 a 416.

PEROLS, JOHAN. "Financial statement fraud detection: An analysis of statistical and machine learning algorithms", *Auditing: A Journal of Practice & Theory*, vol. 30, n.° 2, May 2011, pp. 19 a 50, disponible en [https://www.researchgate.net/publication/256045322_Financial_Statement_Fraud_Detection_An_Analysis_of_Statistical_and_Machine_Learning_Algorithms].

POWER, DAVID J. "'Big Brother' can watch us", en *Journal of Decision Systems*, vol. 25, n.° 51, (sup. 1), 2016, pp. 578 a 588, disponible en [https://www.researchgate.net/publication/304031923_Big_Brother_can_watch_us].

QUICK, DARREN y KIM-KWANG RAYMOND CHOO. "Big forensic data reduction: Digital forensic images and electronic evidence", *Cluster Computing*, vol. 19, n.° 2, 2016, pp. 723 a 740.

SRINIVASA RAO, ARNI S. R. y CALYAMPUDI RADHAKRISHNA RAO (eds.). *Handbook of Statistics*, vol. 43, "Principles and Methods for Data Science", Amsterdam, Elsevier, 2020

RISTEA, ALINA; MOHAMMAD AL BONI, BERND RESCH, MATTHEW GERBER y MICHAEL LEITNER. "Spatial crime distribution and prediction for sporting events using social media",

International Journal of Geographical Information Science, 2020, vol. 34, n.° 9, pp. 1 a 32, disponible en [https://www.tandfonline.com/doi/full/10.1080/13658816.2020.1719495#d1e339].

Roy, Abhimanyu; Jingyi Sun, Robert Mahoney, Loreto Alonzi, Stephen Adams y Peter Beling. "Deep learning detecting fraud in credit card transactions", en *2018 Systems and Information Engineering Design Symposium*, Oxford, ieee, 2018, pp. 129 a 134.

Rutkowski, LLeszek; Maciej Jaworski y Piotr Duda. *Stream data mining: Algorithms and their probabilistic properties*, Cham (Switzerland), Springer, 2020.

Saeed, Umair; Muhammad Sarim, Amna Usmani, Aniqa Mukhtar, Abdul Basit Shaikh y Sheikh Kashif Raffat. "Application of Machine learning Algorithms in Crime Classification and Classification Rule Mining", en Research Journal of Applied Sciences, vol. 4, n.° 3, March 2015, pp. 106 a 114, disponible en [https://www.researchgate.net/publication/344099883_Application_of_Machine_learning_Algorithms_in_Crime_Classification_and_Classification_Rule_Mining].

Shao, Lixu; Yucong Duan, Xiaobing Sun, Honghao Gao, Donghai Zhu y Welkal Miao. "Answering Who/When, What, How, Why through Constructing Data Graph, Information Graph, Knowledge Graph and Wisdom Graph", en *Seke*, July 2017, pp. 1 a 6, disponioble en [https://ksiresearch.org/seke/seke17paper/seke17paper_79.pdf].

Snaphaan, Thom y Wim Hardyns. "Environmental criminology in the big data era", en *European Journal of Criminology*, vol. 18, n.° 5, October 2019, disponible en [https://www.researchgate.net/publication/336352488_Environmental_criminology_in_the_big_data_era].

Song, Guangwen; Wim Bernasco, Lin Liu, Luzi Xiao, Suhong Zhou y Weiwei Liao. "Crime feeds on legal activities: Daily mobility flows help to explain thieves' target location choices", en *Journal of Quantitative Criminology*, vol. 35, n.° 4, December 2019, pp. 831 a 854, disponible en [https://link.springer.com/content/pdf/10.1007/s10940-019-09406-z.pdf].

Srinivas, Konda; B. Kavihta Rani y Aliseri Govrdhan. "Applications of data mining techniques in healthcare and prediction of heart attacks", en *International Journal on Computer Science and Engineering*, vol. 2, n.° 2, 2010, pp. 250 a 255, disponible en [https://www.researchgate.net/publication/49617135_Applications_of_Data_Mining_Techniques_in_Healthcare_and_Prediction_of_Heart_Attacks].

Srinivasa Rao, Arni S. R. y Calyampudi Radhakrishna Rao (eds.). *Handbook of Statistics*, vol. 43, "Principles and Methods for Data Science", Amsterdam, Elsevier, 2020

Steenbruggen, John; Emmanouil Tranos y Peter Nijkamp. "Data from mobile phone operators: A tool for smarter cities?", *Telecommunications Policy*, vol. 39, n.os 3 y 4, May 2015, pp. 335 a 346.

Stewart, Lauren. "Big Data Discrimination: Maintaining Protection of Individual Privacy without Disincentivizing Businesses' Use of Biometric Data to Enhance Security", *Boston College Law Review*, vol. 60, n.° 1, 2019, 349 a 386, disponible en [https://core.ac.uk/reader/200222570].

Sundararajan, Kalaivani y Damon L. Woodard. "Deep Learning for Biometrics: A Survey", ACM Computing Surveys, vol. 51, n.° 3, 2018.

Tassone, Christopher; Ben Martini y Kim-Kwang Raymond Choo. "Forensic visualization: Survey and future research directions", en Kim-Kwang Raymond Choo y Ali Dehghantanha (eds.). *Contemporary Digital Forensic Investigations of Cloud and Mobile Applications*, Cambridge, MA, Syngress, 2017, pp. pp. 163 a 184.

Tilley, Nick y Aiden Sidebottom. *Handbook of crime prevention and community safety*, New York, Routledge, 2017.

Tirgari, Vesal. "Information technology policies and procedures against unstructured data: A phenomenological study of information technology professionals", en *Journal of Management Information and Decision Sciences*, vol. 15, n.° 2, 2012, pp. 87 a 106, disponible en [https://www.abacademies.org/articles/aimsjvol15no22012.pdf].

Tiwari, Shradha; J. N. Chourasia y Vijai S. Chourasia. "A review of advancements in biometric systems", *International Journal of Innovative Research in Advanced Engineering,* vol. 2, n.° 1, pp. 187 a 204, January 2015, disponible en [https://www.ijirae.com/volumes/Vol2/iss1/30.JAEC10091.pdf].

Traunmueller, Martin W.; Giovanni Quattrone y Licia Capra. "Mining mobile phone data to investigate urban crime theories at scale", en *International Conference on Social Informatics*, pp. 396 a 411, Cham (Switzerland), Springer, 2014.

Turkish Ministry of Justice. *Judicial Statistics 2019*, Ankara, Turkish Statistical Institute –tüik–, disponible en [https://data.tuik.gov.tr/Kategori/GetKategori?p=adalet-ve-secim-110&dil=2].

Turing, Allan Mathison. "Computing machinery and intelligence", en *Mind*, vol. lix, n.° 236, October 1950, pp. 433 a 460, disponible en [https://courses.cs.umbc.edu/471/papers/turing.pdf].

Turkish Statistical Institute –tüik–. "The Results of Address Based Population Registration System, 2019", disponible en [https://data.tuik.gov.tr/bulten/index?p=adrese-dayali-nufus-kayit-sistemi-sonuclari-2019-33705#:~:text=Türkiye%20nüfusu%2083%20milyon%20154%20bin%20997%20kişi%20oldu&text=Diğer%20bir%20ifadeyle%20toplam%20nüfusun,531%20bin%20180%20kişi%20oldu.], Ankara, tüik, 2020.

Uliyan, Diaa M.; Somayeh Sadeghi y Hamid A. Jalab. "Anti-spoofing method for fingerprint recognition using patch based deep learning machine", en *Engineering Science and Technology, an International Journal*, vol. 23, n.° 2, 2020, pp. 264 a 273, disponible en [https://www.sciencedirect.com/journal/engineering-science-and-technology-an-international-journal/vol/23/issue/2].

Vaidhyanathan, Siva y Chris Bulock. "Knowledge and dignity in the era of 'big data'", en *The Serials Librarian*, vol. 66, n.os 1 a 4, 2014, pp. 49 a 64, disponible en [https://www.tandfonline.com/doi/epdf/10.1080/0361526X.2014.879805].

WAMBA, SAMUEL FOSSO; SHAHRIAR AKTER, ANDREW JAMES EDWARDS, GEOFFREY CHOPIN y DENIS GNANZOU. "How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study", en *International Journal of Production Economics*, vol. 165, July 2015, pp. 234 a 246, disponible en [https://www.researchgate.net/publication/270276259_How_'big_data'_can_make_big_impact_Findings_from_a_systematic_review_and_a_longitudinal_case_study].

WANG, HONGJIAN; DANIEL KIFER, CORINA GRAIF y ZHENHUI LI. "Crime rate inference with big data", en *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, Association for Computing Machinery, 2016, pp. 635 a 644, disponible en [https://dl.acm.org/doi/pdf/10.1145/2939672.2939736].

WANI, M. ARIF; FAROOQ AHMAD BHAT, SADUF AFZAL y ASIF IQBAL KHAN. "Application of Supervised Deep Learning in Fingerprint Recognition", en *Advances in Deep Learning*, Singapore Springer, 2020, pp. 111 a 132, disponible en [https://www.researchgate.net/publication/331790153_Supervised_Deep_Learning_in_Fingerprint_Recognition].

WHEELER, ANDREW P. y WOUTER STEENBEEK. "Mapping the risk terrain for crime using machine learning", en *Journal of Quantitative Criminology*, vol. 37, n.° 2, June 2021, pp. 445 a 480.

WHITE, MARTIN. "Digital workplaces: Vision and reality", en *Business Information Review*, vol. 29, n.° 4, 2012, pp. 205 a 214, disponible en [https://es.scribd.com/document/886559974/Digital-Workplaces-Vision-and-Reality].

WILLIAMS, GRAHAM J. "Rattle: a data mining GUI for R", en *The R Journal*, vol. 1, n.° 2, 2009, pp. 45 a 55, disponible en [https://journal.r-project.org/archive/2009-2/RJournal_2009-2_Williams.pdf].

WILLIAMS, MATTHEW L.; PETE BURNAP, AMIR JAVED, HAN LIU y SEFA OZALP. "Hate in the machine: Anti-Black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime", en *The British Journal of Criminology*, vol. 60, n.° 1, January 2020, pp. 93 a 117, disponible en [https://academic.oup.com/bjc/article/60/1/93/5537169].

WILSON, DAVID B.; DAVID MCCLURE y DAVID WEISBURD. "Does forensic DNA help to solve crime? The benefit of sophisticated answers to naive questions", en *Journal of Contemporary Criminal Justice*, vol. 26, n.° 4, 2010, pp. 458 a 469.

WIN, KHIN NANDAR; KENLI LI, JIANGUO CHEN, PHILIPPE FOURNIER VIGER y KEQIN LI. "Fingerprint classification and identification algorithms for criminal investigation: A survey", en *Future Generation Computer Systems*, vol. 110, September 2020, pp. 758 a 771.

XU, HONGYAN. "Big data challenges in genomics", en ARNI S. R. SRINIVASA RAO y CALYAMPUDI RADHAKRISHNA RAO (eds.). *Handbook of Statistics*, vol. 43, "Principles and Methods for Data Science", Amsterdam, Elsevier, 2020, pp. 337 a 348.

YAO, SHUYU; MING WEI, LINGYU YAN, CHUNZHI WANG, XINHUA DONG, FANGRUI LIU y YING XIONG. "Prediction of Crime Hotspots based on Spatial Factors of Random Forest", en *The 15th International Conference on Computer Science & Education*, pp. 811 a 815, Oxford, IEEE, 2020.

YAVANOĞLU, ÜYESI URAZ; MEDINE COLAK, BUSRA CAGLAR, SEMRA CAKIR, OZLEM MILLETSEVER y Şeref Sağiroğlu. "Intelligent approach for identifying political views over social networks", en *2013 12th International Conference on Machine Learning and Applications*, vol. 2, pp. 281 a 287, Oxford, IEEE, 2013.

YOO, JIN SOUNG. "Crime data warehousing and crime pattern discovery", en *Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems 2019*, pp. 1 a 6, December 2019.

ZINS, CHAIM. "Conceptual approaches for defining data, information, and knowledge", en *Journal of the American Society for Information Science and Technology*, vol. 58, n.° 4, 2007, pp. 479 a 493, disponible en [https://www.researchgate.net/publication/220432993_Conceptual_approaches_for_defining_data_information_and_knowledge].