

La minería de datos en investigación social. Patrones y reglas de asociación —ciencia social computacional—

Antonio Alaminos-Fernández

(Alicante, Editorial Limencop S.L., 2023)

El libro de ciencia social computacional *La Minería de datos en investigación social. Patrones y reglas de asociación* es un recurso útil y completo para los investigadores, estudiantes y profesionales interesados en aprender las técnicas y aplicaciones de la minería de datos. Esta monografía ha sido realizada durante la estancia de investigación del autor en el GESIS-Leibniz-Institut für Sozialwissenschaften en Colonia (Alemania), fruto del contrato de la convocatoria Margarita Salas en el Departamento de Sociología II de la Universidad de Alicante.

Esta obra presenta un contenido bien estructurado que aborda una reflexión sobre el registro de la actividad social humana, ya sean defunciones, nacimientos o matrimonios en siglos pasados o de tráfico de información en redes sociales en la actualidad, y el desarrollo de métodos y técnicas para el análisis de grandes conjuntos de datos. Como expresa el prologuista de la obra, nuevos tipos de datos implican, casi mecánicamente, métodos de análisis nuevos, y a su vez formas diferentes y alternativas de observar la realidad. Es aquí donde aparece la minería de datos en investigación social desde una nueva mirada para la detección de patrones no secuenciales y estructuras regulares.

En ciencias sociales, la minería de datos se puede utilizar para descubrir patrones y relaciones en datos recopilados de encuestas, estudios de caso, experimentos y otras fuentes. Entre los patrones que más se han estudiado por medio de técnicas de minería de datos en ciencias sociales, caben destacar los siguientes:

1. *Patrones de comportamiento de los consumidores*: la minería de datos puede utilizarse para descubrir patrones en el comportamiento de los consumidores, como las compras realizadas en un sitio web o las respuestas a una encuesta de satisfacción del cliente. Estos patrones pueden utilizarse para mejorar la experiencia del cliente y aumentar las ventas.
2. *Patrones de enfermedades*: la minería de datos puede utilizarse para descubrir patrones en la incidencia de enfermedades en una población determinada. Estos patrones pueden utilizarse para identificar factores de riesgo y desarrollar estrategias para prevenir la enfermedad.
3. *Patrones de delincuencia*: la minería de datos puede utilizarse para descubrir patrones en la delincuencia, como las áreas donde se producen la mayoría de los delitos o el tipo de delitos más comunes. Estos patrones pueden utilizarse para desarrollar estrategias para prevenir la delincuencia y mejorar la seguridad pública.

4. *Patrones comportamiento electoral*: la minería de datos puede utilizarse para descubrir patrones de comportamiento electoral, como las características demográficas de los votantes que tienden a apoyar a un partido político o candidato en particular. Estos patrones pueden utilizarse para desarrollar estrategias de campaña más efectivas.
5. *Patrones de comportamiento en redes sociales*: la minería de datos puede utilizarse para descubrir patrones en el comportamiento de los usuarios de las redes sociales, como las publicaciones más populares o los temas de conversación más comunes. Estos patrones pueden utilizarse para mejorar la eficacia de las campañas de *marketing* y publicidad en línea.

El libro comienza con una introducción que proporciona una visión general de los conceptos básicos de la minería de datos. La minería de datos es una técnica que se utiliza para descubrir patrones y relaciones en grandes conjuntos de datos. El libro comienza explicando los conceptos fundamentales de la minería de datos, incluyendo la definición de patrones y reglas de asociación. El texto se enfoca en los patrones no secuenciales, que son aquellos patrones que no están relacionados con el orden de las transacciones en una secuencia, sino que se enfocan en la presencia simultánea de varios elementos en diferentes transacciones. Estos patrones se pueden descubrir mediante la técnica de minería de datos de «reglas de asociación», que es una técnica comúnmente utilizada en la minería de datos. En este sentido, la minería de datos es una técnica importante en el campo de la computación y se utiliza para el análisis de grandes cantidades de datos, crear modelos matemáticos, simular procesos y extraer información valiosa de los datos para los científicos sociales. Así, el libro proporciona una introducción detallada a la minería de datos y sus aplicaciones en diferentes campos.

A continuación, el autor se centra en los patrones e *itemsets*, conceptos utilizados en el análisis de datos. En el contexto de la minería de datos, los patrones son conjuntos de características o atributos que se encuentran juntos en los datos con cierta frecuencia, y que establecen relaciones entre variables, objetos, eventos o cualquier otro tipo de entidad que se esté analizando. Los patrones se pueden encontrar utilizando técnicas de análisis de datos, como la minería de datos, que buscan identificar patrones estadísticamente significativos en los datos. Los patrones pueden ser simples o complejos. Un patrón simple puede ser una relación directa entre dos variables, como una correlación positiva entre el ingreso y la educación. Un patrón complejo puede ser una serie de relaciones entre varias variables que pueden ser difíciles de detectar sin el uso de técnicas de minería de datos avanzadas. Los patrones pueden ser utilizados para hacer predicciones, tomar decisiones y desarrollar estrategias en diferentes campos, como la sociología, la política, el *marketing*, la salud pública o la seguridad, entre otros.

Los *itemsets*, por otro lado, son conjuntos de elementos que aparecen juntos con cierta frecuencia en un conjunto de datos. Los *itemsets* en sociología pueden ser utilizados para identificar patrones de comportamiento comentados en el párrafo anterior, actitudes, valores, creencias en las personas y en los grupos sociales. Estos *itemsets* pueden ser utilizados para desarrollar políticas públicas, estrategias de *marketing*, programas de intervención y otros proyectos que se basen en una mejor comprensión de las personas y sus comportamientos. A modo de ejemplo ilustrativo, en las compras de los clientes en una tienda, un patrón podría ser que los clientes que compran pan también compran mantequilla y leche. Un *itemset* podría ser que los clientes compran pan, mantequilla y leche juntos con cierta frecuencia. El análisis de patrones e *itemsets* es importante en investigación so-

cial para el descubrimiento de relaciones, validación de hipótesis, identificación de grupos y análisis de tendencias.

Seguidamente, el autor presenta las reglas de asociación o correlación que existe entre diferentes conjuntos de *itemsets*, como una técnica de minería de datos para el descubrimiento de conocimiento dentro de grandes bases de datos. Una regla de asociación se basa en la idea de que ciertas variables tienden a aparecer juntas con más frecuencia que otras en un conjunto de datos. La aplicación de la regla de asociación en minería de datos comienza con la identificación de un conjunto de variables y una métrica para medir la asociación entre ellas. Luego, se utiliza un algoritmo de minería de datos para analizar los datos y descubrir patrones y asociaciones entre las variables. Siguiendo con el ejemplo ilustrativo de antes, supongamos que se tiene un conjunto de datos que incluye información sobre los hábitos de compra de los clientes de una tienda. La regla de asociación podría utilizarse para identificar patrones como «si un cliente compra pan, entonces también es probable que compre leche». Esta regla de asociación sugiere que hay una asociación fuerte entre la compra de pan y la compra de leche en ese conjunto de datos. De esta manera, las reglas de asociación muestran la relación entre diferentes conjuntos de *itemsets* y se presentan en forma de «si..., entonces...». En nuestro ejemplo, «si los clientes compran pan, entonces también compran leche».

Posteriormente, el autor expone la detección de patrones y generación de reglas mediante *softwares* estadísticos y algoritmos. En este sentido se presenta el caso de R, lenguaje de programación ampliamente utilizado en la minería de datos y el análisis estadístico. R cuenta con numerosas herramientas para la detección de patrones en conjuntos de datos grandes. El autor destaca una de las herramientas más utilizadas en R para la detección de patrones, como es la biblioteca de minería de datos llamada «arules». Esta biblioteca implementa el algoritmo «apriori» y otros algoritmos de minería de datos para la generación de reglas de asociación. Asimismo, el autor también menciona otra herramienta útil en R para la detección de patrones, como es la biblioteca «cluster». Esta biblioteca implementa varios algoritmos de *clustering*, como el *k-means* y el *clustering* jerárquico, que permiten agrupar objetos similares en conjuntos y descubrir patrones y relaciones entre ellos. El paquete R permite calcular métricas para evaluar la calidad de los patrones descubiertos y su visualización gráfica. Si seguimos con la ilustración ficticia anteriormente mencionada de la compra en una tienda, R podría detectar que el 50 % de las transacciones incluyen pan y el 80 % de esas transacciones también incluyen leche. El valor de la métrica utilizada indicaría que la compra de pan y leche juntos es un 25 % más probable de lo que se esperaría si los productos se compraran de forma independiente. En suma, utilizando el paquete de R, se pueden detectar patrones y reglas de asociación en los datos.

Después de las explicaciones conceptuales y técnicas realizadas por el autor, la obra se enfoca en un primer apartado práctico con la detección de patrones y asociaciones (reglas) para comprender los hábitos agregados de los clientes a partir de análisis de la cesta de la compra, explorando los artículos que a menudo se compran al mismo tiempo. Este enfoque práctico para la detección de patrones y asociaciones permite que los lectores pueden aplicar directamente lo que aprenden en su propia investigación o práctica. Al centrarse en la cesta de la compra, la obra también ilustra cómo las técnicas de minería de datos pueden ser aplicadas en contextos cotidianos y en la resolución de problemas prácticos. En este sentido, se aplica un ejemplo de cálculo de patrones e *itemsets*, conceptos utilizados en el estudio de datos, presentando los pasos básicos para realizar un análisis técnico. Respecto

al segundo apartado práctico, el autor desarrolla otro ejemplo, basado en el caso de la supervivencia del Titanic, donde la determinación empírica de las reglas puede darse en situaciones en las que la coocurrencia impide establecer una secuencia única, sino que existe una secuencia lógica inserta (superreglas) dentro del patrón evaluado.

En suma, la obra en cuestión presenta una exposición clara y fácil de entender de los conceptos y técnicas computacionales relacionados con la minería de datos, patrones y reglas de asociación. El autor ha logrado explicar de manera detallada y comprensible cada uno de estos temas, lo que permite al lector adquirir un conocimiento profundo sobre ellos. Los ejemplos presentados en la obra son relevantes y realistas, lo que ayuda al lector a comprender cómo aplicar estas técnicas en la práctica. Esto es particularmente valioso para los investigadores sociales que buscan explorar grandes conjuntos de datos, ya que les permite ver cómo estas técnicas pueden aplicarse a sus propias investigaciones. Desde luego, esta obra es altamente recomendable para cualquier investigador/a social interesado/a en la exploración de grandes conjuntos de datos. La finura analítica presentada a lo largo del texto, combinada con la claridad y la facilidad de comprensión de los conceptos y técnicas computacionales, hacen que sea una herramienta valiosa para cualquier persona que busque profundizar su comprensión de la minería de datos en investigación social.

por Javier DE ESTEBAN CURIEL
Universidad Rey Juan Carlos
javier.deesteban@urjc.es

En los márgenes de la democracia liberal. Populismo, nacionalismo y radicalismo en Europa

Ángel Valencia Sáiz y Belén Fernández-García (eds.)
(Granada, Editorial Comares, 2022)

El optimismo de finales del siglo xx y la extendida creencia de que la democracia sería hegemónica y capaz de resolver los problemas más importantes de la humanidad empezaron a ponerse en duda con la crisis financiera de 2007-2008, aunque ya anteriormente podían detectarse algunos síntomas de desafección ciudadana, como el rechazo en 2005, mediante diversos referéndums, a un proyecto de Constitución Europea.

Durante las dos últimas décadas, la progresiva pero constante consolidación electoral de formaciones políticas populistas en Europa, sobre todo aquellas ubicadas en el espec-