

# Trabajo Práctico

Eduardo Raúl Herrero Rivero

Instituto Superior Politécnico Córdoba

Tecnicatura en Ciencia de Datos e Inteligencia Artificial, cohorte 2020.

Materia: Minería de Datos

Docente: Omar Horta Fischer

Córdoba, 2022

## Minería de Datos:

La **minería de datos** o **exploración de datos** (es la etapa de análisis de "knowledge discovery in databases" o KDD) es un campo de la estadística y las [ciencias de la computación](#) referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos.<sup>12</sup> Utiliza los métodos de la [inteligencia artificial](#), [aprendizaje automático](#), [estadística](#) y sistemas de [bases de datos](#). El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior. Además de la etapa de análisis en bruto, supone aspectos de [gestión de datos](#) y de bases de datos, de [procesamiento de datos](#), del modelo y de las consideraciones de inferencia, de métricas de intereses, de consideraciones de la [teoría de la complejidad computacional](#), de post-procesamiento de las estructuras descubiertas, de la visualización y de la actualización en línea.

El término es un concepto de moda, y es frecuentemente mal utilizado para referirse a cualquier forma de datos a gran escala o procesamiento de la información (recolección, extracción, almacenamiento, análisis y estadísticas), pero también se ha generalizado a cualquier tipo de sistema informático de apoyo a decisiones, incluyendo la inteligencia artificial, aprendizaje automático y la inteligencia empresarial. En el uso de la palabra, el término clave es el descubrimiento, comúnmente se define como "la detección de algo nuevo". Incluso el popular libro "La minería de datos: sistema de prácticas herramientas de aprendizaje y técnicas con Java" (que cubre todo el material de aprendizaje automático) originalmente iba a ser llamado simplemente "la máquina de aprendizaje práctico", y el término "minería de datos" se añadió por razones de marketing. A menudo, los términos más generales "(gran escala) el análisis de datos", o "análisis". O cuando se refieren a los métodos actuales, la inteligencia artificial y aprendizaje automático, son más apropiados.

La tarea de minería de datos real es el análisis automático o semi-automático de grandes cantidades de datos para extraer patrones interesantes hasta ahora desconocidos, como los grupos de registros de datos (análisis clúster), registros poco usuales (la detección de anomalías) y dependencias (minería por reglas de asociación). Esto generalmente implica el uso de técnicas de bases de datos como los índices espaciales. Estos patrones pueden entonces ser vistos como una especie de resumen de los datos de entrada, y pueden ser utilizados en el análisis adicional o, por ejemplo, en el aprendizaje automático y [análisis predictivo](#). Por ejemplo, el paso de minería de datos podría identificar varios grupos en los

datos, que luego pueden ser utilizados para obtener resultados más precisos de predicción por un sistema de soporte de decisiones. Ni la recolección de datos, la preparación de datos, ni la interpretación de los resultados y la información son parte de la etapa de minería de datos, pero que pertenecen a todo el proceso KDD como pasos adicionales.

Teniendo en cuenta este contexto, el proyecto presentado anteriormente realiza una clasificación binaria por medio de una CNN entrenada con un set de imágenes de tornillos y de tarugo plásticos, logrando una precisión de más del 95%, basándose en ciertos patrones como la forma y el tamaño.

Los patrones que observamos nosotros, como humanos, son:

MATERIAL: Metal (Tornillos) o Plástico (Tarugos).
USO: Tornillos para chapa, madera o plástico y Tarugos para ladrillo hueco, muro solidos o Durlock.
TAMAÑO: Observamos más variedad de tamaños en los Tornillos que en los Tarugos.
COLOR: Metálico oscuro, brillante o mate en Tornillos y los Tarugos mayoritariamente grises.
FORMA O CONTORNO: Ambos tienen formas propias, definidas y distintivas unos de otros.

Los términos relacionados con la obtención de datos, la pesca de datos y espionaje de los datos se refieren a la utilización de métodos de minería de datos a las partes de la muestra de un conjunto de datos de población más grandes establecidas que son (o pueden ser) demasiado pequeñas para las inferencias estadísticas fiables que se hizo acerca de la validez de cualquier patrón descubierto. Estos métodos pueden, sin embargo, ser utilizados en la creación de nuevas hipótesis que se prueban contra poblaciones de datos más grandes.

Comercialmente, su aplicación podría ser, por ejemplo, una máquina para armar bolsitas con una cantidad determinadas de Tornillos y Tarugos de una medida determinada, para acompañar un producto como un barral para cortina de baño, una estantería para colgar o un ventilador que pueda amurarse a la pared.