



# Actividad evaluable – Actividad 1

**ASIGNATURA:** Minería de Datos

**Título:** *Máster Universitario en Big Data y Ciencia de Datos*

**Materia:** *Ciencia de Datos*

**Créditos:** *6 ECTS*

**Código:** *05MBID*

**Curso:** *Octubre 2023-2024*

1. Actividad.....	3
2. Rúbrica de evaluación.....	4
3. Bibliografía .....	5
3.1. Bibliografía de referencia .....	5
3.2. Bibliografía complementaria.....	5

Fecha de entrega	
1ª Convocatoria	Jueves, 30 de noviembre de 2023 hasta las 23.59h.
2ª Convocatoria	Miércoles, 27 de marzo de 2024 hasta las 23:59h.

# 1. Actividad

DESCRIPCIÓN	
Introducción	El objetivo de esta actividad es introducir conceptos básicos de minería de datos, para ello se va a trabajar en el proceso de extracción de conocimiento: Knowledge Discovery in Databases (KDD).
Objetivo	Consiste en plantear un proceso KDD hasta el preprocesamiento y transformación de datos incluido. En definitiva, en esta actividad se deben tener en cuenta la fase del planteamiento del problema y el objetivo a alcanzar, la elección de la base de datos y selección de datos útiles y la fase de preprocesamiento y transformación de datos.
Trabajo previo	Lectura del material docente de la parte específica que se encuentra disponible desde el comienzo del curso en la carpeta: Recursos y materiales > Material del profesor.
Metodología	En las videoconferencias teóricas (VC) se expondrá al alumnado los conocimientos, material e indicaciones necesarias para que pueda elaborar esta actividad guiada. Además, en la videoconferencia (AG1) se establecerán las pautas concretas y la dinámica que el alumnado debe seguir para realizar la actividad propuesta.
Tarea para el portafolio	<p>Desarrollar de forma clara, enumerando los puntos clave, la metodología y estructura del trabajo a realizar por el alumnado. Es importante la claridad en la definición de la actividad para que no haya interpretaciones incorrectas.</p> <p>En esta actividad se debe hacer una propuesta de un proceso KDD hasta el preprocesamiento y transformación de datos incluido, siguiendo las siguientes pautas:</p> <p><b>1. Planteamiento de problema y objetivo a alcanzar.</b> Se debe definir el contexto y las necesidades que llevan a la realización del proceso KDD. Además, se debe describir el problema que se pretende resolver y la necesidad a la que se busca responder. Es por ello por lo que hay que explicar el objetivo del análisis de datos que se espera afrontar en este proyecto.</p> <p><b>2. Elección de la base de datos y selección de datos útiles.</b> Hay que seleccionar una base de datos que no provenga de repositorios estándar de KDD (como UCI ML Repository). La base de datos escogida debe respetar el tamaño, formato, contenido y las leyes de protección de datos adecuados. En este punto se debe explicar cómo se han filtrado y seleccionado los datos más relevantes para el problema y objetivo planteados en el anterior punto y hay que justificar la exclusión de los datos que no se consideran útiles.</p> <p><b>3. Preprocesamiento y transformación de datos.</b> Se debe describir con detalle que técnicas y métodos se aplicarían para realizar una limpieza y transformación de datos. Además, se debe incluir una discusión sobre por qué y cómo este preprocesamiento y transformación de datos se alinean con el objetivo del proyecto. En esta actividad no es necesaria la implementación de código fuente sobre este proceso.</p> <p>Este trabajo no solo debe ser visto como una actividad puntual, sino también como una posible base para proyectos futuros en otras asignaturas y para el Trabajo de Fin de Máster (TFM).</p> <p>Se valorará especialmente la originalidad, la exactitud y la lógica en la descripción de la base de datos seleccionada y de cada una de las fases del proceso KDD que se realizaría para lograr los objetivos planteados. El documento final deberá tener una extensión de aproximadamente 4 páginas y seguir la estructura sugerida anteriormente.</p>

## 2. Rúbrica de evaluación

	Suspense (< 5)	Aprobado (>= 5)	Sobresaliente (>= 9)
<b>Estilo (30%)</b>	Redacción incoherente gramaticalmente y/o con más de 3 faltas ortográficas.	Redacción coherente, estilo informal o alguna falta ortográfica.	Redacción impecable, con estructura definida, estilo formal y sin faltas ortográficas.
<b>Contenido (40%)</b>	No se realizan todos los pasos del ciclo del proceso KDD, la visión y simplista. Solo se describe teóricamente el proceso KDD y no realiza ningún análisis específico de los datos.	Se realizan todos los pasos del proceso KDD; pero se queda en ámbito superficial.	Se analiza en profundidad cada uno de los pasos y no se utilizan técnicas para ahorrar trabajo de limpieza de datos.  Demuestra que sabe que pasos debe aplicar a su conjunto de datos.  El alumno demuestra conocimiento de los datos y sabe cómo tratarlos.
<b>Originalidad y pasos adicionales (30%)</b>	El alumno se ha centrado en comentar o aplicar el número mínimo de pasos y no explica porque los ha realizado.	Ha realizado todos los pasos para conseguir el objetivo del proceso KDD y explica el motivo de los pasos intermedios.	Ha realizado todos los pasos para conseguir el objetivo del proceso KDD y explica el motivo de los pasos intermedios.  Analiza y lanza hipótesis sobre lo que puede encontrar y que ciclos adicionales realizar.  Crea variables adicionales, investiga y propone técnicas.

## 3. Bibliografía

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data. Communications of the ACM, November 1996/ Vol 39, Nº 11, 27–34.

### 3.1. Bibliografía de referencia

- Model, F. E., Williams, G. J., & Huang, Z. (1996). Modelling the KDD Process.
- Brachman, R. J., & Anand, T. (1994, July). The Process of Knowledge Discovery in Databases: A First Sketch. In KDD workshop (Vol. 3, pp. 1-12).

### 3.2. Bibliografía complementaria

- Siegel, E. (2013). Analítica predictiva. Predecir el futuro utilizando Big Data. Anaya Multimedia-Anaya Interactiva.
- Mayer-Schönberger, V., Cukier, K. (2013). Big data. La revolución de los datos masivos. Turner