



Actividad evaluable – Actividad 2

ASIGNATURA: Minería de Datos

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

Materia: *Ciencia de Datos*

Créditos: 6 ECTS

Código: 05MBID

Curso: *Abril 2024-2025*

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	03/07/2024 hasta las 23:59h
2ª Convocatoria	29/10/2024 hasta las 23:59h

1. Actividad

DESCRIPCIÓN	
Introducción	El objetivo de esta actividad es introducir los conceptos básicos de la minería de datos. Para ello el alumnado trabajará en el desarrollo de un proceso de extracción de conocimiento a partir de bases de datos (KDD del inglés).
Objetivo	Esta actividad consiste en implementar únicamente las etapas de Minería de Datos e Interpretación/Evaluación del proceso de KDD que se comenzó a desarrollar en la Actividad Guiada 1. Recordar, que el modelo debe ser de clasificación binaria (de únicamente dos valores de clase, como puede ser el aprobar o no un examen, otorgar o no un crédito, entre otros ejemplos que fueron dados en clase).
Trabajo previo	Lectura del material docente disponible en Recursos y materiales > Material del profesor y de todos los hilos de conversación del Foro que se generen.
Metodología	En las sesiones se expondrán los conocimientos, materiales e indicaciones necesarias para que el alumnado pueda realizar esta actividad guiada sin dificultades. Se establecerán las pautas concretas y la dinámica que se deberá seguir para realizar la actividad propuesta.
Tarea para el portafolio	<p>En esta actividad se debe completar el proceso KDD que se comenzó en la Actividad Guiada 1, desarrollando las etapas de construcción del modelo de Minería de Datos y de evaluación del modelo/interpretación de los resultados.</p> <p>El dataset no podrá ser cambiado de una actividad a la otra. Es decir, que la vista minable a utilizar en la construcción del modelo, debe ser la salida que produzca el cuaderno .ipynb entregado en la Actividad Guiada 1. E</p> <p>Para aprobar con la nota mínima se deberá:</p> <ul style="list-style-type: none"> - Separar los datos en datos de entrenamiento y testeo - Construir un modelo de clasificación con los datos de entrenamiento - Obtener el accuracy del modelo con los datos de testeo <p>Para alcanzar la nota máxima, se podrá entre otras cosas:</p> <ul style="list-style-type: none"> - Construir la matriz de confusión con los datos de testeo - Obtener el recall y precision de la clase minoritaria con los datos de testeo - Generar un fichero de salida con la respuesta esperada y la generada por el modelo para los datos de testeo - Obtener la importancia de las variables - Construir un segundo modelo y comparar su accuracy

2. Rúbrica de evaluación

	Suspense (< 5)	Aprobado (>= 5)	Sobresaliente (>= 9)
Estilo (30%)	Codificación incoherente y/o con errores.	Codificación coherente y sin errores.	Codificación coherente y sin errores, yendo más allá de lo básico esperado.
Contenido (40%)	No hace lo mínimo solicitado.	Hace lo mínimo solicitado.	Hace lo mínimo solicitado y además realiza los extras.
Originalidad y pasos adicionales (30%)	No utiliza las librerías y métodos adecuados.	Utiliza las librerías y métodos adecuados.	Utiliza las librerías y métodos adecuados, pero también incorpora los no convencionales.

3. Bibliografía

3.1. Bibliografía de referencia

- Jupyter Notebook for data Science: Publishing, Packt. Jupyter Notebook for Data Science. PACKT Publishing, 2018, video.alexanderstreet.com/watch/jupyter-notebook-for-datascience. Accessed 17 May 2021.

3.2. Bibliografía complementaria

- Herrera, F. (2016). Big Data: Preprocesamiento y calidad de datos. novática, 237, 17.