



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 minería de datos. Para ello se va a trabajar en el proceso de extracción de conocimiento: Knowledge Discovery in Databases (KDD).
Objetivo	La actividad consiste en continuar planteando el proceso de KDD que se comenzó en la Actividad Guiada 1. Se debe detallar cuál será el modelo de Minería De Datos que permitirá alcanzar el objetivo y así, resolver el problema. Además, se deberá describir de qué manera podrían discutirse/interpretarse los resultados obtenidos. Por otra parte, se deberá desarrollar un notebook en el cual se implemente todo lo planteado en la Actividad 1.
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	<p>En las videoconferencias teóricas (VC) se expondrá al alumnado los conocimientos, el material y las indicaciones necesarias para que pueda elaborar esta actividad guiada. Además, en la videoconferencia (AG2) se establecerán las pautas concretas y la dinámica que el alumnado debe seguir para realizar la actividad propuesta.</p> <p>En esta actividad se debe completar la propuesta del proceso KDD que se comenzó en la Actividad Guiada 1, siguiendo las siguientes pautas:</p> <p>4. Modelo de Minería de Datos. Se debe indicar de forma precisa cuál será el algoritmo de Minería de Datos más adecuado para resolver el problema que el alumno pretende resolver. Mencionar cuáles adaptaciones en los datos se necesitan para construir el modelo y de qué modo ejecutaría el algoritmo de entrenamiento correspondiente. No es necesario implementarlo, sólo describirlo claramente.</p> <p>5. Discusión e interpretación de los resultados. Describir cómo se podría evaluar el desempeño del modelo, qué tipo de patrones son los que se obtienen del modelo y cómo se los podría utilizar para tomar decisiones.</p> <p>En el caso que el estudiante haya entregado la Actividad Guiada 1, el estudiante deberá también corregir los puntos correspondientes a dicha actividad basándose en la retroalimentación (feedback) que el equipo docente suministre. En caso contrario, se recomienda al estudiante que elabore los puntos correspondientes a la AG1 antes de proceder con los puntos de la documentación de esta actividad. En definitiva, en esta tarea deberá enviarse la documentación del proceso KDD al completo (Actividad Guiada 1 + pautas pertenecientes a la Actividad Guiada 2).</p>
Tarea para el portafolio	<p>Para aprobar con la nota mínima se deberá:</p> <ul style="list-style-type: none"> - Indicar claramente el algoritmo de Minería de Datos elegido. - Describir las adaptaciones necesarias en los datos para construir el modelo. - Detallar el algoritmo de Minería de Datos y el resultado esperado de manera superficial. - Explicar cómo se podrían utilizar esos patrones para tomar decisiones. <p>Para alcanzar la nota máxima, se podrá entre otras cosas:</p> <ul style="list-style-type: none"> - Demostrar un conocimiento profundo del algoritmo elegido y justificar por qué es el más adecuado. - Proponer múltiples métodos de evaluación del desempeño del modelo (e.g., precisión, recall, F1-score, etc.). - Incluir análisis detallados y reflexiones sobre cada etapa del proceso KDD. - Proveer visualizaciones claras y significativas de los resultados del modelo. - Detallar el algoritmo de Minería de Datos con profundidad, mostrando comprensión de los datos y cómo tratarlos. - Proponer técnicas innovadoras o variables adicionales basadas en la investigación.

Por otra parte, se deberá entregar notebook (fichero .ipynb) que incluya la implementación de todo lo propuesto en la Actividad Guiada 1, junto con los datos necesarios para ejecutarlo (fichero .csv) o por consiguiente el dataset preprocesado.

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).

***ACLARACIÓN:** La incorporación de tablas, imágenes o cualquier otro tipo de artefacto en dicho documento contabilizará para la extensión del número de páginas. Por lo tanto, se aconseja que cuando estos artefactos sean extensos, se incorporen en un fichero independiente a modo de Anexo.*

2. Rúbrica de evaluación

	Suspense (< 5)	Aprobado (> = 5)	Sobresaliente (> = 9)
Estilo (30%)	Redacción incoherente gramaticalmente y/o con más de 3 faltas ortográficas. No utilizar librerías estándar.	Redacción coherente, estilo informal o alguna falta ortográfica. Uso adecuado de librerías.	Redacción impecable, con estructura definida, estilo formal y sin faltas ortográficas. Uso adecuado de librerías.
Contenido (40%)	No se especifica el algoritmo de Minería de Datos a utilizar ni el resultado que se obtendría con el mismo. No se indican los datos de entrada y de salida. El dataset final tiene errores.	Se lo detalla, pero de una manera superficial. El dataset final está limpio.	Demuestra que sabe cuál algoritmo de Minería de Datos debe aplicar a su conjunto de datos y para qué puede servir el output del modelo. El alumno demuestra conocimiento de los datos y sabe cómo tratarlos. Se ha hecho un buen análisis de los datos de entrada y se comprende cómo se quiere conseguir el objetivo.
Originalidad y pasos adicionales (30%)	El alumno se ha centrado en comentar lo mínimo y no explica demasiado.	Explica la razón por la cual piensa que el algoritmo elegido resuelve el problema.	Analiza y lanza hipótesis sobre lo que puede encontrar y que ciclos adicionales realizar. Crea variables adicionales, investiga y propone técnicas.

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.