

PEC 1 (20% nota final)

Presentación

En esta Prueba de Evaluación Continuada se trabajan los conceptos generales del ciclo de vida de los datos, y se identifican y revisan sus características. También se trabajan los conceptos fundamentales del Web Scraping.

Competencias

En esta PEC se desarrollan las siguientes competencias del Máster de Data Science:

- Capacidad de analizar un problema en el nivel de abstracción adecuado a cada situación y aplicar las habilidades y conocimientos adquiridos para abordarlo y resolverlo.
- Capacidad para aplicar las técnicas específicas de tratamiento, almacenamiento y administración de datos

Objetivos

Los objetivos concretos de esta Prueba de Evaluación Continua son:

- Conocer el ciclo de vida de los datos y los principales tipos de datos.
- Aprender a aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios o multidisciplinares.
- Desarrollar las habilidades de aprendizaje que permitan continuar estudiando de una manera que tendrá que ser en gran medida autodirigida o autónoma.
- Desarrollar la capacidad de busca, gestión y uso de información y recursos en el ámbito de la ciencia de datos.
- Entender la utilidad, la legalidad y algunas características del web scraping.

Descripción de la PEC a realizar

Ejercicio 1 [70%]

Después de leer el recurso “Calvo, M., Pérez, D., Subirats, L. (2019). Introducción al ciclo de vida de los datos.” contesta las siguientes preguntas con tus propias palabras:

- 1 La clasificación ESCO (European Skills, Competences, Qualifications and Occupations) estandariza los perfiles profesionales. Clasifica en ese estándar los 8 perfiles del ámbito de la ciencia de datos indicando su código. [20%]
- 2 Pon como ejemplo dos ofertas de trabajo (p.ej., LinkedIn o Indeed) que correspondan a diferentes perfiles profesionales dentro del ámbito de la ciencia de datos, y justifica por qué. Indica el enlace de la oferta y una captura de pantalla de la misma. Si es necesario, indica cuáles serían las correcciones pertinentes a las ofertas que has encontrado. (Máximo 200 palabras). [10%]
- 3 Además de las habilidades técnicas, las soft skills o habilidades blandas son muy importantes en la vida laboral. ¿Cómo harías para averiguar cuáles son las soft skills requeridas habitualmente en los perfiles de ciencia de datos? Indica tres de las soft skills que has encontrado. (Máximo 100 palabras). [10%]
- 4 Pon un ejemplo de datos restringidos, privados y públicos. [10%]
- 5 Lista los diferentes factores que influyen en la calidad de los datos y pon un ejemplo que no sea el que se explica en el material de la asignatura. (Máximo 300 palabras). [10%]
- 6 Explica cuáles son los beneficios de las bases de datos no relacionales y pon 3 ejemplos de estas bases de datos (es suficiente con citar el nombre). (Máximo 100 palabras). [10%]

Ejercicio 2 [30%]

Después de leer el recurso “Subirats, L., Calvo, M. (2019). Web Scraping”, capítulos 1 y 7. Contesta las siguientes preguntas con tus propias palabras:

- 1 Cita el nombre de tres librerías útiles para un proyecto de web scraping (Máximo 100 palabras). [10%]
- 2 Explica dos buenas prácticas, a la hora de realizar web scraping, para evitar ser bloqueado (Máximo 100 palabras). [10%]

- 3 Pon un ejemplo donde realizar web scraping podría ser útil para un negocio. (Máximo 100 palabras). [10%]

Recursos

Los siguientes recursos son de utilidad para la realización de la PEC:

Básicos

- Calvo, M., Pérez, D., Subirats, L. (2019). Introducción al ciclo de vida de los datos. Editorial UOC.
- Subirats, L., Calvo, M. (2019). Web Scraping. Editorial UOC.

Complementarios

- Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Capítulo 1. Introduction to Web Scraping.
- Minguillón, J. (2016). Fundamentos de data Science. Editorial UOC.

Criterios de valoración

La ponderación de los ejercicios es la siguiente:

- Ejercicio 1.1: 20%
- Ejercicio 1.2: 10%
- Ejercicio 1.3: 10%
- Ejercicio 1.4: 10%
- Ejercicio 1.5: 10%
- Ejercicio 1.6: 10%
- Ejercicio 2.1: 10%
- Ejercicio 2.2: 10%
- Ejercicio 2.3: 10%

Se evaluará la precisión de los ejemplos así como lo respecto al número de palabras máximo establecido para cada pregunta. La idoneidad y claridad de las respuestas también será evaluada.

Formato y fecha de entrega

Hay que librar un único documento Word, Open Office o PDF (**preferiblemente este último**) con las respuestas a las preguntas.

Este documento se tiene que librar en el espacio de Entrega y Registro de AC del aula antes de las **23:59** del día **18 de octubre**. No se aceptarán entregas fuera de plazo.