

Título de la tarea: Aplicación práctica de tecnologías Big Data.

Ciclo formativo y módulo: Especialización en Inteligencia Artificial y Big Data - Big Data Aplicado.

¿Qué contenidos o resultados de aprendizaje trabajaremos?

Resultados de aprendizaje

- ✓ **RA5.** Valida las técnicas de Big Data para transformar una gran cantidad de datos en información significativa, facilitando la toma de decisiones de negocios.

Contenidos

- 1.- Arquitecturas y modelos de despliegue.
 - 1.1.- Adopción de tecnologías Big Data.
 - 1.2.- Hadoop en la práctica.
 - 1.3.- Modelos de despliegue.
 - 1.4.- Gobierno de datos.
- 2.- Soluciones Hadoop-as-a-Service.
 - 2.1.- Amazon EMR.
 - 2.2.- Microsoft Azure HDInsight.
- 3.- Guía práctica de análisis exploratorio y de aprendizaje automático con Spark

Siguiente »

1.- Descripción de la tarea.



Caso práctico

En esta tarea tendrás que elegir un dataset no utilizado en el curso ni por ningún otro compañero. Para asegurar que esto no ocurre publicarás en el foro de la unidad un mensaje con la descripción y la URL del dataset que hayas elegido. Puedes usar cualquier dataset pública y libremente accesible, por ejemplo en Kaggle o en GitHub.

Una vez elegido el dataset realizarás un análisis exploratorio y un proceso de aprendizaje automático usando Spark en Databricks.

El fichero adjunto incluye el notebook de Databricks que debes usar como plantilla para tu trabajo.



[Pexels](#) (Dominio público)

- [Tarea BDA05 \(Ventana nueva\)](#)

¿Qué te pedimos que hagas?

Elige un dataset pública y libremente accesible de una tamaño mínimo de 10MB y publica en el foro la URL y una breve descripción del contenido del dataset. El dataset que elija cada alumno debe ser diferente de los usados en el curso o por otros alumnos.

Completa los apartados del notebook de Databricks suministrado y sube un archivo comprimido que incluya el notebook con las celdas ejecutadas y el resto de ficheros utilizados (dataset, imágenes, ...).

NOTA IMPORTANTE

Las celdas del notebook deben estar ejecutadas y sin errores. En la guía práctica se realiza un proceso similar con el dataset del Titanic que puedes utilizar para orientarte.

2.- Información de interés.



Recursos necesarios y recomendaciones

Recursos necesarios

- ✓ Conexión a Internet para conectarte a tu cuenta de Databricks.
- ✓ Navegador web.
- ✓ Software para comprimir los archivos de la tarea.

Recomendaciones

- ✓ Antes de abordar la tarea:
 - lee con detenimiento la unidad, consulta los enlaces para saber más, examina el material proporcionado por el profesor y aclara las dudas que te surjan con él.
 - Realiza el examen online de la unidad, y consulta nuevamente las dudas que te surjan. Solo cuando lo tengas todo claro, debes abordar la realización de la tarea.
- ✓ No olvides elaborar el documento explicativo.



Indicaciones de entrega

Una vez realizada la tarea, el envío se realizará a través de la plataforma. El archivo se nombrará siguiendo las siguientes pautas:

Apellido1_Apellido2_Nombre_BDA05_Tarea

Asegúrate que el nombre no contenga la letra ñ, tildes ni caracteres especiales extraños. Así por ejemplo la alumna **Begoña Sánchez Mañas para la quinta unidad del MP de BDA**, debería nombrar esta tarea como...

sanchez_manas_begona_BDA05_Tarea

« Anterior Siguiente »

3.- Evaluación de la tarea.

Criterios de evaluación implicados

Criterios de evaluación RA5

- ✓ a) Se han seleccionado gran cantidad de datos estructurados y no estructurados para reforzar la función de BI.
- ✓ b) Se ha realizado la limpieza y transformación de datos en base a los objetivos predeterminado.
- ✓ c) Se ha comprobado que el Big Data multiplica la relevancia y la utilidad del BI para el negocio.
- ✓ d) Se han conjugado dentro de un modelo de empresa datos de clientes, financieros de ventas, de productos, de marketing, de redes sociales, de la competencia, entre otros, para extraer un análisis valioso y efectivo para el negocio.

¿Cómo valoramos y puntuamos tu tarea?

Rúbrica de la tarea	
Descripción del dataset	2,5 puntos
Análisis exploratorio	2,5 puntos
Preparación de los datos y entrenamiento	2,5 puntos
Predicción, evaluación y conclusiones	2,5 puntos
Redacción clara y correcta, sin errores ortográficos	Se resta 0,1 puntos por cada error ortográfico o expresiones incorrectas.

[« Anterior](#)