



3 PROYECTOS TOP

PARA INICIAR TU CARRERA COMO CIENTÍFICO DE DATOS





¿Estás listo para iniciar tu carrera de ensueño en el mundo del data science?

¡Es hora de reforzar tu portfolio!

Prepárate para optar a un gran puesto de trabajo con un portfolio que muestre las habilidades que no se pueden medir en una entrevista o en un currículum. Aumenta tus opciones de empleo con los siguientes tres proyectos que muestran cómo puedes agregar valor a cualquier negocio o empresa.

1. LIMPIEZA DE DATOS Y PROYECTOS DE INGENIERÍA DE CARACTERÍSTICAS

En el mundo real, los datos se encuentran desordenados, desestructurados y vienen en muchos formatos: audio, video, flujo de clics y texto, ¡lo que sea! Es por eso que la limpieza de datos es tan importante y por qué demostrar esta habilidad en tu portfolio es vital.

La limpieza de datos, también conocida como manipulación de datos, es una de las tareas más importantes que deben realizar los científicos de datos y es una habilidad extremadamente valiosa de dominar.

Después de todo, los científicos de datos dedican hasta el 80% de su tiempo a recopilar y limpiar datos, lo que puede incluir llenar los valores de datos faltantes, eliminar valores atípicos, escalar, codificar los datos y mucho más.

Si eres un programador de Python, ¡te recomendamos que aprendas las librerías Pandas y Numpy!



Y si a tu portfolio le falta un proyecto de limpieza de datos, te sugerimos que comiences a recopilar tus propios datos realizando un web scrapping y limpiando los datos antes de introducirlos en un modelo de aprendizaje automático.

2. DASHBOARD, VISUALIZACIÓN DE DATOS Y ANÁLISIS EXPLORATORIO DE LOS DATOS (EDA)

A los gerentes de contratación les encanta capturar historias desde los datos utilizando diagramas e imágenes y el Análisis exploratorio de datos (EDA) y la creación de dashboards hacen precisamente eso.

EDA es el proceso de analizar datos y generar ideas y conclusiones clave, como tendencias de ventas, efectos estacionales y correlaciones entre variables. Puedes realizar EDA representando diagramas de barras, diagramas de caja y bigotes, nubes de dispersión y gráficos circulares.

Recomendamos aprender Pandas y Matplotlib si eres usuario de Python. Si eres un usuario de R, aprovecha el poder de ggplot2 para esto.

Para diferenciarte, lleva tus proyectos de EDA al siguiente nivel mediante la creación de dashboards interactivos.

¡Imagínate haciendo una presentación sobre el crecimiento de Amazon, frente a Jeff Bezos! No quieres bombardearlo con códigos, matemáticas o ecuaciones, simplemente no le importan.



Lo que debes hacer es contar una historia y desarrollar un dashboard que muestre todas las métricas e indicadores clave de rendimiento (KPI) en un solo lugar.

El objetivo es comunicar un mensaje, resultado o una estrategia de forma clara y sin entrar en detalles técnicos complejos. Puedes usar Bokeh y Plotly si eres un usuario de Python y Shiny si eres un usuario de R.

3. PROYECTO DE DESARROLLO / DESPLIEGUE DE SISTEMAS INTEGRALES

Finalmente, la creación de un proyecto de ciencia de datos desde el inicio hasta el final, demuestra tu comprensión de cada fase de su ciclo de vida y tu capacidad para completar un proyecto de forma independiente.

Recomendamos construir un proyecto que destaque cada una de las siguientes etapas importantes:

- Recolección de datos mediante web scraping.
- Limpieza de datos y realización de ingeniería de características.
- Creación de una visualización de datos interactiva.
- Entrenamiento de tu modelo de Inteligencia Artificial (IA) y Aprendizaje Automático (ML).
- Optimización de la arquitectura del modelo y sus hiperparámetros.

- Probar y evaluar tu modelo utilizando varios KPIs.
- Implementar el modelo en la práctica y usar datos de transmisión procedentes del mundo real.

Por ejemplo, puedes crear un sistema que tome datos financieros de Yahoo en tiempo real, visualizarlos usando un dashboard interactivo y luego entrenar un modelo de red neuronal recurrente de aprendizaje profundo (RNR) o un modelo de pronóstico de series de tiempo como Facebook Prophet en una transmisión entrante de datos reales para predecir los precios futuros de las acciones.

¡Buena suerte!

