

Data Analyst / Data Scientist / Data Engineer

En *DataCompany* queremos hacer crecer nuestro equipo con personas que tengan ganas de desarrollar software basado en datos con dos objetivos: ayudar en la transformación a nuestros clientes y disfrutar del proceso de crear valor a través de la tecnología.

Si quieres ser parte de un equipo que te ofrecerá retos tecnológicos y que te exigirá una mejora y aprendizaje continuo, si no tienes miedo a equivocarte y eres capaz de ser proactivo y tirar del carro cuando te toque, es posible que *DataCompany* sea el sitio donde deberías trabajar. ¿Te interesa? Echa un vistazo a nuestras vacantes en el equipo de Data Science.

¿Sabes lo que vas a hacer además?

El objetivo del proyecto junto con el del equipo es realizar el desarrollo, entrenamiento y productivización de modelos para dar respuesta a necesidades de negocio del cliente. Todos los casos de uso que se llevarán a cabo estarán desarrollados en Cloud (AWS).

Requisitos:

- Mínimo 1 años de experiencia en el desarrollo de piezas de Machine Learning así como en su implementación en entornos de nube pública
- Programación en Python y/o R
- Manejo de librerías de análisis, modelado y visualización de datos.
- Conocimientos y experiencia trabajando con deep learning, especialmente NLP
- Fuerte base de álgebra y estadística.
- Conocimientos en Git.
- Capacidad de síntesis y habilidades comunicativas en presentaciones.

También es deseable:

- Experiencia en despliegue de modelos en producción
- Inglés nivel B2-C1.
- Experiencia y entusiasmo por trabajar con metodologías ágiles (Scrum, Kanban...).

¿Qué te ofrecemos en DataCompany?

- Un buen ambiente de trabajo junto a compañeros que son unos cracks en sus tecnologías y de los que aprenderás cada día.
- Tendrás un presupuesto personal anual de 2.000 euros para que lo dediques a la formación elegida por ti, adicional a cubrir el coste de tus certificaciones.
- Horario flexible, así como teletrabajo y flexibilidad en el almuerzo, para que tu vida personal y profesional encajen sin problema.
- Jornada intensiva los viernes todo el año, y los meses de julio y agosto.
- Retribución flexible para tarjeta transporte, seguro médico para familiares y cheque guardería.
- Tu evaluación y subida salarial la realizarán tus compañeros, nada de jefes porque aquí no encontrarás ninguno.

PRUEBA TÉCNICA

Para evaluar tus conocimientos técnicos, en esta fase del proceso de selección tendrás que realizar un ejercicio práctico del que dispondrás de 24 horas (Lunes 9.15 a Martes 9.15).

Consta de 2 partes:

1. Desarrollo del caso de uso
2. Presentación

Caso de uso

Tenemos el caso de uso de un cliente (@TheBridge_Tech) que desea desarrollar una monitorización de las redes sociales para medir el impacto de su marca y acciones comerciales.

Tendrás que realizar las siguientes tareas:

1. **Recopilar los tweets** donde se mencione la cuenta de @TheBridge_Tech desde comienzo de año hasta el día de ayer incluido. Se recomienda utilizar la [API de Twitter](#). Se deberá recoger:
 - a. Id del mensaje
 - b. Cuerpo del texto del mensaje
 - c. Fecha del tweet
 - d. Id del autor
 - e. Nombre del autor
 - f. Nombre de usuario del autor
 - g. Métricas públicas del tweet (retweet, reply, like, quote)
2. Almacenarlos en una **base de datos SQL** en 2 tablas diferentes a tu elección.
3. Realizar un pequeño **análisis** donde se respondan a las siguientes preguntas de negocio:
 - a. ¿Cuál es el tweet con mayor repercusión social?
 - b. ¿Cuál es el usuario que más menciona a la escuela?
 - c. ¿En qué mes se concentra el mayor número de tweets?
 - d. ¿Qué palabras son más frecuentes?
 - e. ¿Qué tipo de correlación matemática encuentras entre las métricas públicas?
4. Utiliza un **modelo pre entrenado** (ds_thebridge_1_22\3-Machine_Learning\5-NLP\NLTK&CountVectorizer\data\output\finished_model.model) para determinar el sentimiento de los 3 tweets con mayor repercusión. Preguntas:
 - a. ¿Cuáles son las predicciones? Interpreta los resultados.
 - b. ¿Qué variables han podido influir más en las predicciones del modelo?
 - c. ¿Cómo podrías mejorar el modelo?
 - d. ¿Qué otras oportunidades se te ocurren donde se podrían aplicar otros modelos de ML?

5. **Despliega** el modelo (no en local, puedes elegir el proveedor), con un endpoint donde poder enviarle un cuerpo de texto y devuelva la predicción del sentimiento.
6. Realizar una **presentación** de 5 minutos máximo, donde presentes tu solución y la defiendas en una entrevista personal.

Las herramientas o recursos que necesites para el desarrollo de la prueba técnica quedan a tu elección. Deberás subir tanto tu código como la presentación a un repositorio de Git. La entrega será en G Classroom, de un txt con la url de tu repositorio.

¡Mucha suerte!