

Predicción de alquiler de monopatines en Berlín, Alemania

Descripción del proyecto:

Bienvenido al desafío técnico. En este proyecto, tendrás la oportunidad de demostrar tus habilidades en la creación de un modelo de aprendizaje automático, su integración en un pipeline de ML, y la disponibilidad del modelo mediante una API. Trabajarás con un conjunto de datos de la ciudad de Berlín, Alemania, que detalla los alquileres de monopatines por horas a lo largo de varias temporadas.

Contexto:

Berlín cuenta con un sistema de monopatines compartidos, donde los usuarios pueden alquilar monopatines por horas o por días. Estos datos han sido recopilados en un archivo CSV que contiene información sobre 17,380 registros, cada uno representando el número de monopatines alquilados en una hora específica de un día.

El dataset necesario para este proyecto está disponible en el siguiente repositorio de GitHub: [dataset prueba técnica](#)

Tu tarea será predecir el número total de monopatines alquilados en una hora determinada, usando las columnas proporcionadas en el conjunto de datos, con excepción de las columnas que indican el número de usuarios casuales y registrados (`u_casuales` y `u_registrados`). Para lograrlo, deberás construir varios modelos de aprendizaje automático, comparar su rendimiento, y seleccionar el más adecuado. **Es importante destacar que todo el proyecto debe desarrollarse utilizando Python.**

Objetivos del Proyecto:

1. **Construcción del modelo:** Desarrolla y entrena uno o más modelos de aprendizaje automático que puedan predecir el número de monopatines alquilados por hora (columna `total_alquileres`), utilizando las demás columnas como características.
2. **Pipeline de ML:** Implementa un pipeline de ML en Python que automatice el proceso de preprocesamiento de datos, entrenamiento del modelo y evaluación. Este pipeline debe ser modular y fácilmente extensible.
3. **API para predicciones:** Expone el modelo seleccionado a través de una API desarrollada en Python, que permita hacer predicciones en tiempo real sobre nuevos datos.
4. **Funcionalidad extensible (Opcional):** Sumará puntos extra si la API incluye una funcionalidad adicional que permita al usuario seleccionar entre diferentes modelos para realizar las predicciones. Por ejemplo, el usuario podría elegir entre un modelo de regresión lineal y un modelo basado en árboles de decisión.
5. **Repositorio en GitHub:** Presenta tu solución en un repositorio de GitHub correctamente versionado. Se espera que el proyecto pueda ser replicado fácilmente,

sin configuraciones adicionales. Debes proporcionar un archivo README claro y detallado que incluya instrucciones de instalación y uso.

6. **Despliegue en una plataforma gratuita (Opcional):** Ganarás puntos adicionales si despliegas la API en una plataforma gratuita como Render o Railway, demostrando tus habilidades para la implementación y gestión de aplicaciones en la nube.

Entregables:

- Un repositorio de GitHub con el código fuente del proyecto.
- Una API funcional para realizar predicciones.
- Un archivo README que explique cómo configurar y ejecutar el proyecto.
- (Opcional) Enlace al despliegue de la API en una plataforma gratuita.

Cualquier consulta escríbenos al correo: ecusi@datec.com.bo