
Reto aplicación IA en plataforma FIWARE.

“Modelo predictivo de datos de calidad del aire en Córdoba”.

El cálculo de los índice de Calidad del Aire, se basan en la Resolución del 2 de septiembre de 2020, que modifica el Anexo de la Orden TEC/351/2019, del 18 de marzo, por la que se aprueba el Índice Nacional de Calidad del Aire.

El valor del índice lo determinan las concentraciones, en cada estación de medida, de hasta cinco contaminantes clave:

- Partículas en suspensión PM10.
- Partículas en suspensión PM2,5.
- Ozono troposférico (O3).
- Dióxido de nitrógeno (NO2).
- Dióxido de azufre (SO2).

En este conjunto de datos encontrará los valores reales de los parámetros medidos por los diferentes sensores que constituyen la Red de Vigilancia y Control de Calidad del Aire de Andalucía, teniendo en cuenta que, al referirse al índice de Calidad del Aire, los datos proporcionados aplican a estos cinco contaminantes clave.

Los datos proporcionados son comunicados cada hora, y complementados, en caso necesario, por datos modelizados sobre la calidad del aire procedentes del Servicio de Vigilancia Atmosférica del Sistema Copernicus CAMS (Copernicus Atmosphere Monitoring Service) de la Unión Europea.

RETO

El reto-01 ha sido realizado para ingestar los datos de la calidad del aire de manera automatizada en nuestra plataforma FIWARE.local a partir de los datos ofrecidos en el portal de datos abiertos del Ministerio.

Los datos se ofrecen diariamente después del día de su recolección, esto es para el día actual nunca tenemos los datos, sino a partir del día actual - 1. Sería interesante realizar una predicción del día actual 24 horas antes de que se ofrezca a partir de las estaciones físicas que recopilan los datos.

Mecanizando dicho proceso tendremos los datos siempre actualizados desde el año 2021 hasta el día actual, siendo el día actual una predicción aportada a través de un modelo IA. Además, incorporado y mecanizado siempre podremos evaluar de manera gráfica el resultado del modelo aplicado contrastando los datos de previsión con los datos actuales cuando son ingestados.

Para ellos utilizaremos:

Plataforma: Stack de **FIWARE.local** instalado en equipos de investigadores para autoformación, o bien plataforma **FIWARE.voneclod** como servicio definitivo para testear la solución..

- **Objetivo nº 1:** Incorporar un nuevo modelo de datos para sustentar la información de la previsión.
- **Objetivo nº 2:** Análisis gráfico/visual de la evolución como medio de comprobación de la persistencia de los datos predictivos y comparativa con los datos reales de las estaciones..
- **Objetivo nº 3:** Automatización del proceso de generación de la previsión de los datos a un día vista.

Justificación: una vez conocida la forma de gestionar modelos de datos por la plataforma FIWARE, vamos a utilizar utilidades y recursos para proveer un modelo de datos que sustente la previsión a un día vista que sirva para realizar comparativas del modelo IA.

Referencia: Interesa conocer el modelo IA de H2O.ia, una plataforma de inteligencia artificial (IA) líder en el mercado, especializada en aprendizaje automático (machine learning) y automatización del aprendizaje automático (AutoML). Es una herramienta poderosa y versátil que ha sido adoptada por miles de organizaciones en todo el mundo, incluyendo muchas de las empresas Fortune 500.

¿Qué hace H2O.ai?

- **Simplifica la IA:** H2O.ai democratiza la IA al hacer que sea más fácil de usar, incluso para aquellos sin una formación profunda en ciencia de datos.
- **Automatiza tareas:** La plataforma automatiza muchas de las tareas más tediosas y que consumen mucho tiempo en el desarrollo de modelos de machine learning, como la preparación de datos, la selección de características y la optimización de hiperparámetros.
- **Acelera la innovación:** Al agilizar el proceso de desarrollo de modelos, H2O.ai permite a las empresas obtener resultados más rápido y experimentar con nuevas ideas de manera más eficiente.
- **Ofrece una amplia gama de algoritmos:** Incluye una gran variedad de algoritmos de machine learning, tanto supervisados como no supervisados, para abordar una amplia gama de problemas de negocio.
- **Es escalable:** Se puede utilizar en proyectos pequeños y grandes, y se puede escalar fácilmente para manejar grandes conjuntos de datos.

¿Para qué se utiliza H2O.ai?

H2O.ai se utiliza en una amplia variedad de aplicaciones, incluyendo:

- **Predicción:** Predecir el comportamiento de los clientes, el rendimiento de las ventas, la probabilidad de fraude, etc.
- **Clasificación:** Clasificar datos en categorías diferentes, como clientes potenciales, tipos de productos, etc.
- **Detección de anomalías:** Identificar patrones inusuales en los datos que podrían indicar problemas o oportunidades.
- **Segmentación:** Dividir a los clientes en grupos homogéneos para personalizar las campañas de marketing.
- **Optimización:** Optimizar procesos y operaciones para mejorar la eficiencia y reducir costos.

¿Cuáles son las principales características de H2O.ai?

- **H2O Driverless AI:** Esta es la característica más destacada de la plataforma. Utiliza la inteligencia artificial para automatizar el proceso de desarrollo de modelos de machine learning, lo que permite a los usuarios obtener resultados de alta calidad en cuestión de minutos.
- **H2O-3:** La plataforma de código abierto de H2O.ai, que ofrece una amplia gama de algoritmos de machine learning y herramientas para la construcción y despliegue de modelos.
- **H2O Sparkling Water:** Una interfaz que permite a los usuarios de Apache Spark utilizar los algoritmos de H2O.ai.

Enlace a la plataforma e información: <https://h2o.ai/es/>

Tiempo de resolución: Reto satisfactorio 2 semanas, extensible a 3.

Valoración: Rendimiento de las posibles soluciones propuestas. Elección de herramientas. Programación, etc. Investigación sobre el modelo H2O.ia, integración con FIWARE. Representación gráfica de comparación de datos reales/predictivos.

Sugerencia de protocolo de actuación/pautas:

- Creación de nuevo modelo y suscripción.
- Entrenamiento de un modelo predictivo basado en los datos recolectados en el reto-01. En base a este modelo entrenado definir un servicio de respuesta de resultados.
- Creación de script automático, web service, endpoint, etc. (dockerizado o no) que reciba datos del día y hora de la información recopilada por las estaciones y prediga los mismo valores para el día siguiente a la misma hora.
- Incorporación y estudio de la librería “h2o”.