**Documentación del Proyecto de API de Calidad del Aire**

**Introducción**

Este proyecto presenta una API diseñada para proporcionar información sobre la calidad del aire en distintas localidades del País Vasco. La API ofrece servicios que permiten a los usuarios obtener datos actuales y predicciones sobre la calidad del aire basados en un modelo de aprendizaje automático de tipo LSTM (Long Short-Term Memory). El objetivo es facilitar el acceso a información ambiental relevante, promoviendo así la conciencia y el cuidado del medio ambiente entre la ciudadanía.

**Servicios Ofrecidos por la API**

A continuación, se detallan los endpoints disponibles en la API y su funcionalidad:

**1. /stations\_by\_distance**

**Descripción:**

Este endpoint recibe como parámetros la latitud (lat) y la longitud (lon) de una ubicación específica. Retorna una lista de todas las estaciones de monitoreo de calidad del aire ordenadas por proximidad a la ubicación proporcionada. Cada estación incluye detalles como su nombre, código identificador, coordenadas geográficas y la distancia en kilómetros desde la ubicación del usuario.

**Utilidad:**

Permite a los usuarios identificar rápidamente las estaciones más cercanas a su posición, lo cual es útil para acceder a datos relevantes y precisos sobre la calidad del aire en su entorno inmediato.

**2. /nearest\_station**

**Descripción:**

Al proporcionar una latitud (lat) y longitud (lon), este endpoint devuelve información detallada sobre la estación de monitoreo de calidad del aire más cercana. La respuesta incluye el código y nombre de la estación, sus coordenadas geográficas y la distancia exacta desde la ubicación del usuario.

**Utilidad:**

Facilita la obtención de datos específicos de la estación más próxima, permitiendo a los usuarios acceder rápidamente a información relevante sobre la calidad del aire en su ubicación.

**3. /air\_quality**

**Descripción:**

Este endpoint ofrece un resumen de la calidad del aire actual y una predicción para las próximas 24 horas en la estación más cercana a la ubicación especificada por latitud (lat) y longitud (lon). La información proporcionada incluye:

* Código y nombre de la estación más cercana.
* Coordenadas de la estación.
* Distancia desde la ubicación del usuario.
* Resumen actual de la calidad del aire.
* Predicción del resumen de la calidad del aire para las próximas 24 horas.

**Utilidad:**

Brinda a los usuarios información inmediata y predictiva sobre la calidad del aire en su entorno, lo cual es esencial para planificar actividades al aire libre y tomar decisiones informadas en caso de condiciones ambientales adversas.

**4. /air\_quality\_all**

**Descripción:**

Este endpoint permite obtener información sobre la calidad del aire para todas las estaciones disponibles o para una estación específica si se proporciona un identificador (id). La respuesta incluye:

* Código y nombre de cada estación.
* Coordenadas geográficas.
* Resumen actual de la calidad del aire.
* Predicción del resumen de la calidad del aire.

**Utilidad:**

Es una herramienta útil para obtener una visión general del estado de la calidad del aire en diferentes localidades, facilitando análisis comparativos y estudios ambientales a mayor escala.

**Utilidad de los Servicios**

Los servicios ofrecidos por la API tienen diversas aplicaciones prácticas:

* **Salud Pública:** Proporciona información crucial para personas con condiciones de salud sensibles a la contaminación del aire, ayudándoles a evitar zonas con mala calidad del aire.
* **Investigación y Educación:** Facilita el acceso a datos para investigadores, académicos y estudiantes interesados en estudios ambientales y climáticos.
* **Planificación y Gestión Ambiental:** Apoya a autoridades y organizaciones en la monitorización y mejora de la calidad del aire, contribuyendo a políticas y estrategias medioambientales.
* **Desarrollo de Aplicaciones y Servicios:** Permite a desarrolladores integrar datos de calidad del aire en aplicaciones móviles, sitios web y otros servicios digitales, ampliando el alcance y utilidad de la información.

**Detalles sobre la Generación del Modelo LSTM**

El modelo LSTM implementado es fundamental para proporcionar predicciones sobre la calidad del aire. A continuación, se detalla el proceso de creación y entrenamiento del modelo.

**Tipo de Modelo**

Se utilizó una red neuronal recurrente de tipo LSTM (Long Short-Term Memory), adecuada para manejar y aprender de datos secuenciales y temporales. Este tipo de modelo es eficaz para capturar patrones y tendencias en series temporales, como las variaciones en la calidad del aire.

**Preparación de los Datos**

**1. Recolección de Datos:**

Se recopilaron datos históricos de calidad del aire de diversas estaciones en el País Vasco, disponibles en los catálogos de datos abiertos especificados en las bases del concurso.

**2. Procesamiento de Datos:**

* **Normalización de Nombres de Estaciones:** Se desarrolló una función para normalizar los nombres de las estaciones y mapearlos con sus identificadores únicos.
* **Carga y Combinación de Datos:** Los archivos CSV de diferentes estaciones fueron leídos, teniendo en cuenta variaciones en formatos y codificaciones. Se manejaron inconsistencias y se unificaron los datos en un único conjunto.
* **Manejo de Valores Faltantes:** Se aplicaron técnicas como relleno hacia adelante y hacia atrás, y se utilizaron promedios para sustituir valores faltantes, garantizando la integridad de los datos.
* **Mapeo de Variables Categóricas:** Las categorías de calidad del aire (e.g., "Muy bueno", "Malo") fueron mapeadas a valores numéricos para su uso en el modelo.

**Entrenamiento del Modelo**

**1. Preprocesamiento para LSTM:**

* **Escalado de Características:** Se utilizaron técnicas de normalización (MinMaxScaler) para escalar las características entre 0 y 1, mejorando la eficiencia del modelo.
* **Creación de Secuencias de Tiempo:** Se generaron secuencias temporales con ventanas de 24 horas, permitiendo al modelo aprender de patrones diarios.
* **División de Datos:** Los datos se dividieron en conjuntos de entrenamiento y validación, respetando la secuencia temporal para mantener la integridad del aprendizaje.

**2. Arquitectura del Modelo:**

El modelo se compone de:

* Una capa de **enmascaramiento** para manejar valores faltantes.
* Dos capas LSTM con 128 y 64 unidades respectivamente, capaces de capturar dependencias temporales.
* Capas de **Dropout** para prevenir el sobreajuste.
* Una capa densa final con activación **softmax** para la clasificación en las categorías de calidad del aire.

**3. Configuración y Entrenamiento:**

* **Parámetros de Entrenamiento:** Se utilizó el optimizador **Adam** y la función de pérdida **sparse\_categorical\_crossentropy**.
* **Pesos de Clases:** Se calcularon pesos para manejar desequilibrios en las clases, mejorando la capacidad del modelo para predecir categorías menos representadas.
* **Callbacks Utilizados:**
  + **EarlyStopping:** Para detener el entrenamiento si el modelo no mejora, evitando el sobreajuste.
  + **ModelCheckpoint:** Para guardar el mejor modelo obtenido durante el entrenamiento.
  + **ReduceLROnPlateau:** Para ajustar la tasa de aprendizaje en función del rendimiento.

**4. Evaluación del Modelo:**

* Se evaluó el rendimiento utilizando métricas como precisión, matriz de confusión y reporte de clasificación.
* Los resultados demostraron la capacidad del modelo para predecir efectivamente la calidad del aire en base a datos históricos.

**Datos Utilizados para el Entrenamiento**

Los datos provienen de los registros de las estaciones de calidad del aire del País Vasco, incluyendo mediciones de contaminantes y variables meteorológicas como:

* **Contaminantes:** NO, NO₂, NOX, O₃, PM10, PM2.5, SO₂, CO.
* **Variables Meteorológicas:** Temperatura, Humedad, Presión Atmosférica, Velocidad y Dirección del Viento, Radiación Solar.

**Implementación en la API**

El modelo entrenado se integró en la API a través del endpoint /air\_quality, permitiendo generar predicciones en tiempo real basadas en datos actuales. El proceso incluye:

* **Obtención de Datos Recientes:** Se recuperan las mediciones más recientes de la estación más cercana al usuario.
* **Preprocesamiento:** Los datos son escalados y formateados adecuadamente para ser introducidos en el modelo.
* **Predicción:** El modelo genera una predicción de la calidad del aire para las próximas 24 horas.
* **Postprocesamiento:** Los resultados son traducidos a categorías legibles para el usuario (e.g., "Bueno", "Malo").

**Conclusión**

La API desarrollada ofrece un servicio valioso al proporcionar acceso fácil y rápido a datos y predicciones sobre la calidad del aire en el País Vasco. Al combinar datos abiertos con técnicas avanzadas de aprendizaje automático, el proyecto no solo cumple con los requisitos del concurso, sino que también aporta un recurso significativo para la comunidad, fomentando la conciencia ambiental y apoyando iniciativas de salud pública y planificación urbana.

**Nota:** Este proyecto refleja un enfoque integral que abarca desde la recolección y procesamiento de datos hasta el desarrollo y despliegue de modelos predictivos, demostrando habilidades técnicas y un compromiso con la generación de valor público y social