

6 Analítica Prescriptiva

Recientemente, se suscitó la noticia que en Quito los niveles de contaminación atmosférica están muy por sobre el nivel tolerado por la Organización Mundial de la Salud OMS. [Noticia en el portal [Primicias](#)]

El presente proyecto tiene como objetivo predecir la contaminación del aire en Beijing, China, utilizando el conjunto de datos "Beijing Multi-Site Air-Quality Data Data Set", disponible en este [link](#).

Este conjunto de datos incluye datos de contaminantes atmosféricos por hora de 12 sitios de monitoreo de calidad del aire controlados a nivel nacional. Los datos de calidad del aire provienen del Centro de Monitoreo Ambiental Municipal de Beijing. Los datos meteorológicos de cada sitio de calidad del aire están relacionados con la estación meteorológica más cercana de la Administración Meteorológica de China. El período de tiempo va desde el 1 de marzo de 2013 hasta el 28 de febrero de 2017.

Hace unos años, China estableció el Índice de Calidad del Aire (AQI) basado en el nivel de cinco contaminantes atmosféricos, a saber, dióxido de azufre (SO₂), dióxido de nitrógeno (NO₂), partículas suspendidas (PM₁₀), monóxido de carbono (CO) y ozono (O₃) medidos en las estaciones de monitoreo de cada ciudad. A cada nivel de contaminante se le asigna una puntuación individual, y el AQI final es la puntuación más alta de esos cinco contaminantes. Los contaminantes pueden medirse de manera bastante diferente. SO₂, NO₂ y PM₁₀ se miden como un promedio diario. CO y O₃ son más dañinos y se miden como un promedio por hora. El valor final del AQI se calcula por día y tiene la interpretación que se muestra en la siguiente tabla.

AQI	Air Pollution Level	Health Implications
0 - 50	Excellent	No health implications
51 -100	Good	No health implications
101-150	Slightly Polluted	Slight irritations may occur, individuals with breathing or heart problems should reduce outdoor exercise.
151-200	Lightly Polluted	Slight irritations may occur, individuals with breathing or heart problems should reduce outdoor exercise.
201-250	Moderately Polluted	Healthy people will be noticeably affected. People with breathing or heart problems will experience reduced endurance in activities. These individuals and elders should remain indoors and restrict activities.
251-300	Heavily Polluted	Healthy people will be noticeably affected. People with breathing or heart problems will experience reduced endurance in activities. These individuals and elders should remain indoors and restrict activities.
300+	Severely Polluted	Healthy people will experience reduced endurance in activities. There may be strong irritations and symptoms and may trigger other illnesses. Elders and the sick should remain indoors and avoid exercise. Healthy individuals should avoid out door activities.

Este proyecto consiste en el desarrollo de un modelo de Machine Learning que sea capaz de predecir el AQI o el Nivel de Contaminación del Aire para un día determinado. Se debe empezar por el conjunto de datos de uno de los sitios de monitoreo y, luego, si es posible, ampliar el estudio a los demás sitios de monitoreo.

Tareas

Utilizando el conjunto de datos mencionado anteriormente, las siguientes son un conjunto de tareas principales que se debe llevar a cabo, según se describe. Asimismo, los estudiantes tienen la libertad de incluir otras tareas para aumentar el valor de su trabajo.

Tarea 1: Importación de datos, limpieza y preprocesamiento

Se debe centrar en importar los datos proporcionados en un formato adecuado para que el análisis posterior sea más sencillo. También verificar si es necesario realizar alguna limpieza de datos y/o pasos de preprocesamiento.

Tarea 2: Análisis exploratorio de datos

Resumir y visualizar los datos de las formas útiles. Los estudiantes deben pensar en preguntas interesantes que se podrían comprobar con los datos disponibles y dar respuestas textuales o mediante visualización de datos.

Tarea 3: Modelado predictivo

Se debe definir una tarea predictiva que pueda ayudar a predecir la contaminación del aire, a través del valor del AQI o del Nivel de Contaminación del Aire, en función de sus características. Después de definir la tarea, se debe utilizar los datos disponibles para seleccionar y obtener un buen modelo para esta tarea. Se debe justificar el modelo sugerido.

Entregables

El proyecto debe ser realizado por grupos de dos estudiantes y debe ser enviado a través del aula virtual en la Tarea correspondiente, con los siguientes archivos:

- Un informe (de no más de 5 páginas) en formato PDF con la identificación de los miembros del grupo, y con una estructura similar a la siguiente:
 - Introducción;
 - Definición del problema;
 - Preprocesamiento de datos;
 - Análisis exploratorio de datos;
 - Modelado predictivo: configuración experimental y resultados obtenidos;
 - Conclusiones, limitaciones y trabajos futuros;
 - Anexos (opcional).
- Un cuaderno Jupyter/RMarkdown/Colab dinámico listo para ejecutarse que genere tu informe final con todo el código necesario para obtener los resultados que presentas.
- Cualquier archivo complementario necesario para ejecutar tu informe (por ejemplo, archivos de datos, objetos de datos).

Notas importantes

- Cualquier paso de preprocesamiento de datos debe presentarse y justificarse.
- Se deben indicar todos los algoritmos y parámetros utilizados.
- Se premia la organización del texto y la presentación, la claridad del lenguaje y las ideas.
- Se penalizan las largas secuencias de resultados mal formateados.
- El informe también debe hacer referencia a cualquier fuente utilizada y dejar explícito qué parte del trabajo fue influenciada por ella.
- Es importante que tu código no dependa de una ruta absoluta, para que se pueda ejecutar en cualquier computadora.
- El código fuente no debe aparecer en el informe, solo la salida de este.
- Si el cuaderno contiene código que tarda mucho tiempo en ejecutarse, se debe incluir con la opción `eval=False` para que el código no se ejecute. Si se necesita el resultado de tu código, puedes ejecutarlo localmente en tu computadora, guardarlo en un archivo binario y cargarlo en tu informe.

Fecha límite

La fecha límite para enviar el proyecto es el 30 de junio de 2023.