

Taller 3: Modelado y Análisis de Calidad del Aire en Bogotá

Curso Métodos y Modelos

Objetivo General

Aplicar técnicas de análisis de regresión, series temporales, clasificación y clustering para estudiar los factores meteorológicos que influyen en los niveles de $PM_{2,5}$ en Bogotá, su comportamiento en el tiempo y en el espacio, y la posibilidad de clasificarlos o preverlos.

Base de Datos

Se utilizará una base de datos con variables meteorológicas (temperatura, humedad, radiación solar, viento, etc.) y concentraciones de $PM_{2,5}$ entre los años 2016 y 2023, para 19 estaciones de monitoreo de Bogotá.

Pregunta Central

¿Qué factores meteorológicos explican los niveles de $PM_{2,5}$ en Bogotá, cómo varían en el tiempo y el espacio, y cómo pueden preverse o clasificarse los episodios de contaminación?

Instrucciones:

1. Lea cada una de las partes del taller y realice los ejercicios planteados.
2. Responda las preguntas conceptuales de manera argumentada y soportado en soluciones con modelos computacionales como en Python, en el caso de requerirse.

1. Análisis por Temática

1.1. 1. Regresión

- 1.1. ¿Qué variable meteorológica explica en mayor medida las concentraciones de $PM_{2,5}$? Justifique su respuesta aplicando modelos de regresión lineal simple para cada variable.

- 1.2. ¿Qué combinación de variables meteorológicas permite predecir con mayor precisión los niveles de $PM_{2,5}$? Ajuste un modelo de regresión múltiple y evalúe su desempeño utilizando R^2 , MAE y RMSE.
- 1.3. ¿El modelo de regresión es consistente entre estaciones o varía significativamente entre distintas zonas de Bogotá?

1.2. 2. Análisis de Series Temporales

- 2.1. ¿Existen patrones de tendencia o estacionalidad en los niveles de $PM_{2,5}$ para alguna estación representativa o para el promedio ciudad? Realice una descomposición aditiva de la serie diaria.
- 2.2. ¿Qué meses o temporadas concentran los mayores niveles de contaminación? ¿Coincide esto con comportamientos particulares en las variables meteorológicas?

1.3. 3. Clasificación con Árboles de Decisión

- 3.1. Cree una variable categórica de calidad del aire (por ejemplo: Buena, Moderada, Mala) con base en rangos de $PM_{2,5}$ establecidos en la Resolución 2254 del 2017. ¿Qué variables meteorológicas permiten clasificar adecuadamente estas categorías?
- 3.2. ¿Cuál es la regla de decisión más utilizada en el árbol entrenado? Interprete la lógica detrás de las decisiones tomadas por el modelo.

1.4. 4. Clustering (KMeans)

4A. Clustering de días con características similares

- 4A.1. Seleccione variables meteorológicas diarias y aplique KMeans para agrupar los días con patrones similares. ¿Cuántos grupos son apropiados según el método del codo?
- 4A.2. ¿Qué características comunes tienen los días más contaminados según los grupos obtenidos?
- 4A.3. ¿Puede relacionar los grupos con eventos de contaminación identificados en los registros?

4B. Clustering de estaciones

- 4B.1. Resuma por estación el comportamiento promedio de las variables meteorológicas y del $PM_{2,5}$ (promedios anuales o mensuales).
- 4B.2. Aplique clustering sobre las estaciones. ¿Qué grupos se forman? ¿Coinciden con zonas geográficas o características similares?

4B.3. Visualice los grupos en un mapa y discuta posibles implicaciones para la gestión ambiental.

2. Pregunta Integradora Final

¿Qué condiciones meteorológicas y espaciales se repiten sistemáticamente en los episodios de mayor contaminación?

- Integre hallazgos de regresión, series temporales, clasificación y clustering.
- Apoye su análisis con visualizaciones y comparaciones entre estaciones.
- Proponga una breve conclusión con recomendaciones o hipótesis para futuras investigaciones.

Instrucciones de entrega:

- La entrega del taller debe realizarse en un documento en formato PDF, con redacción clara, organizada, concisa y estructurada por secciones.
- Todas las respuestas deben incluir la explicación detallada del razonamiento, no solo el resultado final.
- Se espera que los estudiantes analicen, interpreten y argumenten sus respuestas, especialmente en la formulación de modelos y justificación de decisiones.
- El uso de gráficas, ecuaciones, esquemas y tablas es obligatorio cuando estos elementos apoyen o clarifiquen el análisis.
- Los códigos en Python pueden utilizarse como soporte técnico para resolver los modelos, pero no reemplazan la explicación matemática ni conceptual. Es decir: el desarrollo en Python debe servir como base de análisis, no como único medio de respuesta.
- La entrega final debe realizarse de acuerdo con los grupos previamente definidos y ser subida a la plataforma **AVATA**.