**Proyecto Bedu (Programación y Estadística con R)**

**Modulo 2**

**“Monitoreo y Predicción Inteligente de la Calidad del Aire en México”**

La calidad del aire es uno de los desafíos ambientales más graves que enfrenta México, con profundas implicaciones para la salud pública y el bienestar general. La exposición a contaminantes atmosféricos como el dióxido de nitrógeno (NO2), el ozono (O3), y las partículas en suspensión (PM2.5 y PM10) está asociada con una amplia gama de problemas de salud, incluidas enfermedades respiratorias, cardiovasculares y ciertos tipos de cáncer. A pesar de la gravedad de estos riesgos, muchas ciudades en México carecen de sistemas adecuados de monitoreo continuo y accesible de la calidad del aire. Esta deficiencia limita la capacidad de las autoridades para tomar decisiones informadas y proactivas, impidiendo la implementación efectiva de políticas que podrían mitigar la contaminación.

El reto es aún mayor debido a la falta de datos en tiempo real y accesibles, que son esenciales para predecir y manejar contingencias ambientales. La relación entre el uso de combustibles y los niveles de contaminación tampoco se entiende completamente, lo que dificulta la formulación de estrategias para reducir la exposición a estos contaminantes. Por lo tanto, es crucial desarrollar un sistema avanzado de monitoreo y predicción que no solo detecte contingencias ambientales de manera oportuna, sino que también identifique patrones en el uso de combustibles que contribuyan a la contaminación del aire. Este sistema proporcionará una herramienta poderosa para mejorar la calidad del aire, protegiendo la salud pública y promoviendo un ambiente más limpio y seguro para todos los habitantes de México.

**Problema para Resolver 🚀**

La calidad del aire en México, especialmente en las grandes ciudades como Ciudad de México, es un problema grave que afecta la salud de millones de personas. La exposición continua a contaminantes como NO2, O3, PM2.5 y PM10 está relacionada con enfermedades respiratorias, cardiovasculares y cáncer. Aunque existen estaciones de monitoreo de calidad del aire, la cobertura es limitada y no siempre se tiene acceso a datos en tiempo real, lo que dificulta la detección y manejo oportuno de contingencias ambientales.

**Propuesta para Resolver la Necesidad 💡**

Desarrollar un sistema de monitoreo y predicción de la calidad del aire en tiempo real utilizando técnicas de análisis de datos y aprendizaje automático. Este sistema permitiría detectar, predecir y alertar sobre contingencias ambientales, así como identificar patrones en el uso de combustibles y su relación con los niveles de contaminación. Esto facilitaría la toma de decisiones informadas por parte de las autoridades y la implementación de políticas efectivas para reducir la contaminación del aire.

**¿Cómo Lo Vas a Hacer? 📋**

**1. Elección de Fuente de Datos:**

* SINAICA (Sistema Nacional de Información de la Calidad del Aire): Datos históricos y en tiempo real sobre la calidad del aire en México.

**2. Recolección y Preparación de Datos:**

* Importar datos de las fuentes seleccionadas.
* Limpiar y normalizar los datos para garantizar la consistencia y precisión.
* Integrar datos de diferentes fuentes en un solo data frame.

**3. Análisis Exploratorio:**

* Visualizar tendencias históricas de los niveles de contaminantes.
* Identificar correlaciones entre niveles de contaminación y uso de combustibles.
* Determinar patrones estacionales y diarios en la calidad del aire.

**4.Desarrollo de Modelos de Predicción:**

* Usar técnicas de aprendizaje automático para predecir contingencias ambientales.
* Entrenar modelos en datos históricos y validar su precisión.
* Implementar modelos para la predicción en tiempo real de la calidad del aire.

**5. Desarrollo de un Sistema de Alerta Temprana:**

* Crear un sistema que emita alertas en tiempo real cuando se detecten niveles críticos de contaminación.
* Integrar el sistema con aplicaciones móviles y sitios web para alertar a la población y autoridades.

**6. Implementación de un Dashboard Interactivo:**

* Desarrollar un dashboard que permita visualizar la calidad del aire en tiempo real, las predicciones y las alertas emitidas.
* Incluir gráficos y mapas interactivos que muestren la evolución de la calidad del aire y las áreas más afectadas.

**Desarrollo**

**Desarrollo 1: Análisis de Datos de Contaminantes Atmosféricos**

1. **Importación de datos:**
   * Encuentra y descarga un conjunto de datos de contaminantes atmosféricos. Por ejemplo, puedes usar datos de la calidad del aire proporcionados por el Sistema Nacional de Información de la Calidad del Aire en México (SINAICA) o algún otro recurso disponible en línea.
   * Importa los datos a R usando read.csv() o readr::read\_csv().
2. **Exploración de los datos:**
   * Usa str(), head(), View() y summary() para entender la estructura del data frame y los tipos de contaminantes que se miden.
3. **Extracción de columnas:**
   * Extrae las columnas que contienen los niveles de diferentes contaminantes (por ejemplo, PM10, PM2.5, NO2, O3) y las fechas correspondientes.
4. **Análisis de frecuencias:**
   * Usa la función table() para crear tablas de frecuencias de los niveles de contaminantes en diferentes rangos (por ejemplo, niveles bajos, moderados, altos).
   * Calcula las probabilidades de que un contaminante específico esté en un rango determinado

**Desarrollo 2: Unión de Datos de Múltiples Años**

1. **Importación de datos de diferentes años:**
   * Descarga los datos de varios años, si están disponibles, para observar tendencias a largo plazo.
2. **Exploración y limpieza de datos:**
   * Usa str(), head(), View() y summary() para explorar y limpiar los datos de cada año.
3. **Transformación y unión:**
   * Asegúrate de que las columnas correspondientes tengan el mismo tipo y formato.
   * Une los data frames en uno solo utilizando rbind() o do.call().

**Desarrollo 3: Graficación y Análisis de Probabilidades**

1. **Tablas de frecuencias relativas:**
   * Calcula las probabilidades de que los niveles de cada contaminante caigan en diferentes rangos usando prop.table().
2. **Gráficos:**
   * Crea gráficos de barras para visualizar la distribución de los niveles de contaminantes.
   * Crea un HeatMap para visualizar las combinaciones de niveles de dos contaminantes específicos.

**Desarrollo 4: Bootstrap y Análisis de Independencia**

1. **Bootstrap:**
   * Realiza un procedimiento de bootstrap para analizar la variabilidad de los niveles de contaminantes y su independencia.
2. **Análisis de cocientes:**
   * Calcula cocientes para comparar la probabilidad conjunta de ciertos niveles de contaminantes con las probabilidades marginales.

**Desarrollo 5: Creación y Uso de SmallData**

1. **Creación de SmallData:**
   * Crea un data frame SmallData con columnas relevantes, como Date, PM10, PM2.5, NO2, O3.
2. **Guardado y uso de datos:**
   * Guarda el data frame como un archivo CSV usando write.csv() y posteriormente úsalo para análisis posteriores.

**Desarrollo 6: Análisis de la Suma de Contaminantes**

1. **Cálculo de suma de contaminantes:**
   * Agrega una nueva columna sumaContaminantes para la suma de todos los contaminantes en un día.
2. **Promedio por mes:**
   * Calcula el promedio mensual de la suma de contaminantes.
3. **Serie de tiempo:**
   * Crea y grafica una serie de tiempo del promedio mensual de contaminantes.

**Desarrollo 7: Uso de MongoDB Compass**

1. **Base de datos en MongoDB:**
   * Carga el archivo de datos de contaminantes en MongoDB y realiza consultas similares para extraer información específica, como niveles de contaminantes en días específicos.
2. **Consultas y operaciones:**
   * Realiza consultas para analizar patrones de contaminación, como en días con niveles excepcionalmente altos o bajos.

**Desarrollo 8: Creación de un Dashboard**

1. **Generación de gráficos:**
   * Ejecuta análisis similares a los de momios en datos de contaminantes y guarda los gráficos en formato PNG.
2. **Dashboard en Shiny:**
   * Crea un archivo app.R que contenga un dashboard con pestañas para visualizar diferentes análisis de contaminantes atmosféricos.

**Conclusiones**

El proyecto 'Monitoreo y Predicción Inteligente de la Calidad del Aire en México' demostró la importancia de integrar datos históricos y en tiempo real para analizar la calidad del aire, revelando patrones de contaminación y correlaciones entre contaminantes como PM10 y PM2.5. La visualización de datos y el uso de técnicas de aprendizaje automático permitieron predecir contingencias ambientales y ofrecer herramientas valiosas para la toma de decisiones informadas. Los resultados subrayan la necesidad de políticas más estrictas y sistemas de monitoreo avanzado para mejorar la calidad del aire y proteger la salud pública en México.

**Conclusiones del Proyecto 'Monitoreo y Predicción Inteligente de la Calidad del Aire en México'**

1. **Eficacia en la Recolección y Preparación de Datos:**
   * Se logró integrar de manera efectiva datos históricos de varios años sobre contaminantes atmosféricos, permitiendo un análisis longitudinal que puede mostrar tendencias y cambios en la calidad del aire en México.
   * La limpieza y transformación de datos fue fundamental para asegurar la calidad del análisis, especialmente en el manejo de datos faltantes y la estandarización de formatos.
2. **Análisis de la Calidad del Aire:**
   * El análisis de frecuencias y probabilidades reveló que ciertos contaminantes, como el PM10 y PM2.5, tienen una distribución que varía significativamente con el tiempo y la ubicación, destacando la importancia de un monitoreo constante.
   * Los cocientes calculados para la probabilidad conjunta de niveles elevados de PM10 y PM2.5 sugieren una correlación entre estos contaminantes, lo cual podría indicar fuentes comunes de emisión o condiciones ambientales que favorecen su coexistencia.
3. **Visualización y Comunicación de Resultados:**
   * Las visualizaciones generadas, incluidas las series de tiempo y los gráficos de barras, facilitaron la comprensión de las tendencias en la calidad del aire, mostrando cómo ciertos contaminantes varían a lo largo del tiempo.
   * La creación de un dashboard interactivo con Shiny representa una herramienta valiosa para la comunicación de resultados a un público más amplio, permitiendo la exploración dinámica de los datos.
4. **Uso de Técnicas de Bootstrap:**
   * El análisis de bootstrap proporcionó estimaciones robustas de la media de contaminantes, lo cual es esencial para evaluar la incertidumbre en las mediciones y mejorar la confiabilidad de las conclusiones.
5. **Importancia de la Suma de Contaminantes:**
   * La suma de contaminantes atmosféricos y su análisis en función del tiempo permitió identificar períodos críticos donde la calidad del aire alcanzó niveles peligrosos, sugiriendo la necesidad de intervenciones específicas en ciertos meses del año.
6. **Implicaciones para la Política Pública:**
   * Los resultados de este proyecto subrayan la necesidad de políticas más estrictas para la gestión de la calidad del aire en México, especialmente en zonas urbanas donde los niveles de contaminación tienden a ser más altos.
   * La implementación de sistemas de monitoreo en tiempo real y la predicción basada en modelos pueden ser herramientas efectivas para anticipar contingencias ambientales y tomar medidas preventivas.
7. **Perspectivas Futuras:**
   * La integración de datos adicionales, como información meteorológica y fuentes de emisión, podría mejorar la capacidad predictiva del sistema.
   * La extensión del análisis para incluir contaminantes emergentes y su relación con la salud pública podría ofrecer insights más profundos para la toma de decisiones informadas en la gestión ambiental.

**Referencias**

1. World Health Organization (WHO). (2022). Air Pollution. Retrieved from https://www.who.int/health-topics/air-pollution
2. Instituto Nacional de Ecología y Cambio Climático (INECC). (2021). Informe Nacional de Calidad del Aire 2020. Retrieved from https://www.gob.mx/inecc/
3. Secretaría del Medio Ambiente (SEDEMA), Ciudad de México. (2022). Programa de Contingencias Ambientales Atmosféricas de la Zona Metropolitana del Valle de México. Retrieved from https://www.sedema.cdmx.gob.mx/