Sensor bajo costo:

**flowchart TD**

**B[Investigación tecnologías disponibles en el mercado]**

**B --> C[Desarrollar y programar el código]**

**C--> D[Muestreo]**

**D--> E[Medición simultanea entre los sensores de bajo costo]**

**D--> F[Medición en simultáneo con la estación de referencia]**

**E--> G[Validación del consenso entre las mediciones]**

**F--> H[Validación con respecto a una estación de referencia]**

Desarrollar y programar el código que permitiría la conexión y comunicación entre los diferentes componentes.

Base de datos de mediciones conjuntas entre los diez sensores de bajo costo desarrollados.

Medición en simultáneo con la estación de referencia certificada fueron provistos por el Centro de Investigación y Química Ambiental (CIQA).

Validación del consenso entre las mediciones de los sensores de bajo costo.

Comparación con una estación de referencia.

Exploratorio:

**flowchart TD**

**B[Muestreo simultáneo de concentración de PM<sub>2.5</sub> en los sitios de muestreo]**

**B --> C[Preprocesamiento de datos]**

**C--> D[Análisis de las características temporales de la serie de concentración de PM<sub>2.5</sub>]**

**D--> E[Análisis espacial de la concentración de PM<sub>2.5</sub>]**

**E--> F[Formación de clústeres en función del patrón de emisión de PM<sub>2.5</sub>]**

**F--> G[Análisis espacio-temporal por uso de suelo de la concentración de PM<sub>2.5</sub>]**

Muestreos simultáneos de concentración en diez sitios seleccionados sobre distintos usos de suelo.

Preprocesamiento de datos

Análisis de las características temporales de la serie de concentración de PM<sub>2.5</sub>

Análisis espacial de la concentración de PM <sub>2.5</sub>

Formación de clusteres en función del patrón de emisión

Análisis espacio-temporal por uso de suelo de la concentración de PM<sub>2.5</sub>

Series:

**flowchart TD**

**B[Descarga variables satelitales horarias durante un año de muestreo de la concentración de PM<sub>2.5</sub>]**

**B--> C[Análisis descriptivo de la variabilidad conjunta entre las series horarias de concentración de PM<sub>2.5</sub> y de variables satelitales]**

**C--> D[Transformación de las series temporales en estacionarias]**

**D--> E["Ajuste de modelo teórico a cada una de las series temporales de las variables satelitales (SARIMA)"]**

**E--> ONE**

**subgraph ONE [Pre blanqueamiento]**

**F[Cálculo de las diferencias entre valores observados y predichos por el modelo ajustado para cada una de las series de variables satelitales]**

**F-->G[Cálculo de los residuos de la serie de concentración de PM<sub>2.5</sub> respecto del modelo teórico ajustado para cada una de las series de variables satelitales]**

**G-->H[Análisis del diagrama de correlación cruzada entre ambas series residuales]**

**H--> I([Identificación de los retardos de mayor capacidad predictiva])**

**end**

**ONE --> J([Estimación de un modelo de función de transferencia dinámica lineal entre la concentración de PM<sub>2.5</sub> y cada una de las variables satelitales para estimar la dirección de las relaciones])**

**J--> K([Conclusión acerca del vínculo de la concentración de PM<sub>2.5</sub> con cada una de las variables satelitales])**

Descargar de variables meteorológicas y satelitales durante el periodo de muestreo de la concentración de PM<sub>2.5</sub>

Análisis descriptivo de la variabilidad conjunta entre la concentración de PM<sub>2.5</sub> y variables satelitales

Transformación de las series en estacionarias

Ajuste de modelo teórico a cada una de las series temporales de las variables satelitales (SARIMA)

Preblanqueamiento

Cálculo de las diferencias entre valores observados y predichos por el modelo ajustado para cada una de las variables satelitales

Cálculo de los residuos de la serie de concentración de PM<sub>2.5</sub> respecto del modelo teórico ajustado para cada una de las series de variables satelitales

Cálculo y análisis del diagrama de correlación cruzada entre ambas series filtradas

Identificación de los retardos de mayor capacidad predictiva

Estimación de un modelo de función de transferencia dinámica lineal entre la concentración de PM<sub>2.5</sub> y la variable satelital para estimar la dirección de las relaciones

Modelado:

**flowchart TD**

**B[Concentración de PM<sub>2.5</sub> a partir de sensores de bajo costo]**

**Z[Variables satelitales de público acceso]**

**B -->C[Preparación de la base para el modelado predictivo]**

**Z -->C**

**C--> D[Selección de características para cada uso de suelo]**

**D -->|Todas las variables| E[(BD original)]**

**D -->|Reducción urbano| F[(BD urbana)]**

**D -->|Reducción agrícola y área verde| G[(BD agrícola y área verde)]**

**D -->|Reducción industrial| H[(BD industrial)]**

**E--> TWO**

**F--> TWO**

**G--> TWO**

**H--> TWO**

**subgraph TWO [Modelado y evaluación de algoritmos]**

**K[División datos en entrenamiento y testeo]**

**K-->L[Entrenamiento de algoritmos predictivos]**

**L-->X[Predicción de datos]**

**X-->M[Cálculo métricas desempeño]**

**end**

**TWO-->Y["Selección del mejor modelo"]**

**Y-->S[Validación de predicciones]**

Diagrama dos partes

**flowchart** TD

    A([Inicio]) **-->** B[Descarga bases de datos satelitales]

    B **-->**C[(Base de datos)]

    C**-->**ONE

**subgraph** ONE [Preparación de la base para el modelado predictivo]

    D[Preprocesamiento de datos]

    D**-->**F[Ingenieria de variables]

    F**-->**G[Selección de características]

**end**

    G **-->**|Todas las variables| H[(BD Original)]

    G **-->**|Reducción LASSO| I[(BD LASSO)]

    G **-->**|Reducción Featurewiz| J[(BD Featurewiz)]

Primero y segundo:

H[(BD Original)] **-->** **TWO**

 I[(BD LASSO)] **-->** **TWO**

 J[(BD Featurewiz)] **-->** **TWO**

**subgraph** TWO [Modelado y evaluación de algoritmos]

    K[Dividisión datos en entrenamiento y testeo]

    K**-->**L[Entrenamiento de algoritmos predictivos]

    L**-->**X[Predicción para datos de entrenamiento y testeo]

    X**-->**M[Cálculo métricas desempeño para entrenamiento y testeo]

**end**

    TWO**-->**Y[Sembrar 10 semillas aleatorias]

    Y**-->** N[Repetir diez veces el procedimiento]

    N**-->**S[Cálcular la media de las metricas de desempeño]

    S**-->**T[Determinar el algoritmo predictivo de mayor capacidad]

    T**-->**U([FIN])

El procedimiento que sigo es el siguiente:

1. A partir de las múltiples variables satelitales descargadas creo una base de datos a través de un preprocesamiento, unificación y combinación de las mismas.
2. Luego realizo una extracción de características y transformación de las mismas para crear nuevas características y aumentar la capacidad predictiva de las que ya están involucradas. Además, llevo a cabo el escalado de las variables numéricas involucradas en la base de datos.
3. Una vez con la base de datos completa continúo realizando la selección de características donde busco reducir la base de datos a las variables de mayor importancia para llevar a cabo el modelado predictivo de la variable respuesta, este procedimiento de selección de características lo realizo a partir del algoritmo de LASSO y featurewiz.
4. Una vez realizado este procedimiento cuento con 3 bases de datos: la original sin reducir, la reducida por LASSO y la reducida por featurewiz.
5. Luego para cada una de estas bases divido los datos en entrenamiento y testeo, entreno múltiples algoritmos, entre los cuales se encuentran las redes neuronales recurrentes, bosques aleatorios y regresión lineal múltiple y calculo las métricas de raíz del error cuadrático medio (RMSE), error absoluto medio (MAE), coeficiente de determinación (R^2), coeficiente de correlación de Pearson (r) y coeficiente de concordancia de Lin (CCC) tanto para las predicciones en la base de datos de entrenamiento como en la de testeo.
6. Este procedimiento de la división de los datos, entrenamiento del algoritmo y cálculo de métricas lo realizo 10 veces fijando distintas semillas aleatorias que generen que cada uno de los resultados sean diferentes. Una vez calculadas genero los promedios de las métricas para cada una de las bases de datos y de sus particiones (entrenamiento y testeo).
7. Una vez obtenidas las tablas de las métricas para entrenamiento y testeo para cada una de las bases de datos, determino cual de las combinaciones (selección de características y modelado predictivo) es el que tiene mayor capacidad predictiva.

    C[Extracción y transformación de características]

    C **-->** D[Selección de características]

    D **-->** E[Dividir datos en entrenamiento y prueba]

    E **-->** F[Entrenamiento y evaluación de algoritmos]

    F **-->** G[Repetir 10 veces con distintas semillas aleatorias]

    G **-->** H[Calcular promedios de métricas]

    H **-->** I[Determinar la mejor combinación predictiva]

    I **-->** id1((Fin))

    direction LR

    L**-->** A1[Redes neuronales recurrentes]

    L**-->** A2[Bosques aleatorios]

    L**-->** A3[Regresión lineal múltiple]

    direction TB