

Introducción

En el presente texto se realiza un Seguimiento al Método de la Ingeniería efectuado para solucionar un problema aquí también expuesto. Se parte de la Identificación del problema a solucionarse, pasando por diferentes fases propias de este Método, hasta llegar a propuestas de soluciones y evaluación de estas. Finalmente, todo converge en la escogencia de una de las propuestas, la cual termina en ser implementada para solucionar el problema descrito.

Fase 1: Identificación del Problema

Descripción del Contexto Problemático

En las últimas décadas, la contaminación se ha convertido en un problema de dominio global. Múltiples hechos ocurridos en el pasado han traído consecuencias graves para los ecosistemas de la actualidad, lo que ha empezado a generar preocupaciones en todo el mundo. Es por esto que diversas naciones a lo largo del globo están empezando a tomar medidas con el fin de mitigar los efectos de la contaminación, esperando reducirla.

Colombia, uno de estos países, empezó a reunir datos de la calidad del aire y factores climatológicos a lo largo y ancho del territorio nacional, recogidas por las distintas Autoridades Ambientales del Subsistema de Información de la Calidad del Aire, en el periodo entre 2011 y 2017. Estos datos, además, se tomaron según los parámetros contemplados en el Protocolo para el Monitoreo y Seguimiento de la Calidad del Aire (SISAIRE), lo que garantiza su estandarización y facilidad de manejo como un conjunto.

Definición del Problema

Se busca con estos datos conocer los distintos niveles de contaminación que se han presentado en el territorio nacional, a lo largo del tiempo. Esto con el fin de identificar las zonas con niveles más drásticos, para tomar acciones adecuadas y reducirlos. Se espera, además, estar en capacidad de predecir el comportamiento de estas en un futuro.

Por otro lado, también se busca saber en qué nivel tomarían efecto las medidas que se pudiesen tomar al respecto. Esto es reconociendo que las acciones a tomarse no producirán un efecto inmediato, y sus resultados se empezarán a notar pasado un tiempo de realizarse.

Identificación de necesidades

Las siguientes son las necesidades identificadas para la solución de este problema:

- Se requiere leer y gestionar la información recogida por las Autoridades Ambientales del SISAIRE.
- Se necesita identificar las zonas más contaminadas del país a partir de los datos recogidos.
- Se necesita encontrar las zonas que más han aumentado sus niveles de contaminación según los datos recogidos.
- Se requiere mostrar los cambios a lo largo del tiempo en los niveles de contaminación de una zona.
- Se necesita estar en la capacidad de predecir los cambios en los niveles de contaminación de una zona, tanto si se interviene en ella como si no.
- Se necesita permitir una visualización de los datos a través de gráficas que comparan diversas variables estudiadas.
- Se requiere permitir una visualización geográfica de los datos de contaminación a través de mapas de calor del territorio nacional.

Fase 2: Recopilación de la Información Necesaria:

Para encontrar una solución óptima para este problema, es necesario conocerle de antemano. Tras una investigación intensiva, aquí se resume la información recopilada que puede llegar a resultar útil a la hora de solucionar el problema:

GMaps.NET:

Es un control del framework .NET de código abierto, gratuito y multiplataforma, que usa enrutamiento, geo codificación, direcciones y mapas de Google, Yahoo!, Bing, entre otros. Es utilizado en diversas aplicaciones desarrolladas en el framework para visualizar mapas geográficos de diferentes zonas, dadas las coordenadas geográficas de la zona deseada a conocer. Se utiliza en aplicaciones C# de tipo Windows Forms y Mobile.

IDEAM:

“El IDEAM es una institución pública de apoyo técnico y científico al Sistema Nacional Ambiental, que genera conocimiento,

produce información confiable, consistente y oportuna, sobre el estado y las dinámicas de los recursos naturales y del medio ambiente, que facilite la definición y ajustes de las políticas ambientales y la toma de decisiones por parte de los sectores público, privado y la ciudadanía en general.” (IDEAM, s.f.)

Subsistema de Información Sobre Calidad del Aire – SISAIRE:

El SISAIRE es la principal fuente de información para el diseño, evaluación y ajuste de las políticas y estrategias nacionales y regionales de prevención y control de la calidad del aire. Es un sistema bajo ambiente Web que permite al IDEAM la captura y revisión constante de los datos que las autoridades ambientales toman sobre calidad del aire y reportan al Sistema. (SISAIRE, s.f.)

Autoridad Ambiental

Son actores de orden institucional que, de acuerdo con la Ley 99 de 1993, se encargan de la gestión integral del medio ambiente en el país. Incluyen, pero no se limitan a, Corporaciones Autónomas Regionales, el Ministerio de Medio Ambiente y Desarrollo Sostenible, y la Autoridad Nacional de Licencias Ambientales. (ANI, s.f.)

Tiempo de exposición

Periodo normativo en el cual se toma la medición de una variable para un estudio.

Tecnología:

Se refiere al conjunto de técnicas utilizadas mediante diferentes métodos para obtener un resultado deseado.

Monitoreo de aire:

“Un Sistema de Vigilancia de la Calidad del Aire – SVCA se puede definir como un conjunto de equipos de monitoreo de los contaminantes atmosféricos, que se encuentran instalados en un lugar de interés con un propósito determinado; esta infraestructura debe estar acompañada de todas las actividades necesarias para su correcto funcionamiento, dentro de las cuales se puede mencionar la operación por personal calificado, programas de mantenimiento preventivo y correctivo, un sistema de administración de información que permita una correcta validación de los datos, entre otras. Su implementación se realiza bien sea cuando se presenta una problemática específica relacionada con la calidad del aire, o en zonas con población por encima de los cincuenta mil habitantes.” (SIAC, s.f.)

Formato JSON:

Es un formato de texto sencillo para el intercambio de datos. Se trata de un subconjunto de la JavaScript Object Notation, y es utilizado en general como una alternativa a XML.

Estado del arte:

El problema ya descrito ha sido abordado en ocasiones posteriores. Aquí se listan algunas de las alternativas ya propuestas como solución a este problema:

Calidad del Aire AirVisual (iOS y Android):

Esta aplicación muestra datos globales de la calidad del aire, así puedes saber cómo está la contaminación en distintos países y prepararte si vas a viajar. Cuenta con pronóstico de 7 días y puedes activar alertas específicas para la zona o alcaldía en la que te encuentras. (IQAir AG, s.f.)



Aire (iOS y Android):

La aplicación oficial del Gobierno de la Ciudad de México envía notificaciones con información sobre la contaminación en la zona, la radiación solar ultravioleta y el pronóstico ambiental para el día siguiente. (Dirección de Monitoreo Atmosférico, s.f.)



Aire Bogota (iOS y Android):

Esta aplicación permitirá conocer el IBOCA (Índice Bogotano de calidad del aire) para los ciudadanos de la ciudad de Bogotá, además de generar reportes y pronósticos del comportamiento de los contaminantes (PM2.5, PM10, O3) cada hora, así como las respectivas recomendaciones y acciones voluntarias que se deben tener en cuenta para prevenir complicaciones de salud. (Secretaría Distrital de Ambiente Bogotá, s.f.)



Plume:

Plume Air Report proporciona en tiempo real los niveles de contaminación en su área, y pronostica cómo evolucionará la calidad del aire cada hora, durante las próximas 24 horas; al igual que un pronóstico meteorológico.

Mide la calidad del aire en tu ciudad, el nivel de contaminación, el nivel de cada contaminante, un índice de calidad del aire internacional según tu zona del mundo y

unas recomendaciones básicas sobre actividad física al aire libre. (Plume Labs, s.f.)



Se dan de ejemplo estas aplicaciones por el uso de sus datos georreferenciados, que permiten conocer la ubicación y distribución de las zonas contaminadas. Además, pronostican cómo evolucionará la calidad de aire en un periodo de tiempo. Estas aplicaciones son tenidas en cuenta porque tienen una relación con el problema que se desea desarrollar, permitiendo una visualización más amplia y fácilmente trabajable de lo que se espera resolver.

Fase 3: Búsqueda de Soluciones Creativas:

Teniendo en cuenta las distintas necesidades que se identificaron en la primera parte de este documento, se reconocen dos principales problemas: leer y analizar los datos recogidos por el SISAIRE, y predecir el comportamiento a futuro de estos. Para ello, se realizaron dos procesos de Brainstorming, para cada uno de los problemas a solucionar.

Primeramente estas fueron las ideas propuestas para leer y analizar los datos recogidos:

- Organizar los datos recogidos mediante una estructura de datos conocida previamente (Hashtable, Árbol Binario de Búsqueda, etc)
- Organización de los datos a través de objetos y clases que almacenan y clasifican la información recogida (Clase Ciudad, Contaminante, Entidad Ambiental, etc)
- Análisis de los datos mediante fórmulas y herramientas estadísticas propias de Microsoft Excel.
- Visualizar los datos a través de mapas en el framework .NET:
- Desarrollar una aplicación que utilice los mapas proporcionados por GMaps, utilizando el lenguaje de programación C#
- Clasificar los datos reunidos de acuerdo con el área de cuyos datos se desean conocer.
- Plasmar los datos clasificados en un mapa de calor, en donde las zonas más fuertes indican las zonas más contaminadas.
- Permitir la visualización de los datos recogidos históricamente mediante un control que modifique la fecha a mostrarse.
- Leer y analizar los datos recogidos mediante el lenguaje de programación estadístico R
- Realizar una lectura parcial de la base de datos mediante la utilización del formato JSON, consultando únicamente la información necesitada en el momento de cada operación realizada.

Por otro lado, estas son las ideas propuestas para realizar predicciones y análisis a futuro de los datos:

- Utilización de modelos de análisis predictivo de datos:
- Modelos descriptivos
- Modelos de decisión
- Modelo ensemble
- Modelo uplift
- Analizar las series de tiempo mediante distintos métodos estadísticos:
- ARIMA (Modelo autorregresivo integrado de promedio móvil).
- ARMA (Modelo autorregresivo integrado de media móvil).
- Método de Holt
- ANOVA (Análisis de la varianza).
- Aplicación de técnicas de predicción

- Técnicas de regresión (lineal, árboles de clasificación, etc)
- Técnicas de aprendizaje computacional (redes neuronales, algoritmo de vecino más próximo, etc)
- Realización de predicciones aleatorias sin fundamento estadístico
- Realización de predicciones de acuerdo con criterios propios de los ingenieros (tal persona opina que tal zona incrementará su contaminación en tanto por ciento, por dar ejemplo).

Paso 4: Transición de ideas a diseños preliminares:

En esta fase se realiza un pensamiento algo más profundo y con cuerpo de las ideas propuestas anteriormente. Esto conlleva, inicialmente, a que algunas ideas no se encuentren factibles como una solución al problema, por lo que se listan las razones por las que estas se descartan. Más adelante, se recuenta aquellas ideas que se mantienen, y se realiza un diseño preliminar de cómo podrían ser estas implementadas.

Filtro Inicial de Propuestas

Luego de un ejercicio de lluvia de ideas para generar posibles soluciones a los problemas expuestos, se encontró que algunas de las propuestas no son realmente óptimas para este cometido. A continuación se exponen las ideas que se descartaron y los motivos para ello.

Respecto a la lectura y análisis de los datos recogidos:

- La propuesta 1 se descarta porque se hará uso de un conjunto de datos con un tamaño considerablemente grande. Construir y organizar estos datos en alguna de las estructuras de datos expuestas no resulta concebible, pues la cantidad de datos que hay exige un tamaño de memoria bastante grande, con el que no resulta óptimo realizar operaciones, lo que destruye los conceptos con los que las estructuras de datos fueron diseñadas. Es por esto que se descarta la opción 1.
- En la propuesta 2 se plantea organizar y dividir los datos a través de estructuras de datos propias, las cuales representan entidades del entorno que entrarían en juego para analizar estos datos. Sin embargo, cómo se encontró que se puede utilizar el formato JSON para intercambiar datos entre la aplicación desarrollada y la base de datos almacenada en la nube, se pueden realizar consultas directamente sobre ésta, sin haber una necesidad de organizar la totalidad de datos encontrados. Es por esto que se descarta esta opción.
- Microsoft Excel, el gestor de hojas de cálculo de Microsoft Office, ha probado ser capaz de realizar buenos análisis estadísticos y matemáticos sobre conjuntos de datos. Sin embargo, la aplicación posee un límite al tamaño de datos con los que puede trabajar, y la base de datos que se está utilizando

supera este límite, lo que impide el uso de esta herramienta. Por esto, la opción 3 queda descartada.

- La propuesta 6 plantea utilizar R, un lenguaje y entorno de programación enfocado al análisis estadístico, el cual resultaría bastante útil a la hora de describir y predecir el comportamiento de los datos con los que se trabajará. Sin embargo, como parte del proceso de aprendizaje con el que se desarrolla el curso de Proyecto Integrador I, se está aprendiendo a utilizar el lenguaje C# en el entorno de desarrollo Visual Studio, por lo que se optará por desarrollar la aplicación en su totalidad en estas condiciones, y se descarta el uso de R con este objetivo.

Respecto a las predicciones y el análisis a futuro de los datos:

- Realizar predicciones aleatorias sin fundamento estadístico no tiene cabida dentro de los parámetros de calidad entre los que se espera se encuentre la solución desarrollada. Además, de llegarse a utilizar esta solución como una herramienta para identificar áreas en las que se deben concentrar las medidas para reducir la contaminación, un análisis erróneo y producto del azar podría desembocar en la mala utilización de recursos para el Estado, lo que resulta ser una problemática grave. Es por esto, y por diversas razones aquí no listadas, que la alternativa 3 es descartada.

Transición a Diseños Preliminares

Tras el filtro anterior, se procede a explicar modelos preliminares con los que se podría dar solución al problema descrito a partir de las ideas anteriores:

Respecto a la lectura y análisis de datos

- **Analizar series de tiempo mediante distintos métodos estadísticos:**
Existen diferentes métodos estadísticos que pueden realizar análisis de datos. Basta con encontrar uno que permita encontrar la información requerida por la solución de este problema. Y, al ser estos un conjunto grande de simples datos, el cálculo de los parámetros requeridos por estos métodos no resulta tan pesado para la máquina que trabaje esta solución, lo que facilita su uso e implementación.
- **Visualizar los datos a través de mapas en el framework .NET:**
Se puede desarrollar una aplicación de Windows Forms en el framework .NET que utilice los mapas proporcionados por GMaps. Bastará entonces con clasificar la información en la base de datos de acuerdo con el área de cuyos datos se desee conocer, y luego proceder a mostrarlos en forma de mapas de calor, dependiendo del nivel de contaminación encontrado. Se puede, incluso, agregar controles que permitan controlar en mayor medida la información mostrada en el mapa, cómo una barra deslizadora que muestre el cambio de los datos en el tiempo.
- **Realizar una lectura parcial de la base de datos mediante la utilización del formato JSON:**

A través de este formato, es significativamente más fácil manejar el volumen de datos encontrado en la base de datos dada, pues basta con realizar búsquedas sobre esta misma de acuerdo con diferentes parámetros que la misma solución puede definir. Así, un software manejador de bases de datos puede encargarse de las tareas de consultas y ordenamientos, devolviendo únicamente un subconjunto de datos significativamente menor, con el que se pueden realizar tareas más complejas en un tiempo bastante menor.

Respecto a las predicciones y el análisis a futuro de los datos:

- **Utilización de modelos de análisis predictivo de datos:**
Existen diversos modelos de análisis predictivo de datos que pueden realizar análisis a futuro encontrando patrones estadísticos entre la información reunida del pasado. Si se logra sintetizar la información ya reunida de forma tal que el modelo de análisis escogido pueda trabajar con esta, entonces se puede realizar predicciones precisas y concretas para la solución de este problema.
- **Aplicación de Técnicas de Predicción:**
Algo más complejas, pero ya bastante utilizadas en la industria, una técnica de aprendizaje computacional puede realizar predicciones y conjeturas respecto al comportamiento futuro de la contaminación basándose en los datos que ya se hayan reunido, de manera similar a los algoritmos de recomendación en plataformas de streaming como Spotify o Netflix.

Fase 5: Evaluación y Selección de la Mejor Solución:

Definición de criterios de evaluación de soluciones

En esta sección se definieron los criterios relevantes en la toma de decisión de una solución, los criterios son:

Criterio A: Accesibilidad a los datos.

En este criterio se evaluará, en una escala continua del 1 al 5, qué tan fácil es acceder a la información solicitada.

Calificaciones de 5 se darán a soluciones que retornen información en menos de 10 segundos, 4 a soluciones que tarden entre 10 y 20 segundos en retornar datos, 3 a soluciones que tarden entre 20 y 30, 2 a las que tarden entre 30 y 50 y 1 a toda solución que retorne información en tiempos mayores a 50 segundos.

Criterio B: Los datos son actualizados.

Este criterio define la actualidad con la que los datos se muestran. Es decir, que la solución utilizada pueda retornar datos que sean inmediatos, tomados de los últimos valores reunidos de calidad del aire en Colombia. Se evalúa en una escala discreta con valores posibles 1, 3 y 5.

La solución se califica con 1 si la solución utiliza un archivo precargado de memoria, y este es leído sin modificación alguna durante toda ejecución.

La solución se califica con 3 si los datos se actualizan automáticamente cada que se ejecute.

La solución se califica con 5 si permite que los datos sean actualizados en tiempo real sin afectar en gran medida su funcionamiento.

Criterio C: Análisis de los datos.

Este criterio evalúa el tipo de análisis que la solución pueda realizar sobre los datos. Así, una solución que realice análisis profundos y con precisión recibe calificaciones mejores. Se evalúa en una escala discreta del 1 al 5.

- [1] Representa que la solución analiza menos del 10% de los datos dados.
- [2] Representa que la solución analiza entre el 10 y el 30% de los datos.
- [3] Representa que la solución analiza entre el 30 y el 50% de los datos.
- [4] Representa que la solución analiza entre el 50 y el 90% de los datos.
- [5] Representa que la solución analiza más del 90% de los datos, o su completitud.

Criterio D: Facilidad para mostrar un análisis de forma gráfica.

Cada solución implica el uso de una herramienta de programación que provee las funcionalidades para graficar los datos. Este criterio evalúa, en una escala continua del 1 al 5, cada solución según la cantidad de funcionalidades que incluye.

- [1] La solución no provee ninguna herramienta necesaria para visualizar la información.
- [2] La solución provee algunas, pero menos de la mitad, de las funcionalidades necesarias para visualizar la información.
- [3] La solución provee más de la mitad, pero no todas, de las funcionalidades necesarias para visualizar la información.
- [4] La solución provee todas las funcionalidades necesarias.
- [5] La solución provee todas las funcionalidades necesarias y otras no necesarias pero complementarias y adecuadas.

Resumen de las valoraciones de cada solución:

Nombre de la solución / Criterio	Criterio A (Lectura)	Criterio B (Actualización)	Criterio C (Análisis)	Criterio D (Visualización)	Total
Mapas en .NET	0	0	0	4	4
Lectura JSON	5	5	0	0	10

Nombre de la solución / Criterio	Criterio C (Análisis)	Total
ARIMA	5	5
Modelos de análisis predictivo	4	4
Técnicas de predicción	4	4

Dados los resultados de la tabla anterior, se eligen las opciones de lectura y acceso de datos por medio del formato JSON y el uso del modelo ARIMA para el análisis y, para la visualización de la información, se elige la opción de mapas en el framework .NET`

Referencias

ANI. (s.f.). *Autoridad Ambiental | Portal ANI*. Obtenido de Portal ANI:

<https://www.ani.gov.co/glosario/autoridad-ambiental>

Dirección de Monitoreo Atmosférico. (s.f.). *Aire*. Obtenido de Google Play:

https://play.google.com/store/apps/details?id=mx.gob.df.aire&hl=es_CO

IDEAM. (s.f.). *ACERCA DE LA ENTIDAD*. Obtenido de IDEAM:

<http://www.ideam.gov.co/web/entidad/acerca-entidad>

IQAir AG. (s.f.). *Calidad del Aire | AirVisual*. Obtenido de Google Play:

https://play.google.com/store/apps/details?id=com.airvisual&hl=es_CO

Plume Labs. (s.f.). *Plume Labs: Contaminación del Aire*. Obtenido de Google Play:

https://play.google.com/store/apps/details?id=com.plumelabs.air&hl=es_CO

Secretaria Distrital de Ambiente Bogotá. (s.f.). *Aire Bogotá*. Obtenido de Google Play:

https://play.google.com/store/apps/details?id=co.gx7.innovation.ambientebogota&hl=es_CO

SIAC. (s.f.). *Monitoreo aire - IDEAM*. Obtenido de SIAC - Sistema de Información

Ambiental de Colombia: <http://www.siac.gov.co/monitoreoaire>

SISAIRE. (s.f.). *¿Qué es el SISAIRE?* Obtenido de :::::Subsistema de Información sobre

Calidad de Aire: <http://aircolombia.sisaire.gov.co/portal/MainPage.col>