

Luis Gómez
Elias Estupiñan
Mateo Ramírez
Manuel Balanta

Método de la Ingeniería.

1. Identificación del Problema.

1.1. *Descripción del Contexto del Problema.*

Colombia es uno de los países con mayor biodiversidad del mundo y cuenta en su territorio con un gran número de ecosistemas que constituyen nuestro patrimonio natural y nuestros diversos recursos: el agua, el suelo, la flora, la fauna y el aire. Con el crecimiento de las ciudades y de la población colombiana, se ha acelerado el uso y degradación de estos recursos incorporando al ambiente sustancias extrañas cuya permanencia ha afectado la salud y bienestar de la sociedad y el de los ecosistemas.

Teniendo en cuenta lo anterior, el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) ha construido una base de datos con reportes desde comienzos del año 2011 hasta finales del año 2017, con ayuda de los Sistemas de Vigilancia de Calidad de Aire instalados por Corporaciones Autónomas Regionales y las Autoridades Ambientales de los Grandes Centros

Urbanos, con el fin de tener una base sólida de donde sacar información para futuros proyectos ambientales y sanitarios.

Por lo tanto, para una mejor administración y análisis de dichos datos, es fundamental la información que en ellos se muestra, a través de las variables meteorológicas estudiadas, en la dispersión, concentración y transporte de los contaminantes para conocer la población expuesta y las consecuencias de la contaminación.

1.2. Definición del Problema.

Debido a la alta demanda en transformación territorial, social y ambiental del país, el IDEAM desea implementar un software que le ayude a predecir cambios meteorológicos y conocer la población afectada y las consecuencias de la contaminación, teniendo en cuenta la información recopilada por el Sistema de Vigilancia de Calidad del Aire.

1.3. Especificación de Requerimientos.

1.3.1. Requerimientos Funcionales.

- La aplicación debe estar en la capacidad de leer de manera eficiente los datos. Para que esto suceda, los datos no tienen que haberse leído anteriormente.
- La aplicación debe estar en capacidad de realizar predicciones en cambios meteorológicos, teniendo en cuenta variables como la temperatura estudiada, el material particulado, la dirección y velocidad del viento y la radiación solar global
- La aplicación debe estar en la capacidad de mostrar el mapa de calor, con base a la calidad del aire, que hay en Colombia. Para esto, se necesita haber cargado los datos con anterioridad y, además, haber realizado las predicciones.
- La aplicación debe estar en capacidad de filtrar la información resultante de las predicciones. Para esto, se debe de haber hecho con anterioridad el análisis y la lectura de la información.
- La aplicación debe estar en capacidad de mostrar la información de cada lugar estudiado en las predicciones meteorológicas. Para esto, se debió haber cargado la información y haberse realizado las predicciones con anterioridad. Además, se debió haber “pintado” en pantalla el mapa sobre el cual se mostrará la información.

2. Recopilación de la Información.

2.1. Marco Teórico

Con el objetivo de obtener una claridad absoluta de los conceptos involucrados, es necesaria cierta investigación sobre las definiciones de los términos relevantes para el contexto del problema.

2.1.1. Predicción

Es una expresión que anticipa aquello que, supuestamente, va a suceder, pudiéndose predecir a partir de revelaciones de algún tipo, hipótesis o indicios. Teniendo en cuenta la naturaleza de los datos que se usarán para la implementación del problema descrito con anterioridad, se podría vincular más este término al campo de las ciencias dado el origen de la base de datos y la organización responsable de haberlos recogidos. Por lo tanto y, teniendo en cuenta lo anterior, predicción en el ámbito de la ciencia se definiría como “un anticipo de lo que ocurrirá de acuerdo con el análisis de las condiciones existentes”, donde no siempre se cumplen dada la existencia de variables desconocidas u otras cuyos valores no se pueden anticipar, a día de hoy, con precisión.

2.1.2. Unidad

“Referencia convencional que se usa para medir la magnitud física de un determinado objeto, sustancia o fenómeno”, fijándose de acuerdo con estándares internacionales, donde permite el cálculo o la medición de variables como la longitud, la masa, el volumen, la temperatura, el tiempo etc. Actualmente, el

sistema más extendido y por lo tanto usado en el mundo es el llamado Sistema Internacional de Unidades (SI), basado en el métrico decimal.

2.1.3. Meteorología

“Disciplina científica y técnica que se encarga de estudiar y predecir los diversos fenómenos que se producen en la atmósfera, para comprender por un lado su funcionamiento, composición, estructura y evolución, y por otro lado para tener importantes predicciones diarias”. Todo lo anterior, con el fin de generar “reportes” que ayuden en las diferentes actividades humanas tales como la agricultura, la aeronáutica, la navegación, actividades militares y actividades relacionadas con la administración de recursos ambientales. Todo lo anterior, gracias al estudio de variables meteorológicas en particular tales como la temperatura, humedad, radiación solar, entre otras.

2.1.4. Bases de Datos

“Colección de datos almacenados de cierta forma y con una estructura específica”, la cual, gracias a su estructura, posee operaciones optimizadas tales como guardar, eliminar, consultar y gestionar la información que en ella se encuentra. Las bases de datos utilizadas hoy en día están conformadas por modelos o tablas, donde cada tabla tiene N atributos, que pueden ser de diferentes datos, donde se guardan como registros.

3. Búsqueda de Soluciones Creativas

Dadas las funcionalidades del programa, se decidió dividir en tres partes la búsqueda de las soluciones: visualización de los datos en pantalla, lectura de la base de datos y la realización de las predicciones meteorológicas.

Adicionalmente, se utilizó el método de lluvias de ideas para realizar la búsqueda de soluciones.

3.1. Visualización de los datos en pantalla

- Idea 1: Pintar el mapa de Colombia y dibujar las ciudades que se utilizaron en el estudio y las predicciones meteorológicas.
- Idea 2: Implementar un grafo donde los nodos sean los lugares estudiados y las aristas, representaría la contaminación que hay en dicho lugar.
- Idea 3: Implementar un mapa de calor con tres colores, donde el color rojo, representa las ciudades con los índices más críticos en la calidad del aire, amarillo, para un índice no tan crítico y finalmente, el color verde que represente a las ciudades con la mejor calidad de aire.

3.2. Lectura de la base de datos

- Idea 1: Leer el archivo según criterios de búsqueda previamente establecidos por el usuario final.
- Idea 2: Establecer una conexión a través de la API, (usando SODA.NET) del sitio de origen de los datos, para obtenerla de manera fragmentada y categorizada; sin necesidad de tener todo el archivo que contiene los datos.
- Idea 3: Leer el archivo en formato csv, como archivo de texto plano.

3.3. Predicciones Meteorológicas

- Idea 1: Obtener estimaciones o pronósticos de valores futuros de una serie temporal a partir de la información histórica contenida en la serie observada, esto con técnicas de predicción.
- Idea 2: Utilizar el modelo autorregresivo integrado de media móvil.
- Idea 3: Usar el modelo matemático de regresión lineal.
- Idea 4: Árboles de decisión regresión.
- Idea 5: Bosques aleatorios de regresión.

4. Transición de las Ideas a los Diseños Preliminares

4.1 Visualización de los Datos en la Pantalla

Se decide prescindir de la idea 2, dado que modelar un grafo para este problema descrito, no sería lo adecuado, pues carecería de sentido representar vértices como ciudades y aristas como índices de contaminación. Por esto, sería imposible realizar los algoritmos básicos de grafos, pues no habría una forma de establecer una relación.

Así, se procede a explicar con mayor profundidad las ideas 1 y 3:

- Idea 1: La idea consiste en tener un mapa de Colombia en una ventana de un formulario de la aplicación, en dicho mapa aparecerán todas las

ciudades que aparecen en el dataset y se mostrará el dato de predicción de la ciudad que el usuario escoja.

- Idea 3: Esta idea consiste en tener un Slider que permita desplazarse desde el año 2011 hasta el año que se calculó la predicción. A medida que el usuario pase por los años el programa tendrá implementado un mapa de calor con tres colores: rojo, amarillo y verde, en el cual cada uno representa los índices de la calidad del aire, donde rojo es mayor concentración, amarillo mediana concentración y verde menor concentración.

4.2 Lectura de la Base de Datos

Sin lugar a duda, la idea 3 se descarta, ya que leer el dataset como un archivo de texto con la cantidad de datos que cuenta este, sería inviable, porque, consumiría mucho espacio en memoria y haría que el programa fuese muy lento.

A continuación las ideas que pasaron el filtro explicadas más a detalle.

- Idea 1: Leer el JSON del API con los datos que el usuario final seleccione.
- Idea 2: Con SODA.NET y el método de búsqueda GET, del JSON traer toda la información del dataset.

4.3 Predicciones Meteorológicas

Puesto que para este caso los algoritmos a implementar para estimar una predicción se buscan que sean fáciles de interpretar y que a la hora de aplicarlos tengan un sentido dentro del contexto del problema y soporten la complejidad de los datos. Se decidió omitir las ideas 1, 4 y 5, siendo las dos primeras (idea 1 y 4) sencillas, pero no soportaría la complejidad de los datos y no competen al contexto del problema; finalmente, la idea 5 es difícil de interpretar. Se procede a explicar las ideas 2 y 3.

- Idea 2: Ya que el conjunto de datos lo permite se puede el modelo ARIMA que consiste en utilizar variaciones y regresiones de datos estadísticos con el fin de encontrar patrones para una predicción hacia el futuro. Se trata de un modelo dinámico de series temporales, es decir, las estimaciones futuras vienen explicadas por los datos del pasado y no por variables independientes.
- Idea 3: Este modelo consisten en aplicar el algoritmo de regresión lineal que consiste en entrenar la mejor línea a través de todos los puntos de datos. Este algoritmo puede ser Regresión Lineal Simple, en donde se realiza la predicción con una sola variable, Regresión Lineal Múltiple, en donde se crea un modelo para la relación entre múltiples variables de entradas independientes.

5. Evaluación o Selección de la Mejor Solución

5.1. Visualización de los datos

Rúbrica:

<i>Criterio</i>	<i>Calificación</i>	<i>Valor</i>
Relación de los datos con lo que se observa en pantalla	[1] No hay relación. [2] La relación de los datos con lo que se observa en pantalla es escueta y no permite entender la solución del problema. [3] Hay una relación evidente con los datos y lo que se observa en pantalla que permite obtener una mejor solución.	50%
Interacción usuario-programa	[1] La visualización de los datos es estática y no le permite al usuario interactuar con ellos. [2] El usuario puede interactuar con al menos el 50% de los datos.	30%
Reto para el equipo de desarrolladores	[1] La solución a implementar no requiere de nuevos conocimientos. [2] El equipo de trabajo implementa medianos nuevos conocimientos para poder llegar a una solución. [3] El equipo de trabajo tiene que aplicar nuevos conocimientos que representan unos retos para poder llegar a la solución.	20%

Evaluación:

	Relación de los datos con lo que se observa en pantalla	Interacción usuario-programa	Reto para el equipo de desarrolladores	Total
Mapa de Colombia con las ciudades que aparecen en el dataset y se muestra la predicción de la ciudad que el usuario escoja.	2	1	1	1,5

Slider que permite desplazarse desde el año 2011 hasta el año que se calculó la predicción. A medida que el usuario pase por los años el programa tendrá implementado un mapa de calor con tres colores: rojo, amarillo y verde, en el cual cada uno representa los índices de la calidad del aire, donde rojo es mayor concentración, amarillo mediana concentración y verde menor concentración.	3	2	3	2,7
--	---	---	---	-----

5.2. Lectura de la Base de Datos

Rúbrica:

Criterio	Calificación	Valor
Velocidad a la hora de cargar los datos	[1] El método utilizado para cargar los datos demora más de 5 minutos y no requiere de una conexión a internet. [2] El método utilizado para obtener los datos demora menos de 3 minutos. [3] El método utilizado carga los datos al instante, siempre y cuando se disponga de conexión a internet.	60%
Alojamiento de los datos	[1] Los datos se encuentran en la aplicación. [2] Los datos se obtienen del API apenas inicia la aplicación. [3]	30%

	Los datos se obtienen del API cuando sea necesario.	
--	---	--

Evaluación

	Velocidad a la hora de cargar los datos	Alojamiento de los datos	Total
Leer el JSON del API con los datos que el usuario final seleccione.	2	2	2
Con SODA.NET y el método de búsqueda get, del JSON traer toda la información del dataset.	3	3	3

5.3. Predicciones meteorológicas

Rúbrica:

Criterio	Clasificación	Valor
Evaluación sobre una serie de tiempo	[1] El método utilizado no trabaja sobre una serie de tiempo. [2] El método utilizado trabaja sobre una serie de tiempo	50%
Conocimiento previo del equipo de	[1] El equipo de desarrolladores no tiene	50%

desarrolladores	un conocimiento previo del modelo. [2] El equipo de desarrolladores tiene un conocimiento previo del modelo.	
-----------------	--	--

Evaluación:

	Evaluación sobre una serie de tiempo	Conocimiento previo del equipo de desarrolladores	Total
Modelo ARIMA	2	2	2
Regresión lineal	1	1	1