

Método de ingeniería

Paso 1. Identificación del problema

- *Descripción del contexto problemático*

A medida que las industrias se expanden se logran avances tecnológicos increíbles los cuales aportan de manera significativa al ser humano desde tareas cotidianas hasta mandar al ser humano al espacio, pero a medida que la tecnología avanza también lo hacen los niveles de contaminación los cuales se han comenzado a convertir en una problemática más alarmante para el ser humano.

La contaminación es medidas en diferentes tipos pero los más comunes son las partículas PM10(partículas gruesas) y las partículas PM2,5(partículas finas), Estas últimas son más peligrosas que cualquiera debido a que son de tamaño microscópico y se filtran y adhieren fácilmente en los seres vivos y terrenos(afectando sus nutrientes).

Daño en la salud.

Son capaces de filtrarse en el cuerpo humano y afectar tanto los pulmones como el corazón, múltiples estudios científicos vincularon la exposición a la contaminación por partículas a una variedad de problemas, que incluye:

- muerte prematura en personas con enfermedades cardíacas o pulmonares.
- infartos de miocardio no mortales.
- latidos irregulares.
- asma agravada.
- función pulmonar reducida.
- síntomas respiratorios aumentados, como irritación en las vías respiratorias, tos o dificultad para respirar.
 - La exposición a la contaminación por partículas tiende a afectar en su mayoría a personas con enfermedades cardíacas o pulmonares, niños y adultos mayores.

Daño ambiental

El viento puede transportar las partículas a través de largas distancias y luego, estas pueden instalarse en el suelo o el agua. Según la composición química, los efectos de esta sedimentación pueden provocar:

- que los lagos y arroyos se vuelven ácidos.
- cambio en el balance nutricional de las aguas costeras y de las grandes cuencas fluviales
- reducción de los nutrientes del suelo

- daño en los bosques sensibles y cultivos agrícolas
- efectos perjudiciales sobre la diversidad de ecosistemas
- contribución a los [efectos de la lluvia ácida](#).

En Colombia, las autoridades ambientales durante el periodo de 2011 a 2016 implementaron y renovaron sus sistemas de vigilancia con tecnología automática para realizar la evaluación de los contaminantes atmosféricos, en detrimento de los equipos manuales. Provocando que se mejorará la confiabilidad, temporalidad y oportunidad de las mediciones, permitiendo a las autoridad ambiental adoptar medidas de control precisas.

El Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM fue el encargado de recolectar datos relacionados a la contaminación del IRE y almacenarlos en una base de datos a partir de los reportes de realizados por cada sistema de vigilancia, en el subsistema sobre calidad de aire- SISAIRE. la validación final se dio a través de un proceso conjunto de y constructivo con cada autoridad ambiental, lo cual permite la solución de consistencias en los datos.

A partir de lo anterior, se plantea una problemática sobre cómo intervenir por medio de medidas preventivas y soluciones reales a los territorios del país más afectados por la contaminación del aire para hacer un llamado tanto a las entidades públicas locales como a los ciudadanos que transitan y habitan esas zonas. Sin embargo, es necesario que todos los datos recolectados se reúnan en conjunto para un acceso más fácil y directo para ayudar a la toma de decisiones.

- Identificación de necesidades
 - Identificar las zonas más afectadas y hacer predicciones en cada unas de las estaciones.
 - Mostrar un mapa con toda la información necesaria de un departamento o ciudad específica.
 - Ilustrar mediante el mapa cuáles pueden ser las zonas más afectadas por la contaminación.
- Definición del problema

Los datos recolectados sobre la cantidad de partículas PM 2.5 que se encuentran en el aire de las zonas estudiadas dentro del territorio colombiano se deben organizar de manera tal que se pueda ilustrar gráficamente cuáles son las zonas con mayor presencia de estas partículas, y se realice una predicción con técnicas de minería de datos.

Paso 2. Recopilación de la información necesaria

La contaminación atmosférica es el principal riesgo ambiental para la salud en las Américas (WHO, 2016a). La Organización Mundial de la Salud estimó que una de cada nueve muertes en todo el mundo es el resultado de condiciones relacionadas con la contaminación atmosférica (WHO, GBoD 2016). Los contaminantes atmosféricos más relevantes para la salud son material particulado (PM) con un diámetro de 10 micras o menos, que pueden penetrar profundamente en los pulmones e inducir la reacción de la superficie y las células de defensa. La mayoría de estos contaminantes son el producto de la quema de combustibles fósiles, pero su composición puede variar según sus fuentes. Las directrices de la OMS sobre la calidad del aire recomiendan una exposición máxima de 20 µg/m³ para las PM₁₀ y **una exposición máxima de 10 µg/m³ para las PM_{2.5}** (WHO, 2005), basado en las evidencias de los efectos sobre la salud de la exposición a la contaminación del aire ambiente.

Las partículas de contaminación se pueden clasificar en 2 principales PM₁₀ y PM_{2.5}, las partículas PM₁₀ son denominadas partículas “gruesas” y las PM_{2.5} son partículas “finas” las cuales suponen un mayor riesgo para la salud en comparación a las primeras ya que estas alcanzan un nivel microscópico que pueden estar y entrar fácilmente a cualquier lugar.

Daño en la salud.

Son capaces de filtrarse en el cuerpo humano y afectar tanto los pulmones como el corazón, múltiples estudios científicos vincularon la exposición a la contaminación por partículas a una variedad de problemas, que incluye:

- muerte prematura en personas con enfermedades cardíacas o pulmonares.
- infartos de miocardio no mortales.
- latidos irregulares.
- asma agravada.
- función pulmonar reducida.
- síntomas respiratorios aumentados, como irritación en las vías respiratorias, tos o dificultad para respirar.
 - La exposición a la contaminación por partículas tiende a afectar en su mayoría a personas con enfermedades cardíacas o pulmonares, niños y adultos mayores.

Daño ambiental

El viento puede transportar las partículas a través de largas distancias y luego, estas pueden instalarse en el suelo o el agua. Según la composición química, los efectos de esta sedimentación pueden provocar:

- que los lagos y arroyos se vuelven ácidos.

- cambio en el balance nutricional de las aguas costeras y de las grandes cuencas fluviales
- reducción de los nutrientes del suelo
- daño en los bosques sensibles y cultivos agrícolas
- efectos perjudiciales sobre la diversidad de ecosistemas
- contribución a los [efectos de la lluvia ácida](#).

¿Que producen las partículas PM?

Las partículas gruesa pueden ser generadas de manera natural por fenómenos naturales tales como la erupción de volcanes, incendios forestales, nubes de polvo, etc y por actividades humanas ejemplos de estas son las actividades industriales, trituración de canteras, labores agrícolas o de construcción.

Índice de medición ICA:

El Índice de calidad del aire (ICA) permite comparar los niveles de contaminación del aire de las estaciones de monitoreo que conforman un Sistema de Vigilancia de Calidad del Aire (Unidades espaciales de referencia), en un tiempo t , que corresponde al período de exposición previsto en la norma para cada uno de los contaminantes que se está midiendo.

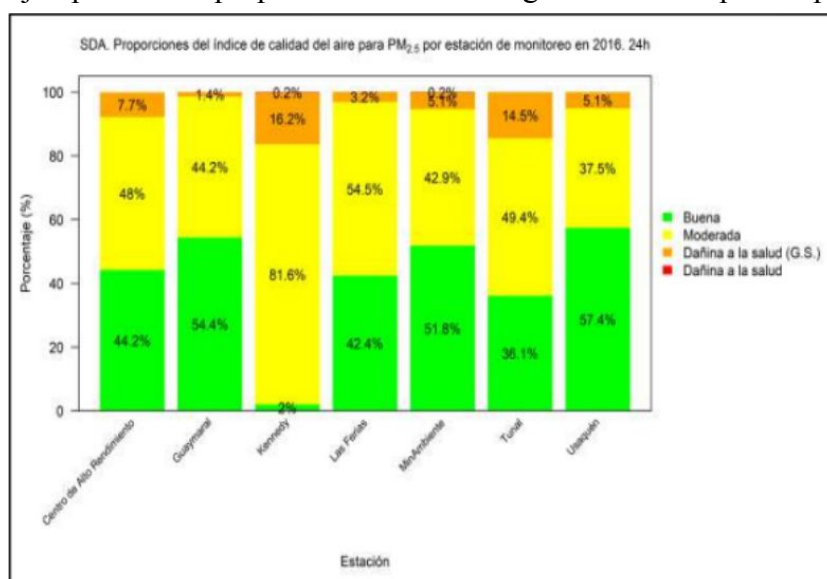
Teniendo en cuenta los contaminantes que son monitoreados en el país, las características de los combustibles que se distribuyen y los equipos que actualmente se encuentran en las SVCA, el Índice se calcula de manera independiente para cada uno de los seis principales contaminantes, los cuales son PM₁₀, PM_{2,5}, SO₂, NO₂, O₃ y CO.

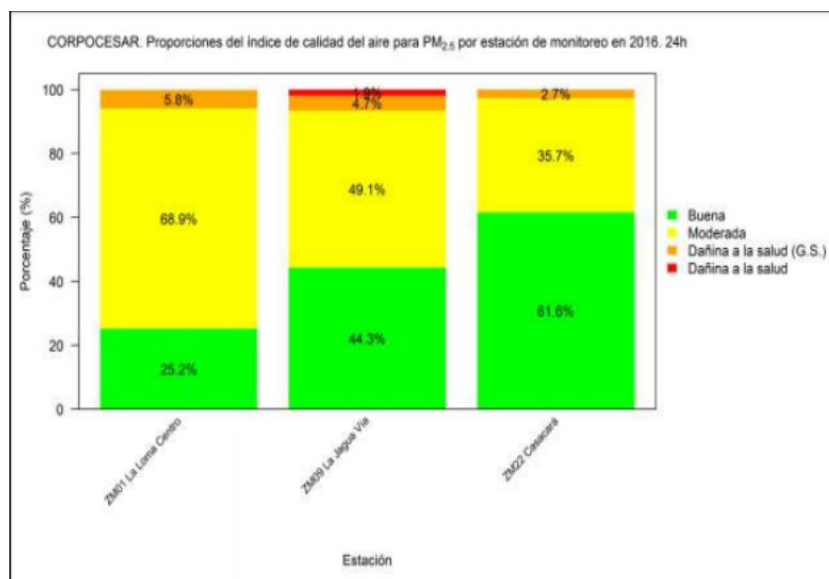
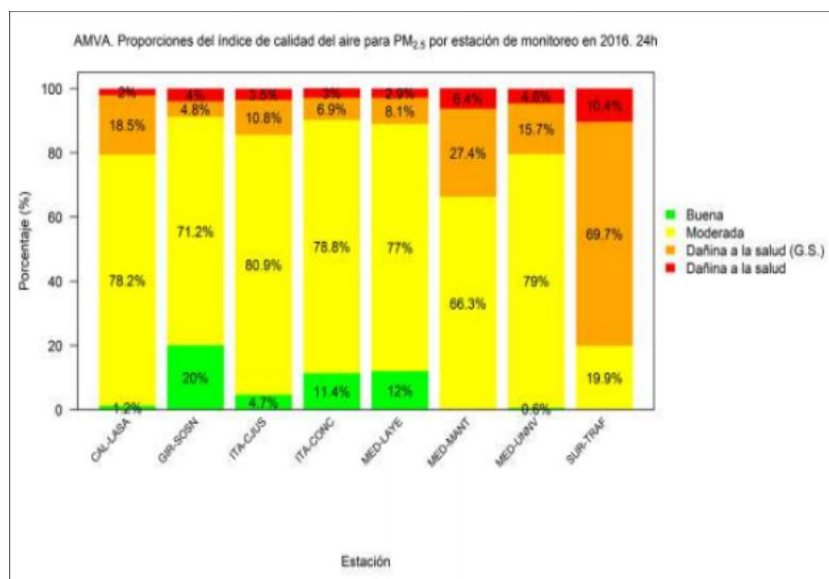
Los valores del ICA se ubican en una escala adimensional de 0 a 500, que han sido agrupados en 6 rangos que guardan estrecha relación con la amenaza que, a la salud humana, representan dichos niveles de contaminación del aire. (Ideam, 2017)

Rango (ICA)	Colores	Clasificación	$PM_{2,5} 24h$ $\frac{\mu g}{m^3}$
$0 \leq ICA \leq 50$	Verde	Buena	0,0 - 15,4
$51 \leq ICA \leq 100$	Amarillo	Moderada	15,5 - 40,4
$101 \leq ICA \leq 150$	Anaranjado	Dañina para la salud en grupos sensibles	40,5 - 65,4
$151 \leq ICA \leq 200$	Rojo	Dañina para la salud	65,5 - 150,4

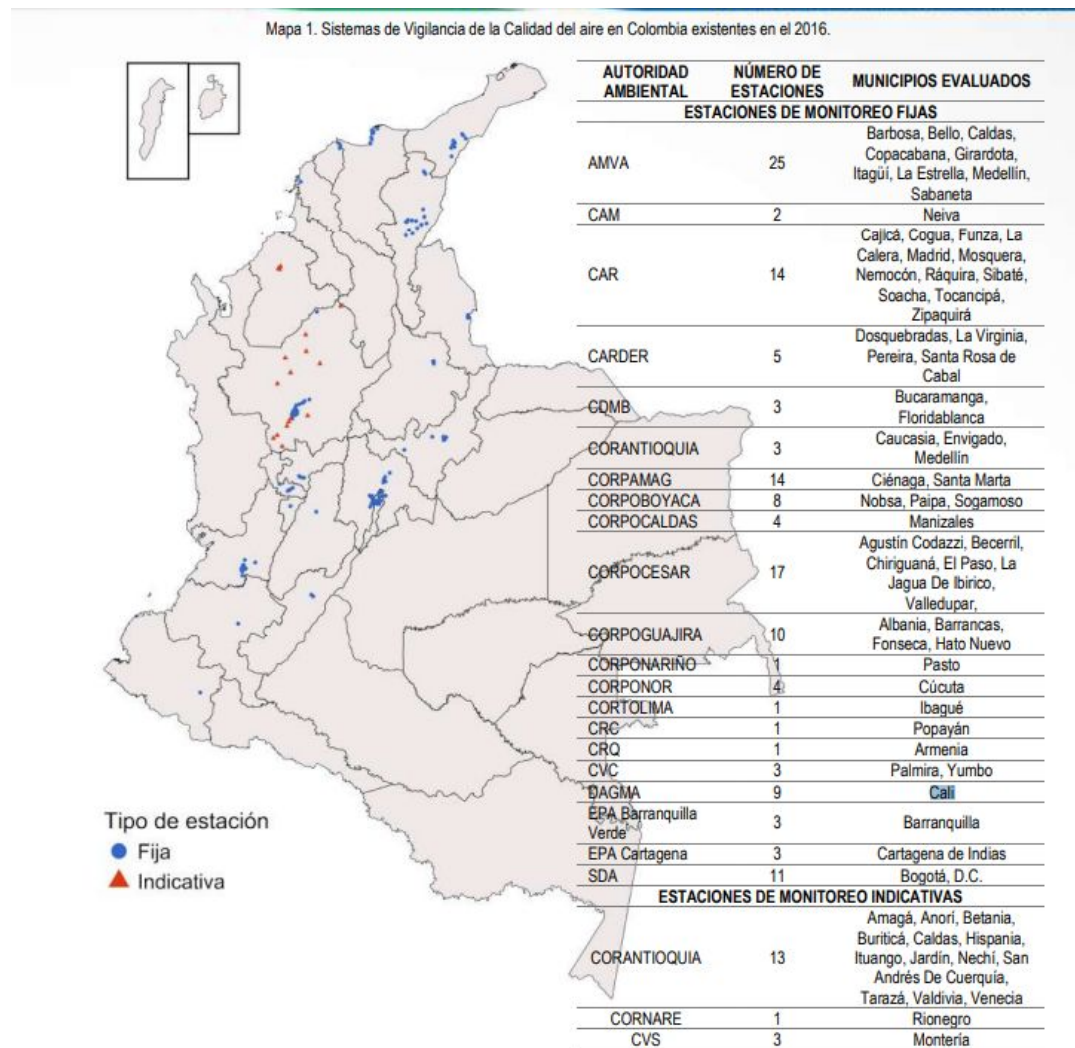
$201 \leq ICA \leq 300$	Morado	Muy dañina para la salud	150,5 - 250,4
$301 \leq ICA \leq 400$	Marrón	Peligrosa	250,5 - 350,4
$301 \leq ICA \leq 500$	Marrón	Peligrosa	350,5 - 500,4

Ejemplos de las proporciones de las categorías del ICA para el pm_{2,5} en el ejemplo hecho





Mapa de las zonas donde se realizó el estudio con sus respectivas autoridades ambientales.



Herramientas utilizadas:

C#: es un lenguaje de programación multiparadigma desarrollado y estandarizado por Microsoft como parte de su plataforma .NET, C# es un de los lenguajes de programación diseñados para la infraestructura de lenguaje común.

Su sintaxis básica deriva de C y C++ y utiliza el modelos de objetos de plataforma .NET parecido al de java.

Microsoft Visual Studio: es un entorno de desarrollo integrado(IDE) para las plataformas windows, linux y macOS. Es compatible con múltiples lenguajes de programación y permite a los desarrolladores crear sitios y aplicaciones web, así como servicios web en cualquier entorno compatible con la plataforma .NET .

GMaps:

Esta es una extensión del paquete de aplicaciones nuget de la IDE visual studio para el lenguaje de programación C#, esta extensión permite al programador desarrollar aplicaciones relacionados con ubicaciones puntuales cualquier parte del mundo mediante mapas a tiempo real.

Geocodificación: Es el proceso mediante el cual se convierten direcciones(como las direcciones de las calles) en coordenadas geográficas(latitud y longitud), en las cuales se pueden usar marcadores en el mapa o posiciones en el mapa.

Geocodificación inversa: Es el proceso de convertir coordenadas geográficas en direcciones legibles.

Proceso de descubrimiento de conocimiento en bases de datos: Por sus siglas es KDD(knowledge discovery in databases) Es el proceso para descubrir el conocimientos es interactivo e iterativo, y consiste principalmente en 9 pasos.

- **Paso 1: desarrollando el entendimiento del dominio de la aplicación.**

Se busca entender y definir las metas del usuario final y el ambiente en el cual el proceso se va a llevar a cabo. Teniendo un entendimiento de las metas del KDD, a partir de aquí el proceso de búsqueda de conocimiento empieza definiéndolo en los siguientes tres pasos.

- **Paso 2: seleccionando y creando un conjunto de datos en los cuales el descubrimiento va a ser realizado.**

Cuando las metas se han establecido se determina que tipo de información es necesaria para hallar conocimiento, buscando qué información está disponible, cual se necesita adicionar y después integrar en el conjunto de datos utilizados para el descubrimiento de conocimiento incluyendo atributos que pueden ser considerados para el proceso. Si se dejan pasar atributos que son importantes para la investigación es probable que esta falle. La compensación que representa el aspecto interactivo y iterativo en el proceso empieza de hallar el mejor y disponible conjunto de datos y luego expandirlo y observar los efectos en el proceso de búsqueda de conocimiento.

- **Paso 3:Preprocesamiento y limpieza.**

En este caso, la fiabilidad es mejorando mediante la limpieza de datos, en esta se manejan valor olvidadas y removidos los casos atípicos(los procesos para esta limpiezas se definen en el tiempo consumido), esto puede llevar a usar métodos estadísticos complejos o mineo de datos. Como ejemplo tenemos que si un atributo no tiene la suficiente confiabilidad o la suficiente información en el momento que se haga la predicción estos datos se tomarán en cuenta y es probable que la predicción sea errónea.

- **Paso 4: transformación de datos.**

En esta paso la generación para mejor información para el minado de datos ya ha sido preparada y desarrollada, hay diferentes métodos como lo son la reducción de dimensión(tales como selección y extracción de características y muestreo de registros) y transformación de atributos(como las volver discretos datos numéricos y transformaciones funcionales), este paso puede ser crucial para el éxito del proyecto el cual es muy específico del proyecto. Aún sin

utilizar el mejor método de transformación de datos los resultados de estas nos pueden insinuar la transformación necesaria. Así, el proceso KDD reflexiona sobre sí mismo y conduce a una comprensión de la transformación necesario.

Después de realizar estos pasos, los siguientes 4 pasos son relacionados con el minado de datos, donde nos enfocamos en que algoritmo emplea cada proyecto.

- **Paso 5: Escogiendo la tarea apropiada para el minado de datos.**

Los diferentes tipos son clasificación, regresión, o cápsula miento, cual tomar depende de las metas del proceso para encontrar el conocimiento. Hay 2 meta grandes en el minado de datos, predicción y descripción. La predicción a menudo se denomina minería de datos supervisada, mientras que la minería de datos descriptiva incluye los aspectos de supervisión y supervisión sin supervisión. La mayoría de técnicas de minado son basados en aprendizaje inductivo, donde un modelo es construido explícito o implícito generalizando suficientes números de entrenamiento.

- **Paso 6: escogiendo el algoritmo de minado de datos.**

Teniendo una estrategia, ahora decidimos las técnicas, en este escenario también se seleccionan diferentes métodos para hallar patrones. Por ejemplo considerando precisión vs entendimiento es mejor utilizar redes neuronales, mientras que para la última es mejor árboles de decisiones. El meta aprendizaje es aquel que causa que los algoritmos de minado sean exitosos o no en un problema particular, así este enfoque intenta hacer ver en qué condiciones se debe escoger el algoritmo de minado es el más apropiado.

- **Paso 7: Utilizando el algoritmo de minado.**

Después de implementado el algoritmo se necesitará que se ejecute varias veces para tener un resultado satisfactorio. por ejemplo, ajustando los parámetros de control del algoritmo, como el número mínimo de instancias en una sola hoja de un árbol de decisión.

- **Paso 8: evaluación.**

En este paso comparamos los resultados con las metas planteadas en el primer paso, además de hacer un seguimiento de los pasos para ver cómo afectaron en el resultado del minado de datos, en este paso se centra en la comprensibilidad y uso del modelo inductivo. En este último paso el conocimiento es documentado para ser usado.

- **Paso 9: usando el conocimiento hallado.**

En este paso incorporamos la información hallada en el sistema o tomamos una decisión mas haya, el conocimiento se vuelve activo en el sentido de hacer cambios en el sistema y aquí es cuando se comprueba que tan efectivo fue el proceso de descubrimiento de conocimiento.

Minería de datos: Conocida de inglés como “data mining” es un campo de la estadística y las ciencias de la computación referida al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos (Es parte de la etapa de análisis en el proceso de descubrimiento de conocimiento en bases de datos).

Paso 3. Búsqueda de Soluciones Creativas (Relación forzada):

Se piensa usar el método de la relación forzada para encontrar las diferentes posibles soluciones, utilizando las siguientes palabras escogidas:

- Lámpara.
- Camaleón.
- Arena.
- Mancha.

1. **Lámpara/Camaleón:** Representar con pequeñas luces de colores los departamentos en los cuales se presenta mayor cantidad de contaminación. Utilizando una nueva instancia de GMarkerGoogle (que es una especialización de GMapMarker) y le proporcionamos una ubicación (un PointLatLng) y un tipo de marcador.
2. **Arena/Camaleón:** Colocar en el mapa concentraciones de puntos coloridos (como granos de arena) equivalentes a la concentración de contaminación.
El cual se podría implementar utilizando el framework Asp.net MVC que es un patrón de diseño utilizado para desacoplar la interfaz de usuario (vista), los datos (modelo) y la lógica de aplicación (controlador) junto al ViewBag se utiliza para pasar datos del controller a la vista. De esta manera, se utilizarían las herramientas para ilustrar en el mapa los marcadores de distintos colores en donde cada color representa un nivel específico de contaminación.
3. **Camaleón/Mancha:** Marcar con colores determinados toda el área en la que se presenta una misma cantidad de contaminación, dejando el mapa en su totalidad de color. Se puede crear un formato JSON en el cual se especificará las modificaciones deseadas para el mapa, utilizando también el API de JavaScript de Maps que permite crear cuatro tipos de mapas básicos (hoja de ruta, satélite, híbrido y terreno) que puede modificarse utilizando capas y estilos, controles y eventos, y varios servicios y bibliotecas.

Paso 4. Transición de las Ideas a los Diseños Preliminares

Se descarta la alternativa 1 porque es muy sencilla, se necesitan herramientas más completas para poder hacer las modificaciones deseadas en el programa y así cumplir con los requisitos planteados a partir de la definición del problema.

Alternativa 2. Creación del ViewBag con Asp.NET MVC.

Al utilizar este concepto, las personas pueden obtener fácilmente la información sobre tipos de lugares como los departamentos o ciudades en donde quieran conocer los datos de contaminación de las partículas. Con el color del marcador especificado, los lugares pueden

ser fácilmente conocidos por el usuario. De esta manera, se asignan todos los detalles del marcador de lugares, coordenadas y tipos.



Alternativa 3. Creación del archivo JSON con el API de JavaScript de Google Maps.

Es posible utilizar los mapas de calor, los cuales facilitan que los espectadores entiendan la distribución de los terremotos, según informa USGS. En lugar de colocar un marcador en cada epicentro, los mapas de calor usan color y forma para representar la distribución de los datos.

Paso 5. Evaluación y selección de la mejor solución.

Criterios

- *Criterio A.* Facilidad de implementación, se refiere al nivel de dificultad que podría presentarse en el momento de la implementación del código para la solución. Teniendo en cuenta factores que depende de los desarrolladores como el conocimiento previo de las herramientas.
 - [3] Alta (no hay desconocimiento de las herramientas)
 - [2] Media (se conoce un poco sobre las herramientas con las cuales se trabaja, sin embargo se debe buscar más literatura)
 - [1] Baja (se desconoce totalmente cómo funcionan las herramientas que se planean utilizar para el planteamiento de la solución)
- *Criterio B.* Completitud, se evalúa este criterio con el fin de definir si las herramientas utilizadas dentro de cada solución específica ofrecerá cumplir con todas o la mayor cantidad posible de necesidades del problema satisfechas.
 - [3] Todas las necesidades.
 - [2] Alrededor del 90% de las necesidades satisfechas.
 - [1] Menos del 70% de las necesidades satisfechas

	Criterio A	Criterio B	Total
Alternativa 2. Creación del ViewBag con Asp.NET MVC.	1	3	4
Alternativa 3. Creación del archivo JSON con el API de JavaScript de Google Maps.	2	3	5

Selección

De acuerdo con la evaluación anterior, se decide que la mejor opción es la alternativa de archivo JSON pues facilita la implementación de la solución y ofrece una solución completa.

Referencias

Ideam (2017). Índice de calidad del aire (ICA). Tomado de
<http://www.ideam.gov.co/documents/11769/641368/2.01+HM+Indice+calidad+aire.pdf/5130ffb3-a1bf-4d23-a663-b4c51327cc05>