



## Proyecto Integrador I

2020-1

Christian Flor - A00355624, Cesar Canales - A00345026, Carlos Restrepo - A00355028,  
Daniel Fernández - A00354694, Felipe Sanchez - A00355727

### REQUERIMIENTOS FUNCIONALES

#### 1. Introducción

##### 1.1 Propósito

El propósito de este documento es dar a conocer las funcionalidades, restricciones, objetivos y alcance del nuevo sistema de software para el análisis de los efectos de la calidad del aire en diferentes cultivos de los municipios del Valle del Cauca a todas aquellas personas del ámbito académico o de la industria agrícola a quien pueda serle útil.

##### 1.2 Ámbito del sistema

Gas impact analyzer on crops in the Valle del Cauca, es un nuevo software cuyo propósito es determinar si la concentración de diversos contaminantes en la atmósfera tienen relación con la diferencia entre las hectáreas sembradas y las hectáreas cultivadas con distintos cultivos transitorios del departamento.

Con la información obtenida a partir del análisis de varios registros proporcionados por tres bases de datos se espera obtener patrones que expliquen los comportamientos presentados, en aras de poder aplicar dichos patrones para alcanzar una mayor eficacia y eficiencia en los cultivos. Sin embargo, este software no constituye un sistema apropiado para predecir el comportamiento futuro de los cultivos ya que los datos de los que se disponen no tienen las características para realizar los cálculos pertinentes.

Como bien dice una parte del nombre del futuro sistema, este solo se centrará en los cultivos del Valle del Cauca, no se podrán realizar consultas para otros departamentos de Colombia.

A Partir de los datos ya almacenados en las bases de datos se analizan los contaminantes en la atmósfera desde el sistema, por ello no utilizaremos elementos de hardware.

### 1.3 Definiciones, acrónimos y abreviaturas

Acrónimo del nombre del proyecto: GIAOC-Valle del Cauca

### 1.4 Referencias

- [DATOS DE CALIDAD DEL AIRE EN COLOMBIA 2011-2017.](#)
- [Superficie Cosechada en Hectáreas con Cultivos Transitorios en el Departamento del Valle del Cauca del Año 2000 al 2018.](#)
- [Superficie Sembrada en Hectáreas con Cultivos Transitorios en el Valle del Cauca del año 2000 al 2018.](#)

### 1.5 Visión general del documento

A partir de este punto se encuentran dos apartados de importancia. En el primero de estos se describen todos aquellos factores que afectan al software en cuestión y a las precondiciones que este necesita para desempeñar sus funciones de manera adecuada. Así mismo, también se describirán a grandes rasgos dichas funciones y los usuarios hacia los cuales va dirigido este nuevo sistema.

En la siguiente sección se explicarán a detalle los requisitos, funciones, restricciones y atributos del sistema con el objetivo de que los diseñadores puedan desarrollarlo y así el programa cumpla con los requerimientos solicitados por el usuario.

## 2. Descripción general

### 2.1 Perspectiva del producto

Ya que este software realiza cálculos sobre bases de datos de terceros es altamente dependiente de estas. El programa extrae los datos de tres diferentes bases de datos, una que contiene información sobre diferentes contaminantes y su concentración en el aire, otra que contiene información de las hectáreas de tierra sembradas y una última que ofrece datos sobre dichas hectáreas cultivadas. Las tres bases de datos pertenecen al conjunto de datos abiertos de Colombia.

Se requiere que estos sets continúen siendo de carácter abierto para acceder a ellos y poder realizar funciones propias de la aplicación, como identificar patrones, mostrar gráficas, etc.

## 2.2 Funciones del producto

GIAOC-Valle del Cauca será capaz de llevar a cabo las siguientes operaciones:

- Determinar si existe relación entre la concentración de distintos tipos de contaminantes y el crecimiento de diversos cultivos en diferentes municipios del Valle del Cauca.
- Generar gráficas que muestren el comportamiento de un cultivo Y en función de la concentración de algún gas X.
- Encontrar las mejores combinaciones de factores para el crecimiento de cualquiera de los cultivos bajo estudio.
- Generar gráficas que muestran el cambio de los cultivos y de los contaminantes a través del tiempo.

## 2.3 Características de los usuarios

Se espera que aquellos que manejen este software pertenezcan ya sea al contexto académico o que estén directamente involucrados con los procesos siembra.

Si el producto cae en el sector productivo, los usuarios que interactuarían de primera mano con él serían aquellos encargados de la planeación de los cultivos y, por lo tanto, personas con la formación académica necesaria para comprender los datos mostrados por el aplicativo.

## 2.4 Restricciones

GIAOC-Valle del Cauca presenta las siguientes restricciones que limitan su desarrollo o funcionamiento:

- Debido a la cantidad de datos que se manejan, es posible que los equipos disponibles para su desarrollo no sean los más adecuados.
- Está atado al contenido de las tres bases de datos sobre las cuales se están trabajando. Además, no se pueden usar todos los datos ya que estos están restringidos a valores de ubicación específicos.
- Se debe usar el lenguaje de programación C# para desarrollar este proyecto.
- Las características de los datos no permiten usar la inferencia estadística. En su lugar se usará la estadística descriptiva.
- Se debe tener cierto conocimiento sobre la estadística descriptiva para el correcto uso de las fórmulas matemáticas necesarias.

- La comunicación entre los desarrolladores se ve limitada a reuniones virtuales debido a la contingencia por el Covid-19.
- Se debe tener conocimiento básico sobre el lenguaje de consulta de bases de datos.

## 2.5 Suposiciones y dependencias

Como se ha dicho previamente, el software depende totalmente de las bases de datos sobre las que trabaja. Si estas llegan a cambiar en su estructura, contenido o su carácter abierto es muy probable que se deban hacer ajustes pertinentes a la situación en el programa.

## 2.6 Requisitos futuros

Entre las posibles mejoras futuras se encuentran las siguientes:

- Ampliación del número de cultivos bajo análisis.
- Extensión del Valle del Cauca a otros departamentos de Colombia.
- Aumento en el número de gases estudiados.
- Generar una mayor cantidad de informes.
- Aplicar técnicas de estadística que permitan hacer predicciones.

## 3. Requisitos específicos

Para diferenciar entre funciones, requisitos, restricciones y atributos se usará la siguiente notación compuesta por dos letras y tres números: la primera letra denota la el tipo característica (F : función, R: requisito, C: restricción, A: atributo). La segunda denota su importancia (E: esencial, O: opcional). Los tres números siguientes tienen un rango desde 001 hasta 999, usados para diferenciar entre los mismos tipos.

### 3.1 Interfaces externas

A nivel de software se debe tener instalada la última versión de .Net para evitar posibles errores. Además, se hará uso de librerías de terceros para llevar a cabo las funciones correspondientes a mostrar gráficos. Se considera el uso de librerías que presten funciones de cálculo estadístico, a su también es factible el uso de herramientas externas que faciliten la consulta en las bases de datos.

En términos de hardware no se requiere de elementos extras para realizar el desarrollo y se espera que funcione correctamente en equipos de cómputo que no posean especificaciones de alta gama.

Por último, se debe disponer de una conexión estable a internet debido a los volúmenes de datos que deben ser leídos.

### 3.2 Funciones

- **FE001:** Determinar si existe relación entre la concentración de distintos tipos de contaminantes y el crecimiento de diversos cultivos en diferentes municipios del Valle del Cauca.

Dada la información de las bases de datos pertinentes, el programa usa el proceso de minería de datos, específicamente la técnica de clustering para determinar si existe relación alguna entre la concentración de contaminantes en el aire y la la cantidad de hectáreas cosechadas.

Se tiene previsto el uso del algoritmo K-means para realizar agrupamiento entre datos que compartan características comunes.

- **FE002:** Generar gráficas que muestren el comportamiento de un cultivo Y en función de la concentración de algún gas X.

El usuario podrá seleccionar el cultivo y el contaminante de su interés, y al cambiar valores en el segundo atributo el programa le mostrará a través de varios tipos de gráficos cómo se afecta el primero.

Los gráficos a generar son los siguientes:

- Gráfico lineal.
- Gráfico circular.
- Vista general de los datos de los datasets.
- Sistema de filtro de datos.
- Gráfico de dispersión.
- Gráfico de barras.
- Gráfico stacked area

- **FO001:** Generar gráficas que muestran el cambio de los cultivos y de los contaminantes a través del tiempo.

Con el fin de mostrar datos históricos, el programa desplegará una línea de tiempo en la cual se podrá apreciar como cambia la calidad del aire y los cultivos conforme pasa el tiempo.

### 3.3 Requisitos de rendimiento

El programa debe ser capaz de manejar volúmenes significativos de datos. Se estima que el tamaño de las entradas es del orden de  $10^4$ , aunque es posible que este factor crezca hasta  $10^6$ .

Para el hardware disponible el problema no recae en la cantidad de datos ya que estos pueden almacenarse fácilmente, sin embargo la cantidad de operaciones a realizar pueden representar un desafío para la capacidad de cómputo disponible.

### 3.4 Restricciones de diseño

Este proyecto cuenta con dos restricciones principales en su diseño.

- **CE001:** El programa debe ser desarrollado en un lenguaje de programación específico (C#), ya que así fue solicitado por el cliente.
- **CE002:** Esta restricción, también definida por el cliente, es que el software debe hacer uso de la base de datos del siguiente enlace, [clic aquí](#).

### 3.5 Atributos del sistema

- **AE001 (Fiabilidad):** Determinar si existe relación entre la información de dos bases de datos y se hará por medio del **algoritmo K-means**, el cual es ampliamente usado en minería de datos. Este algoritmo agrupa un conjunto de datos en **K** diferentes grupos, de los cuales los datos de cada grupo comparten características comunes e interesantes que los han hecho distanciarse de los demás grupos.
- **AE002 (mantenibilidad):** Mantener en funcionamiento este software se reduce principalmente a hacer seguimiento de los cambios en las bases de datos ya que se pueden dar aquellos que provoquen un funcionamiento inadecuado del producto. Así mismo, es pertinente

seguir las versiones más recientes de las librerías usadas como también posibles optimizaciones del algoritmo a priori.

- **AE003 (portabilidad):** Se espera que el producto final no tenga un peso significativo debido a que los datos sobre los que trabaja están almacenados en la nube. De este modo el programa no representará una carga significativa para la memoria de los equipos en los que se almacena.
- **AE004 (seguridad):** Por ahora, el programa no dispone de una diferenciación de usuarios y no presenta funciones que deban ser restringidas a un grupo selecto de personas.

### 3.6 Otros requisitos

Sin información por ahora.