

UNIVERSIDAD ICESI

FACULTAD DE INGENIERÍA
DEPARTAMENTO DE TIC

PROYECTO INTEGRADOR I

MÉTODO DE LA INGENIERÍA PARA LA SOLUCIÓN DE UN PROBLEMA CON
MANIPULACIÓN DE DATOS

NATALIA ISABEL GONZÁLEZ MURILLO

JUAN DAVID LECTAMO CAICEDO

LINA JOHANNA SALINAS DELGADO

JUAN JOSÉ VALENCIA JARAMILLO

CALI, VALLE DEL CAUCA

2020

Tabla de contenidos

1. Definición del problema.....	pág. 2-3
2. Recopilación de la información.....	pág.
3. Búsqueda de soluciones creativas.....	pág.
4. Transición de las ideas a los diseños preliminares.....	pág.
5. Evaluación y selección de la mejor solución.....	pág.
6. Preparación de los informes y las especificaciones.....	pág.
7. Implementación.....	pág.
Referencias.....	pág.

1. Definición del problema

La economía de un país en parte se ve reflejada no solo por su administración, sino por sus importaciones y exportaciones, las cuales le permiten interactuar en el mercado con el objetivo de crear relaciones con otros países y obtener ingresos de los cuales se pueda sacar provecho, por ejemplo para aumentar su desarrollo interno. Las últimas dependen en gran medida de la oferta disponible con la que cuente el país y el sector específico donde se realicen, pues esto a su vez determina el tipo de producción que permite obtener los productos finales.

La producción en un sector presenta distintos retos dentro de los cuales no solo la oferta y la demanda juegan un papel importante, sino el mercado laboral, los costos, las condiciones del entorno (como la tecnología empleada o el ambiente donde se desarrollan), entre otros. El medio ambiente es quizás una de las condiciones menos controlables a la hora de producir, especialmente en campos como la agricultura y a pesar de que existen métodos para reducir riesgos como las plagas, existen otros factores relacionados que definitivamente no pueden ser controlados, como el clima.

El clima juega un rol importante a la hora de determinar una situación en la cual haya un ciclo natural de crecimiento óptimo y deseado para las semillas que se cultivan, y por eso conocer su comportamiento es vital para proporcionar una producción que en lo posible reduzca las pérdidas por malas condiciones en las cosechas. No obstante, a pesar de ser un factor de condición general para todo el sector de producción agrícola, su comportamiento afecta de manera distinta a determinados tipos de cultivos y en consecuencia los resultados son diferentes según la situación.

En Colombia, el sector agrícola hace parte de sus exportaciones y se ha hecho popular especialmente por ofrecer sus productos representativos como un impulso a posicionarse en el mercado global de los mismos. Ahora, si su producción se pudiera impulsar de tal forma que se pudiera garantizar las mejores condiciones climáticas para hacerlo, su economía se

vería entonces estimulada y sus beneficios podrían aumentar de tal forma que su provecho sería mayor, tal como se había mencionado ya antes.

El IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales) ha recopilado datos de la calidad del aire en Colombia desde el 1 de enero del 2011 hasta el 31 de diciembre del 2017. En dichos datos se encuentran mediciones periódicas a lo largo de todos sus departamentos con variables de interés relativas a condiciones ambientales. Por lo tanto, *con el objetivo de ayudar a mejorar y estimular la producción de algunos de los cultivos del sector agrícola del país, es pertinente utilizar la información de la base de datos del IDEAM y realizar un estudio sobre ella que permita darle un seguimiento al clima que pueda llegar a determinar patrones considerables y significativos que puedan ser de ayuda para conocer su comportamiento durante los últimos años y estimar su posible desempeño para el futuro.*

2. Recopilación de la información

Una vez definido el problema es importante realizar una especificación de los requerimientos funcionales pertinentes para el desarrollo de la solución buscada. (Esta especificación se lleva a cabo en un documento aparte el cual se puede encontrar adjuntado en la carpeta de documentos del proyecto donde se encuentra también el presente informe) también puede ser consultado en:

<https://docs.google.com/document/d/1Y8yXifhRCbZd2cjOa9ajSHvqCnCoAQ26d9to0WjM7k4/edit?usp=sharing>

A continuación se presenta toda la información adicional necesaria relacionada a la definición del problema con el objetivo de que las soluciones sean más significativas.

Importaciones y exportaciones en Colombia

En Colombia tanto las importaciones como las exportaciones han presentado un decrecimiento durante los últimos años, y a pesar de encontrarse en el cuarto puesto de todo américa del sur, la nación cuenta con 184 países socios a los cuales destina sus exportaciones (principalmente en productos como : petróleo, combustibles y café), como se puede evidenciar en [1].

Mercado agrícola en Colombia

Algunos de los mercados más importantes para Colombia en cuanto a exportaciones es el mercado agrícola. El ministro de Agricultura y Desarrollo Rural, Aurelio Iragorri Valencia, destacó la forma en que el agro y los alimentos ayudaron al repunte de las exportaciones con un crecimiento de 29,2% en marzo. “Aunque se siguen consolidando los tradicionales como café, banano y flores, también empiezan a repuntar nuevos productos como cacao, aguacate, y frutas exóticas” [2].

Por otro lado además, según el DANE [3] , “Las exportaciones de Combustibles y productos de las industrias extractivas participaron con 60,9% del valor FOB total de las exportaciones; así mismo, Manufacturas con 16,1%, Agropecuarios, alimentos y bebidas 19,7%, y otros sectores con 3,2%.”

Agricultura

La agricultura de Colombia es variada y distinta en las diferentes regiones del país, como lo menciona Bell [4] :

Los productos agrícolas de Colombia son tan variados como su clima y su topografía. Toda clase de tierras y climas se encuentran aquí, desde los tropicales extremos hasta los de zonas

templadas; existen las zonas cálidas donde se dan los plátanos, la caña de azúcar y el tabaco, y la zona fría, tierra de papas, trigo y cebada, comunes en la meseta del interior.

Colombia puede dividirse en cuatro zonas primordiales: la primera es la zona costera que incluye los dos océanos, el Atlántico y el Pacífico; la segunda, las mesetas y los valles del interior del país con una altura de 3.000 pies sobre el nivel del mar, la tercera zona las estribaciones de la cordillera de los Andes, con alturas de hasta 6.000 pies (el reino del café), y el cuarto y último, las grandes montañas y mesetas de más de 6.000 pies de altura sobre el nivel del mar.

Las condiciones climáticas son muy diferentes en cada una de las zonas. La diferencia entre las costas del Caribe y del Pacífico es también muy marcada; en la primera, las temporadas están claramente definidas; de diciembre a mayo, estación seca y de junio a diciembre, la lluviosa, mientras que en la costa Pacífica, en especial en la parte norte de Buenaventura, llueve todos los días y la precipitación es tan fuerte como en los trópicos. Las dos costas son cálidas y húmedas, las temperaturas altas registraron 95° F. Las condiciones a lo largo del valle del Magdalena, hasta Girardot, son muy similares a las de la costa Atlántica, aunque la parte alta del valle recibe más lluvias que la parte baja del valle, siendo más abundante el caudal del río, y las temporadas son menos exactas a medida que se aproximan a la costa.

En la segunda zona, que incluye el valle del Cauca y toda Antioquia etc., hay dos temporadas húmedas y secas en el año que se alternan así: enero, febrero y marzo, estación seca; abril, mayo y junio, estación húmeda; julio, agosto y septiembre, seca y octubre, noviembre y diciembre, estación húmeda. La temperatura promedio varía entre los 64° y los 84°F, con una temperatura promedio de 76°F. La zona más tórrida del año es durante julio, agosto y septiembre, aunque el clima durante todo el año es semi-tropical.

En la tercera zona climática las divisiones de húmedo y seco son similares, aunque la precipitación es mayor. Desde los 5.000 hasta los 7.000 pies el clima es ideal, ni muy caliente ni muy frío, con una variación menor en la temperatura durante el año, siendo los promedios, alto y bajo, de 72° y 58° respectivamente.

En la cuarta zona, comprendida entre los 6.000 y los 9.000 pies de altura sobre el nivel del mar, hay heladas ocasionales en los niveles más altos y las temperaturas son 40°F la baja y 64°F la alta. Las temporadas no están claramente definidas y las lluvias son frecuentes. El mejor café es originario de los niveles más bajos de esta zona; en las altas mesetas, como la de Bogotá, productos como el trigo y la cebada son comunes a pesar de la ausencia de estaciones definidas.

Productos de las diversas zonas

En la primera zona, la parte norte de la costa Pacífica, sólo es apropiada para la tagua y el caucho debido a las incesantes lluvias. El plátano, el maíz y los bananos se cultivan a baja escala. La población, casi toda de raza negra, es reducida y no parece haber un futuro inmediato de oportunidad para esta región, que es quebrada debido a las estribaciones de la cordillera Occidental de los Andes y a las formaciones de mesetas bajas cercanas a la costa Pacífica. Al sur de Buenaventura el clima cambia, con menor precipitación en la región de Tumaco y Barbacoas y se encuentra ganado en el valle del río Patía, en donde la tierra es buena también para la siembra de algodón, caña de azúcar, cacao, tabaco, caucho, etc., pero por causa de la población tan reducida y las condiciones climáticas no se prevén desarrollos de alguna importancia. Esta última región -y se aplica a toda la costa Pacífica- produce cantidades considerables de tagua (marfil vegetal) y caucho.

La costa Caribe y los valles de los ríos Sinú y el Magdalena producen bananos, maíz, tabaco, algodón, caña de azúcar, tagua y caucho y en esta región la ganadería se ha convertido en una industria de importancia. Los bananos se exportan en grandes cantidades desde Santa Marta. El azúcar y el arroz prometen convertirse en fuentes de riqueza para el país en un futuro cercano y el tabaco también se está cultivando para la exportación.

La segunda zona, aquella de las mesetas y valles del interior, es adecuada para los cultivos de maíz, algodón, caña de azúcar, cacao, tabaco, café y una variedad menor de vegetales, entre los que se encuentra la yuca.

En la tercera zona abunda el maíz, el café, los frijoles y la papa.

La cuarta zona produce trigo, cebada y otros cereales, así como las frutas propias de la zona templada, los vegetales y la papa que se dan mejor en estas tierras frías. Existen cuatro productos de consumo doméstico que se encuentran en todo el territorio nacional: maíz, caña de azúcar, plátano y yuca; todos se consiguen en abundancia en todos los climas de Colombia y se constituyen en productos básicos de la dieta nacional, junto con la carne de res, que es de consumo de todas las clases. El maíz se siembra en todos los rincones del país. La caña de azúcar se encuentra también en las cordilleras, y crece en parches pequeños, a lado y lado de las montañas, y sólo en el Valle del Cauca y Bolívar se produce en grandes cantidades y se vende a los ingenios azucareros que cuentan con equipo moderno para la producción del azúcar blanco. El plátano es uno de los alimentos esenciales en la dieta del país, como lo es la papa para los norteamericanos.

El plátano crece en todas las regiones de Colombia, aun en las montañas de más de 7.000 pies. Los plátanos y los bananos se utilizan para el intercambio de productos en los mercados de las localidades, y en las áreas rurales apartadas toda casa, rancho y aun la más humilde choza, tiene más de una mata de plátano para el consumo. La producción total de plátano y banano, como la de caña de azúcar y el maíz, es muy grande, pero solo los bananos de Santa Marta se exportan a través de la Compañía United Fruit, a pesar del intento de una compañía alemana, hace ya algunos años, de establecer una gran bananera en el golfo de Urabá, sin éxito alguno. Hay excelentes áreas para la producción del banano en la costa Caribe, en el Departamento de Bolívar, hacia el sur de Cartagena, donde se incrementa día tras día la producción para suplir la demanda del mercado local y el fácil acceso al puerto de Cartagena. Aún más, este tipo de industria está muy adaptado a las condiciones de la mano de obra de los nativos, pero infortunadamente la industria bananera para la exportación, que depende en su totalidad del transporte marítimo rápido y el éxito del negocio, radica en la organización de las grandes compañías.

El maíz también ha sido producto de exportación. Los Estados Unidos importaron un total de 6.420 dólares colombianos en 1917, aunque en 1919 las exportaciones de maíz representaron un total de 32.720 kilos, avaluados en 1.360 dólares, por exportaciones a Cuba a través de goletas.

El azúcar refinado se exporta de Colombia primordialmente a la Zona del Canal desde Cartagena y Buenaventura. El azúcar que sale por Cartagena procede de cañaduzales de Sincerín y las exportaciones desde Buenaventura provienen de “La Manuelita” en el Valle del Cauca, cerca de Palmira, a 30 millas de Cali. Las exportaciones de azúcar refinada del ingenio “La Manuelita”, alcanzaron en 1917 la cifra de 1.774.743 kilos, valuados en 161.736 dólares colombianos y su destino final es el Canal de Panamá. Este valor se incrementó drásticamente entre 1918 y 1919 y se pagaron precios más altos. Las exportaciones desde Cartagena en 1915 llegaron a los 134.037 dólares y en 1916 a 249.239, y toda la producción se envió a la Zona del Canal. En 1917, las exportaciones de azúcar alcanzaron las 2.026.164 libras, y reportaron 274.341 dólares.

Resumen de las condiciones

Cada uno de los cultivos importantes de cada distrito ha sido analizado por separado en las explicaciones por región, en donde se dan los detalles de la producción, métodos, precios, etc. Al analizar de forma general la agricultura en Colombia, no se puede afirmar que se encuentre desarrollada. Esta condición se debe principalmente a la falta de carreteras, que no se han construido por la naturaleza quebrada e irregular de la tierra y las grandes distancias que separan unos sitios poblados de otros. Por este motivo, con mucha frecuencia algunos productos importados le hacen fuerte competencia a artículos domésticos, a pesar de los altos aranceles para la protección del producto nacional. Un ejemplo de esto es el trigo que puede importarse de los Estados Unidos para ser molido en el Caribe o en el Pacífico por un precio inferior que el traído por el río Magdalena desde la región de Bogotá.

Grandes porciones de las cordilleras Oriental, Occidental y Central están casi despobladas. En particular en la sección del Valle del Magdalena y otras regiones del país. Los centros de la población se encuentran en Cundinamarca, Boyacá, Antioquia, Caldas, el Valle del Cauca, algunas partes de Santander y de la costa Atlántica, algunos sectores de Nariño al sur del país, cerca de Pasto.

Los principales productos de los departamentos y regiones del sur son los siguientes:
Boyacá: trigo, cebada, papa, fríjol, maíz, vegetales, ganado y caballos, unas pocas ovejas, poca azúcar y algo de café.

Cundinamarca: café, trigo, maíz, ganado bovino, ovejas y azúcar. La agricultura es más avanzada en Cundinamarca que en cualquier otro lugar del país y la mano de obra es abundante siendo casi toda indígena.

Antioquia: café, algo de algodón, caña de azúcar en la forma de panela y plátanos.

Tolima: cacao, azúcar, café, arroz y ganado (aunque no de buena calidad).

Costa Atlántica: bananos en Santa Marta, maíz, azúcar, tabaco, cacao (no en abundancia) cerca de Barranquilla y arroz al sur de Cartagena.

Nariño: trigo, anís, papa, vegetales y poco ganado. La población es casi toda indígena.

Caldas: café, maíz, frijol y papa. Buena mano de obra, casi toda blanca.

Valle del río Cauca: ganado, azúcar, arroz, maíz y frijol. Mano de obra: mulatos

Departamento del Cauca: café, ganado, trigo y maíz. Mano de obra: indios.

Santander: café, tabaco, cacao y maíz. Mano de obra: blanca pero de calidad muy regular.

Costa Pacífica: prácticamente todo es selva tropical. Ganado, maíz, caucho y algo de azúcar en el valle del Patía y en los alrededores de Tumaco.

Sierra Nevada, región del Departamento de Magdalena: se siembra café en pequeñas extensiones, debido a la escasez de mano de obra. Los indios también siembran trigo y papa

Cultivos

Entre algunos de los cultivos en impulso discutidos en [5] para el sector agrícola en Colombia, se encuentran los siguientes : soya, trigo, chontaduro, cacao, tomate, aguacate, granadilla, yuca. Las condiciones climáticas deseadas para la manutención adecuada de los mismos se pueden encontrar resumidas en la siguiente hoja de cálculo. (también se puede encontrar adjuntado en la carpeta de documentos del proyecto donde se encuentra también el presente informe).

<https://docs.google.com/spreadsheets/d/188xy-H7awWeWpGhIOjyTxdhVRLM-xWJw4MljHxhKm9o/edit?usp=sharing>

Datos abiertos

Como se puede encontrar en [6]:

El gobierno de Colombia cuenta con una red de datos abiertos que ofrece de manera pública y unificada, el acceso a todos los datos publicados por las entidades públicas de Colombia en formato abierto, con el fin de que éstos puedan ser usados por cualquier persona para desarrollar aplicaciones o servicios de valor agregado, hacer análisis e investigación, ejercer labores de control o para cualquier tipo de actividad comercial o no comercial.

Data set

El grupo de datos recolectado por el IDEAM puede consultar de manera directa en [7].

Minería de datos

La minería de datos también conocida como Data Mining, es el conjunto de técnicas que se usan para la selección y extracción de datos, respecto a variables específicas, de una gran cantidad de datos, también analiza los grandes volúmenes de datos, sintetiza, identifica y

agrupa patrones de comportamiento entre estos. El modelo de Data Mining cumple una secuencia de pasos que son enunciados a continuación. Como se puede encontrar en [8]:

1. Definición del problema: en este paso se define claramente el problema y se considera la forma en que los datos contribuyen a la solución de este, también se analizan los requerimientos, se define el alcance del problema, entre otros. En este paso es importante plantearse diferentes interrogantes como: ¿qué se está buscando? ¿qué tipo de relaciones se está tratando de encontrar? con ello se espera tener una síntesis clara de qué es lo que se espera encontrar al profundizar en los datos.

2. Preparación de los datos: debido a la dispersión de los datos, se necesita que haya una selección previa de los datos, esto no significa que los datos deban ser eliminados, sino de verificar que las fuentes sean confiables e identificar cuáles son las variables más apropiadas para su uso en la investigación.

3. Exploración de datos: en este paso se debe conocer los datos para tomar las decisiones adecuadas al crear los modelos de minería de datos. Identificar los datos que no son relevantes para el objetivo del estudio, también se usan métodos estadísticos (tales como intervalos de mínimos y máximos, media, desviación estándar y distribución de datos) para una medición de los datos que permita conocer datos atípicos, o datos defectuosos que puedan alterar la precisión de los modelos y que por lo tanto deben ser descartados.

4. Construcción de modelos: para la construcción del modelo de Data Mining, es necesario identificar las columnas de datos que se desean usar, en este paso se usan los datos obtenidos en el punto inmediatamente anterior, y se aplican a ellos las distintas técnicas de Data Mining. Se debe tener en cuenta que al momento de cambiar los datos si estos se actualizan, se debe reestructurar el modelo de la Data Mining.

5. Exploración y validación de modelos: antes de hacer la implementación del modelo de Data Mining se debe validar el correcto funcionamiento de los modelos diseñados, en la validación se deben tomar datos de prueba para evaluar la precisión del rendimiento de los modelos al momento de realizar la predicción de los datos.

6. Implementación y actualización de modelos: esta es la fase final de la secuencia, es el momento de llevar a cabo la implementación de uno o más modelos que solucionen el problema definido.

Los pasos del modelo de Data Mining no son secuenciales, por consiguiente si el diseñador encuentra una falla en un paso avanzado como por ejemplo que el problema identificado en el paso uno debe ser reestructurado para obtener una mejor producción de los datos, se debe poder modificar y rediseñar la estructura del Data Mining.

Estado del arte

IDEAM: El Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), que forma parte del sistema mundial de observación para la producción de datos meteorológicos en tiempo real, mantiene un programa operativo para el pronóstico del tiempo y cuenta con las corridas de modelos globales para resoluciones espaciales que varían entre los 80 a 190 km, con dos tipos de modelos de mesoescala de cobertura nacional y regional para resoluciones espaciales de 5 y 25 km. Los pronóstico, avisos y alertas oportunos de fenómenos meteorológicos, como tormentas, vientos fuertes, mareas de tempestad, entre otros, son valiosísimos para salvar vidas humanas, al proporcionar la información necesaria para sacar a la población de la zona de peligro o ayudar a las personas a tomar rápidamente las medidas requeridas, además permite a las empresas, a los funcionarios del gobierno local y al público en general modificar sus operaciones y proteger sus bienes. Para conocer el reporte de pronósticos generados por el IDEAM, consulte [9].

AEMET: La Agencia Estatal de Meteorología hace uso de aplicaciones específicas de software. Entre las que cabe resaltar las estaciones de trabajo meteorológicas, el editor gráfico de predicciones para la generación automática de productos y los modelos numéricos de predicción del tiempo, de la calidad del aire y del clima. Para conocer más a detalle sobre estas aplicaciones consulte [10].

3. Búsqueda de soluciones creativas

Ahora que se tiene la información necesaria para el problema, es pertinente realizar la búsqueda de soluciones creativas que puedan dar solución al problema con las condiciones requeridas.

Consideraciones: cabe mencionar que el problema definido puede ser dividido en dos momentos, primero el manejo y análisis de los datos (1) y el segundo el uso y presentación de los resultados obtenidos anteriormente (2), y por lo tanto para ambas es pertinente encontrar ideas creativas que permitan decidir la mejor manera de hacerlo.

1. La minería de datos es un proceso que debe llevarse a cabo de manera indiscriminada pues debido a la magnitud de los datos disponibles este es el proceso estructurado más efectivo para trabajar sobre tal información, sin embargo, las técnicas que deben ser definidas en uno de sus pasos (paso 4) pueden variar y por lo tanto es pertinente nombrar todas las posibilidades, para luego hacer un análisis y una posterior evaluación y selección (paso 4 y 5 respectivamente del presente informe)
2. El uso de los resultados obtenidos luego de aplicar el proceso de minería de datos puede variar y por lo tanto es donde debe tomarse una decisión a partir de un planteamiento de ideas, un análisis y una posterior evaluación y selección (paso 3,4 y 5 respectivamente del presente informe)

Por lo anterior, a continuación se plantean dos técnicas de generación de ideas, el arte de preguntar y la relación forzada, que serán usadas únicamente en la segunda sección (uso de los resultados).

Técnica 1: El arte de preguntar, esta técnica consiste en plantear interrogantes al rededor del problema para luego tratar de dar respuesta a los mismos con la intención de poder conseguir una idea estructurada que responda a todas las necesidades. (Aquí se hará énfasis en el cómo)

Técnica 2: Relación forzada, el objetivo de esta técnica es poder tomar un componente (entidad, producto, etc) relacionado al problema y un componente totalmente ajeno para luego realizar una conexión forzada entre ellos a partir de palabras que desarrollan los componentes en sí mismos.

Ideas para las técnicas de minería de datos.

1. **Técnicas de verificación:** permiten realizar un análisis sobre la población de la cual los datos fueron extraídos haciendo uso de las técnicas estadísticas de inferencia, (pruebas de hipótesis, análisis de varianza (ANOVA)).
2. **Técnicas de descubrimiento:** permiten obtener resultados significativos a partir de los datos.
 - 2.1. Técnicas de descripción: permiten describir los datos según un criterio, clustering (agrupamiento), resumen automático (resumen de datos importantes), visualización (representaciones gráficas).
 - 2.2 **Técnicas de predicción**
 - 2.2.1 Técnicas de clasificación: permiten realizar predicciones a través de técnicas especializadas, redes neuronales, redes bayesianas, árboles de decisión, etc.
 - 2.2.2 Técnicas de regresión: permiten encontrar información a través de modelos de regresión.

Ideas para el uso de los datos.

Técnica 1 : El arte de preguntar.

¿Por qué? Porque se quiere contribuir al sector agrícola, identificando municipios y/o departamentos con las mejores condiciones climatológicas para un mayor rendimiento de los cultivos.

¿Para cuándo? Para los años próximos, de tal manera que los lugares sugeridos puedan ser explotados y puestos a producir durante este periodo de tiempo.

¿Cómo? A través de un mapa que presente los posibles departamentos y sus condiciones climáticas.

¿Cómo? A través de un reporte estadístico que presente los resultados de manera resumida por cada año.

La composición de los puntos anteriores genera las siguientes ideas:

Porque se quiere contribuir al sector agrícola, identificando municipios y/o departamentos con las mejores condiciones climatológicas para un mayor rendimiento de los cultivos para los años próximos, de tal manera que los lugares sugeridos puedan ser explotados y puestos a producir durante este periodo de tiempo a través de:

1. **Mapa de cultivos** : presenta todos los departamentos para la selección y los cultivos más apropiados.
2. **Reporte de datos:** presenta por año los datos más relevantes con respecto al clima de manera resumida.

Técnica 2 : Relación forzada.

CLIMA	VASO
-------	------

Variable	Fluido
No controlable	Llenado
Alterar	Límite

Variable y Límite, nos llevan a la idea de:

3. Mapa delimitado: un mapa que nos permita delimitar una zona seleccionada según una variable de interés proporcionada.

Alteración y Llenado, nos llevan a la idea de:

4. Mapa de colores : Un mapa que nos permita llenar las zonas de interés a través de un código de colores.

CULTIVO	HIELO
Cosechas	Temperatura
Crecimiento	Frío
Plantación	Duro

Cosechas y temperatura, nos llevan a la idea de:

5. Mapa de temperatura: Un mapa que permita visibilizar las zonas con la temperatura adecuada para el tipo de cosecha seleccionado.

REGIÓN	AGUA
Lugar	Líquida
Ubicación	Abundante
Terreno	Humedad

Lugar y Humedad nos lleva a:

6. Mapa de humedad: Un mapa que resalte los lugares con la humedad adecuada para las plantaciones de interés.

REGLA	CLIMA
Específico	Calidad
Medidor	Temperatura
Escala	Humedad

Medidor y Temperatura nos lleva a:

7. Medidor de temperatura: Una escala que ayude a informar el clima en respectivos lugares.

GRILLO	TEMPERATURA
Natural	Medible
Predicción	Saturación
Ruido	Constante o Variable

Ruido y Saturación nos lleva a:

8. Gráfico de variables: Un gráfico de la variable seleccionada que nos permita medir los picos o momentos críticos del clima en los últimos años.

4. Transición de las ideas a los diseños preliminares

Ya que las ideas han sido generadas, es momento de ampliarlas y realizar un análisis sobre los aspectos extra que conlleva llevarlas a cabo, con el objetivo de poder realizar un descarte de aquellas ideas que no son del todo útiles.

Ideas para las técnicas de minería de datos.

A partir de todas las posibles técnicas de minerías de datos se eligen las siguientes para hacer un uso total de ellas.

1. **Técnicas de verificación:** el uso de estas técnicas es indispensable para realizar un estudio y análisis previos sobre el comportamiento de las variables de diseño y así poder lograr un mejor aprovechamiento de los mismos (en este caso las variables de interés son aquellas relativas al clima : Fecha, Latitud, Longitud, Departamento, Variable (que incluye los siguientes valores : Precipitación Líquida, Velocidad del tiempo, Radiación Solar Global, Presión atmosférica, Temperatura, Humedad Relativa, Radiación UVB), Unidades y Concentración además de ser complementario para algunas ideas propuestas como el reporte de datos, las técnicas elegidas son : Pruebas de hipótesis, Intervalos de confianza y ANOVA, siendo útil para obtener una conclusión de la población a partir de los datos de muestra.

2. **Técnicas de descubrimiento:** permiten obtener resultados significativos a partir de los datos.

2.1. Técnicas de descripción: el uso de estas técnicas es complementario a algunas ideas propuestas como el reporte de datos, además que permiten la presentación de los datos de manera más específica. Las técnicas elegidas son Clustering que permite agrupar los datos de manera que sus características sean similares, siendo útil para generar resultados de interés como climas comunes, etc. y Visualización que permite representar de manera gráfica grupos particulares de datos, siendo útil también para el estudio de las variables de diseño anteriormente mencionadas.

- 2.2 Técnicas de predicción:

2.2.1 Técnicas de clasificación: el uso de estas técnicas es importante para la predicción de los posibles resultados a partir de unos datos

datos, sin embargo para esto primero se deben estudiar los datos para decidir cuál es la más adecuada según su carácter.

Ideas para el uso de los datos.

1. **Mapa de cultivos:** la idea con este mapa es listar los departamentos del país con un indicador con el que se pueda interactuar y que permita conocer cuáles son los cultivos más convenientes a partir de sus características climáticas, mostrando así todas las descripciones de desempeño de todas las variables de interés. sin embargo no es tan visual para el usuario lo que reduce un poco su experiencia en el uso de la misma.
2. **Reporte de datos:** la idea con este reporte es mostrar un resumen de datos en tablas y gráficas que permita mostrar al usuario indicadores con el desempeño de las variables de interés en los últimos años con el objetivo de poder hacer la información más fácil de interpretar. sin embargo esta idea es limitada para aquellos usuarios que no saben cómo leer estos elementos descriptivos.
3. **Mapa delimitado:** la idea con este mapa es parecida a la del mapa de climas, pero en este caso la presentación es más específica pues teniendo en cuenta una zona única seleccionada permite conocer una descripción para una única variable de interés, y al contrario del mapa de climas el usuario debería proporcionar una consulta particular en vez de interactuar con el mapa que ya contiene todas las consultas.
4. **Mapa de colores:** la idea con este mapa es clasificar cada zona por una escala de colores de manera que permita a través de la misma verificar visualmente dado un cultivo cuál sería la mejor zona esperada o más apta para producirlo.

5. **Mapa de variables:** la idea con este mapa es una fusión entre las ideas 5 y 6 (mapa de temperatura y mapa de humedad) que permita a través de una escala de colores mostrar cual el desempeño de las distintas variables seleccionadas para todo el país haciendo más evidente en cuáles zonas se evidencian valores menores, normales o irregulares.
6. **Medidor de temperatura:** la idea con este medidor es crear una escala de temperaturas que permitan clasificar el comportamiento de las zona a partir de su temperatura y por lo tanto, sí sería pertinente realizar una producción de cultivos allí. Esta idea se descarta porque solo hace parte de la solución del problema, por lo tanto es incompleta.
7. **Gráfico de variables:** la idea con este gráfico es describir visualmente los datos de manera que se pueda evidenciar su comportamiento durante los últimos años para así evidenciar sus posibles cambios de manera visual.

5. Evaluación y selección de la mejor solución

Con el objetivo de poder seleccionar la mejor solución posible para el problema definido es considerable realizar un método de evaluación que permita decidir sobre cuál es la mejor opción posible para ser desarrollada. Para realizar la evaluación de ideas, se establecen los siguientes criterios basados en los costos diversos que conllevan aplicar la solución planteada y el nivel de significancia que tendrá la solución con un puntaje de 1 a 3 de la siguiente manera:

- Costo de implementación del desarrollador: este costo implica la facilidad de implementación para el desarrollador en un lenguaje de programación. Puede ser bajo (si la solución es fácilmente traducible), intermedio (si la solución es parcialmente traducible), o alto (si la solución es difícilmente traducible)

1. Bajo

2. Intermedio
 3. Alto
- Costo de lectura y comprensión para el usuario: este costo implica la facilidad del usuario para leer e interpretar los resultados mostrados. Puede ser baja (si la solución es fácil de leer y entender), intermedio (si la solución es parcialmente legible y entendible), o alto (si la solución es difícil de leer y entender)
 1. Bajo
 2. Intermedio
 3. Alto
 - Nivel de significancia de la información al usuario: este criterio implica el nivel de significancia y sentido de la información que presenta la solución a elegir. Puede ser bajo (si la información que presenta la solución es incompleta, vaga y sin sentido), intermedio (si la información que presenta la solución es parcialmente completa y parcialmente significativa), o alto (si la información que presenta la solución es completa, significativa)
 1. Bajo
 2. Intermedio
 3. Alto

Basado en esta consideración de costos la solución con menos costo será la seleccionada y la mejor a implementar. Consecuentemente, las respectivas soluciones posibles y los criterios de juicio desarrollados son agrupados en la siguiente tabla para poder llevar a cabo la evaluación.

Idea	Costo de implementación	Costo de lectura	Total	Nivel de significancia
Mapa de cultivos	2	1	3	2
Reporte de datos	2	2	4	3

Mapa delimitado	2	2	4	3
Mapa de colores	3	1	4	2
Mapa de variables	3	2	5	3
Gráfico de variables	1	3	4	1

A partir de la evaluación anterior, la idea con mejor reducción de costo posible con respecto a las demás es el mapa de cultivos, puesto que primero implica un costo promedio para el desarrollador y segundo implica un costo de lectura y comprensión bajo lo que permite mostrar la solución de una forma más fácil para el usuario. Por otro lado, a pesar de que su nivel de significancia es intermedio, sigue representando un costo menor por lo que será la que se usará para solucionar el problema definido.

Teniendo en cuenta lo anterior, también se implementará la idea del reporte de datos y el mapa de colores como complementos junto a la idea ganadora para lograr mayor especificación en la información brindada al usuario; los cuales estarán apoyados en gráficos cuyo objetivo será resumir la información y brindar un nivel de detalle mayor para lograr el nivel de significancia deseado.

6. Preparación de los informes y las especificaciones

Finalmente una vez elegida la solución definitiva a ser desarrollada es importante realizar un diseño que permita el direccionamiento de la misma hacia la implementación, para esto:

1. se especifica un diccionario de datos para mejorar la comprensión del contenido de los datos con los que se pretende trabajar (se puede encontrar adjunto a la carpeta de documentos del proyecto donde se encuentra también el presente informe).

2. se desarrolla un diagrama de clases (se puede encontrar adjunto a la carpeta de documentos del proyecto donde se encuentra también el presente informe).
3. Además para poder tener una concepción general de la interacción de los objetos dentro del sistema se desarrollan tres diagramas de secuencia que modelan las funcionalidades más importantes (estos se pueden encontrar adjunto a la carpeta de documentos del proyecto donde se encuentra también el presente informe).

Ahora como parte del desarrollo lógico del presente informe se realiza una *síntesis reflexiva* que permita aportar una perspectiva autoconsciente sobre los autores y su proceso de aprendizaje y trabajo a la hora de escribirlo:

7. Implementación

La debida implementación de la aplicación solicitada, con las funciones requeridas y la solución seleccionada se llevarán a cabo en el lenguaje de programación (C#) con la lista de las siguientes tareas a desarrollar:

La implementación del proyecto se incluye en el siguiente proyecto de GitHub:

<https://github.com/backtojuan/HarvestingForTheFuture>

Referencias

[1] O.V. Fernández Rodríguez, “América del sur en el comercio global (2007-2016)”, *Agenda Internacional* 2018, Vol. 25 Issue 36, p 169-190. 22p , pp. 173 , 2018.

[2] Ministerio de Agricultura y Desarrollo Rural, “Crece potencial exportador del agro”, 2017, [En línea] <https://bit.ly/38vRTm9> [Accedido: 7-mar-2020]

[3] DANE, “Exportaciones”, 2020, [En línea] <https://bit.ly/32XkX4K> [Accedido: 7-mar-2020]

- [4] P. L. Bell. , “Colombia : manual comercial e industrial”, 2012, pp. 207 - 212 [En línea] <http://repositorio.banrep.gov.co/handle/20.500.12134/9310?show=full> [Accedido: 7-mar-2020]
- [5] Imprenta nacional de Colombia, “SENADO DE LA REPUBLICA PROYECTOS DE LEY ESTATUTARIA”, 2019, [En línea] http://www.secretariasenado.gov.co/legibus/legibus/gacetas/2019/GC_0728_2019.pdf [Accedido : 1-may-2020]
- [6] Ministerio de Tecnologías de la Información y las Comunicaciones, “Datos Abiertos”, 2020, [En línea] <https://www.mintic.gov.co/portal/inicio/7244:Datos-Abiertos> [Accedido: 7-mar-2020]
- [7] Ministerio de Tecnologías de la Información y las Comunicaciones, “DATOS DE LA CALIDAD DEL AIRE EN COLOMBIA 2011-2017”, 2018, [En línea] <https://bit.ly/32WLNKq> [Accedido: 7-mar-2020]
- [8] Microsoft Docs, “Data Mining Concepts”, 2019, [En línea] <https://docs.microsoft.com/es-es/analysis-services/data-mining/data-mining-concepts?view=allproducts-allversions> [Accedido: 7-mar-2020]
- [9] IDEAM, “Pronósticos IDEAM” , 2020 [En línea] <http://www.pronosticosyalertas.gov.co/pronosticos> [Accedido: 15-mar-2020]
- [10] AEMET, “Aplicaciones específicas de software meteorológico”, 2020 [En línea] http://www.aemet.es/es/idi/tic/aplicaciones_meteorologicas [Accedido: 15-mar-2020]