

PROYECTO FINAL - CIENCIA DE DATOS APLICADA

Presentado por:

Oscar Bosigas

Código: 202220008

o.bosigas@uniandes.edu.co

David Martínez

Código: 202224801

d.martiez@uniandes.edu.co

Universidad de los Andes

Maestría en Ingeniería de la Información - MINE

2022

TABLA DE CONTENIDO

1. Entendimiento del problema.....3

2. Ideación.....3

3. Implicaciones4

4. Enfoque analítico5

5. Entendimiento de los datos.....5

6. Entendimiento del producto.....6

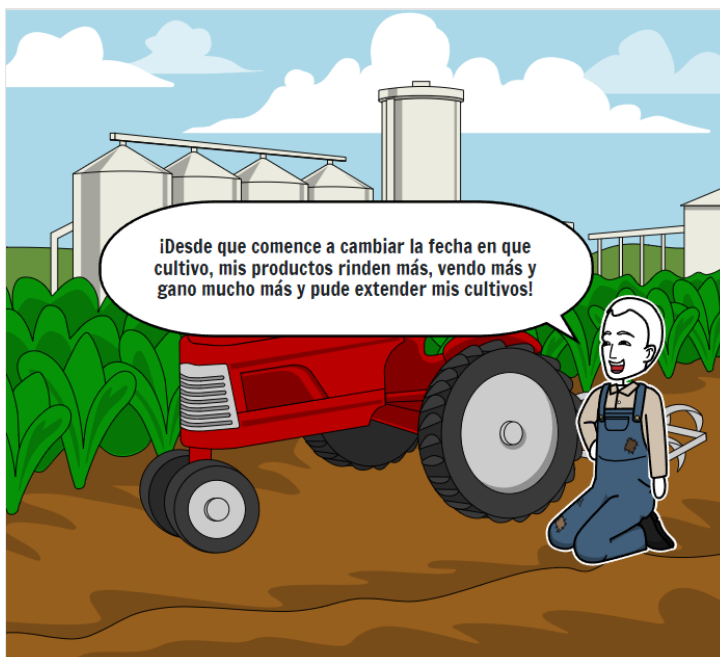
7. Conclusiones6

1. Entendimiento del problema

La agricultura es una actividad económica relevante para el gobierno colombiano y dadas las ventajas geográficas y naturales con las que cuenta el país, es un enfoque priorizado en el desarrollo productivo. En el caso de la agricultura colombiana, se requiere el desarrollo una serie de sistemas de información que permitan que entidades territoriales y los mismos agricultores tomen decisiones y se informen sobre la actividad. Este proyecto propone un Sistema de Información predictiva para la agricultura en Colombia, donde mediante la utilización de técnicas avanzadas se pueda crear un modelo que permita predecir el comportamiento de los cultivos en los siguientes años igualmente, presentando esto visualmente en una herramienta que nos permita crear un dashboard de como los cultivos se han comportado en años anteriores y su tendencia en el futuro, con el fin de identificar oportunidades y datos que ayuden a la toma de decisiones en el país.

2. Ideación





Solución DFMC = Tomar todas las medidas pertinentes para determinar un posible futuro respecto a en que zonas y que cultivo pueda ser el que tenga un rendimiento mejor en el futuro dependiendo de cuando se cultiva (se toma referencia de fechas por año y por semestre)

3. Implicaciones

Los riesgos se pueden centrar en que cultivadores de cierto cultivo comiencen a perder ventas respecto al análisis que se ha realizado igualmente zonas donde tengan un rendimiento no esperado en su cultivo puedan estar en declive frente a esto.

Al no tener una medida de ventas de cultivos se puede caer en el error en dar información no acertada respecto a donde ir y a que cultivo tiene mejor ganancia respecto al comprador como al vendedor.

4. Enfoque analítico

Algunas de las preguntas que se buscan responder son las siguientes:

- Los municipios que más producen un determinado producto.
- El producto que más se produce en cada municipio.
- Producto más sembrado y producido en Colombia.
- Municipios o departamentos tienen más áreas sembradas y cosechadas.
- Municipio o departamentos con más producción por año.
- Total de hectáreas sembradas en Colombia por año.
- Total de hectáreas cosechadas en Colombia por año.
- Cultivo que se pueda sembrar en un futuro con un margen de productividad mayor.
- Zonas donde no se deban cultivar en ciertas fechas al no ser aprovechadas en un margen mínimo.

Esto con el fin de determinar tanto los municipios y departamentos como los productos y su producción más representativos de Colombia y así poder buscar una optimización o mejoras en estos casos puntuales, debido a que representan una parte importante de la economía y su análisis nos puede ayudar a un nivel económico y social aportando información a las organizaciones gubernamentales del agro que puedan ayudar a los involucrados.

Algunas de las técnicas de análisis que esperamos aplicar son principalmente clasificación, agrupación y regresión.

5. Entendimiento de los datos

Los datos que vamos a utilizar para hacer el diseño de nuestra solución fueron tomados de datos abiertos en el siguiente link : [Evaluaciones Agropecuarias Municipales EVA | Datos Abiertos Colombia](#), en donde se tienen cerca de 200000

registros en los cuales se ha tomado información desde el año 2006 hasta el año 2018. La información que se nos brinda de allí es anualizada y por semestres, por lo que los datos y predicciones que se van a realizar se van a mostrar por semestres.

6. Entendimiento del producto

Al realizar el análisis exploratorio (se encuentra en GitHub) se notaron varias cosas importantes, como es que el cultivo que más se le tomaron datos fue el maíz esperando encontrar que fuera la papa ya que somos un territorio que se centra en este cultivo, igualmente como era de esperarse las zonas con más información recolectada resulto ser por Boyacá. Se determina que la relación entre área sembrada y área cosechada está estrechamente relacionada, pero algo curioso es que la correlación del área sembrada con el rendimiento es de casi 0 por lo que se encuentra una relación inexistente entre estas dos variables.

7. Conclusiones

Al realizar la correlación de variables se demostró que hay campos en los que no se encuentra una relación tan efectiva como se esperaría.

El promedio del rendimiento total no es tan bueno, aunque somos un país tropical con una gran cantidad de hectáreas donde se pueda sembrar se tiene un rendimiento más bajo de lo esperado.

Repositorio:

<https://github.com/OscarBosigas/Proyecto-Final-Ciencia-de-Datos-Aplicada>