**Título del Proyecto:**  
**Desarrollo y despliegue de un Modelo analítico para seguros agrícolas indexados en café (Boyacá – Quindío)**

**Autores:**  
Diego Dayan Niño Pérez

Lizeth Daniela Ortiz Perdomo

Camilo Andrés Flórez Esquivel  
Miguel Mateo Sandoval Torres

**Grupo:** DSA – G23  
**Docente:** Karen  
**Fecha:** Octubre 2025

**Problema y contexto**

La caficultura colombiana constituye uno de los pilares económicos y culturales más importantes del país, con más de 550 000 familias cafeteras y una destacada participación en el PIB agrícola nacional. Sin embargo, su sostenibilidad enfrenta una amenaza creciente: la variabilidad climática provocada por los fenómenos El Niño y La Niña, así como por las alteraciones en los patrones de lluvia, temperatura y humedad.

Estos fenómenos generan floraciones irregulares, pérdidas de cosecha y una marcada volatilidad en los rendimientos, lo que se traduce en inestabilidad de ingresos para los productores. Entre 2008 y 2013, por ejemplo, la producción nacional de café cayó cerca de un 33 % por efectos climáticos adversos. Estudios recientes advierten que el 80 % de las áreas cafeteras latinoamericanas podrían sufrir pérdidas significativas si las tendencias actuales continúan.

A pesar de esta vulnerabilidad, la mayoría de los pequeños y medianos caficultores no cuentan con herramientas financieras modernas que les permitan protegerse frente a estos riesgos. En este contexto, los seguros agrícolas indexados surgen como una alternativa innovadora: compensan automáticamente al productor cuando variables medibles como la lluvia, la temperatura o el índice de vegetación NDVI superan o descienden de ciertos umbrales críticos.

No obstante, su implementación efectiva requiere modelos analíticos calibrados al contexto local, que integren fuentes de datos climáticas, satelitales y productivas. Actualmente, Colombia carece de un modelo multivariado que relacione de forma precisa los eventos climáticos con las pérdidas de rendimiento a nivel regional.

Por ello, este proyecto busca **construir una base de datos y un modelo predictivo supervisado** que permita cuantificar el riesgo climático en las zonas cafeteras de **Boyacá y Quindío**, regiones con condiciones contrastantes:

* Boyacá: clima más seco y variable, con mayor amplitud térmica.
* Quindío: clima húmedo y más estable, característico del Eje Cafetero.

El análisis comparativo entre ambas zonas permitirá evaluar la portabilidad regional del modelo y validar su potencial como herramienta de apoyo para la gestión del riesgo agrícola, tanto para aseguradoras como para productores.

En síntesis, la problemática combina riesgo climático, incertidumbre económica y brecha tecnológica, y demanda soluciones de analítica aplicada que vinculen datos ambientales y productivos con mecanismos financieros automatizados, alineados con los objetivos del Fondo Nacional del Café y los planes de sostenibilidad del sector agroindustrial.

**Pregunta de negocio y alcance del proyecto**

**Pregunta de negocio**

¿Es posible estimar de forma confiable el riesgo o pérdida esperada del cultivo de café a partir de variables climáticas y satelitales (NDVI), para respaldar el diseño de un seguro agrícola indexado?

**Objetivo general**

Desarrollar un **modelo analítico multivariado** que relacione indicadores climáticos (precipitación, temperatura, humedad) y satelitales (NDVI/EVI) con el rendimiento productivo del café, permitiendo cuantificar la probabilidad de pérdida y generar índices de activación para seguros agrícolas.

# ****Alcance del Proyecto****

El proyecto abarca el **diseño, desarrollo y despliegue** de un modelo analítico multivariado que permita **estimar el riesgo climático y productivo del cultivo de café** como base para la valoración de **seguros agrícolas indexados** en Colombia.

Su desarrollo comprende cinco componentes principales:

### ****1. Integración de datos****

Recolección, limpieza y consolidación de datos históricos de **fuentes oficiales**:

* **IDEAM:** variables climáticas (precipitación, temperatura, humedad).
* **MODIS:** índices satelitales NDVI y EVI.
* **FNC:** rendimiento, producción y área cultivada.
* **DANE:** contexto económico y censos agropecuarios.

Los datos serán versionados con **DVC** para garantizar trazabilidad y reproducibilidad.

### ****2. Análisis exploratorio****

Identificación de patrones y correlaciones entre variables climáticas, satelitales y productivas.  
Comparación entre **Boyacá (zona seca)** y **Quindío (zona húmeda)** para validar la portabilidad regional del modelo.

### ****3. Modelado predictivo****

Entrenamiento de un **modelo supervisado multivariado** (p. ej., Random Forest o Gradient Boosting) que relacione las condiciones climáticas con los rendimientos del café.  
Evaluación mediante métricas como **MAE**, **RMSE** y **R²** para estimar la pérdida esperada y definir un **índice de riesgo climático**.

### ****4. Empaquetamiento y despliegue****

Empaquetamiento del modelo entrenado, desarrollo de una **API** para consultas y un **tablero interactivo** (Streamlit/AWS) para visualizar resultados y métricas.  
Todo el código será gestionado en **GitHub** y los datos en **DVC**.

### ****5. Validación y documentación****

Elaboración de un **informe técnico** con resultados, métricas y recomendaciones para su aplicación práctica en aseguradoras y cooperativas agrícolas.

**Descripción de los conjuntos de datos a emplear**

El desarrollo del modelo analítico requiere integrar diversas fuentes de información que reflejen las condiciones **climáticas, satelitales, productivas y económicas** del sector cafetero colombiano.  
Los datos seleccionados garantizan cobertura temporal (2010-2025), espacial (departamental y municipal) y temática suficiente para modelar la relación entre **clima – vegetación – rendimiento**.

A continuación, se describen los principales conjuntos de datos:

| **Fuente / Institución** | **Tipo de información** | **Variables clave** | **Cobertura y frecuencia** | **Formato / Acceso** | **Uso principal en el modelo** |
| --- | --- | --- | --- | --- | --- |
| **IDEAM** (Instituto de Hidrología, Meteorología y Estudios Ambientales) | Climática | Precipitación, temperatura media, humedad relativa, brillo solar, evapotranspiración | Nacional (departamental / estaciones). Periodicidad diaria y mensual. | CSV, XLS. Descarga pública IDEAM y repositorios institucionales. | Variable explicativa principal para evaluar la influencia del clima en la producción. |
| **MODIS – NASA / Google Earth Engine** | Satelital (índices de vegetación) | NDVI (Normalized Difference Vegetation Index), EVI (Enhanced Vegetation Index) | Cobertura nacional con resolución espacial de 250 m. Series mensuales 2015–2025. | CSV exportado desde Google Earth Engine. | Indicadores de vigor vegetal y respuesta del cultivo a la variabilidad climática. |
| **FNC – Federación Nacional de Cafeteros** | Productiva / Agrícola | Producción total, rendimiento por hectárea, área cultivada, precios internos del café. | Desagregada por departamento y año (2010–2025). | XLS o CSV. Solicitud institucional / bases históricas públicas. | Variable objetivo (target) del modelo predictivo y base para cálculo de pérdidas. |
| **DANE – Departamento Administrativo Nacional de Estadística** | Económica / Agropecuaria | PIB agropecuario, área sembrada, censos agrícolas, exportaciones. | Regional / nacional, anual. | XLS / API DANE. | Variables de contexto y control para la calibración económica del modelo. |

**Tratamiento general de los datos**

* **Unificación temporal y espacial:** todos los registros se agregan a nivel mensual por departamento o municipio, usando un identificador geográfico común.
* **Limpieza y normalización:** eliminación de duplicados, control de valores atípicos, interpolación de faltantes y conversión de unidades.
* **Control de versiones:** los conjuntos se almacenan bajo un esquema de **versionamiento con DVC**, lo que permite rastrear modificaciones y garantizar la reproducibilidad del pipeline.
* **Estructura final del dataset:**  
  Cada fila representa una observación mensual por región, con columnas para variables climáticas, índices NDVI/EVI y rendimiento productivo del café.