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

**Autores:**  
Camilo Andrés Flórez Esquivel

Diego Dayan Niño Pérez

Lizeth Daniela Ortiz Perdomo

Miguel Mateo Sandoval Torres

**Grupo:** DSA – G23  
**Docente:** Juan Fernando Pérez  
**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 climáticas 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.**

Esta herramienta 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 nivel de riesgo o pérdida esperada de un 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 predecir 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 del departamento de Boyacá, Colombia.

Su desarrollo comprende cinco componentes principales:

### ****1. Integración de datos****

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** con una granularidad mensual.

| **Variable** | **Descripción** | **Uso** | **Fuente** |
| --- | --- | --- | --- |
| Precipitación | suma de precipitaciones acumuladas en mm | Permiten la identificación de exceso / déficit hídrico y cálculo de índices SPI/SPEI | IDEAM |
| Temperatura máxima | Valores máximos de temperatura en °C | Identificación de oleadas de calor y periodos de sequía críticos | IDEAM |
| Temperatura mínima | Valores mínimos de temperatura en °C | Identificación de enfriamientos extremos y periodos de heladas | IDEAM |
| Humedad relativa | Promedio de humedad en % | Calibración de índices de estrés hídrico y correlación con rendimiento de cultivo | IDEAM |
| NDVI (Normalized Difference Vegetation Index) | Estimación de la biomasa y vigor de la vegetación. [-1,1] | Apoyo para el rendimiento y condición de cultivo | NASA |
| EVI (Enhanced Vegetation Index) | Variante del NDVI | Seguimiento a la productividad de cultivo y calibración del riesgo base | NASA |
| Rendimiento | Valor promedio de producción por área cultivada en t/ha | Variable dependiente para modelación de la pérdida | FNC |
| Producción | Producción total en toneladas | Referencia económica | FNC |
| Área cultivada | Superficie total de café cultivado y/o en producción en ha | Normalización y cálculo de pérdidas | EVA |
| PIB agropecuario departamental | Indicador departamental del sector agrícola en MM $ | Indicador de referencia para cálculo de impacto económico del seguro | DANE |
| Costo fijo de cultivo | Costos asociados a cultivo en $ / t | Pertinente para el cálculo del costo de aseguramiento relativo | DANE |
| Índice de precios del café | Variación del precio interno del café en % | Útil para conversión de pérdidas físicas a económicas. | DANE |
| Productores | Censo Nacional Agropecuario, productores, tipo de productores, tamaños de cultivo. | Caracterización socioeconómica de los potenciales beneficiarios | FNC DANE |

\* IDEAM. Instituto de hidrología, meteorología y estudios ambientales

\* NASA. National Aeronautics and Space Administration

\* FNC. Fondo Nacional de Cafeteros

\* EVA. Evaluación Agropecuaria Municipal

\* DANE. Departamento Administrativo Nacional de Estadísticas.

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.

**Exploración de los datos**

El propósito del análisis está en familiarizarse con los datos, identificar patrones, correlaciones y anomalías, que puedan influir sobre la ejecución del proyecto. Estos datos dentro de las posibilidades han sido ajustados al mismo nivel de granularidad establecida (series mensuales en un periodo entre el 2010 y 2025) de tal manera que se pueda manejar uniformidad y correlación entre los mismos.

**Variables climatológicas.**

Se emplean datos provenientes de las estaciones metereológicas del IDEAM que se encuentran circundantes a la región cafetera en Boyacá, caracterizada por 11 municipios, se identifican un total de 129 estaciones útiles que se encuentran clasificadas de la siguiente manera:

|  |  |
| --- | --- |
| **Categoría** | **Cantidad** |
| Pluviométrica | 46 |
| Climatológica Principal | 27 |
| Limnimétrica | 16 |
| Climática Ordinaria | 13 |
| Limnigráfica | 10 |
| Pluviográfica | 9 |
| Metereológica Especial | 6 |
| Agrometereológica | 1 |
| Sinóptica principal | 1 |

* Precipitación. La región cafetera de Boyacá presenta precipitaciones promedio mensuales de 155 mm/mes. Suelen presentar un comportamiento climatológico bimodal, es decir dos picos estacionales de altas precipitaciones (abril-mayo y octubre-noviembre) intercalado con periodos secos (enero-febrero y mayo-agosto).
* Temperatura máxima. La región de estudio cuenta con una temperatura máxima promedio de 24.8 °C (± 2.0 °C).
* Temperatura mínima. La región de estudio cuenta con temperaturas mínimas promedio de 14.2° C (± 1.8 °C). Principalmente marcadas entre julio y septiembre. Lo que puede representar un riesgo de heladas en las zonas de mayores elevaciones.
* Humedad relativa. Se presenta una humedad relativa de 78% (± 9%). Con un comportamiento inversamente proporcional a la temperatura máxima.

**Índices de vegetación.**

Se extrajeron las series temporales OD13Q1 (250 m) para zonas cafeteras de Boyacá, analizadas a escala mensual, mediante promedios compuestos de 16 días.

**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.