

# Pronóstico de pluviosidad y clasificación de riesgo de pérdida de producción en los municipios de Colombia

**Carlos Andrés Rocha Ruiz** *Contratista - Departamento Nacional de Planeación*

**Eduardo Escobar Gutiérrez** *Contratista - Departamento Nacional de Planeación*

**Mariana Quevedo Hernández** *Pasante - Departamento Nacional de Planeación*

---

## Resumen

La gran variabilidad del clima en los diferentes territorios nacionales y las consecuencias que este puede tener en la producción agrícola hace necesaria la implementación de diferentes instrumentos financieros, como seguros, que mitiguen el riesgo de pérdida de producción a causa de cambios en los niveles de lluvia o pluviosidad. Este proyecto busca contribuir en la reducción de la afectación en la producción agrícola a partir de señales y alertas de riesgo de pérdida de producción por lluvias sobre los territorios que requieran cubrirse con instrumentos financieros. Por tal motivo, se realiza el análisis de series históricas e información de pluviosidad de las diferentes estaciones meteorológicas instaladas alrededor del territorio nacional, permitiendo, a través de un modelo, pronosticar el clima en los territorios de interés para luego clasificar dichos territorios en función del riesgo de pérdida a causa del clima en el periodo de tiempo pronosticado. De tal modo, este trabajo es un insumo clave para generar alertas en los territorios y poder cubrir con anticipación el riesgo de pérdida de producción a causa del clima.

Para el pronóstico de los niveles de lluvia en los centros de producción agrícola, se tomó información triangulada a las cabeceras municipales a partir de los datos reportados por las estaciones meteorológicas. Una vez triangulada la información, se utilizaron metodologías de aprendizaje de máquina para la creación de una herramienta útil que permita pronosticar los niveles de lluvia en las diferentes cabeceras municipales.

Para la medición del nivel de riesgo de pérdida de producción, se tomaron datos de producción agrícola para cada grupo de cultivos producidos en Colombia de la Evaluación Agropecuaria Municipal (EVA) y el Censo Nacional Agrícola (CNA) y con metodologías de inteligencia artificial y aprendizaje automático, generar una herramienta que permita calcular la probabilidad de pérdida de producción a causa del clima.

*Keywords:* r, shiny, markdown, aprendizaje automático, inteligencia artificial, clasificador

---

## Introducción

La producción agrícola representa una parte importante la economía de Colombia y la variabilidad del clima dentro del territorio nacional tiene importantes consecuencias para la esta, razón por la cual el cambio en los niveles de pluviosidad puede representar un riesgo de pérdida de producción para los diferentes cultivos. Por esta razón es importante establecer medidas de prevención o cubrimiento a este riesgo, para lo cual es necesario tener datos de los niveles de pluviosidad dentro de los diferentes territorios del país donde se presenta producción agrícola y así mismo tomar los datos de producción para todos los grupos de cultivo estableciendo un cálculo de la probabilidad de pérdida de producción como consecuencia del clima.

El objetivo de este proyecto es contribuir en la reducción del riesgo de pérdida de producción a la que están expuestos los diferentes cultivos y municipios del país, por medio de señales de alerta y una clasificación del riesgo a causa de los cambios en los niveles de pluviosidad para lograr mitigar estas adversidades a través de instrumentos financieros como seguros.

A continuación, aparece la metodología utilizada para la realización del proyecto, así como los resultados obtenidos.

## Metodología

El proyecto está dividido en dos etapas, la primera corresponde al pronóstico de lluvias y la segunda a la medición del nivel de riesgo de pérdida de producción agrícola.

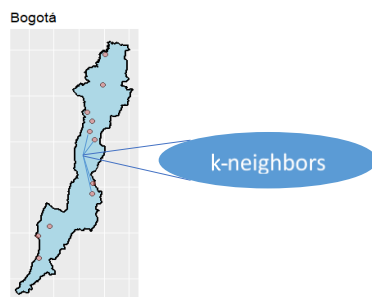
Para el pronóstico de lluvias se utilizaron los datos de pluviosidad recolectados por las diferentes estaciones meteorológicas instaladas en el país que permiten tener un registro de la cantidad de lluvias acumulada por metro cuadrado en cierto momento del tiempo. Sin embargo, estas estaciones meteorológicas no necesariamente se encuentran ubicadas en un territorio de alta producción agrícola. Por tal motivo, se requiere triangular la información reportada las diferentes estaciones meteorológicas en un determinado momento del tiempo en aquellos territorios de interés donde se concentre una alta cantidad de territorios productivos.

En un principio, la información recolectada por las estaciones meteorológicas se encontraba en un formato de texto con unos patrones de procesamiento singulares, de tal forma, se programó un algoritmo que permitió limpiar y procesar los datos de manera automática por medio de una función con la cual podía organizarse la información de cualquier archivo que cumpliera con la adecuada estructura de información.

De igual manera, existe un porcentaje general en las bases correspondiente a valores faltantes, esto debido a que la estación meteorológica instalada reporta un valor, pero existen ocasiones en que no reporta valor alguno, ya sea por problemas en ese momento en la infraestructura o en la estación meteorológica. Por tal motivo, se utilizó una metodología conocida como *Predictive mean matching* que consiste en, a partir de distintas iteraciones, encontrar el valor más probable de acuerdo con la distribución propia de los datos. La figura 1 muestra los resultados de la distribución estimada para los datos reales frente a los datos imputados de los meses con menor ajuste (marzo) y mayor ajuste (noviembre).

La distribución de las lluvias por mes se centra en una menor cantidad de pluviosidad relativa. De igual manera, las distribuciones de los datos imputados y de los datos reales dentro de la serie, por mes, son bastante parecidas. Al imputar los datos únicamente comparando los mismos meses a través de los años, se garantiza tener en cuenta el componente estacional marcado en el clima, imputando de acuerdo con la distribución de meses con mayores niveles de lluvia, sin interferencia de meses con bajos niveles de lluvia y viceversa.

Así mismo, las diferentes cabeceras municipales concentran a su alrededor un alto nivel de territorios productivos medido por UPA (Unidad de Producción Agropecuaria), Por consiguiente, la triangulación de la información reportada en las estaciones meteorológicas se realizó en el centroide del polígono de dichos territorios por medio de la metodología de *cluster* conocida como *k-neighbors* para unos puntos dados. Este algoritmo identifica la cercanía de unos puntos (estaciones meteorológicas) frente a otros puntos en el plano (centroide del polígono del municipio) y por medio de una función, se definen los pesos que cada estación meteorológica, aporta al punto a triangular como se muestra en el mapa 1.



Mapa 1: Triangulación de los puntos de interés a partir de información de las estaciones meteorológicas

## Resultados

Una vez consolidada la información, se implementaron diferentes metodologías de pronóstico de series de tiempo que permitan capturar la volatilidad estocástica del proceso. Por una parte, se utilizaron modelos ARIMA convencionales con resultados regulares, pero con dificultad de procesamiento. Por otro lado, se emplearon enfoques de aprendizaje supervisado para el pronóstico del clima a un periodo adelante, para ello, se han utilizado bosques de decisión, máquinas de soporte vectorial y modelos de ensamble boosting.

Tabla 1: Comparativa en el error de pronóstico al cuadrado de los modelos

Modelo	Media MSE	Desviación estándar MSE
ARIMA	757,5035	1397,9938
Random Forest	582,7345	933,9274
<b>SVM</b>	<b>465,5302</b>	<b>744,2542</b>
Boosting	608,8974	923,0372

La tabla 1 muestra la comparativa entre los modelos, para la media y la desviación estándar del error de pronóstico de todos los modelos calculados de la misma clase para todos los municipios. Entre menor error de pronóstico, el modelo tendrá un mejor desempeño y con una desviación estándar baja, el error no tenderá a ser muy variable. El mejor modelo es la máquina de soporte vectorial con un menor error de pronóstico y una menor desviación estándar, lo cual implica que este modelo es el mejor de los modelos explorados para la estimación del clima en un principio.

Dentro del motor de pronóstico, se utilizaron modelos de umbrales (*TAR models*) para lograr hacer pronósticos a una ventana mayor de tiempo, capturando la variabilidad a largo plazo de fenómenos de cambio climático en determinado periodo del tiempo (como el fenómeno del niño o la niña) para así ensamblar un modelo con máquinas de soporte vectorial (SVM) para hacer pronósticos con ventanas de tiempo más lejanas.

Así mismo, se generó una herramienta que permite visualizar los resultados de las triangulaciones y pronósticos de lluvias en las diferentes cabeceras municipales a partir de mapas y series de tiempo. La herramienta consta de dos mapas, el primero muestra la ubicación de las diferentes cabeceras municipales y las estaciones meteorológicas alrededor del territorio nacional. La figura 1 muestra la visualización del primer mapa, en azul las cabeceras municipales y en rojo las estaciones meteorológicas.

El segundo mapa por su parte, muestra la cantidad de lluvia por metro cuadrado en forma de círculos a través del tiempo en cada cabecera municipal, dicho mapa cuenta con un círculo cuyo radio identifica la cantidad de lluvia. Adicional cuenta con una barra de tiempo, la cual permite visualizar el comportamiento a través del tiempo. La figura 2 muestra la visualización del segundo mapa.

Adicional, la herramienta cuenta con un visor de serie de tiempo el cual tiene las 1120 cabeceras municipales en el territorio continental del país con su respectiva serie de tiempo de lluvias y su pronóstico. La figura 3 muestra dicha gráfica donde la parte azul es el dato triangulado real y la parte naranja es la parte pronosticada.

Y, por último, la herramienta también cuenta con una visualización de los municipios cuyo promedio de lluvias en un mes específico fue el más alto en todo el país, la figura 4 muestra la visualización de dicha tabla.

Figura 1. Mapa de ubicación de estaciones meteorológicas y cabeceras municipales

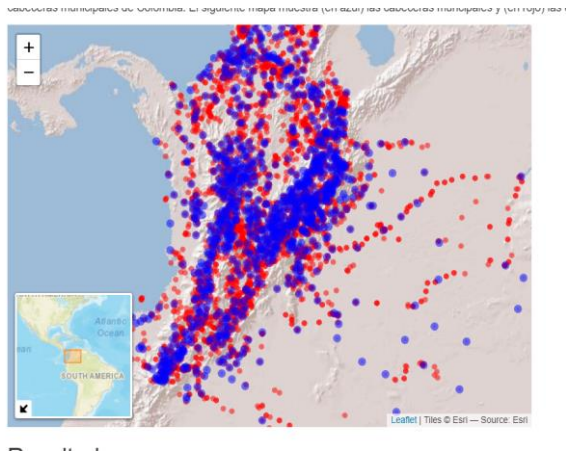


Figura 2. Mapa de triangulación de lluvias y pronósticos en las diferentes cabeceras municipales



Figura 3. Serie de tiempo de la herramienta de pronósticos de lluvias

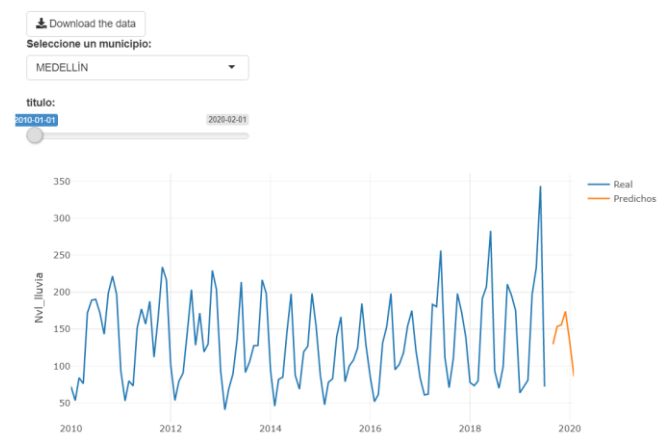


Figura 4. Visualización de municipios con mayores lluvias promedios en un periodo de tiempo determinado (2010-01)

Show	25	entries	Search:	
Nvl_lluvia	municipio	periodos		
1180.6301	GUACHENÉ	2010-01-01		
1124.3082	LÓPEZ	2010-01-01		
1045.3033	EL CARMEN	2010-01-01		
996.1492	SAMANA	2010-01-01		
935.6013	MORALES	2010-01-01		
843.1568	JURADÓ	2010-01-01		
808.7943	EL LITORAL DEL SAN JUAN (Dacordó)	2010-01-01		
731.2191	PADILLA	2010-01-01		