

Dirección de Desarrollo Digital

Unidad de Científicos
de Datos



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación

PRONÓSTICO DEL CLIMA PARA LA IMPLEMENTACIÓN DE SEGUROS A LA PRODUCCIÓN AGRÍCOLA A NIVEL TERRITORIAL

Entidad

Departamento Nacional de Planeación

- Dirección de Desarrollo Digital.
- Dirección de Desarrollo Rural Sostenible.

Sector

Planeación

Lenguaje

R.

Fuente de datos

IDEAM

Presentación

La gran variabilidad del clima en los diferentes territorios nacionales y las diferentes consecuencias que este puede tener en la producción agrícola hace necesario 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 generar 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 al 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.

High level of volatility in weather around several national territories comes with consequences that this may have in farming production makes necessary implement several financial instruments, as insurances, in order to mitigate the losing production risk due changes in raining level and weather conditions. This project seeks contribute in reduction of the affectation in farming production by generating signals and alerts about the risk of losing productions for rainfall growth in national territories that require cover with financial instruments.

For this reason, the use and analyse historial series of rainfall among several weather stations installed around the national territories, allows through a model to, forecast the weather in interest territories to then classify them territories by those with lose risk of production for weather change in a specific period of time. On that way, that allows generate alarms in territories with higher risk and cover them in advance the risk of losing production for weather change.

Objetivo general

Generar alarmas que permitan prevenir y mitigar posibles riesgos de pérdida o de sufrir efectos adversos en la producción agrícola en los territorios a causa de la variabilidad del clima.

Objetivos específicos

1. Identificar, imputar y triangular la información reportada por las estaciones meteorológica ubicadas en diferentes puntos en el territorio nacional con respecto a las cabeceras municipales identificadas como centros de producción agrícola.



2. Generar un motor de pronóstico que se adapte a las series de tiempo del clima en las diferentes cabeceras municipales y haga predicciones un periodo de tiempo determinado de acuerdo con la necesidad.

Metodología

Los datos de pluviosidad recolectados por las diferentes estaciones meteorológicas instaladas en el país permiten tener un registro de la cantidad de lluvia acumulada por metro cuadrado en diferente momento del tiempo. La información de dichas estaciones se encuentra disponible desde el año 1900 hasta el año 2019. 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. Como lo muestra la Figura 1, el porcentaje de valores faltantes en el agregado por meses de los datos reportados por la estación meteorológica oscila entre el 2% y el 6%, lo que implica alrededor de entre 1 a 3 años de información faltante por mes.

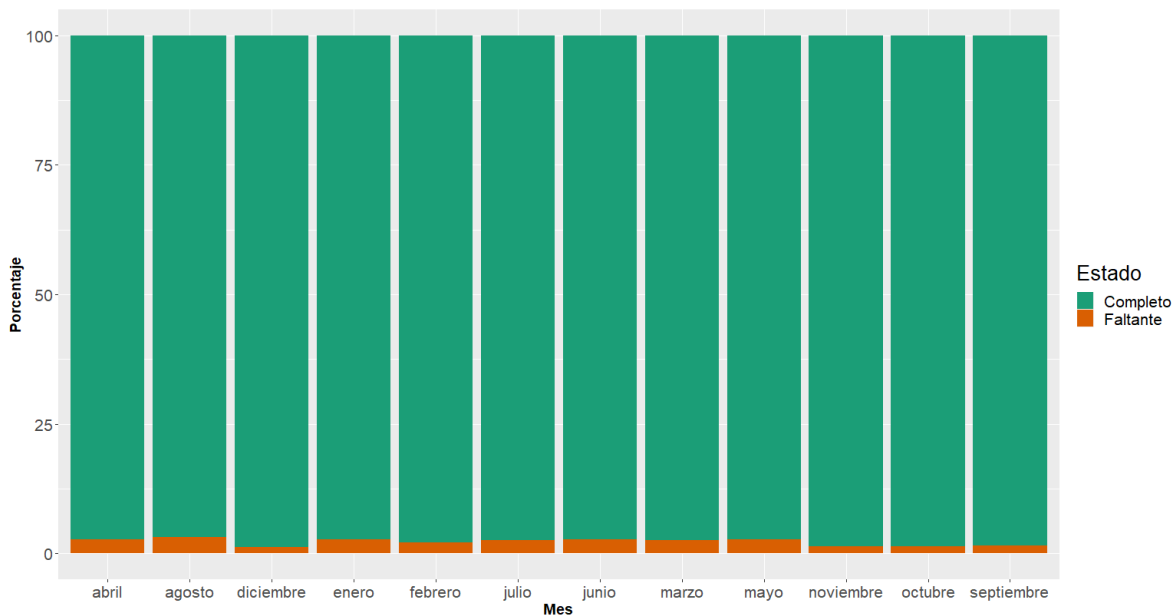


Figura 1: Porcentaje de datos faltantes (agregado de meses en el periodo 1980 – 2019)



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 2 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).

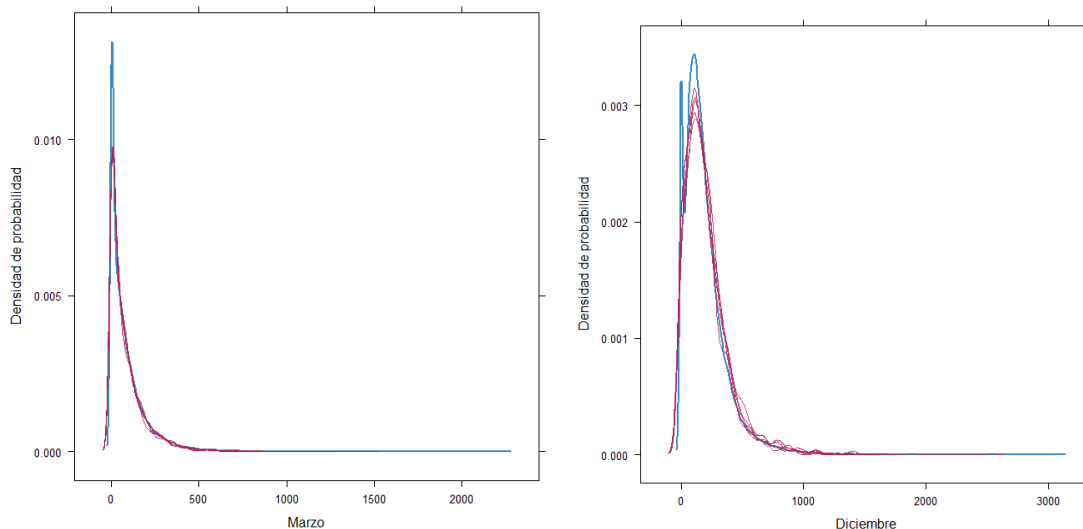
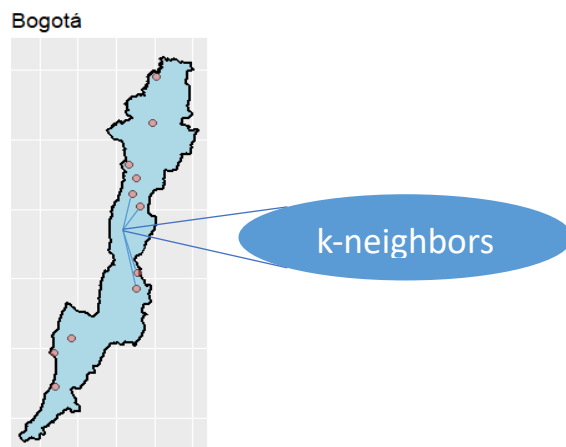


Figura 2: Distribución de datos reales (azul) frente imputados (rojo) de marzo y noviembre

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

Igualmente, con el algoritmo se calculan unos pesos que “aportan” a partir de una función *gaussiana*, definida en la ecuación 1, al nuevo valor triangulado en el mapa.

$$D(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \quad (1)$$

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 (Figura 3) 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 (Figura 4), máquinas de soporte vectorial (Figura 5) y modelos de ensamble boosting (Figura 6).

Resultados

Para la estimación de los modelos, se entrenó uno por cada cabecera municipal con diferentes metodologías. Los resultados expuestos en las gráficas corresponden al pronóstico de los diferentes modelos en los municipios que registran mayor y menor cantidad de lluvia promedio, para la comparación de los modelos, se compararon los pronósticos en el periodo 1980-2016 debido a que en el 2016, Colombia fue afectada por el fenómeno del niño, además dichas regiones se eligieron en principio, por que presentan periodos de lluvias altas con mayor frecuencia tienen un componente aleatorio que es más complicado de pronosticar. Así mismo, las regiones que presentan periodos de lluvias bajas tienden a ser relativamente más sencillo de pronosticar por su baja volatilidad.



El futuro es de todos

DNP Departamento Nacional de Planeación

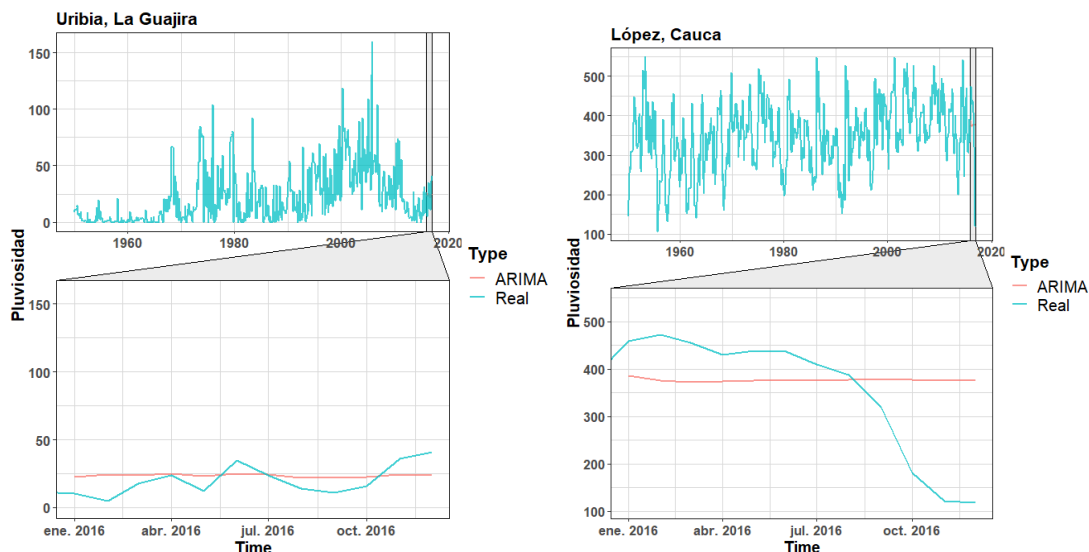


Figura 3: Comparativa de los modelos ARIMA en la estimación en las cabeceras con menor (Uribia) y mayor (López) cantidad de lluvia promedio en Colombia

Los pronósticos realizados por los modelos ARIMA no tienen los resultados esperados, el pronóstico se está centrando en la media, mas no logra capturar los cambios de nivel, ni los choques, ni la volatilidad dentro de la serie. Los municipios tomados, corresponden a aquellos con menor cantidad de lluvia promedio frente al municipio con mayor cantidad de lluvia promedio. En ambos casos, el pronóstico del modelo ARIMA se centra únicamente en la media.

Hay que tener en cuenta que, para la estimación de más de un periodo, se utilizan valores predichos previos, lo que conlleva a un aumento en el error en periodos lejanos a los datos reales. Sin embargo, de acuerdo con la necesidad, en un principio no es necesario pronosticar más periodos al inmediatamente posterior.



El futuro es de todos

DNP
Departamento
Nacional de Planeación

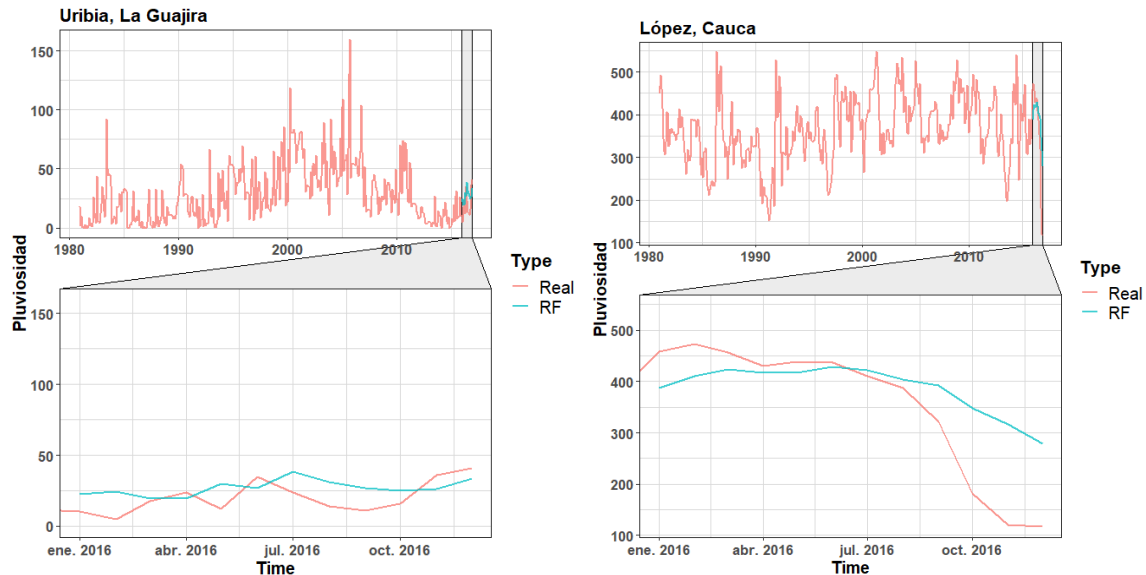


Figura 4: Comparativa del modelo supervisados de bosques de decisión en la estimación en las cabeceras con menor (Uribia) y mayor (López) cantidad de lluvia promedio en Colombia

Por su lado, los bosques de decisión tienen un mejor rendimiento que los modelos ARIMA, sin embargo, los cambios de nivel no los captura de manera satisfactoria (como pasa con el municipio de López, Cauca). Sin embargo, el costo computacional del entrenamiento de este modelo es significativamente inferior al modelo ARIMA, lo cual implica a una ganancia sustancial.

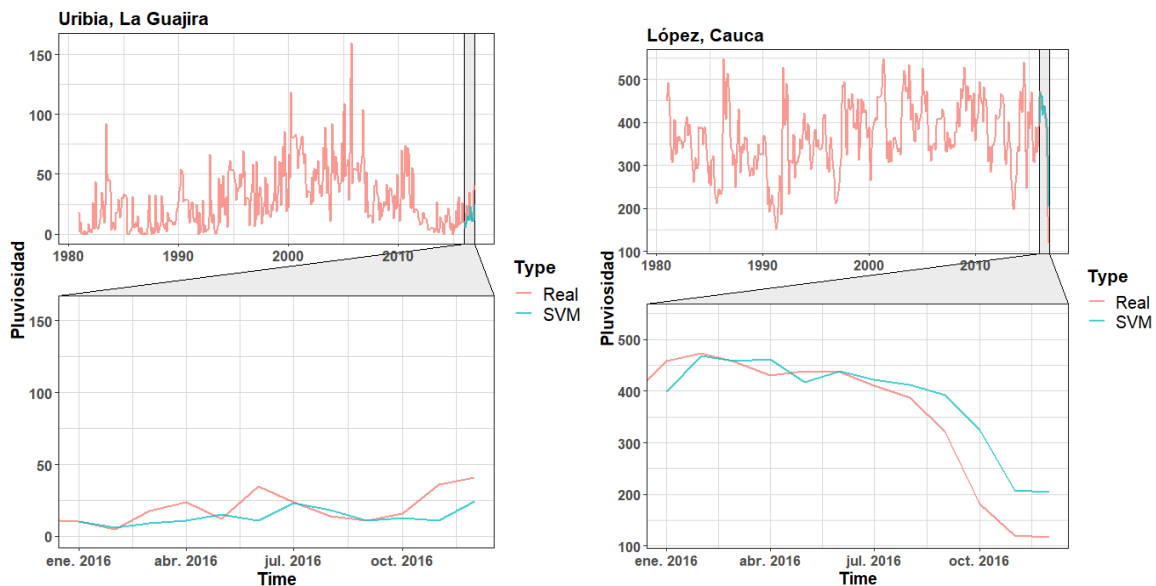


Figura 5: Comparativa del modelo supervisados de máquinas de soporte vectorial en la estimación en las cabeceras con menor (Uribia) y mayor (López) cantidad de lluvia promedio en Colombia



Por su parte, las máquinas de soporte vectorial con un *kernel* lineal tienen una ganancia sustancial con un menor error de pronóstico y menor volatilidad frente a los dos anteriores. Sin embargo, el costo computacional es superior al *Random Forest*, pero sin superar al ARIMA. A pesar de ello, dicho costo computacional adicional es tolerable, teniendo en cuenta las diferencias significativas en la calidad del pronóstico de un modelo a otro.

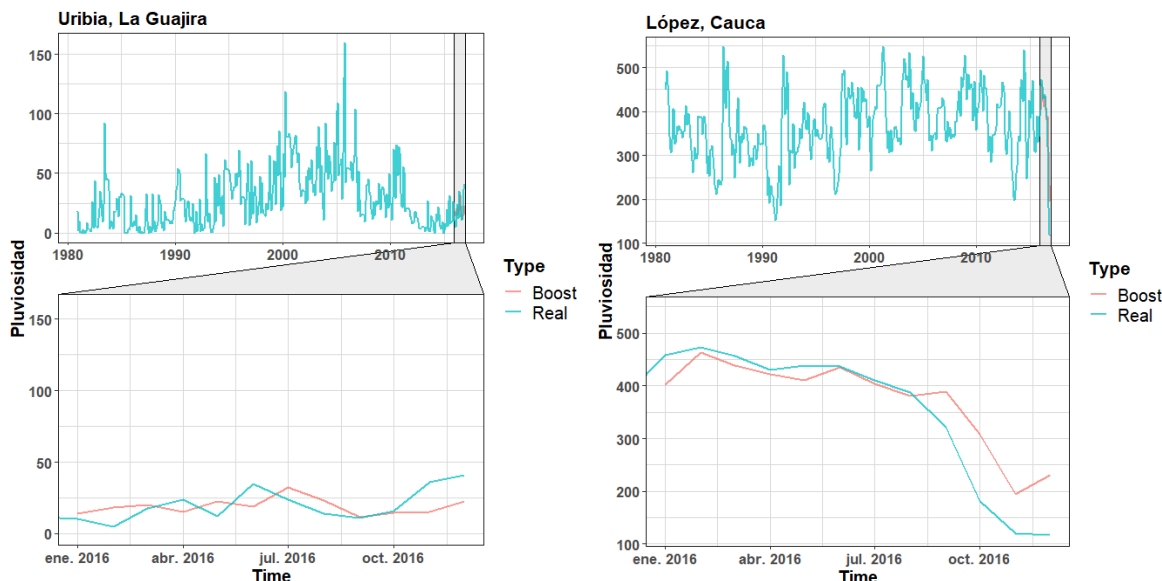


Figura 6: Comparativa del modelo supervisado de boosting en la estimación en las cabeceras con menor (Uribe) y mayor (López) cantidad de lluvia promedio en Colombia

Por último, el modelo de *boosting* se ajusta mejor, pero tiene mayor varianza en su pronóstico, lo cual hace difícil saber cuál modelo tiene un mayor rendimiento y cuál no. Adicionalmente, este modelo demanda un alto nivel de recursos computacionales, por lo que no es práctico su entrenamiento y (dependiendo la situación) puede que no sea óptimo. Por tal motivo, se calculó el error de pronóstico promedio de todos los modelos estimados de cada cabecera municipal para cada tipo de modelo.

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

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

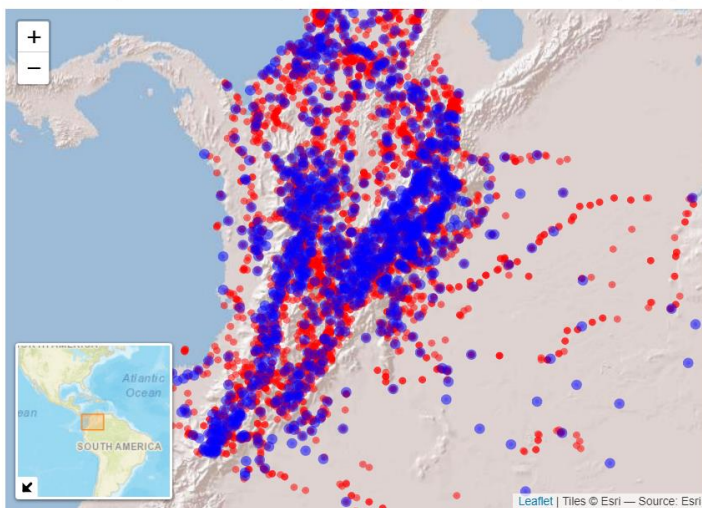
La tabla 1 muestra la comparativa entre los modelos, para la media y la desviación estándar del error de pronóstico de todos los modelos calculados de la misma clase para todos los municipios. Entre menor error de pronóstico, el modelo tendrá un mejor desempeño y con una desviación estándar baja, el 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.



De igual manera, dentro del motor de pronóstico, se utilizó modelos de umbrales (*TAR models*) lograr hacer pronósticos a una ventana mayor de tiempo cuyo objetivo es capturar 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 maquinas de soporte vectorial (SVM) para hacer pronósticos con ventanas de tiempo más lejanas. El motor de pronóstico permite hacer predicciones de 1 a 12 meses de acuerdo con cada necesidad (cabe resaltar que entre mayor sea la ventana de pronóstico, el error de pronóstico se va a acumular por lo que la precisión se podría ver afectada a largo plazo en el pronóstico)

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, dicho mapa es con el objetivo de visualizar y tener en cuenta dichas cabeceras municipales que pueden contar con algún tipo de problemas en cuanto a su triangulación debido a la lejanía del punto en el mapa con respecto a la ubicación de las estaciones meteorológicas más cercanas. La figura 7 muestra la visualización del primer mapa, en azul las cabeceras municipales y en rojo las estaciones meteorológicas.

cabeceras municipales de Colombia. El siguiente mapa muestra (en azul) las cabeceras municipales y (en rojo) las estaciones meteorológicas.



Resultados

Figura 7. Mapa de ubicación de estaciones meteorológicas y cabeceras municipales

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, es decir que a mayor radio en el círculo, mayor es la cantidad de lluvia triangulada en x periodo de tiempo determinado. Adicional cuenta con una barra de tiempo, la cual permite visualizar el comportamiento a través del tiempo. La figura 8 muestra la visualización del segundo mapa.



Resultados

El siguiente mapa muestra los datos triangulados y pronósticos de los siguientes tres meses en los diferentes municipios de Colombia.

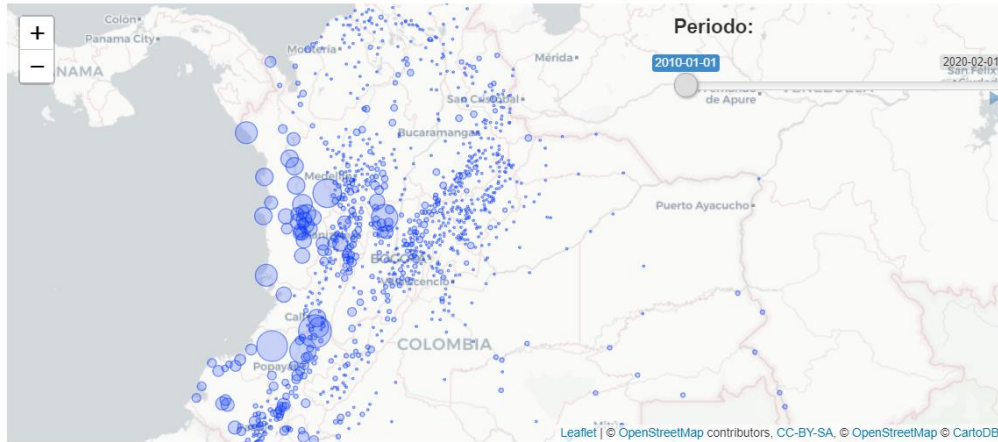


Figura 8. Mapa de triangulación de lluvias y pronósticos en las diferentes cabeceras municipales

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 9 muestra dicha gráfica donde la parte azul es el dato triangulado real y la parte naranja es la parte pronosticada.

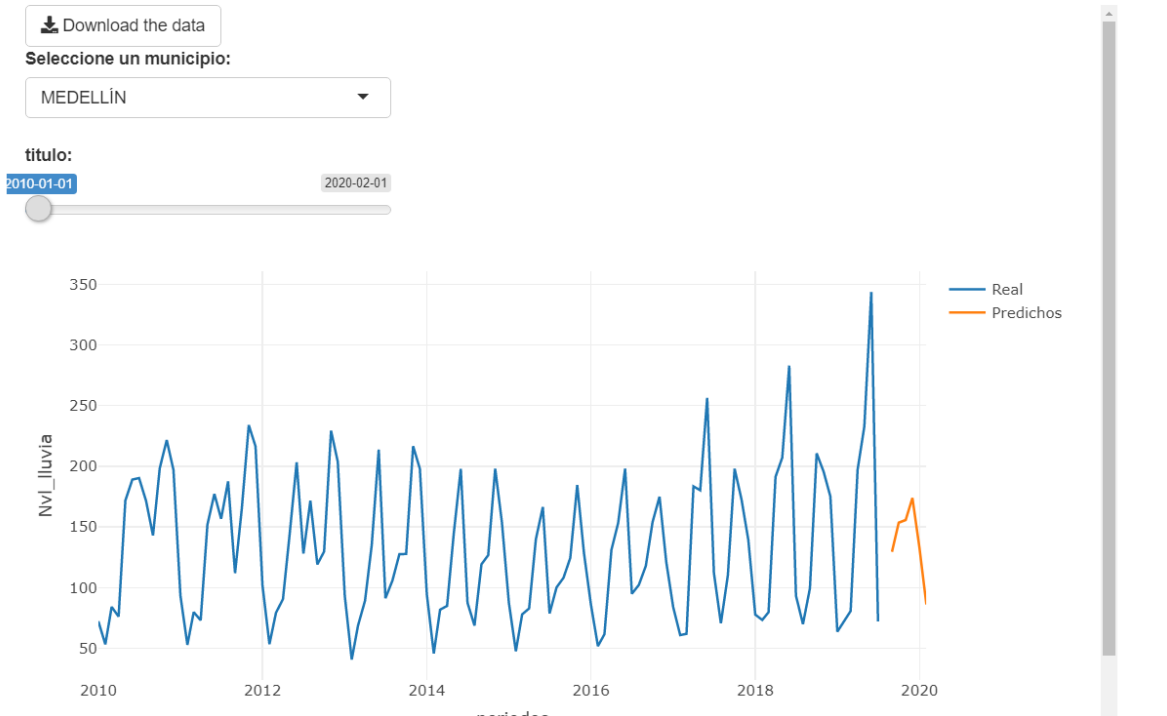


Figura 9. Serie de tiempo de la herramienta de pronósticos de lluvias



El futuro es de todos

DNP Departamento Nacional de Planeación

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 10 muestra la visualización de dicha tabla.

Show 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 JUÁN (Docordó)	2010-01-01
731.2191	PADILLA	2010-01-01

Figura 10. Visualización de municipios con mayores lluvias promedios en un periodo de tiempo determinado (2010-01)

Por último, se habilitó también un botón de descarga que permite descargar todos los datos triangulados y pronosticados para todas las cabeceras municipales en el territorio nacional.

Esta herramienta sirve como insumo para hacer una clasificación de riesgo de pérdida de producción a causa del clima con el fin de implementar seguros territoriales a la producción y mitigar el riesgo de pérdida.

Conclusiones

1. Se generó un desarrollo que procesa y triangula la información de las estaciones meteorológicas a las cabeceras municipales (o a cualquier serie de puntos en el mapa que se le suministra).
2. Se generó un motor de pronóstico con modelos de ensamble que permite predecir el nivel de lluvia promedio a nivel mensual para cada cabecera municipal.
3. Se generó una herramienta de visualización que muestra en mapas y series de tiempo el comportamiento de las lluvias en las diferentes cabeceras municipales.
4. Esta herramienta sirve como insumo para clasificar los municipios por riesgo de pérdida de producción a causa de cambios en los niveles de lluvia.

Socialización

Se hizo la socialización con la DDRS