

Universidad Internacional de La Rioja

**Escuela Superior de Ingeniería y
Tecnología**

**Máster Universitario en Análisis y Visualización
de Datos Masivos**

Predicción niveles de precipitación para la identificación de zonas agrícolas

Trabajo Fin de Máster

Tipo de trabajo: Piloto experimental

Presentado por: Guerrero Álvarez, Jorge Alejandro

Director: Cortés Cabezas, Albeiro

Resumen

Las condiciones climáticas representan un factor primordial en el desarrollo de los cultivos, fenómenos climáticos como lluvias intensas y sequías simbolizan una amenaza a la producción agrícola. La dificultad para estimar los niveles de lluvias de los próximos meses representa una limitante a fin de evaluar si una zona es apropiada para la agricultura, además de aumentar el riesgo de pérdida parcial o total de los cultivos como consecuencia de sequías o inundaciones. En el presente trabajo se plantea la construcción de un modelo predictivo capaz de estimar los niveles de precipitación para los próximos doce meses utilizando el historial climático de la región de Cundinamarca, en Colombia. Posteriormente, analizando las predicciones del modelo se propone la identificación de las zonas y meses propicios para la agricultura. El modelo predictivo utiliza técnicas de aprendizaje automatizado (*machine learning*) que permitirán al modelo aprender de los datos climáticos históricos con el fin de generar predicciones. Los resultados generados por el modelo preservan la tendencia de los niveles de lluvia registrados en los históricos de las zonas examinadas. Finalmente, analizando las predicciones se identificaron las zonas y meses adecuados para actividades agrícolas.

Palabras Clave: precipitación, predicción climática, zonas agrícolas, aprendizaje automatizado, redes neuronales artificiales, big data

Abstract

Weather conditions represent a key element in the development of crops, climatic events such as heavy rains and droughts symbolize a threat to agricultural production. The complexity in forecasting the levels of rainfall for the coming months represents a limitation in order to assess whether an area is suitable for agriculture and additionally rising the risk of partial or total loss of crops as a result of droughts or floods. The aim of this paper is the development of a predictive model capable of forecasting the rainfall levels in the next twelve months using the historical weather data of the Cundinamarca region, in Colombia. Subsequently, analyzing the model forecasting outputs, it is proposed the identification of areas and seasons suitable for agriculture. The predictive model use machine learning techniques that enable the model to learn from the weather historical data in order to generate rainfall forecast. The outputs generated by the predictive model preserve the trend of the rainfall levels recorded in the historical data of the evaluated areas. Finally, reviewing the forecasting data the process was able to locate the areas and months suitable for agricultural activities.

Keywords: rainfall, weather forecast, farming zones, machine learning, artificial neural networks, big data

Índice de contenidos

1. Introducción.....	8
1.1 Justificación	8
1.2 Planteamiento del trabajo	10
1.3 Estructura de la memoria.....	11
2. Contexto y estado del arte.....	12
2.1 Orígenes de la predicción climática	12
2.2 Bases teóricas de la predicción climática.....	13
2.3 Origen predicciones climáticas computarizadas.....	15
2.4 Técnicas de Machine Learning para predicciones climáticas	16
2.5 Predicción climática en la agricultura	20
2.6 Trabajos relacionados.....	22
2.7 Futuro de la predicción climática.....	24
2.8 Conclusiones	25
3. Objetivos concretos y metodología de trabajo	26
3.1. Objetivo general.....	26
3.2. Objetivos específicos	26
3.3. Metodología del trabajo	27
4. Desarrollo específico de la contribución	29
4.1 Fuente de datos	29
4.1.1 Tratamiento de datos	29
4.1.2 Descripción datos	29
4.2 Muestreo	31
4.3 Exploración	34
4.4 Ajustes	38
4.5 Modelamiento.....	41
4.6 Evaluación.....	47

4.6.1 Evaluación fase entrenamiento	47
4.6.2 Evaluación fase predicciones.....	48
4.6.3 Generación predicciones	51
4.7 Identificación zonas agrícolas.....	55
5. Conclusiones y trabajo futuro	65
5.1. Conclusiones	65
5.2. Líneas de trabajo futuro	66
6. Bibliografía	67
Anexos	71
Anexo I. Catálogo estaciones hidrometeorológicas.....	71
Anexo II. Predicciones precipitaciones estaciones hidrometeorológicas Cundinamarca ...	78
Anexo III. Clasificación zonas agrícolas Cundinamarca 2021	93

Índice de tablas

Tabla 1. Fases framework SEMMA	28
Tabla 2. Análisis valores muestra.....	33
Tabla 3. Análisis valores conjunto de datos.....	35
Tabla 4. Estaciones hidrometeorológicas seleccionadas.....	40
Tabla 5. Fases construcción modelo	41
Tabla 6. Evaluación MSE - fase entrenamiento.....	48
Tabla 7. Registro precisión predicciones	51
Tabla 8. Análisis valores predicciones precipitación.....	53
Tabla 9. Clasificación zonas agrícolas 2021.....	59
Tabla 10. Clasificación meses actividades agrícolas 2021	62
Tabla 11. Identificación zonas precipitaciones elevadas - Cundinamarca 2021.....	64

Índice de figuras

Figura 1. Representación SVM (Support Vector Machine) (elaboración propia)	17
Figura 2. Arquitectura redes neuronales artificiales (elaboración propia).....	19
Figura 3. Red neuronal recurrente y representación extendida (elaboración propia)	19
Figura 4. Calendario agrícola Hispanoamérica (Goloubinoff et al., 1997)	21
Figura 5. Mapa probabilístico de predicciones modelo MetNet (Sønderby et al., 2020).....	23
Figura 6. Comparación predicciones y valores reales de niveles de lluvia modelo LSTM (Poornima & Pushpalatha, 2019).....	24
Figura 7. Esquema framework SEMMA (elaboración propia)	27
Figura 8. Contenido archivo precipitaciones mensuales (elaboración propia).....	30
Figura 9. Mapa estaciones hidrometeorológicas (elaboración propia)	31
Figura 10. Archivo consolidado (elaboración propia)	32
Figura 11. Histogramas características climáticas (elaboración propia).....	33
Figura 12. Histograma características climáticas conjunto de datos total (elaboración propia)	34
Figura 13. Diagramas de dispersión características climáticas (elaboración propia)	36
Figura 14. Matriz de correlación y mapa de calor características climáticas (elaboración propia)	37
Figura 15. Conversión formato y estructura archivo precipitación (elaboración propia)	38
Figura 16. Transformación valores atípicos (elaboración propia).....	39
Figura 17. Procedimiento ajuste valores nulos (elaboración propia)	39
Figura 18. Comparativa precisión modelos (elaboración propia)	42
Figura 19. Representación conceptos época y lote (elaboración propia).....	44
Figura 20. Fórmula MSE (Fandango, 2018)	44
Figura 21. Proceso general predicción (elaboración propia).....	46
Figura 22. Escenarios de underfitting, overfitting y óptimo (elaboración propia)	46
Figura 23. Error cuadrático medio - fase entrenamiento (elaboración propia).....	47
Figura 24. Evaluación predicciones precipitación (elaboración propia).....	49

Figura 25. Comparación valores reales y predicciones niveles de precipitación (elaboración propia).....50

Figura 26. Evaluación precisión predicciones - dataset pruebas (elaboración propia)50

Figura 27. Histórico y predicciones de precipitación - estación 2120646 (elaboración propia)52

Figura 28. Histórico y predicciones de precipitación - estación 2306517 (elaboración propia)52

Figura 29. Histórico y predicciones de precipitación - estación 2120557 (elaboración propia)53

Figura 30. Histograma predicciones niveles precipitación (elaboración propia)54

Figura 31. Zonas analizadas Cundinamarca55

Figura 32. Identificación zonas agrícolas - enero 2021 (elaboración propia)56

Figura 33. Identificación zonas agrícolas - febrero 2021 (elaboración propia)57

Figura 34. Identificación zonas agrícolas - marzo 2021 (elaboración propia)58

Figura 35. Clasificación zonas agrícolas Cundinamarca 2021 (elaboración propia)60

Figura 36. Comparación clasificación zonas agrícolas (elaboración propia)61

Figura 37. Clasificación mensual zonas agrícolas Cundinamarca 2021 (elaboración propia)62

Figura 38. Comparación clasificación meses agricultura (elaboración propia).....63

1. Introducción

Desde los inicios de la humanidad el hombre ha intentado comprender los fenómenos naturales como la lluvia, tormentas eléctricas y granizadas debido a que este tipo de acontecimientos han representado una amenaza permanente a su supervivencia. La necesidad de entender los fenómenos naturales impulsó al hombre a realizar numerosos experimentos, como la observación del comportamiento del cielo y las nubes, el análisis de la fuerza y dirección de los vientos, la observación de las estrellas. Estas actividades intentaban explicar el origen de los fenómenos naturales y predecir futuras ocurrencias.

La agricultura es una de las labores con mayor dependencia de las condiciones climáticas, los cambios climáticos drásticos representan una amenaza para los cultivos. Factores externos como el cambio climático constituyen un reto adicional para los agricultores. Por este motivo, analizar las condiciones climáticas de los terrenos que serán utilizados para actividades agrícolas, es un requisito fundamental para garantizar el éxito de las cosechas.

1.1 Justificación

La agricultura tiene un rol determinante en la supervivencia humana, representa el principal proveedor de alimentos. Además, es una de las actividades más relevantes en la economía mundial, generando empleos y materias primas para la elaboración de bienes de consumo.

Los fenómenos climáticos como lluvias torrenciales o sequías representan un riesgo para las actividades agrícolas. El aumento desproporcionado de las precipitaciones en los cultivos genera inundaciones, terrenos inestables y suelos fangosos, ocasionando distintas enfermedades que perjudican el crecimiento de las plantaciones. Del mismo modo, la escasez de agua en temporadas de sequía origina afectaciones en el desarrollo de los cultivos. Las plantas como todos los seres vivos requieren agua para el funcionamiento apropiado de sus procesos metabólicos; la oportuna distribución de agua durante las diferentes etapas del cultivo garantizará su crecimiento y desarrollo. La ocurrencia de estos fenómenos climáticos afecta la producción agrícola, generando pérdidas parciales o totales de los cultivos.

La selección de terrenos es una de las actividades primordiales a fin de garantizar el éxito en la agricultura. En la elección de zonas agrícolas es recomendable considerar aspectos como topografía, clima, composición del suelo con los nutrientes apropiados (nitrato, fósforo, potasio, amonio, calcio entre otros). Actualmente los agricultores tradicionales realizan la selección de terrenos usualmente siguiendo tradiciones culturales o familiares. Las tradiciones o costumbres familiares sugieren el tipo de productos y temporadas para actividades agrícolas, convirtiendo la agricultura en un legado que se transmite de generación en

generación; por otra parte, la prorrogación de las tradiciones en el sector agrícola implica la omisión de información relevante para el desarrollo de los cultivos como pronósticos y periodicidad de fenómenos climáticos (precipitaciones, vientos fuertes, etc.), el análisis de estos fenómenos puede determinar el éxito o fracaso de las cosechas.

El seguimiento de tradiciones familiares en la agricultura tradicional obstaculiza la adopción de nuevas técnicas y procesos en los cultivos. Además, la baja tasa de penetración de internet en zonas rurales es otra de las limitantes para la adquisición de nuevos procesos y técnicas agrícolas. La falta de innovación en la agricultura tradicional limita su competitividad frente a organizaciones multinacionales. Las empresas del sector agrícola utilizan la tecnología como apoyo fundamental para mejorar sus procesos y optimizar sus niveles de producción.

El crecimiento de la población mundial posiciona a la agricultura como una actividad fundamental a fin de garantizar la creciente demanda de alimentos y bienes de consumo. En las condiciones actuales, cada día es más relevante la innovación en los procesos agrícolas, la necesidad de combinar naturaleza y tecnología es fundamental a fin de dar soluciones a problemáticas actuales de la agricultura. Las inundaciones o sequías son fenómenos climáticos que afectan numerosas hectáreas de cultivos cada año, con el propósito de afrontar estos desafíos climáticos es relevante pronosticar la ocurrencia de estos fenómenos naturales. El conocimiento previo de fenómenos como las precipitaciones ayudaría a optimizar la selección de terrenos para actividades agrícolas, con el propósito de prevenir inundaciones que afecten el desarrollo de los cultivos. En definitiva, la utilización del pronóstico de precipitaciones como criterio de selección de terrenos agrícolas, representaría una alternativa a fin de reducir las pérdidas económicas generadas por eventos climáticos.

1.2 Planteamiento del trabajo

La selección de terrenos para actividades agrícolas es una de las principales decisiones que el agricultor debe realizar. La elección tradicional de terrenos considera aspectos como calidad del suelo, temperatura de la zona, características de la vegetación y fuentes de agua cercanas; la omisión de los pronósticos climáticos en particular de las lluvias representa un factor de riesgo en el éxito de las cosechas. Con el objetivo de evitar pérdidas de cultivos por eventos naturales como las inundaciones y sequías, el agricultor debería adicionar el análisis de pronósticos de precipitación a los criterios de selección de terrenos. El conocimiento previo de los niveles de lluvia para los próximos meses representa una información esencial que permitiría al agricultor tomar decisiones basándose en datos. Además, los pronósticos también proporcionarían una forma de identificar las zonas adecuadas para realizar actividades agrícolas.

En el presente trabajo se propone la construcción de un modelo predictivo, capaz de realizar el pronóstico de los niveles de precipitación mensual en el departamento de Cundinamarca, en Colombia. El pronóstico se realizará analizando los datos históricos registrados en las estaciones hidrometeorológicas ubicadas en el territorio de Cundinamarca, analizando los resultados de las predicciones generadas por el modelo se realizará un proceso de clasificación de zonas utilizando las categorías (adecuada, inadecuada). Finalmente, combinando la clasificación obtenida y la ubicación geográfica de las estaciones hidrometeorológicas, se realizará una representación gráfica del territorio de Cundinamarca que permita visualizar las zonas analizadas, el nivel de lluvias estimado y su correspondiente categorización para actividades agrícolas. La visualización permitirá identificar de forma sencilla y eficiente los territorios propicios para actividades agrícolas.

1.3 Estructura de la memoria

El presente documento se encuentra estructurado de la siguiente forma:

En el **Capítulo 1** se realiza la presentación del tema, la descripción del problema a tratar, sus posibles causas y la relevancia del problema. Adicionalmente, se elabora una propuesta para solucionar la problemática planteada.

En el **Capítulo 2** se expone el estudio del contexto de aplicación relacionado a las metodologías y procedimientos utilizados para realizar predicciones climáticas, revisión de bases teóricas analizando trabajos previos, estudios actuales y su correspondiente asociación con la propuesta realizada para resolver la problemática.

En el **Capítulo 3** se presenta el objetivo principal y los objetivos específicos derivados del análisis de la temática expuesta y la metodología de trabajo propuesta para cumplir los objetivos planteados.

En el **Capítulo 4** se expone el desarrollo de las actividades propuestas orientadas a generar una contribución específica en el campo de las predicciones de los niveles de lluvia y su impacto en la selección de zonas agrícolas.

Finalmente, en el **Capítulo 5** se presentan las conclusiones obtenidas a partir del desarrollo de las actividades planteadas y se describen las posibles líneas futuras de trabajo.

2. Contexto y estado del arte

En este capítulo se realiza un estudio de la evolución en la predicción climática que permitirá conocer las definiciones y conceptos utilizados en el análisis y resolución de la problemática expuesta previamente.

2.1 Orígenes de la predicción climática

Desde los inicios de la humanidad el hombre ha intentado comprender los fenómenos naturales como lluvias y granizadas, principalmente para establecer estrategias y mecanismos de defensa que garanticen su supervivencia. Inicialmente, el hombre intentó predecir la ocurrencia de fenómenos naturales basándose en comportamientos como la observación del cielo, las nubes, la intensidad del viento y otras manifestaciones naturales que precedían a las tormentas y otros fenómenos naturales. El surgimiento de actividades como la agricultura y la navegación generaron en el hombre una necesidad constante de conocer e intentar predecir las condiciones climáticas. El conocimiento de los fenómenos climáticos era esencial para el desarrollo de los cultivos y la planeación de viajes marítimos.

En la antigua Babilonia los astrólogos relacionaban los halos solares y lunares a los cambios climáticos que se presentaban en su territorio. Tenían la convicción de que los eventos relacionados a la luna rodeada por un halo oscuro, producirían nubes y lluvias, en el caso de visualizar un pequeño halo alrededor del sol representaría una inminente llegada de las lluvias (Erickson, 1991).

La transmisión de conocimiento y costumbres de forma oral entre generaciones facilitó la propagación de una serie de creencias, principalmente relacionadas a métodos para predecir el clima a largo plazo como lo son las “cabañuelas”, un método popular sin bases científicas, basado en la observación del clima de días específicos del año utilizados para predecir las condiciones climáticas de los próximos meses. El origen de la palabra cabañuela algunos señalan proviene del norte de Europa. En estos territorios sus pobladores construían pequeñas cabañas a las orillas de los ríos con el objetivo de observar los niveles de humedad y la climatología de días específicos de la etapa estival, con el propósito de realizar predicciones meteorológicas para los siguientes meses (Fuentes & Fuentes, 2015). Tradiciones como las cabañuelas han sobrevivido al paso del tiempo y son utilizadas actualmente en regiones de España y América.

Las predicciones climáticas con un enfoque científico iniciaron en el siglo XIX de la mano del matemático francés Urbain J. J. Le Verrier (1811-1877), época en la cual se empezaron a realizar las primeras observaciones meteorológicas en territorio europeo.

En la guerra de Crimea los países de Francia, Inglaterra, Austria y Turquía formaron una coalición para combatir el avance de Rusia hacia los territorios de oriente medio. En la tarde del 13 de noviembre de 1854 una gran tormenta con fuertes lluvias y vientos afectó la zona del puerto de Balaclava en las orillas del mar negro, este fenómeno climático generó grandes afectaciones; las tiendas de campaña de las tropas francesas e inglesas quedaron destruidas. Luego de la tormenta en la zona se presentaron climas fríos y nieve, generando numerosas bajas en las tropas de la coalición. Las consecuencias de este fenómeno climático hicieron que el emperador francés Napoleón III encargara a Le Verrier el estudio de lo sucedido en el puerto de Balaclava. El principal propósito del análisis era determinar si era posible predecir las condiciones climáticas presentadas en Balaclava. El pronóstico de este tipo de fenómenos naturales ayudaría a evitar nuevas pérdidas en el campo de batalla.

Le Verrier empezó su investigación visitando múltiples observatorios en Europa con el objetivo de recolectar y analizar los datos climáticos registrados en los observatorios. Analizando los datos Le Verrier detectó que una borrasca se había desplazado por distintos territorios europeos, concluyendo que era posible anticipar este tipo de fenómenos climáticos. El éxito de la predicción dependía del envío de información meteorológica en tiempo real, utilizando el telégrafo como medio de comunicación (Cox, 2002). Una de las propuestas realizadas por Le Verrier posterior al estudio consistía en la construcción de múltiples estaciones interconectadas por telégrafo, con el propósito de comunicar en tiempo real cambios en las condiciones climáticas. El conocimiento previo de la ocurrencia de estos fenómenos climáticos permitiría anticipar la llegada de tormentas a territorio francés.

2.2 Bases teóricas de la predicción climática

La palabra meteorología proviene del griego *meteoros*, significa (elevado, que está en el aire) y está relacionada a fenómenos físicos que ocurren en la atmósfera que envuelve a nuestro planeta (Vicuña, 2009). Analizando el origen de la palabra meteorología se puede definir de una manera más formal como “la ciencia que estudia los procesos y fenómenos físicos en la atmósfera y que forman el tiempo meteorológico” (Petterssen, 1968).

La meteorología como método de observación de fenómenos naturales ha estado desde los inicios de la humanidad, el filósofo griego Aristóteles uno de los primeros expositores registró sus ideas y observaciones de los fenómenos climáticos en su publicación titulada

Meteorología (Taub, 2004). El paso de la observación empírica de los fenómenos naturales hacia la utilización de métodos más formales fue dado por el matemático noruego Vilhelm Bjerknes (1862-1951), conocido como uno de los pioneros de la meteorología moderna. Bjerknes elaboró el teorema de la circulación en el cual expone la forma en la cual la circulación se desarrolla en fluidos geofísicos definiendo la transición de los estudios de Bjerknes de la hidrodinámica hacia la física atmosférica (Thorpe et al., 2003). Acorde a las estimaciones de Bjerknes “llevaría unos meses calcular los cambios de estado en la atmósfera para un par de horas, pero de tener éxito en las predicciones calculadas frente a los comportamientos observados sería un gran triunfo para la ciencia” (Friedman, 1993).

En el año 1904 Bjerknes publicó el artículo *‘El problema de la predicción climática desde el punto de vista de la física y la mecánica’* en el cual expone como una solución racional a un problema de predicción climática debería cumplir dos condiciones: tener el conocimiento exacto del estado actual de la atmósfera en el momento inicial y poseer el conocimiento preciso de las leyes físicas por las cuales un estado atmosférico se desarrolla a partir de otro (Bjerknes, 1999).

El matemático británico Lewis Richardson (1881- 1953) inspirado por las publicaciones de Bjerknes que buscaban la forma de calcular los estados futuros de la atmósfera desde la perspectiva de la física, realizó múltiples investigaciones recopilando sus observaciones y descubrimientos en el libro *‘Predicciones climáticas con procesos numéricos’*, en la publicación Richardson expone las ecuaciones y cálculos requeridos para realizar predicciones climáticas analizando variables como el tiempo, longitud, latitud, altura del terreno sobre el mar, temperatura, velocidad del viento entre otros factores (Richardson, 2007). El manejo de múltiples variables en los cálculos de las predicciones climáticas de unas pocas horas eran complejos de resolver y no tenían un grado de precisión elevado. Estas complicaciones llevaron a Richardson a imaginar una fábrica del clima. La invención consistía en una enorme organización de más de 64000 personas trabajando simultáneamente, cada persona estaba encargada de calcular los pronósticos para una región específica. La coordinación de las personas era realizada por un matemático ubicado en un punto central de la edificación, desde este puesto de control se solicitaban las predicciones en el mismo instante que sucedían los fenómenos climáticos, realizando los cálculos en tiempo real. Richardson al final de su análisis expresó la siguiente idea: “Quizás algún día en un futuro cercano sea posible con el progreso de la computación realizar los cálculos climáticos con mayor rapidez a la que avanza el clima” (metrics et al., 1981).

2.3 Origen predicciones climáticas computarizadas

El punto de partida de las predicciones climáticas ejecutadas desde un computador lo realizó una máquina electrónica de cálculos denominada ENIAC por sus siglas en inglés (*Electronic Numerical Integrator and Automatic Calculator*). El único computador capaz de realizar cálculos de forma automática para la época. Esta máquina fue la invención de John W. Mauchly (1907-1980) y Presper J. Eckert (1919-1995), diseñada y construida en la escuela Moore de ingeniería eléctrica de la Universidad de Pensilvania, finalizando su construcción en 1946. La maquinaria de ENIAC estaba conformada aproximadamente de 18000 válvulas electrónicas, 5000 interruptores. La velocidad de operación de ENIAC permitía ejecutar cientos de cálculos numéricos, facilitando las labores de investigación a los científicos, sin la asistencia de ENIAC los investigadores tendrían que emplear una gran cantidad de horas para realizar estas operaciones matemáticas (Hartree, 1946). La máquina fue diseñada con propósitos militares, principalmente para calcular la trayectoria de las municiones. Aunque la mayoría de las tareas de construcción de la máquina fueron realizadas por hombres, es importante resaltar el gran aporte de las mujeres, ellas realizaron las tareas de configuración requeridas para ejecutar los cálculos matemáticos. Estos aportes significativos llevaron a las mujeres a ser catalogadas como las pioneras de la programación (Fritz, 1996).

En 1950 ENIAC realizó la primera predicción climática numérica ejecutada por un computador de la mano del matemático John Von Neumann (1903-1957). El matemático quería utilizar las capacidades de cálculo de ENIAC para resolver problemas de matemática aplicada, entre los cuales se encontraba la predicción del clima. El primer cálculo del clima fue realizado el día 5 de marzo de 1950 y se prolongó por 33 días seguidos, el objetivo de este experimento era realizar predicciones climáticas para un segmento de 24 horas. Al final del experimento se obtuvieron dos modelos de predicción para las 24 horas, el primer modelo generado no era tan preciso, aunque tenía algunos puntos donde las predicciones se asemejaban a las condiciones reales. El segundo modelo titulado "modelo barotrópico" consistía en un modelo de la atmósfera en el cual los movimientos del aire solamente eran de carácter horizontal. El modelo barotrópico generó resultados prometedores para un sistema de predicciones a largo plazo comparado con las observaciones realizadas (Platzman, 1979).

2.4 Técnicas de Machine Learning para predicciones climáticas

En el transcurso de los años la predicción climática ha pasado por varias etapas, desde la astrología con la observación de los ciclos lunares para pronosticar lluvias, la predicción basada en el comportamiento de las nubes y el cielo, la creación de mapas de clima basados en las observaciones realizadas en zonas cercanas transmitidas por telégrafo, la invención de predicciones climáticas con enfoque numérico utilizando fórmulas para modelar el comportamiento de la atmósfera.

El incremento en los volúmenes de información generado por los sensores climáticos y estaciones hidrometeorológicas requiere un nuevo enfoque para el procesamiento de estos grandes volúmenes de datos. Una de las alternativas para abordar la gestión de grandes volúmenes de datos es el uso de big data. La implementación de tecnologías big data permite la obtención de datos de múltiples fuentes estructuradas, no estructuradas y el posterior manejo distribuido de los datos para su alistamiento y procesamiento. La tendencia de las tecnologías actuales se ha centrado en el uso de la computación en la nube y tecnologías como Apache Hadoop como soluciones big data, es relevante también analizar previamente si los algoritmos a utilizar están optimizados para este tipo de ambientes y tecnologías (Sahasrabuddhe, 2015).

Las predicciones climáticas pueden combinarse con distintas fuentes de datos y utilizarse en diferentes ámbitos. Las predicciones de retrasos en vuelos es uno de los ejemplos más conocidos, analizando de forma combinada datos históricos de vuelos y registros históricos del clima es posible pronosticar si un vuelo puede tener retraso en su llegada debido a condiciones climáticas (Mayer-Schönberger & Cukier, 2013).

La capacidad de procesamiento es uno de los aspectos más importantes en el cálculo de las predicciones climáticas; pero, también es importante identificar algoritmos o procedimientos que permitan procesar grandes volúmenes de datos sin la intervención humana. Enfoques como el '*machine learning*' rama de la inteligencia artificial enfocada a que aplicaciones y sistemas aprendan de los datos o de sus experiencias pasadas y sean capaces de interpretar la información de forma autónoma (Alpaydin, 2020). Este enfoque en el cual los sistemas utilizan sus experiencias previas para mejorar su eficiencia es una de las alternativas más utilizadas en la actualidad en diferentes sectores de la industria como la manufactura, servicios financieros, transporte, agencias de gobierno entre otros. La búsqueda de métodos para el aprendizaje en las máquinas se ha estudiado desde los inicios de la computación y fue concebida originalmente como la forma en que los computadores pudieran aprender de la

experiencia y que este aprendizaje mejorara automáticamente la eficiencia de los propios programas durante sus ejecuciones (Michie, 1968).

En el campo de las predicciones climáticas existen múltiples modelos de *machine learning*, estos modelos son capaces de realizar diferentes tipos de pronósticos a corto y largo plazo. Los modelos más relevantes son aquellos capaces de interpretar y procesar correctamente variaciones aleatorias en los datos climáticos. Existen dos tipos de modelos que cumplen estas características: SVM (*Support Vector Machine*) y ANN (*Artificial Neural Network*) (Koshkarov, 2018).

SVM (*Support Vector Machine*): Modelo de aprendizaje supervisado que contiene algoritmos orientados a resolver problemas de clasificación múltiple de datos y series de tiempo. SVM es una herramienta de clasificación y predicción basada en regresiones, como se observa en la Figura 1.

El modelo utiliza la teoría de machine learning para optimizar la precisión de las predicciones realizadas, evitando automáticamente el sobre ajuste a los datos. Este tipo de modelos ha sido utilizado en problemas de reconocimiento de texto, reconocimiento facial entre otras aplicaciones (Jakkula, 2006). Particularmente modelos SVM han sido utilizado para realizar estudios de predicciones climáticas de lluvias extremas (Nayak & Ghosh, 2013).

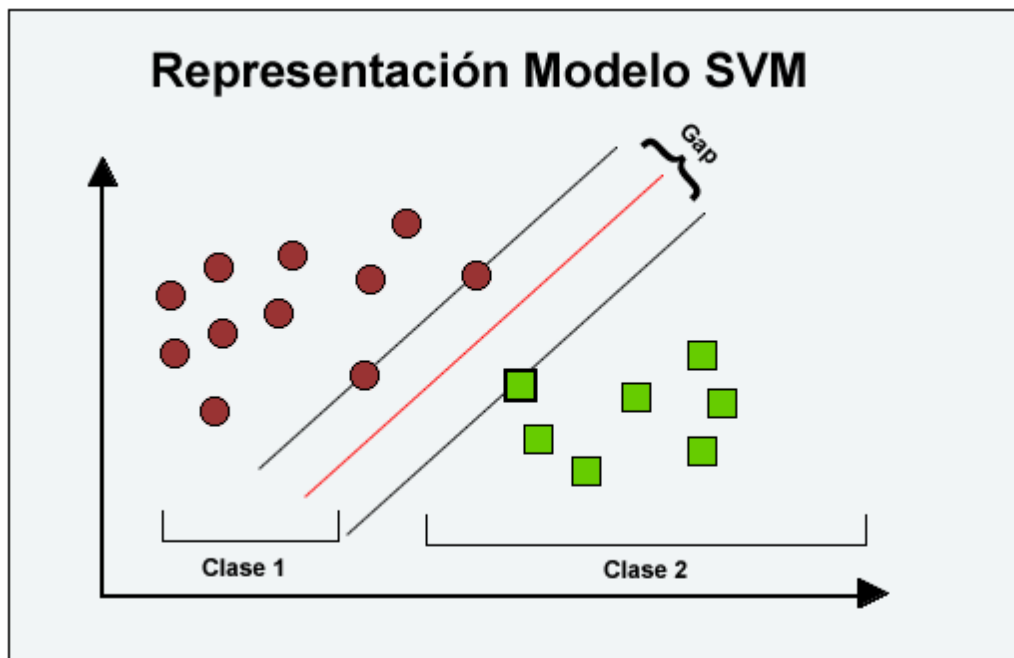


Figura 1. Representación SVM (*Support Vector Machine*) (elaboración propia)

ANN (Artificial Neural Network): Modelo computacional no lineal capaz de procesar información de forma similar a las redes neuronales biológicas de ahí su nombre. Las redes neuronales artificiales pueden tener varios niveles de complejidad debido a su densamente interconectada red de unidades de procesamiento, las conexiones son similares a la interconexión entre neuronas en el cerebro humano. La característica más importante de este tipo de redes es su naturaleza adaptativa donde se realiza aprendizaje por ejemplo reemplazando la forma tradicional de programación para la resolución de problemas (Hassoun & Hassoun, 1995).

La arquitectura de una red neuronal artificial está conformada por una capa inicial denominada (*input layer*), esta capa tiene la función de recibir la información de entrada que será procesada en el modelo. Posterior a la recepción de los datos, se encuentran un conjunto de capas ocultas (*hidden layers*) encargadas de realizar los cálculos computacionales sobre los datos de entrada, las capas ocultas son capaces de reconocer patrones presentes en los datos por este motivo también son conocidas como capas de aprendizaje. Finalmente se tiene una capa de salida (*output layer*) encargada de obtener los resultados generados por las capas ocultas.

Existen dos tipos de flujos de información en las redes neuronales, el primer tipo de flujo conocido como redes prealimentadas (*feed-forward*) en esta clase de flujos no se presentan ciclos entre las unidades de procesamiento, como se observa en la Figura 2. Las unidades reciben la información, realizan el procesamiento y transmiten los datos a la siguiente capa de forma progresiva. El segundo tipo de flujo de datos se presenta en las redes retroalimentadas (*feed-back*), caracterizadas por presentar ciclos entre las unidades de procesamiento. El objetivo de las redes retroalimentadas es la retención de conocimiento y son ampliamente utilizadas para resolver problemas de series temporales. En su forma más general las redes retroalimentadas están compuestas por un grupo de unidades de procesamiento. La salida generada por cada unidad de procesamiento es utilizada como entrada para todas las unidades de procesamiento incluyendo a la misma unidad generadora de información (Yegnanarayana, 2009).

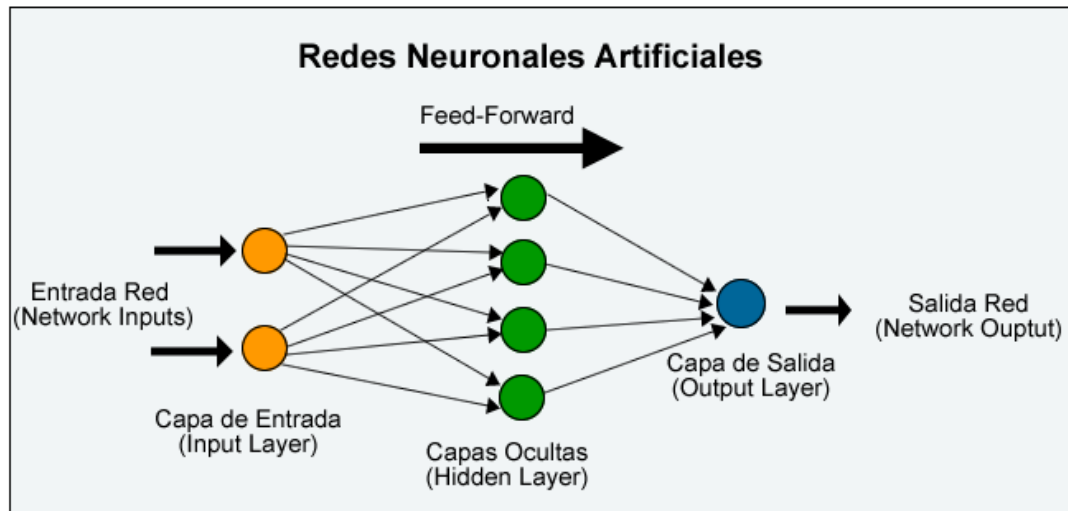


Figura 2. Arquitectura redes neuronales artificiales (elaboración propia)

RNN (Recurrent Neural Network): Las redes neuronales recurrentes son un tipo de red que no tiene una estructura tradicional de capas, permitiendo la conexión aleatoria entre las unidades de procesamiento. Este tipo de redes son capaces de crear un tipo de memoria temporal que les permite a resolver tipos de problemas de pronósticos, asociación de patrones y problemas de optimización (Mehrotra et al., 1997). Las conexiones en la serie temporal se modelan como un grafo dirigido, como se puede observar en la Figura 3. Las redes LSTM (Long Short Term Memory) son un tipo particular de redes neuronales recurrentes que tienen la particularidad de almacenar información entre las conexiones previas (memoria a corto plazo) y son ampliamente utilizadas para el cálculo de predicciones en series temporales en las cuales es relevante el orden de ocurrencia de los eventos (Kratzert et al., 2018).

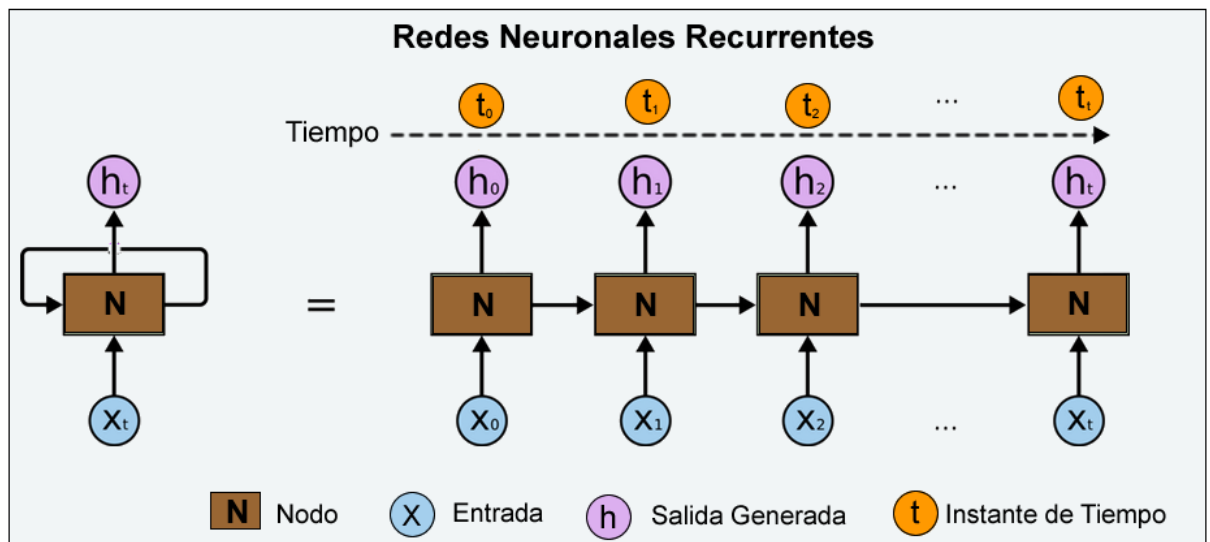


Figura 3. Red neuronal recurrente y representación extendida (elaboración propia)

La evolución de los algoritmos y métodos de programación ha permitido la creación de frameworks y librerías especializadas en redes neuronales artificiales. El principal propósito de estas librerías es simplificar y acelerar la implementación de modelos basados en redes neuronales. En el caso particular del lenguaje de programación multipropósito Python (Python Foundation, 2020) se tienen un conjunto de librerías orientadas a la construcción de modelos de machine learning; la librería más conocida por su amplia variedad de algoritmos es Scikit-learn (Pedregosa et al., 2011). La librería permite realizar tareas como: preprocesamiento de datos, selección y ejecución de modelos machine learning. Los modelos generados son utilizados para resolver problemas de clasificación, predicción, agrupamiento, regresión entre otras opciones. Particularmente, para la implementación de redes neuronales artificiales se encuentra la librería Keras (Gulli & Pal, 2017). La interfaz de alto nivel de Keras permite utilizar todo el potencial de plataformas de machine learning como TensorFlow, plataforma encargada de realizar operaciones matemáticas a través de un grafo de flujo de datos (Hope et al., 2017) facilitando el mantenimiento y evolución de los modelos.

2.5 Predicción climática en la agricultura

El clima es uno de los factores determinantes en el éxito o fracaso de actividades agrícolas. Desde los inicios de la humanidad las comunidades han realizado distintos acercamientos a fin de predecir la ocurrencia de fenómenos climáticos. En el ámbito religioso la falta de lluvia era asociada a la ira de entidades supremas, en el caso particular de la cultura Maya el dios Chaac era el encargado de traer la lluvia a los territorios (West, 2014).

Posterior a la atribución de lluvias a entidades sobrenaturales, la agricultura se convirtió en una tradición transmitida de generación en generación. Las tradiciones orales y escritas ayudaron a la creación de calendarios agrícolas basados esencialmente en las festividades religiosas, en este calendario se especificaban las fechas y actividades agrícolas, como se observa en la Figura 4. Algunas de estas tradiciones siguen vigentes en determinadas zonas rurales.

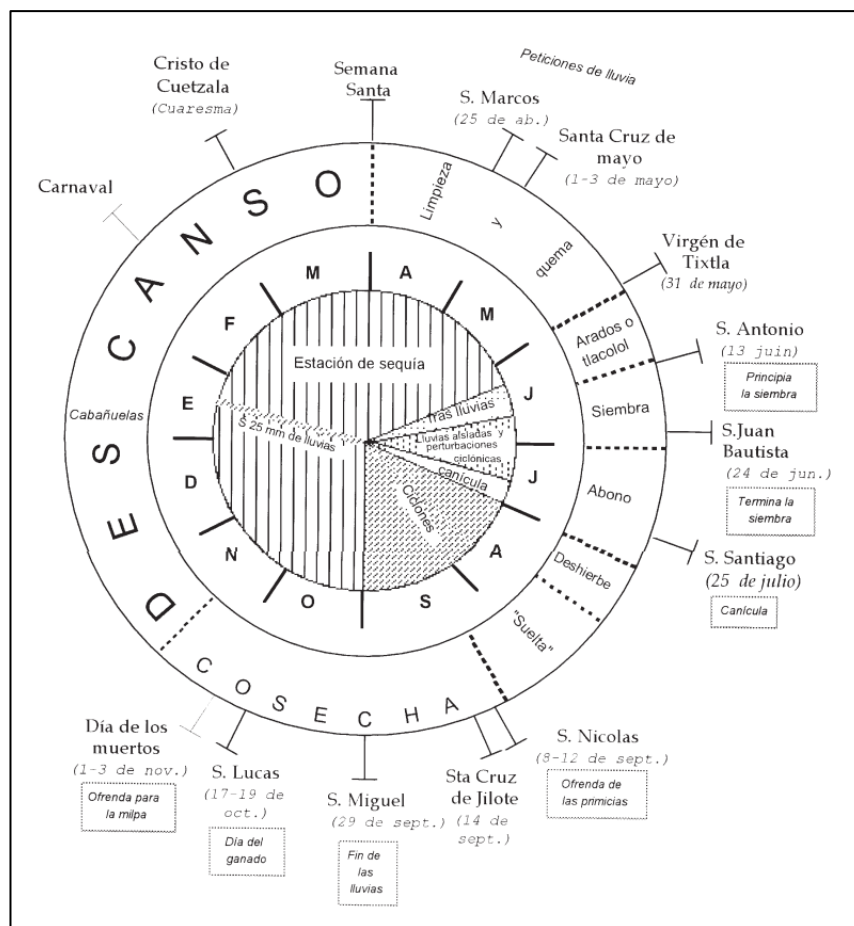


Figura 4. Calendario agrícola Hispanoamérica (Goloubinoff et al., 1997)

En la agricultura como en otros ámbitos dependientes del clima existen diferentes tipos de predicciones climáticas, las más relevantes son: pronósticos a corto plazo (12-72 horas), pronósticos a medio plazo (hasta once días) y pronósticos a largo plazo (desde 12-30 días y hasta dos años) (World Meteorological Organization, 2010). Los diferentes tipos de pronósticos son requeridos en las diferentes etapas de los cultivos, desde las actividades diarias realizadas en las cosechas para su crecimiento hasta la planeación de nuevos cultivos. El grado de precisión de las predicciones climáticas es dependiente del período pronosticado, generalmente las proyecciones a largo plazo tienen un grado de precisión menor comparadas con las predicciones a corto plazo. Esta situación se produce como resultado de pronósticos a largo plazo basados en modelos matemáticos y no en la constante variabilidad de las condiciones de la atmósfera que si son registradas en los pronósticos de corto plazo.

Los avances en la predicción climática representan una oportunidad para mejorar la gestión de riesgos asociados al clima en los cultivos. La planeación de las cosechas se realiza generalmente con meses de anticipación, la incertidumbre sobre las condiciones climáticas proyectadas respecto a las condiciones reales implica que el sector agrícola debe permanecer alerta y reaccionar rápidamente a los cambios climáticos que se presentan en sus cultivos a

fin de mitigar los daños y proteger sus cultivos. Es fundamental para el agricultor comprender la importancia del conocimiento previo de las predicciones climáticas en sus territorios, con el propósito de incorporar esta información en la toma de decisiones relacionadas a los cultivos, pero también es indispensable la habilidad de anticipar los posibles resultados asociados a cada decisión bajo diferentes situaciones de pronósticos climáticos (Hansen, 2002).

La implementación de soluciones *Big Data* en el sector de la agricultura enfocadas al análisis de las condiciones climáticas es una tendencia reciente en los países de Latinoamérica. En el territorio colombiano una alianza entre la Federación de Productores de Arroz FEDEARROZ y un grupo de científicos de la CCAFS (*Climate Change Agriculture and Food Security*) examinaron la información histórica de los cultivos de arroz analizando su relación con el histórico del clima para generar un sistema de apoyo a decisiones estratégicas que estuviera disponible a los agricultores de la federación (FAO, 2019).

Los desafíos para la predicción climática enfocada a las actividades agrícolas radican en mejorar la adopción de este tipo de sistemas o tecnologías en las zonas rurales. Adicionalmente es necesario identificar las razones por las cuales la comunidad en algunos casos se resiste a la adopción de nuevos enfoques tecnológicos en la agricultura. La estrategia de acercamiento a las comunidades debe iniciar con la identificación de los comportamientos y motivaciones de los agricultores (Sivakumar, 2006).

2.6 Trabajos relacionados

En la actualidad existen diversos estudios relacionados a las predicciones de niveles de precipitación. A continuación, se describirán algunos modelos relacionados a la temática del presente trabajo. El modelo *RainNet* utiliza una red neuronal artificial para predecir la intensidad de las precipitaciones, el análisis es realizado en períodos de cinco minutos utilizando datos climáticos históricos obtenidos del servicio climático alemán (DWD) (Ayzel et al., 2020).

MetNet está definido como un modelo neuronal del clima, diseñado para estimar niveles de precipitación medidos en milímetros de lluvia por hora (mm/h), el proceso analiza intervalos de hasta 8 horas (Sønderby et al., 2020). El modelo utiliza los datos climáticos registrados por los satélites de Estados Unidos. Examinando los datos *MetNet* es capaz de generar un mapa probabilístico de las precipitaciones, como se puede observar en la Figura 5 donde se comparan los resultados obtenidos por el modelo y los datos climáticos reales.

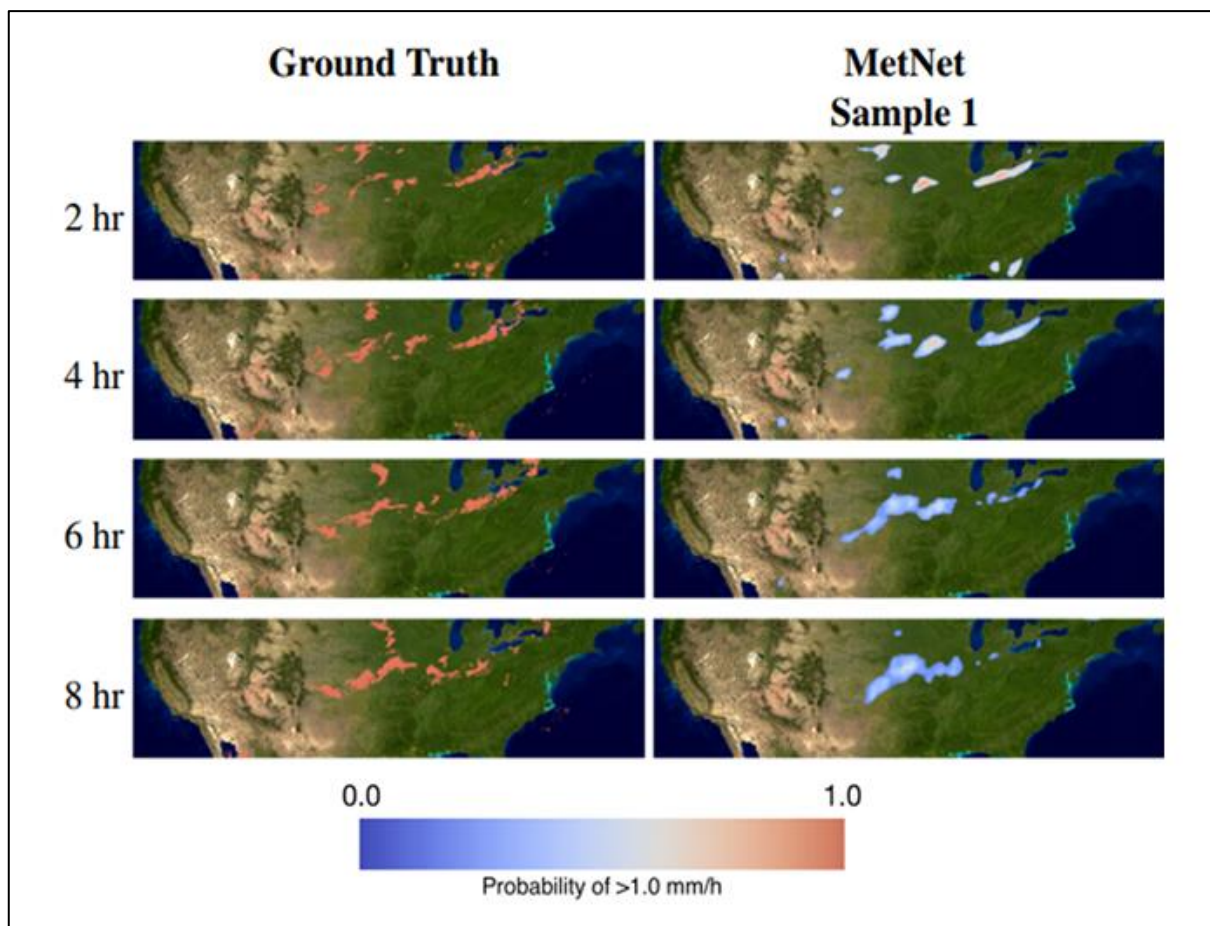


Figura 5. Mapa probabilístico de predicciones modelo MetNet (Sønderby et al., 2020)

Los modelos basados en redes neuronales recurrentes son ampliamente utilizados en el estudio de las predicciones climáticas. Este tipo de redes neuronales permiten el cálculo de pronósticos climáticos en determinados instantes de tiempo, las predicciones son posibles debido a la capacidad de aprender de los eventos previos en la serie temporal analizada. Una de las arquitecturas específicas para estos modelos es LSTM (*Long Short Term Memory*), la cual está diseñada para modelar clasificaciones y predicciones basadas en secuencias temporales de datos. Un estudio realizado en India diseñó un modelo LSTM para la predicción de lluvias. En el análisis realizado se evidenció la precisión de este tipo de modelos frente a un modelo tradicional de redes neuronales recurrentes (Poornima & Pushpalatha, 2019). El modelo utilizó el histórico de datos climáticos generados desde el año 1980 hasta 2014 de la ciudad Hyderabad en India. En el estudio se analizaron variables como: temperatura, nivel de humedad, velocidad del viento a fin de realizar las predicciones de las lluvias. En la Figura 6 se puede observar como las predicciones generadas por el modelo en un período de 97 días (julio – septiembre) se acercan a las cifras reales de niveles de lluvia, las diferencias más significativas se presentan en los valores extremos y se encuentran atribuidas a la

consideración de algunos registros del conjunto de datos como valores atípicos omitidos por el modelo.

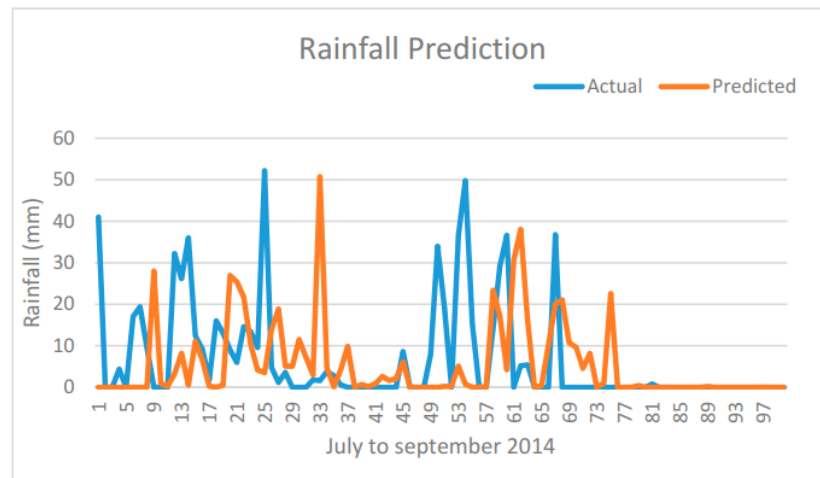


Figura 6. Comparación predicciones y valores reales de niveles de lluvia modelo LSTM (Poornima & Pushpalatha, 2019)

2.7 Futuro de la predicción climática

La constante evolución de la capacidad y velocidad de procesamiento en los computadores hace pensar que probablemente en los próximos años se generarán pronósticos climáticos con mayor precisión y en menor tiempo. El futuro de la predicción climática en términos tecnológicos no parece tener limitantes y es posible pensar en la generación de pronósticos climáticos de los próximos meses o años en cuestión de segundos. Sin embargo, el meteorólogo Edward Lorenz (1917 - 2008) pionero de la teoría del caos expone una serie de argumentos que explican como la capacidad del pronóstico del clima no está limitada por los avances tecnológicos, sino por las condiciones caóticas de la atmósfera. “La dependencia sensible sobre las condiciones iniciales de un sistema no lineal están relacionadas con el comportamiento aleatorio del sistema” (Lorenz, 1963). A partir de los estudios expuestos por Lorenz el futuro de la predicción climática debería enfocarse en generar predicciones a corto plazo, las cuales tienen un menor grado de variabilidad sobre las condiciones atmosféricas iniciales.

2.8 Conclusiones

La predicción climática es un campo de estudio que está en constante evolución, siempre en la búsqueda de pronósticos más precisos y generados en el menor tiempo posible. Esta evolución es generada principalmente por la invención de nuevas técnicas y procedimientos utilizados en el ámbito de la climatología.

Utilizando las librerías especializadas en *machine learning* que existen actualmente es posible la construcción de modelos predictivos para estimar los niveles de lluvias. Estos modelos orientados a la asistencia de la sociedad agrícola permitirán mejorar la toma de decisiones referentes a la planeación y desarrollo de los cultivos.

La implementación de soluciones tecnológicas en el ámbito agrícola no solamente involucra la construcción de software, el éxito de este tipo de soluciones depende en mayor medida de la adopción que tenga la solución en las comunidades agrícolas. El conocimiento de las costumbres y motivaciones de los agricultores es un factor determinante en la adopción de tecnologías de la información y de esta forma lograr una cooperación entre la tecnología y la agricultura.

3. Objetivos concretos y metodología de trabajo

En este capítulo se registra el objetivo general, los objetivos específicos y la metodología utilizada para cumplir los objetivos.

3.1. Objetivo general

El objetivo general del trabajo es: **Identificar las zonas más apropiadas para la agricultura a partir de un modelo de predicción de los niveles de precipitación.** Para el desarrollo de este objetivo se realizará el análisis del histórico de las condiciones climáticas del territorio de Cundinamarca, esta información se utilizará como entrada del modelo predictivo que permitirá generar un pronóstico de los niveles de lluvia para las zonas analizadas y de esta forma determinar los territorios adecuados para actividades agrícolas. A partir de los pronósticos generados se realizará un mapa de la región que permita referenciar geográficamente las zonas analizadas y su respectiva clasificación con el propósito de identificar las zonas apropiadas para actividades agrícolas de una forma sencilla y concreta.

3.2. Objetivos específicos

Los objetivos específicos para este trabajo son los siguientes:

- Elaborar el conjunto de datos que será utilizado como entrada del modelo predictivo a partir de los registros de las estaciones hidrometeorológicas de Cundinamarca.
- Determinar la técnicas o metodologías de predicción que serán utilizadas para la construcción del modelo predictivo de precipitaciones.
- Construir el modelo predictivo encargado de pronosticar los niveles de lluvia de las zonas del departamento de Cundinamarca.
- Evaluar la precisión del modelo respecto a los datos reales de los niveles de precipitación en las zonas evaluadas y generar las conclusiones respectivas.
- Elaborar el mapa de Cundinamarca con la ubicación de las zonas analizadas y su respectiva clasificación para la realización de actividades agrícolas.

3.3. Metodología del trabajo

La metodología seleccionada para el desarrollo de este trabajo es el framework SEMMA acrónimo de (*Sample, Explore, Modify, Model, Assess*) creada por el instituto SAS (SAS Institute, 2017). El framework originalmente diseñado para proyectos de minería de datos tiene aplicabilidad a diferentes tipos de industrias y proyectos que involucren análisis de datos. La selección de SEMMA como metodología de trabajo se realizó analizando el tipo de actividades requeridas para cumplir los objetivos planteados. En la Figura 7 se puede observar las distintas actividades que componen cada una de las fases de SEMMA.

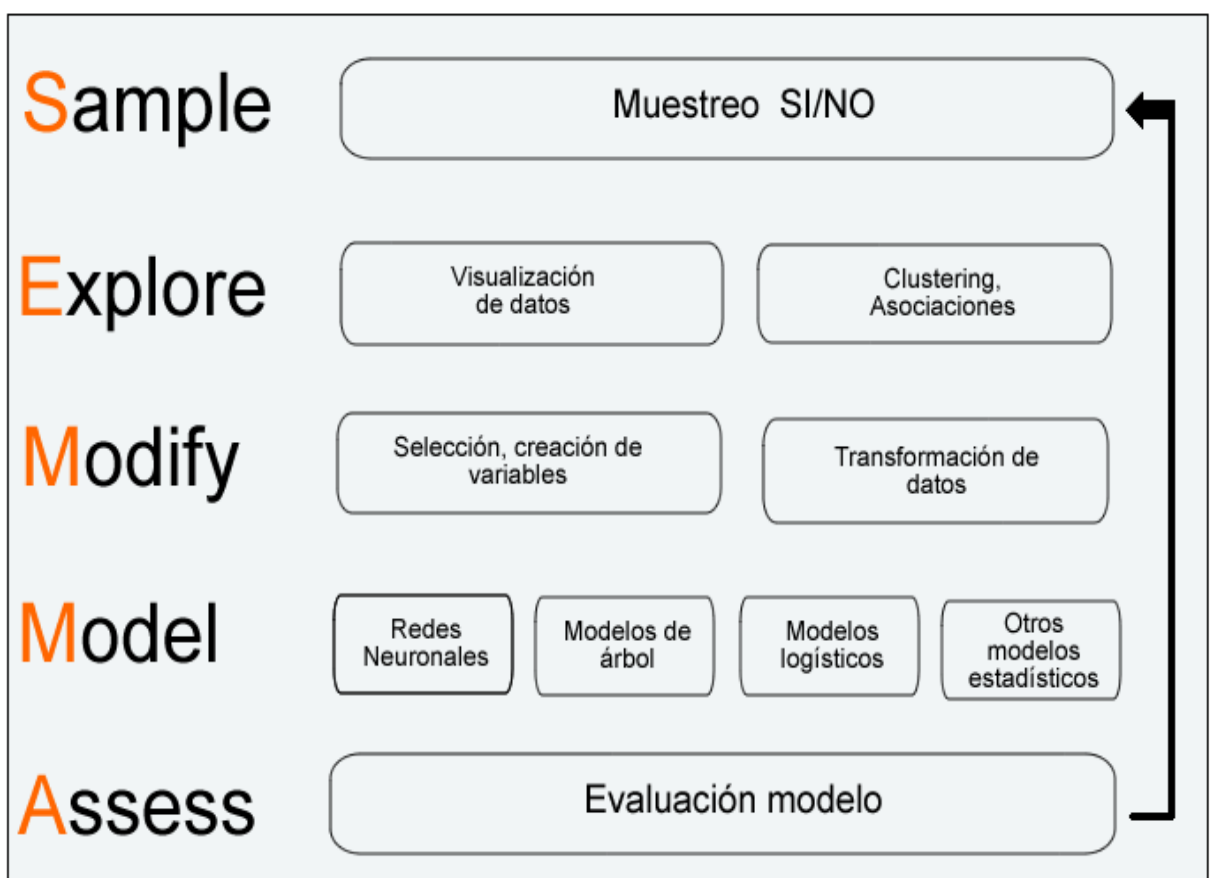


Figura 7. Esquema framework SEMMA (elaboración propia)

A continuación, se realizará una breve explicación de las fases que componen el framework SEMMA.

Tabla 1. Fases framework SEMMA

Fase	Objetivo
Sample (Muestreo)	Realizar un análisis preliminar de los datos sobre una muestra del conjunto de datos.
Explore (Exploración)	Búsqueda de patrones y relaciones desconocidas entre los datos, datos anómalos con el objetivo de obtener un mayor entendimiento del conjunto de datos utilizado.
Modify (Ajustes)	Transformación y selección de variables, rangos de valores y agrupamientos que permitan analizar los datos de una forma adecuada en el modelo a implementar.
Model (Modelamiento)	Construcción del modelo requerido para satisfacer los requerimientos o necesidades establecidas inicialmente.
Assess (Evaluación)	Evaluar la calidad de las salidas generadas por el modelo y generar las conclusiones correspondientes.

Fuente: elaboración propia.

Es importante mencionar que si los criterios de aceptación para el modelo implementado no se cumplen, es recomendable realizar los ajustes correspondientes y volver a realizar las actividades contenidas en las fases de SEMMA, ya que la naturaleza del framework está basada en un proceso de experimentación iterativo (Olson & Delen, 2008) .

4. Desarrollo específico de la contribución

En esta sección se describirán las actividades realizadas en el desarrollo del presente trabajo. Iniciando con la descripción de la fuente de datos utilizada y continuando con las actividades propias de cada una de las fases de la metodología SEMMA.

4.1 Fuente de datos

4.1.1 Tratamiento de datos

En el desarrollo del presente trabajo se utilizó como fuente de datos el histórico de series hidrometeorológicas expuesto en el sitio web de la Corporación Autónoma Regional ([CAR](#)). En este sitio se encuentran los registros climáticos de las estaciones hidrometeorológicas localizadas en la región de Cundinamarca, ubicada en Colombia. Esta fuente de datos es de acceso público y puede ser utilizada de forma libre acorde la iniciativa de [datos abiertos](#) liderada por el gobierno de Colombia, por esta razón no es requerido realizar un tratamiento específico para la utilización de los datos climáticos.

4.1.2 Descripción datos

Los datos de las estaciones se encuentran clasificados en diferentes temáticas dependiendo de las mediciones registradas. Las temáticas contenidas en el sitio corresponden a registros de radiación solar, punto de rocío, temperatura media, velocidad y dirección del viento, precipitaciones totales mensuales, humedad relativa media mensual, evaporación total mensual, caudales medios mensuales, brillo solar total mensual entre otras mediciones climáticas. Cada una de las características descritas se encuentra almacenada en un archivo independiente de Microsoft Excel con extensión(.xlsx). Los archivos tienen la siguiente estructura: la pestaña inicial llamada 'Catálogo' contiene el listado de estaciones hidrometeorológicas en las cuales se realizó la medición, las pestañas restantes corresponden a las mediciones realizadas para cada una de las estaciones. En la Figura 8 se puede observar el contenido del archivo de precipitaciones mensuales.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
2	SICLICA - Sistema de Información Climatológica e Hidrológica																									
3																										
4	VALORES TOTALES MENSUALES DE PRECIPITACIÓN (mm)																									
5																										
6	ESTACIÓN : 2120074 ZIPAQUIRÁ																									
7																										
8	Latitud	5° 1' 26.6" N	X=N=1047320	Departamento	CUNDINAMARCA	Corriente R. NEGRO	Categoría																		P G	
9	Longitud	73° 59' 56.6" W	Y=E=1008700	Municipio	ZIPAQUIRA	Cuenca R. NEGRO	Fecha Instalación																		1/2/1960	
10	Elevación	2813 m.s.n.m		Oficina Provincial	9 SABANA CENTRO		Fecha Suspensión																			
11																										
13	AÑO	ENERO	FEBRE	MARZO	ABRIL	MAYO	JUNIO	JULIO	AGOST	SEPTI	OCTUB	NOVIE	DICIE													
15																										
16	1932	59.8	19.4	40.3	49.6	138.8	34.2																			
17	1933			67.7	88.4	155.4	163.1	56.7	90.3	58.5	150.1	147	66.7													
18	1934	33.2	111.5	94.1	29.6	184.9	75	53	44.2	44.6	110.1	145.9	22													
19	1935	24.1	49.3	50.7	90.4	206.2	45.1	76.2	87	82.7	89.7	246.2	3.8													
20	1936	1.2	12.6	5.2			15.7	71																		
Catalogo		2120074	2401051	2120046	2120637	...																				

Figura 8. Contenido archivo precipitaciones mensuales (elaboración propia)

En el departamento de Cundinamarca, la entidad regional CAR tiene alrededor de 95 estaciones hidrometeorológicas destinadas a la medición de diferentes factores climáticos como: niveles de precipitación, temperatura, humedad, brillo solar entre otras mediciones. Utilizando la información referente a la ubicación de las estaciones hidrometeorológicas (latitud, longitud) se realizó un mapa referenciando la distribución de las estaciones en el territorio de Cundinamarca, como se puede observar en la Figura 9.

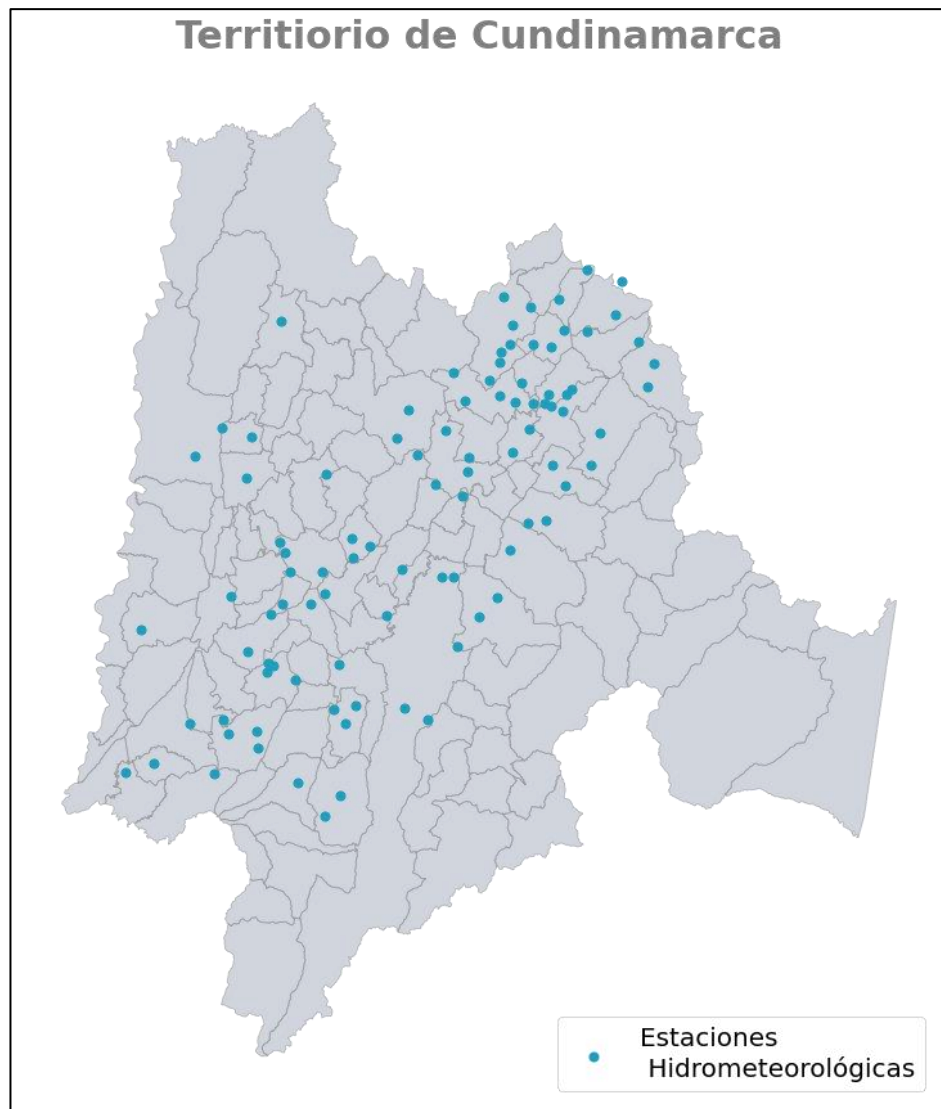


Figura 9. Mapa estaciones hidrometeorológicas (elaboración propia)

Utilizando los datos de las estaciones hidrometeorológicas registrados en la fuente de datos, se realizó la construcción de un catálogo de estaciones, para cada estación se registraron las siguientes propiedades: código estación, nombre estación, latitud, longitud, ciudad. El catálogo puede ser consultado en el Anexo I del presente trabajo.

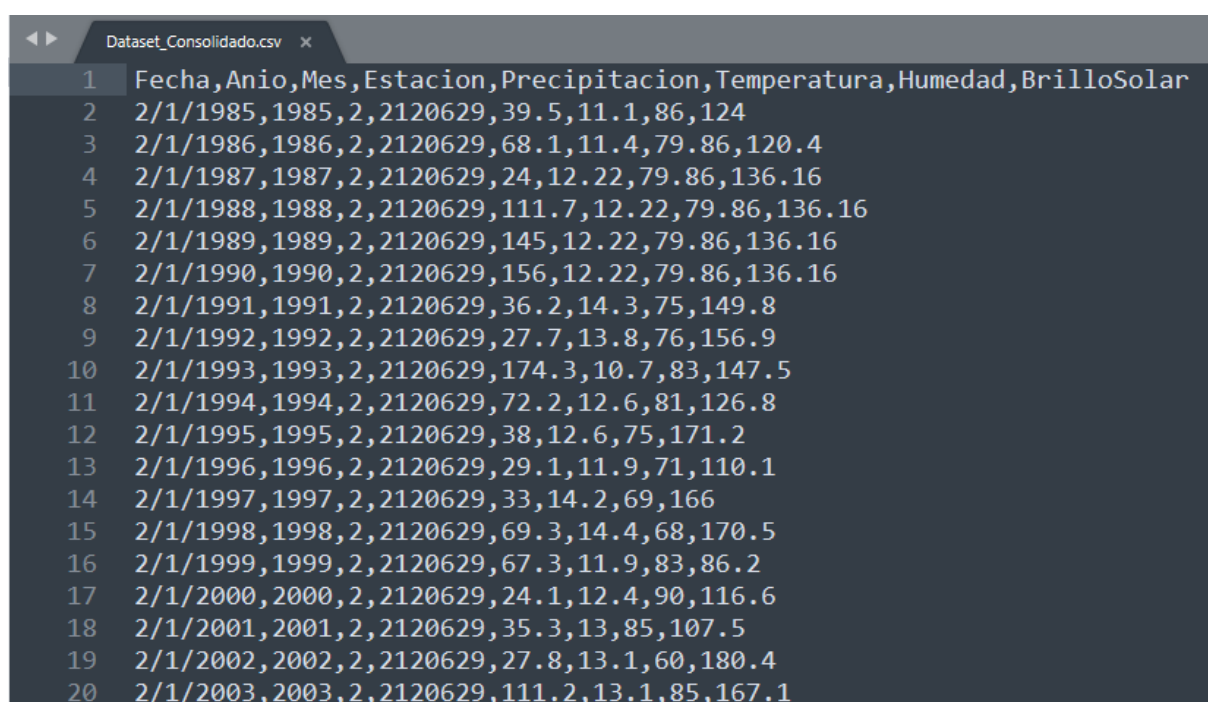
A continuación, se describirán las actividades realizadas en las fases de la metodología SEMMA.

4.2 Muestreo

Previo a desarrollar las actividades de muestreo es necesario realizar actividades de transformación en la fuente de datos, principalmente con el propósito de unificar los datos de las estaciones para cada una de las mediciones registradas y establecer un formato para su posterior análisis. El formato seleccionado es valores separados por comas, mejor conocido

como (csv) por sus siglas en inglés. El formato csv permitirá la lectura y análisis de los datos utilizando lenguajes de programación que tienen librerías especializadas en machine learning como Python o R.

La transformación inicial de la fuente de datos permitirá obtener un archivo de formato csv para cada una de las mediciones climáticas (precipitaciones, temperatura, humedad, brillo solar). Posterior a la generación de los archivos csv, se realizará un proceso de unificación de las mediciones generando un archivo consolidado. La estructura del archivo generado contendrá en la primera línea los encabezados (fecha, año, mes, estación, precipitación, temperatura, humedad, brillo solar), las filas restantes registran los datos de las mediciones mensuales. En la Figura 10 se puede observar la estructura del archivo consolidado.



```
Dataset_Consolidado.csv x
1 Fecha,Anio,Mes,Estacion,Precipitacion,Temperatura,Humedad,BrilloSolar
2 2/1/1985,1985,2,2120629,39.5,11.1,86,124
3 2/1/1986,1986,2,2120629,68.1,11.4,79.86,120.4
4 2/1/1987,1987,2,2120629,24,12.22,79.86,136.16
5 2/1/1988,1988,2,2120629,111.7,12.22,79.86,136.16
6 2/1/1989,1989,2,2120629,145,12.22,79.86,136.16
7 2/1/1990,1990,2,2120629,156,12.22,79.86,136.16
8 2/1/1991,1991,2,2120629,36.2,14.3,75,149.8
9 2/1/1992,1992,2,2120629,27.7,13.8,76,156.9
10 2/1/1993,1993,2,2120629,174.3,10.7,83,147.5
11 2/1/1994,1994,2,2120629,72.2,12.6,81,126.8
12 2/1/1995,1995,2,2120629,38,12.6,75,171.2
13 2/1/1996,1996,2,2120629,29.1,11.9,71,110.1
14 2/1/1997,1997,2,2120629,33,14.2,69,166
15 2/1/1998,1998,2,2120629,69.3,14.4,68,170.5
16 2/1/1999,1999,2,2120629,67.3,11.9,83,86.2
17 2/1/2000,2000,2,2120629,24.1,12.4,90,116.6
18 2/1/2001,2001,2,2120629,35.3,13,85,107.5
19 2/1/2002,2002,2,2120629,27.8,13.1,60,180.4
20 2/1/2003,2003,2,2120629,111.2,13.1,85,167.1
```

Figura 10. Archivo consolidado (elaboración propia)

Inicialmente, en las actividades de muestreo se realizará la extracción de un conjunto aleatorio de 100 registros, utilizando la muestra se desarrollará una inspección sobre los valores (valores frontera, valores faltantes, valores atípicos) con el propósito de determinar la calidad de los datos registrados por las estaciones climáticas.

En la Figura 11 se puede observar la representación gráfica de los datos contenidos en la muestra. Los histogramas generados permiten identificar características como: rango de valores y distribución de los datos para cada una de las características climáticas registradas.

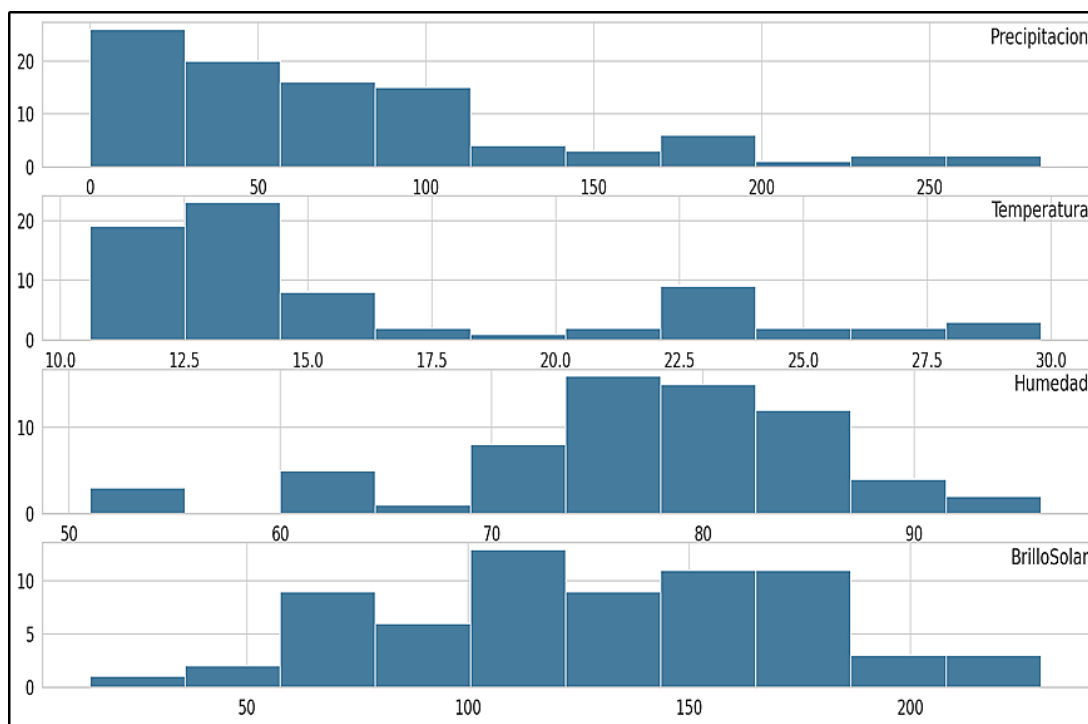


Figura 11. Histogramas características climáticas (elaboración propia)

Utilizando la información de la Figura 11, se obtienen los resultados registrados en la Tabla 2.

Tabla 2. Análisis valores muestra

Característica Climática	Descripción
Precipitación	Unidad de Medida: milímetros de agua (mm) o litros de agua por unidad de superficie (m ²). Es decir 1 milímetro de agua (mm) de lluvia equivale a 1 L de agua por m ² .
	Rango Valores: [0- 282]
	Valores atípicos: La muestra no tiene
	Cantidad valores nulos: 5%
Temperatura	Unidad de Medida: grados centígrados (°C)
	Rango Valores: [10.6 - 39.8]
	Valores atípicos: La muestra no tiene
	Cantidad valores nulos: 29 %
Humedad Relativa	Unidad de Medida: porcentaje (%)
	Rango Valores: [51 - 96]
	Valores atípicos: La muestra no tiene
	Cantidad valores nulos: 34%
Brillo Solar	Unidad de Medida: horas (hrs)
	Rango Valores: [14.6 - 229.5]
	Valores atípicos: La muestra no tiene
	Cantidad valores nulos: 32%

Fuente: elaboración propia.

Acorde a la información generada en la Tabla 2, los valores de cada una de las características se encuentran en rangos válidos, no se encuentran valores negativos ni valores atípicos en la muestra. En la siguiente fase de la metodología se continuará el análisis con la totalidad del conjunto de datos.

4.3 Exploración

En la fase de exploración se realizarán inspecciones utilizando la totalidad de los datos, el conjunto de datos analizado contiene 10656 registros. Inicialmente se realizará una inspección de calidad de datos (valores nulos, valores atípicos, rango de valores). Posteriormente, se crearán gráficos de dispersión y un análisis de correlación con el propósito de identificar las variables más influyentes respecto a la variable a predecir. En este caso particular la variable dependiente corresponde a los niveles de precipitación mensual.

La inspección de calidad del conjunto de datos se realizará utilizando histogramas, este tipo de gráfico permitirá visualizar los rangos de los valores y la distribución de los datos.

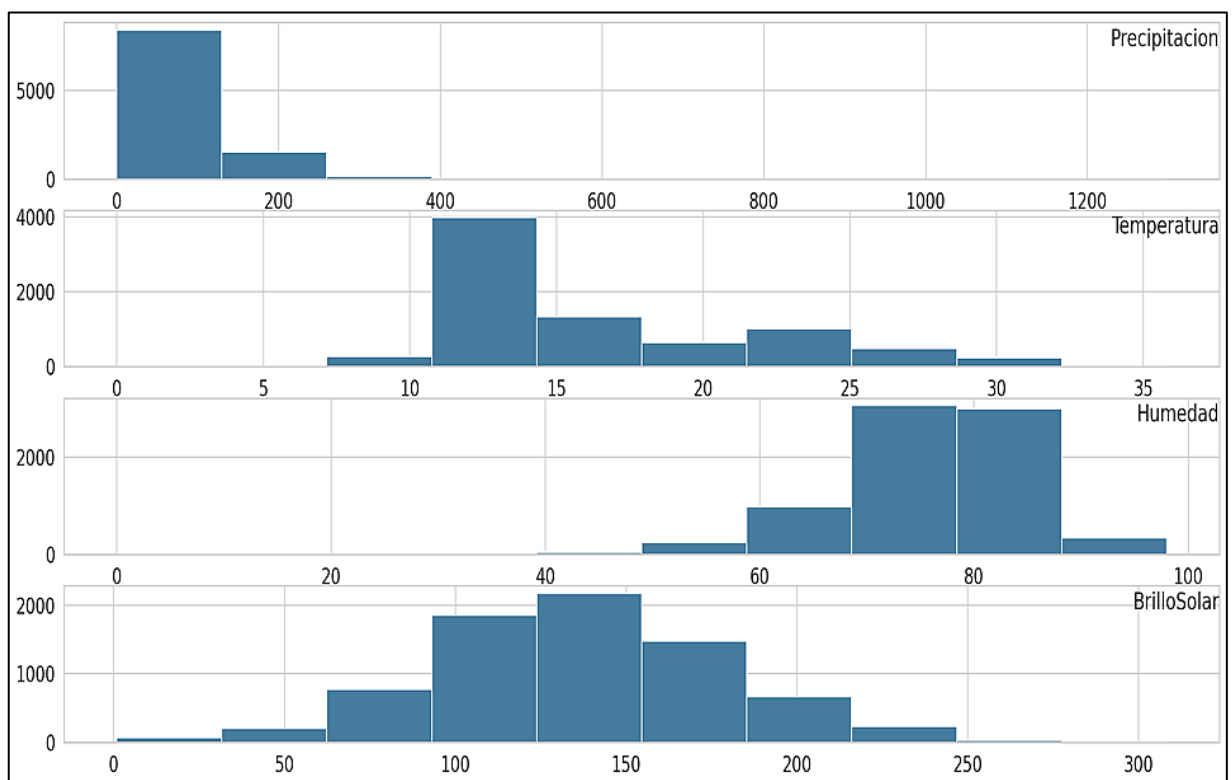


Figura 12. Histograma características climáticas conjunto de datos total (elaboración propia)

En el histograma de la Figura 12, se puede observar la distribución de los datos para cada una de las características climáticas, particularmente se observa como la distribución de los datos en la gráfica de precipitación presenta valores atípicos correspondientes a registros superiores a 1000 mm mensuales, los ajustes correspondientes a este tipo de valores

anómalos serán revisado en la siguiente fase de la metodología. Utilizando la información generada en la Figura 12, se obtienen los siguientes resultados registrados en la Tabla 3.

Tabla 3. Análisis valores conjunto de datos

Característica Climática	Descripción
Precipitación	Unidad de medida: milímetros de agua (mm)
	Rango valores: [0 - 1298]
	Valor promedio: 78.16
	Valores atípicos: Se encuentra el valor 1298 mm, esta cifra será considerada como atípica ya que su valor es ampliamente superior respecto a la media de 78.16 mm
	Cantidad valores nulos: 4.65%
Temperatura	Unidad de medida: grados centígrados (°C)
	Rango valores: [0 - 35.8]
	Valor promedio: 16.41
	Valores atípicos: No presenta
	Cantidad valores nulos: 25.11%
Humedad Relativa	Unidad de medida: porcentaje (%)
	Rango valores: [0 - 98]
	Valor promedio: 76.09
	Valores atípicos: No presenta
	Cantidad valores nulos: 27.91%
Brillo Solar	Unidad de medida: horas (hrs)
	Rango valores: [0.9 - 308.3]
	Valor promedio: 136.68
	Valores atípicos: No presenta
	Cantidad valores nulos: 29.99%

Fuente: elaboración propia.

Acorde a la información generada en la Tabla 3, los valores de cada una de las características se encuentran en rangos válidos, en la siguiente fase de la metodología se realizarán los ajustes correspondientes para tratar los valores atípicos y los valores nulos identificados en esta etapa de la metodología.

A continuación, se elaborarán gráficos de dispersión entre las características climáticas con el objetivo de identificar posibles correlaciones entre las variables analizadas.

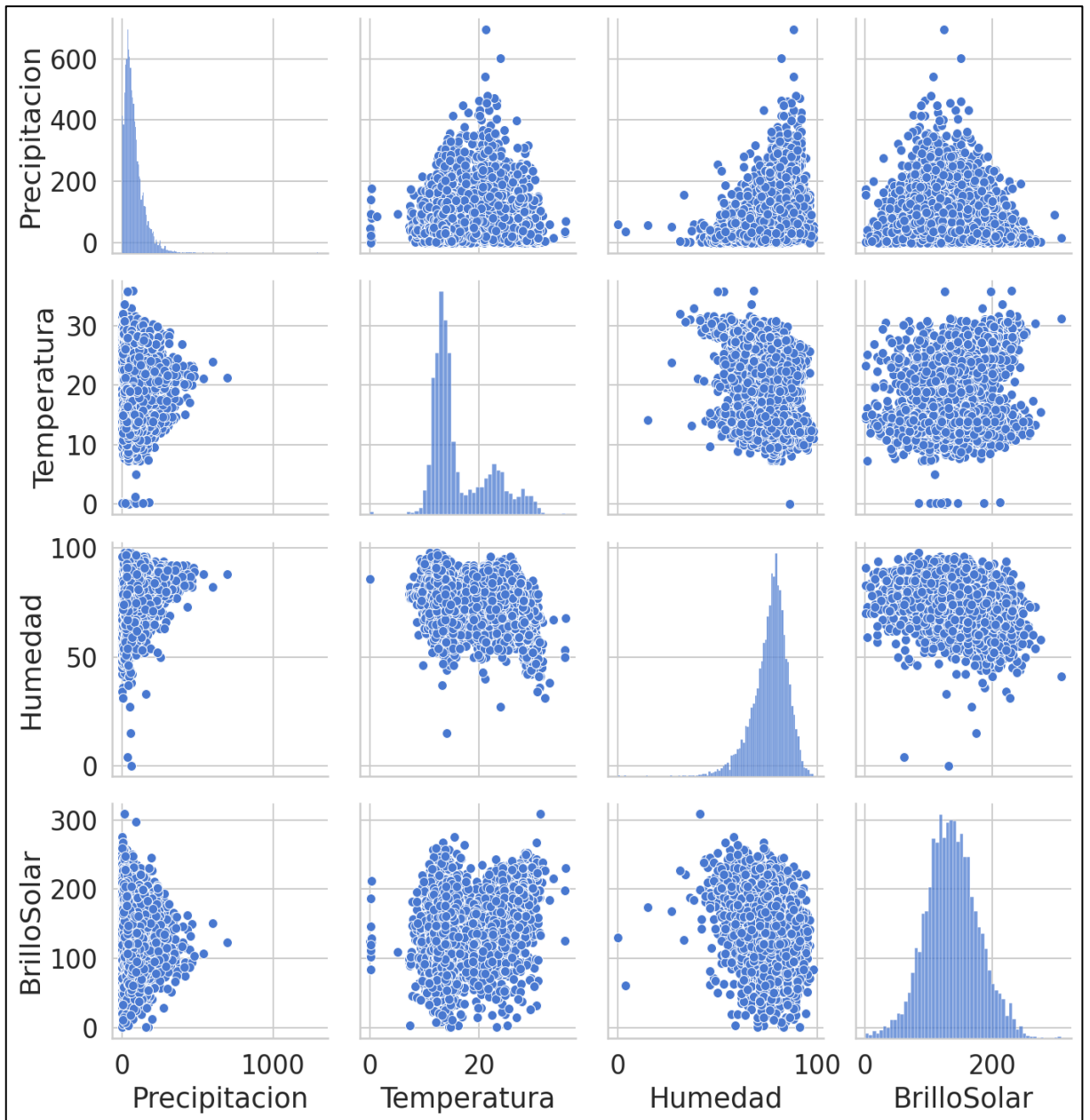


Figura 13. Diagramas de dispersión características climáticas (elaboración propia)

En la Figura 13, se puede observar como en los gráficos de dispersión entre las características (precipitación, humedad) y (precipitación, temperatura) probablemente presenten una correlación positiva. En el caso de las características (precipitación, brillo solar) se observa una aparente correlación negativa. Con el propósito de verificar las correlaciones mencionadas se realizará el cálculo de la matriz de correlación para las características climáticas y de esta forma analizar e identificar las variables más relevantes para el cálculo de la variable precipitación. La matriz de correlación se combinará con un gráfico de mapa de calor, en el cual se asociará el grado de correlación a la intensidad del color, representando un grado de correlación mayor las tonalidades más oscuras.

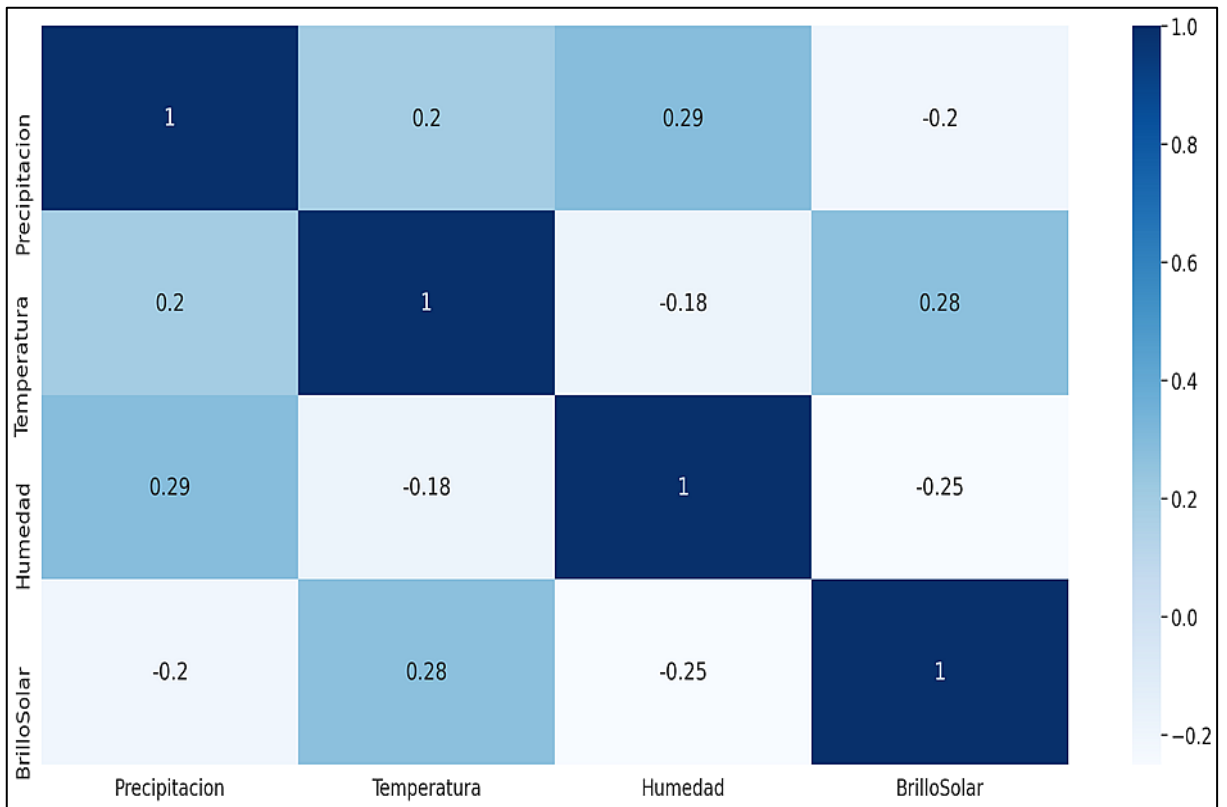


Figura 14. Matriz de correlación y mapa de calor características climáticas (elaboración propia)

En la Figura 14 se puede observar los coeficientes asociados a la matriz de correlación generada para las características climáticas. Los coeficientes de correlación lineal tienen asociado una escala de color, representando las tonalidades más oscuras los mayores valores de correlación. Analizando los resultados obtenidos de la matriz de correlación respecto a la precipitación se observa que las características temperatura y humedad relativa generan una correlación positiva baja, el aumento en la temperatura y en la humedad relativa generará un aumento en el nivel de precipitación. El brillo solar generó una correlación negativa respecto a la precipitación, un aumento en el valor de la variable brillo solar (horas de sol mensuales) producirá un descenso en los niveles de precipitación (mm de agua mensuales). La característica climática más representativa respecto a la precipitación es la humedad relativa con un coeficiente de correlación positivo de 0.29. La relación entre los niveles de precipitación y la humedad relativa se ha estudiado en múltiples escenarios con el objetivo de mejorar los sistemas de predicción climática (Singh et al., 2019)

4.4 Ajustes

En la fase de ajustes se realizarán las transformaciones correspondientes en el conjunto de datos con el objetivo de preparar y estructurar los datos para la ejecución del modelo en la siguiente fase de la metodología. Las transformaciones en los datos comprenden tareas como: corrección de valores atípicos, asignación de valores nulos, creación de columnas adicionales, filtrado de datos, combinaciones de datos.

A continuación, se realizará la descripción de las transformaciones y ajustes realizados en el conjunto de datos de características climáticas utilizado en el presente trabajo.

- a) Conversión de formato:** La transformación de los datos inició con la conversión de formato Microsoft Excel (.xlsx) utilizado en la fuente de datos hacia un formato de valores separados por comas (.csv). La tarea de conversión también incluyó la transposición de valores con el objetivo de convertir el conjunto de datos en una serie temporal, cada registro del conjunto de datos representará un instante de tiempo (t) en un rango de fechas. Esta actividad se realizó para cada uno de los archivos que almacenaban las mediciones de las características climáticas (precipitación, temperatura, humedad, brillo solar) generando para cada una el correspondiente archivo (.csv). En la Figura 15 se puede observar el archivo origen (izquierda) y el archivo generado en este caso para la precipitación (derecha).

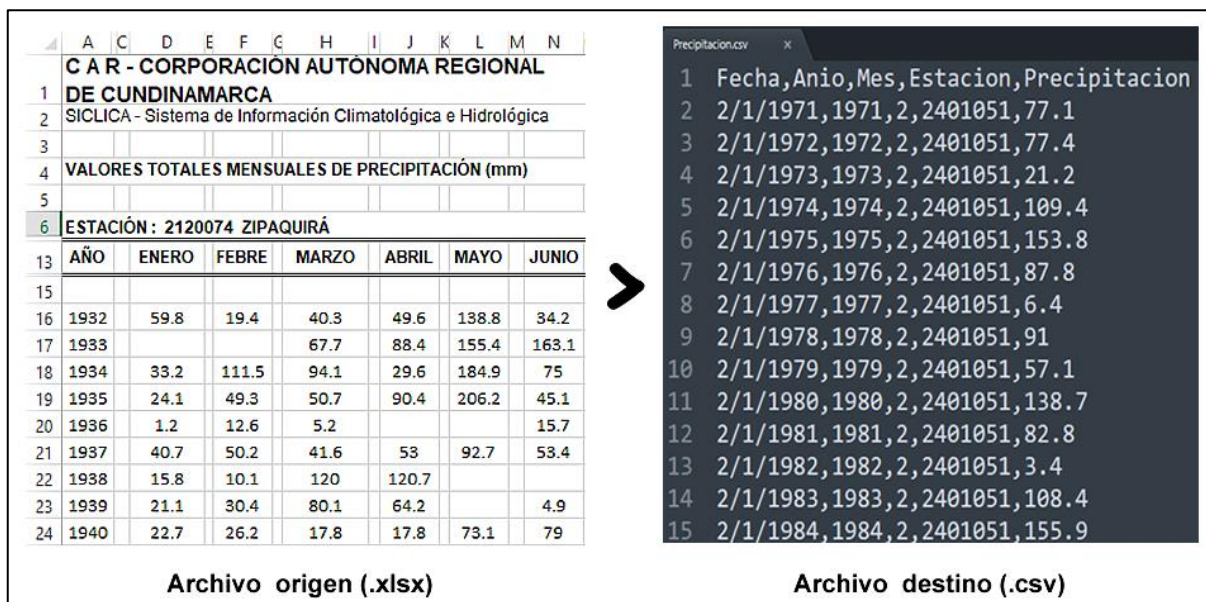


Figura 15. Conversión formato y estructura archivo precipitación (elaboración propia)

- b) Combinación archivos:** La actividad tiene como propósito la unificación de los archivos (.csv) de características climáticas generados en el paso anterior. Esta

unificación de datos es requerida para consolidar el conjunto de datos. La consolidación es un prerrequisito para la ejecución del modelo que será utilizado en la siguiente fase de la metodología. El proceso de combinación realizará la unificación de los archivos de precipitación, temperatura, humedad relativa y brillo solar. La estructura del archivo consolidado contendrá en la primera línea los encabezados (fecha, año, mes, estación, precipitación, temperatura, humedad, brillo solar), las filas restantes registran los datos de las mediciones mensuales.

- c) **Tratamiento valores atípicos y nulos:** Los valores identificados como atípicos o nulos en la fase 4.3, serán ajustados con el propósito de finalizar el alistamiento del conjunto de datos para la ejecución de los modelos predictivos en la siguiente fase de la metodología.

Los valores atípicos serán reemplazados por el valor máximo o mínimo del rango definido de valores legítimos que tenga la característica climática analizada, dependiendo la magnitud del valor será el ajuste correspondiente, como se observa en la descripción del procedimiento en la Figura 16.

Transformación Valores Atípicos

Rango Valores: [0,120]

Valor atípico: 300 → El valor se ajusta a la cantidad máxima del rango → **Nuevo Valor :** 120

Valor atípico: -100 → El valor se ajusta a la cantidad mínima del rango → **Nuevo Valor :** 0

Figura 16. Transformación valores atípicos (elaboración propia)

Los valores nulos serán reemplazados por el valor del promedio de la característica climática para la estación y mes correspondiente, como se observa en la descripción del procedimiento en la Figura 17.

Transformación Valores Nulos

Muestra conjunto de datos

Fecha	Año	Mes	Estacion	Precipitacion	Temperatura	Humedad	BrilloSolar
2/1/1985	1985	2	2120629	30	11.1	86	124
2/1/1986	1986	2	2120629	20	11.4	79.86	120.4
2/1/1987	1987	2	2120629	•	12.22	79.86	136.16
2/1/1988	1988	2	1112828	111.7	12.22	79.86	136.16
2/1/1989	1989	2	2120629	10	12.22	79.86	136.16

Estacion 2120629, en el mes 2 presenta un valor nulo • en la precipitación
Se utilizarán los datos de la estación 2120629, mes 2 para calcular el promedio

Valor Promedio: $[30+ 20+10]/3 = 20$

Reemplazar valor nulo con el promedio

Fecha	Año	Mes	Estacion	Precipitacion	Temperatura	Humedad	BrilloSolar
2/1/1987	1987	2	2120629	20	12.22	79.86	136.16

Figura 17. Procedimiento ajuste valores nulos (elaboración propia)

d) Selección datos para el modelo: La completitud de los datos es un factor clave para la construcción y evaluación de un modelo de *machine learning*. Por este motivo, se realizará el análisis del conjunto de datos, seleccionando las estaciones que contengan datos para todas las características climáticas (temperatura, humedad, brillo solar, precipitación) y de esta forma finalizar el alistamiento de los datos.

Realizando la inspección de la totalidad del conjunto de datos, se encontró que, de las 95 estaciones hidrometeorológicas ubicadas en el departamento de Cundinamarca, 19 estaciones contienen datos para todas las características climáticas. En la Tabla 4 se registran las estaciones seleccionadas que serán utilizadas como fuente de datos en la elaboración del modelo.

Tabla 4. Estaciones hidrometeorológicas seleccionadas

Código estación	ID zona	Nombre estación	Latitud	Longitud	Altura (m.s.n.m)	Municipio
2119514	19	Universidad fusagasugá udec	4° 20' 8,484" N	74° 22' 12,900" W	1725	Fusagasugá
2120540	4	Checua	5° 7' 0,696" N	73° 51' 41,184" W	2603	Nemocón
2120557	10	La primavera	4° 51' 28,296" N	74° 12' 42,984" W	2623	Subachoque
2120562	8	Guatavita	4° 54' 31,897" N	73° 51' 56,196" W	2679	Guatavita
2120629	9	Venecia	4° 50' 38,904" N	74° 23' 58,308" W	2686	Facatativá
2120630	15	Doña juana	4° 30' 3,097" N	74° 8' 14,496" W	2800	Bogotá, d.c.
2120641	16	Esc samper madrid	4° 28' 23,303" N	74° 32' 8,592" W	541	Apulo
2120644	18	Argelia	4° 21' 27,397" N	74° 45' 2,808" W	315	Ricaurte
2120646	14	Mesitas	4° 34' 43,787" N	74° 26' 20,616" W	1057	El colegio
2306507	3	Inst. Agric. Esc. Vocacional	5° 9' 31,212" N	74° 7' 43,788" W	1940	Pacho
2306517	6	Guaduas	5° 3' 24,299" N	74° 35' 53,808" W	1023	Guaduas

2401519	2	Novilleros	5° 20' 8,412" N	73° 47' 5,712" W	2590	Villa de san diego de ubate
2120548	5	Iberia la	5° 2' 13,4" N	73° 43' 27,5" W	2721	Choconta
2120559	11	Apto guaimaral- usta	4° 47' 27,600" N	74° 3' 12,492" W	2603	Bogotá, d.c.
2120561	17	Munal el	4° 33' 2,6" N	74° 1' 5,4" W	2565	Sibaté
2401513	1	Simijaca	5° 30' 43,597" N	73° 51' 28,908" W	2622	Simijaca
2120639	13	La mesa	4° 37' 28,704" N	74° 28' 54,300" W	1215	La mesa
2120647	12	La esperanza	4° 42' 21,888" N	74° 25' 49,008" W	1391	La mesa
2306516	7	El acomodo	5° 1' 3,685" N	74° 18' 34,812" W	1394	La vega

Fuente: elaboración propia.

4.5 Modelamiento

En la fase de modelamiento se realizarán una serie de actividades orientadas a seleccionar, configurar, entrenar y probar el modelo de machine learning que será implementado para dar solución a la problemática de predicción de niveles de precipitación expuesta en el presente trabajo.

En la Tabla 5 se realiza una descripción de las actividades desarrolladas durante el proceso de modelamiento.

Tabla 5. Fases construcción modelo

Actividad	Descripción
Selección modelo	Acorde a los modelos de redes neuronales analizados en las secciones 2.4 y 2.6 del presente documento, el modelo más apropiado para realizar predicciones en series temporales de eventos son las redes neuronales LSTM (<i>Long Short Term Memory</i>), este tipo de redes se caracterizan por tener una memoria temporal que les permite recordar dependencias de pasos previos e incorporarlas como entradas para las siguientes etapas de procesamiento de la red neuronal.

Elección software	Se realizará la implementación del modelo LSTM utilizando el lenguaje de programación Python, incluyendo las librerías pandas, SciKit-Learn, TensorFlow 2.0 y Keras. Estas librerías son ampliamente utilizadas para la construcción de modelos de machine learning. En el componente visual se utilizarán las librerías matplotlib, geopandas, descartes, shapely con el propósito de graficar el mapa del territorio y la librería altair a fin de adicionar comportamientos interactivos como las etiquetas de descripción emergente (<i>tooltips</i>).																				
Ajustes conjunto de datos	Representa las modificaciones realizadas en el conjunto de datos como: normalización de variables, eliminación variable resultado (niveles de precipitación) en los conjuntos de datos de entrenamiento y pruebas.																				
Evaluación configuración inicial modelo	<p>Con el objetivo de seleccionar la configuración inicial del modelo, se realizó la definición de tres modelos candidatos con diferentes configuraciones. Posteriormente, se evaluaron los resultados de los modelos respecto a los datos del conjunto de pruebas. El análisis de precisión realizado corresponde al porcentaje de estaciones climáticas que presentan un diferencial en los niveles de precipitación menor a 40 mm respecto a los datos reales. La Figura 18 contiene los resultados de la comparación de modelos realizada, en la cual se puede evidenciar que el modelo A registra el mayor nivel de precisión (44.2%). Por este motivo, los hiperparámetros del modelo A serán utilizados para la construcción del modelo predictivo del presente trabajo.</p> <div data-bbox="555 1292 1417 1608" style="border: 1px solid black; padding: 10px; margin: 10px auto; width: fit-content;"> <p style="text-align: center;">Comparación precisión modelos</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr style="background-color: #4a7ebb; color: white;"> <th>Hiperparámetros</th> <th>Modelo A</th> <th>Modelo B</th> <th>Modelo C</th> </tr> </thead> <tbody> <tr> <td>Épocas</td> <td style="text-align: center;">100</td> <td style="text-align: center;">200</td> <td style="text-align: center;">150</td> </tr> <tr> <td>Tamaño lote</td> <td style="text-align: center;">32</td> <td style="text-align: center;">12</td> <td style="text-align: center;">20</td> </tr> <tr> <td>% datos entrenamiento</td> <td style="text-align: center;">80%</td> <td style="text-align: center;">80%</td> <td style="text-align: center;">70%</td> </tr> <tr style="background-color: #d9d9d9;"> <td>% Precisión predicciones</td> <td style="text-align: center;">44.2%</td> <td style="text-align: center;">41.4%</td> <td style="text-align: center;">41.2%</td> </tr> </tbody> </table> </div> <p style="text-align: center; color: #4a7ebb; font-style: italic;">Figura 18. Comparativa precisión modelos (elaboración propia)</p>	Hiperparámetros	Modelo A	Modelo B	Modelo C	Épocas	100	200	150	Tamaño lote	32	12	20	% datos entrenamiento	80%	80%	70%	% Precisión predicciones	44.2%	41.4%	41.2%
Hiperparámetros	Modelo A	Modelo B	Modelo C																		
Épocas	100	200	150																		
Tamaño lote	32	12	20																		
% datos entrenamiento	80%	80%	70%																		
% Precisión predicciones	44.2%	41.4%	41.2%																		
Definición división conjunto de datos	Con el propósito de entrenar y probar el modelo se requiere dividir el conjunto de datos en 2 partes (entrenamiento y validación). La selección de estos porcentajes se realizó verificando que el modelo no presentara un comportamiento de sobreajuste (overfitting), con el propósito garantizar la capacidad de generalización del modelo en los cálculos de las predicciones.																				

	<p>En este trabajo se realizó la siguiente definición de porcentajes sobre el conjunto de datos para las fases entrenamiento y prueba del modelo:</p> <ul style="list-style-type: none"> • Datos entrenamiento: 80% • Datos validación: 20%
<p>Definición hiperparámetros modelo</p>	<p>En las tareas de aprendizaje automático los valores de configuración utilizados para iniciar la fase de entrenamiento reciben el nombre de hiperparámetros.</p> <p>Los hiperparámetros utilizados para la construcción del modelo LSTM son:</p> <ul style="list-style-type: none"> • Épocas: 100 • Tamaño Lote: 32 • Optimizador: ADAM (<i>Adaptive Moment Estimation</i>) • Función Objetivo: MSE (<i>Mean Squared Error</i>) • Porcentaje Validación Entrenamiento: 10% <p>A continuación, se realizará una descripción de los conceptos de los hiperparámetros:</p> <p>Épocas: Representa la cantidad de iteraciones en las cuales el modelo procesa la totalidad del conjunto de datos en la red neuronal. El número de épocas fue determinado comparando la precisión de los valores generados por el modelo respecto a los valores reales.</p> <p>Lote o Batch: Subdivisiones del conjunto de datos utilizadas para realizar el proceso de aprendizaje del modelo.</p> <p>Tamaño de Lote: Número total de muestras del conjunto de datos de entrenamiento contenidas en un lote.</p> <p>Porcentaje Validación Entrenamiento: Representa la fracción del conjunto de datos de entrenamiento que será utilizado como datos de validación. Esta porcentaje de datos no es utilizado en las actividades de entrenamiento, solamente será utilizado para validación del modelo utilizando la función objetivo (MSE).</p> <p>En la Figura 19 se puede observar la representación de los conceptos de época y lote.</p>

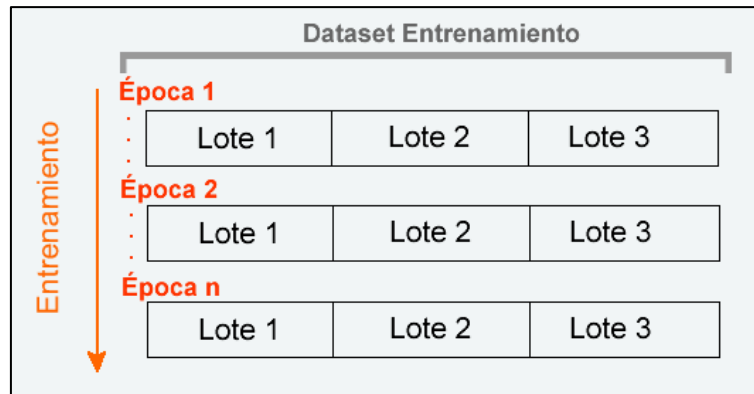


Figura 19. Representación conceptos época y lote (elaboración propia)

Optimizador: Procedimiento encargado de mejorar iterativamente los valores de los parámetros con el propósito de reducir el error generado por la red neuronal. El optimizador se apoya en la función objetivo para determinar si el error aumenta o disminuye. El optimizador ADAM (*Adaptive Moment Estimation*) es un algoritmo de optimización basado en gradientes de primer orden. ADAM es computacionalmente eficiente, no requiere grandes cantidades de memoria y es capaz de procesar problemas de grandes cantidades de datos y parámetros (Kingma & Ba, 2017). La selección de ADAM como optimizador se realizó examinando sus grandes capacidades de procesamiento y sus escasos requerimientos de cómputo.

Función Objetivo: Es el procedimiento utilizado para estimar la diferencia entre la predicción y el dato real. En este caso se utilizó la función MSE (*Mean Squared Error*) o error cuadrático medio, MSE mide el promedio de los errores al cuadrado es decir la diferencia entre los datos reales y los datos generados (predicción), la fórmula para el cálculo de la función MSE se puede observar en la Figura 20.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2$$

Figura 20. Fórmula MSE (Fandango, 2018)

<p>Compilación modelo</p>	<p>Proceso encargado de realizar la configuración del modelo para la fase de entrenamiento. La compilación utiliza los hiperparámetros de función objetivo y optimizador.</p>
---------------------------	---

<p>Entrenamiento modelo</p>	<p>Proceso encargado de realizar el entrenamiento, aprendizaje y evaluación del modelo. En esta etapa se utiliza el subconjunto de datos designados para el entrenamiento, los datos son divididos en 2 partes: X(características), Y(variable objetivo). En esta caso particular X representa el conjunto de las características climáticas (temperatura, humedad y brillo solar) y la sección Y es el nivel de precipitación registrado para las características climáticas mencionadas. Los hiperparámetros utilizados en la fase de entrenamiento son: épocas y tamaño de lote.</p> <p>Posterior a la finalización de la fase de entrenamiento el modelo creará parámetros para realizar el aprendizaje y evaluación de los resultados generados, este proceso será ejecutado para cada una de las épocas definidas en el modelo. La evaluación entre las salidas generadas por el modelo y los datos reales (Y) es realizada utilizando la función objetivo definida (MSE).</p>
<p>Predicciones Modelo</p>	<p>En la fase de predicción se realizan 2 actividades:</p> <ol style="list-style-type: none"> 1) Predicciones utilizando datos de prueba: Utilizando los datos designados para las pruebas (20% del conjunto de datos) se realizará el proceso de predicción y comparación de los datos reales frente a los resultados del modelo. En este caso particular se realizará la predicción del nivel de precipitación y posterior comparación del resultado obtenido respecto al nivel de precipitación real registrado. 2) Predicciones siguiente año: Procedimiento para calcular el nivel de precipitación mensual del año 2021. Para realizar el proceso predictivo el modelo tomará como entrada los datos de los últimos diez años para estimar la predicción correspondiente. El proceso predictivo será realizado para cada una de las 19 estaciones hidrometeorológicas definidas en la sección 4.4 del documento. Para cada estación se ejecutará el modelo predictivo utilizando los datos de cada mes de forma independiente, con el objetivo de mejorar la calidad de las predicciones realizadas. En la Figura 21 se puede observar el proceso general de predicción.

Fuente: elaboración propia.

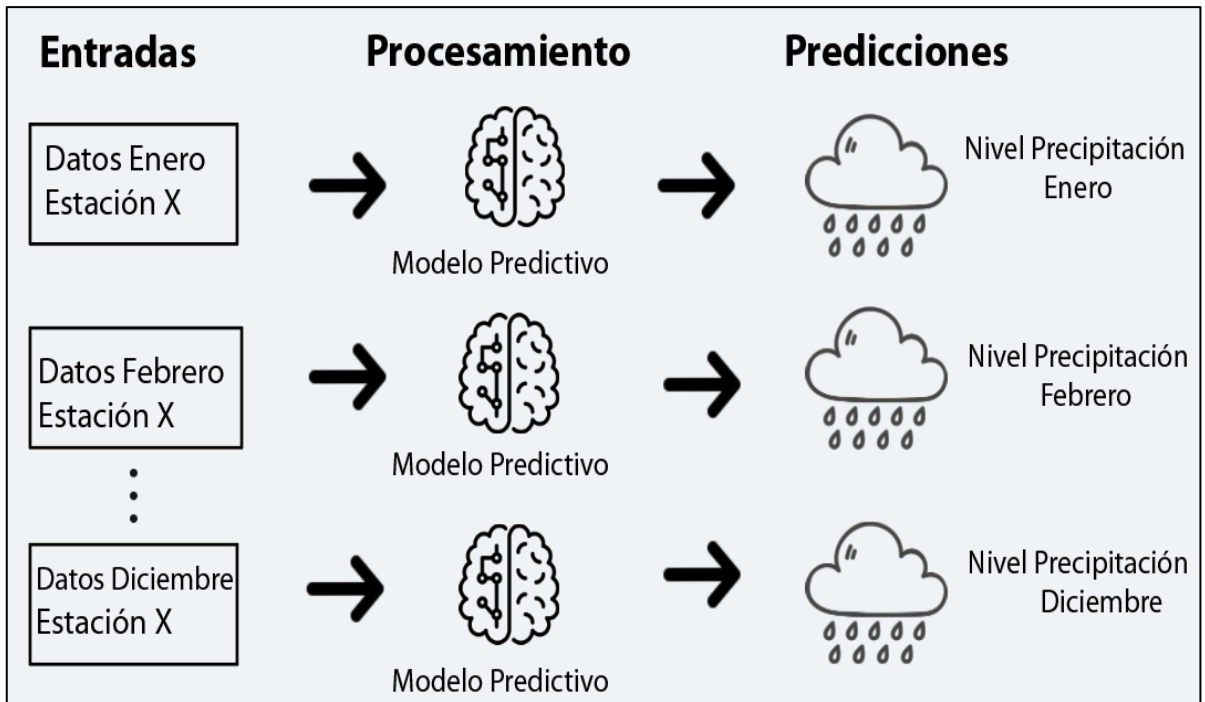


Figura 21. Proceso general predicción (elaboración propia)

En las actividades correspondientes a las fases de configuración de hiperparámetros y entrenamiento del modelo se analizaron los resultados generados con el propósito de estimar una predicción óptima, evitando escenarios de overfitting (modelo ajustado a los datos, incapaz de generalizar) y underfitting (modelo simple para explicar variaciones en los datos). En la Figura 22 se puede observar una representación de los escenarios de overfitting, underfitting y el escenario óptimo para un modelo machine learning.

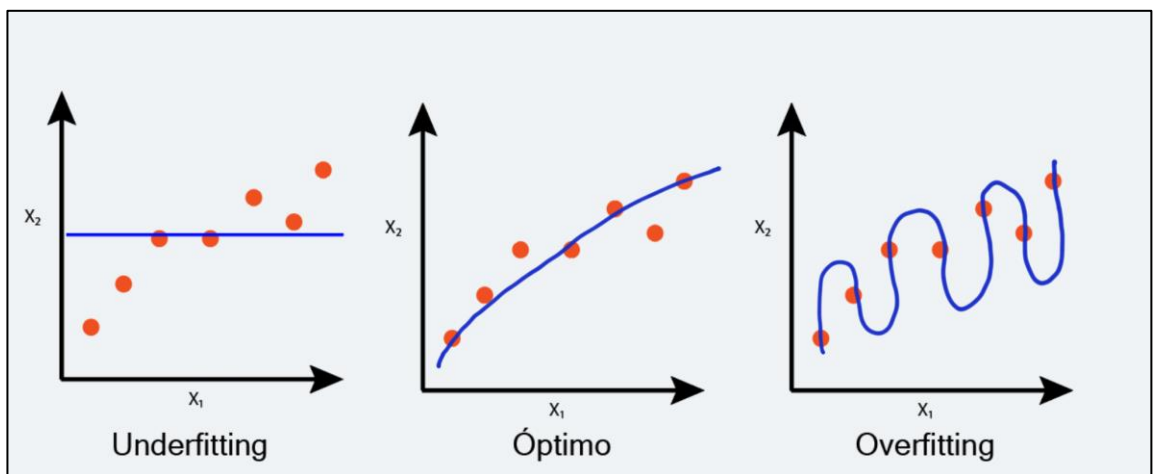


Figura 22. Escenarios de underfitting, overfitting y óptimo (elaboración propia)

4.6 Evaluación

La evaluación del modelo se realiza en las fase de entrenamiento y predicción. A continuación, se describirán los detalles de la evaluación del modelo en las fases mencionadas.

4.6.1 Evaluación fase entrenamiento

Durante la ejecución del proceso de entrenamiento se definió destinar el 10% del conjunto de datos de entrenamiento para realizar la evaluación de aprendizaje del modelo. El método utilizado para validar la calidad de las predicciones es la función objetivo (MSE), con esta función se calculará el diferencial entre las predicciones y los datos reales para determinar el grado de aprendizaje del modelo y la precisión de las estimaciones generadas. El procedimiento de evaluación es ejecutado para cada época definida en los hiperparámetros del modelo.

En la Figura 23 se puede observar el registro del error cuadrático medio durante las 100 épocas de entrenamiento definidas para el modelo utilizando los datos de la estación climática 2120630. En la gráfica se evidencia una reducción gradual del error a medida que incrementan las épocas procesadas. La disminución del error es consecuencia de la mejora progresiva en el aprendizaje del modelo. Las predicciones generadas en la época 100 son cercanas a los valores reales utilizados para la validación del modelo en la etapa de entrenamiento ya que el error cuadrático medio es 0.13.



Figura 23. Error cuadrático medio - fase entrenamiento (elaboración propia)

En la Tabla 6 se realizó el consolidado del cálculo de la función objetivo (MSE) que representa la diferencia entre las predicciones y los datos reales utilizados en la fase de entrenamiento para la validación del modelo.

Tabla 6. Evaluación MSE - fase entrenamiento

Estación	Función MSE
2119514	0.14
2120540	0.25
2120548	0.09
2120557	0.11
2120559	0.17
2120561	0.08
2120562	0.25
2120629	0.18
2120630	0.13
2120639	0.06
2120641	0.07
2120644	0.10
2120646	0.08
2120647	0.09
2306507	0.11
2306516	0.10
2306517	0.24
2401513	0.09
2401519	0.22
Promedio SME	0.13

Fuente: elaboración propia

En la evaluación del modelo realizada en la fase de entrenamiento se encontró que el promedio del error cuadrático para las 19 estaciones analizadas es 0.13. Esta cifra indica que las estimaciones realizadas por el modelo son muy cercanas a los datos reales de la etapa de entrenamiento.

4.6.2 Evaluación fase predicciones

La evaluación del modelo en la fase de predicciones se realizó utilizando el 20% del conjunto de datos reservado para evaluar las predicciones generadas. Inicialmente se ejecutó el modelo para cada una de las 19 estaciones y posteriormente se realizó la comparación de los valores reales de precipitación frente a los valores generados por el modelo. En la Figura 24 se observa los valores de la precipitación en los meses analizados, el gráfico contiene los datos históricos (cifras reales dataset entrenamiento), datos reales (cifras reales dataset pruebas) y las predicciones (datos generados por el modelo). En términos generales se puede

observar que las estimaciones realizadas por el modelo de los niveles de precipitación siguen la tendencia general de los datos reales registrados en la estación hidrometeorológica.

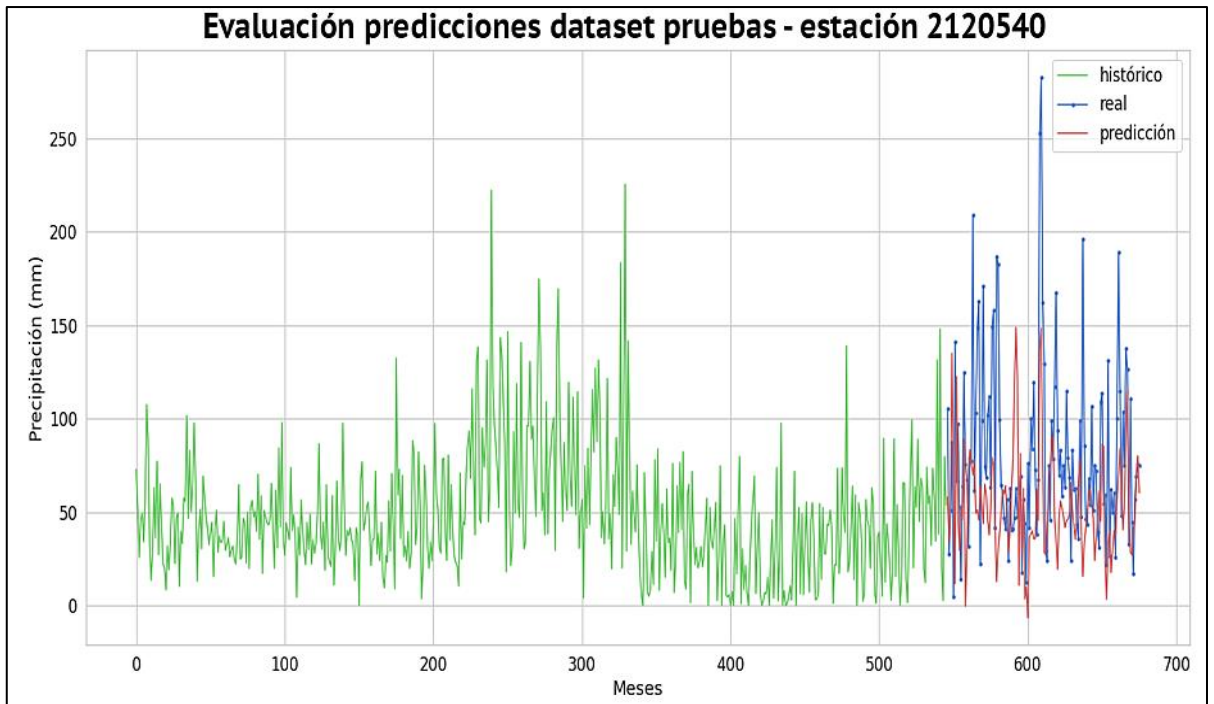


Figura 24. Evaluación predicciones precipitación (elaboración propia)

Realizando un acercamiento de los valores reales y las predicciones en la Figura 25 se observa como las predicciones generadas por el modelo registran valores de precipitación aproximados y sigue la tendencia general de los datos reales. Las diferencias más notables en las predicciones se presentan en los valores extremos en los cuales el modelo genera valores con diferencias relevantes respecto a los datos reales.

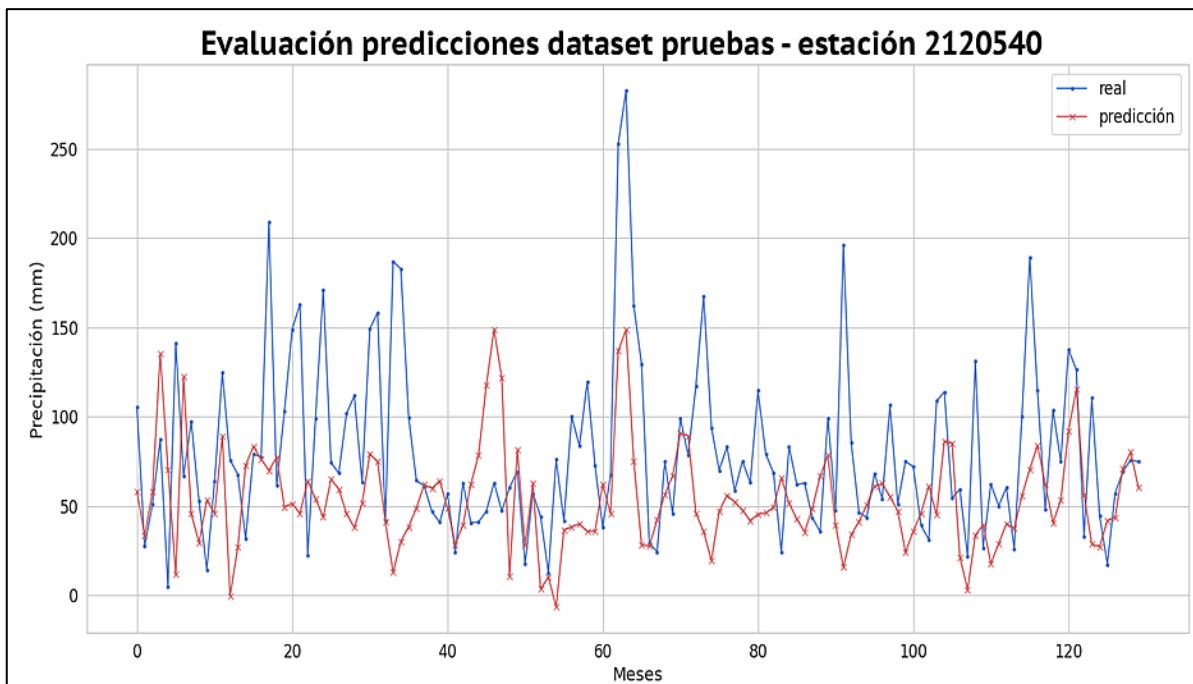


Figura 25. Comparación valores reales y predicciones niveles de precipitación (elaboración propia)

Con el objetivo de comparar los valores generados por el modelo y los datos reales se realizó un proceso de clasificación en el cual se evaluaron las diferencias en los niveles de precipitación, en la clasificación se definieron dos categorías: valores con una diferencia absoluta menor a 40 mm y valores con diferencia absoluta mayor a 40mm. Se estableció el valor 40 ya que es un valor intermedio en los datos de precipitación registrados. En la Figura 26 se observa la evaluación de la precisión en las predicciones realizadas sobre el conjunto de datos de prueba en la cual el 59.2% de las predicciones tienen una proximidad al valor real inferior a 40 mm de agua, mientras que un 40.8% de las predicciones tienen un diferencial mayor a 40 mm respecto a los datos reales de precipitación.

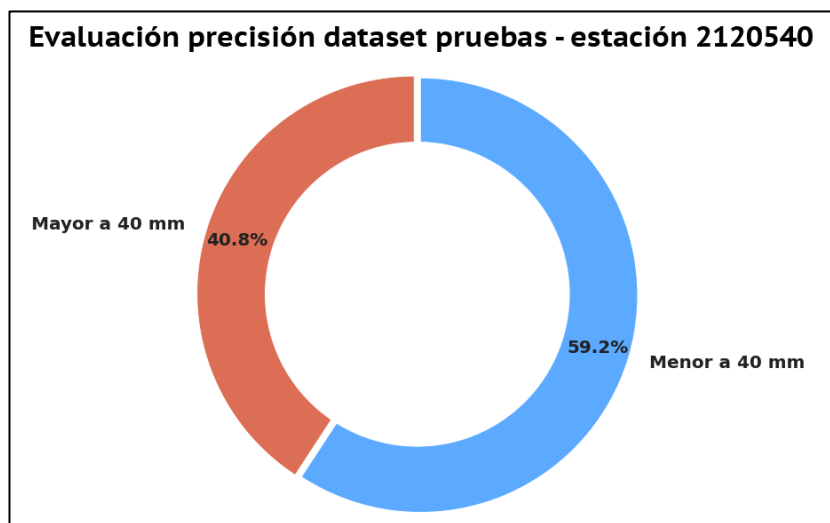


Figura 26. Evaluación precisión predicciones - dataset pruebas (elaboración propia)

En la Tabla 7 se registran los porcentajes de precisión para el total de las estaciones analizadas. Acorde a la cálculos de precisión realizados para cada una de las estaciones la precisión de las predicciones generadas con un diferencial menor a 40 mm corresponde a 44.2% de los datos analizados y el restante 55.7% de los datos presentan un diferencial mayor a 40 mm. Las diferencias mayores a 40mm se generan en mayor medida en los valores elevados de precipitación en los cuales el modelo no realiza una predicción aproximada al valor real generando un diferencial de valores significativo como se evidenció en las predicciones de la Figura 25.

Tabla 7. Registro precisión predicciones

Estación	Menor a 40 mm	Mayor a 40 mm
2119514	58.7%	41.3%
2120540	59.2%	40.8%
2120548	53.1%	46.9%
2120557	51.7%	48.3%
2120559	48.7%	51.3%
2120561	65.9%	34.1%
2120562	60.9%	39.1%
2120629	54.1%	45.9%
2120630	53.3%	46.7%
2120639	27.4%	72.6%
2120641	27.6%	72.4%
2120644	41.4%	58.6%
2120646	28.3%	71.7%
2120647	38%	62%
2306507	36.7%	63.3%
2306516	16.1%	83.9%
2306517	28.9%	71.1%
2401513	37.1%	62.9%
2401519	53.4%	46.6%
Promedio	44.24 %	55.7 %

Fuente: elaboración propia

4.6.3 Generación predicciones

Utilizando como fuente de datos los registros mensuales de las características climáticas (temperatura, humedad y brillo solar) de los últimos diez años pertenecientes a las estaciones hidrometeorológicas se realizó la ejecución del modelo con el propósito de generar las predicciones de niveles de precipitación para el año 2021. En las Figuras 27-29 se puede observar los niveles de precipitación históricos (120 meses) y los valores de precipitación generados por el modelo (últimos doce meses).

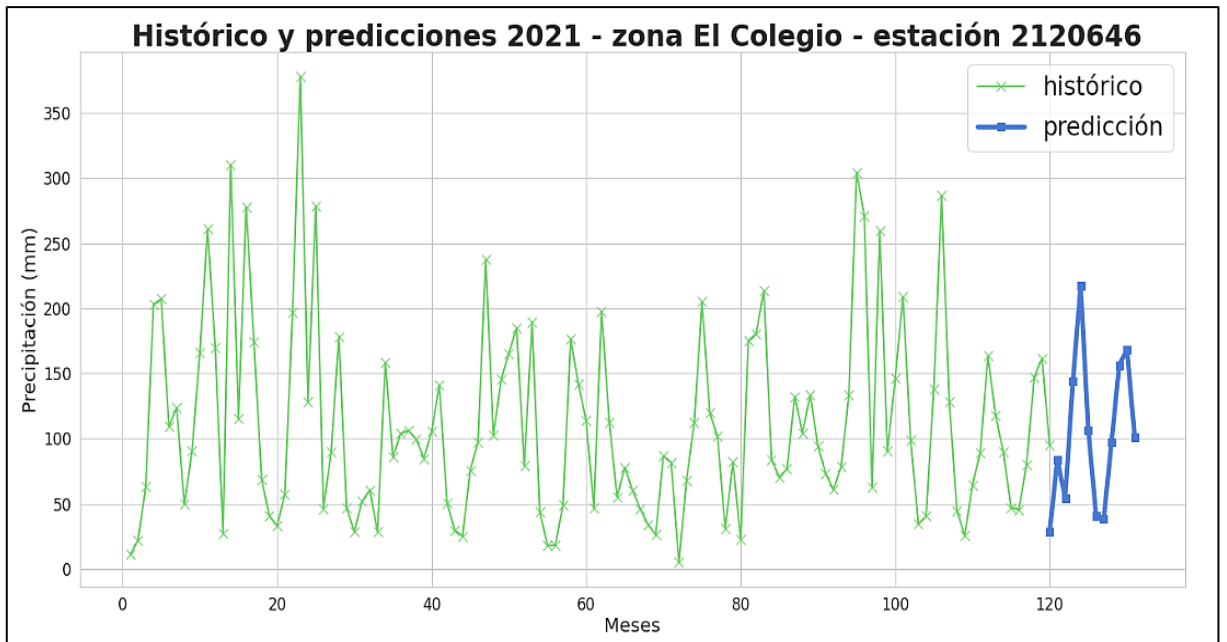


Figura 27. Histórico y predicciones de precipitación - estación 2120646 (elaboración propia)

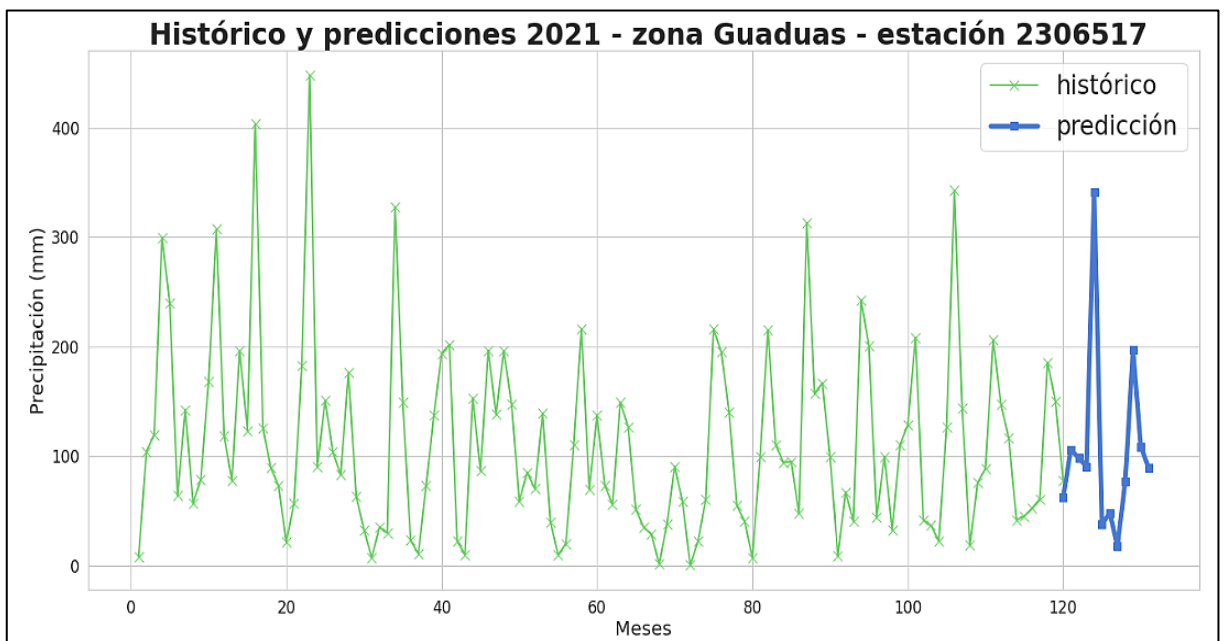


Figura 28. Histórico y predicciones de precipitación - estación 2306517 (elaboración propia)

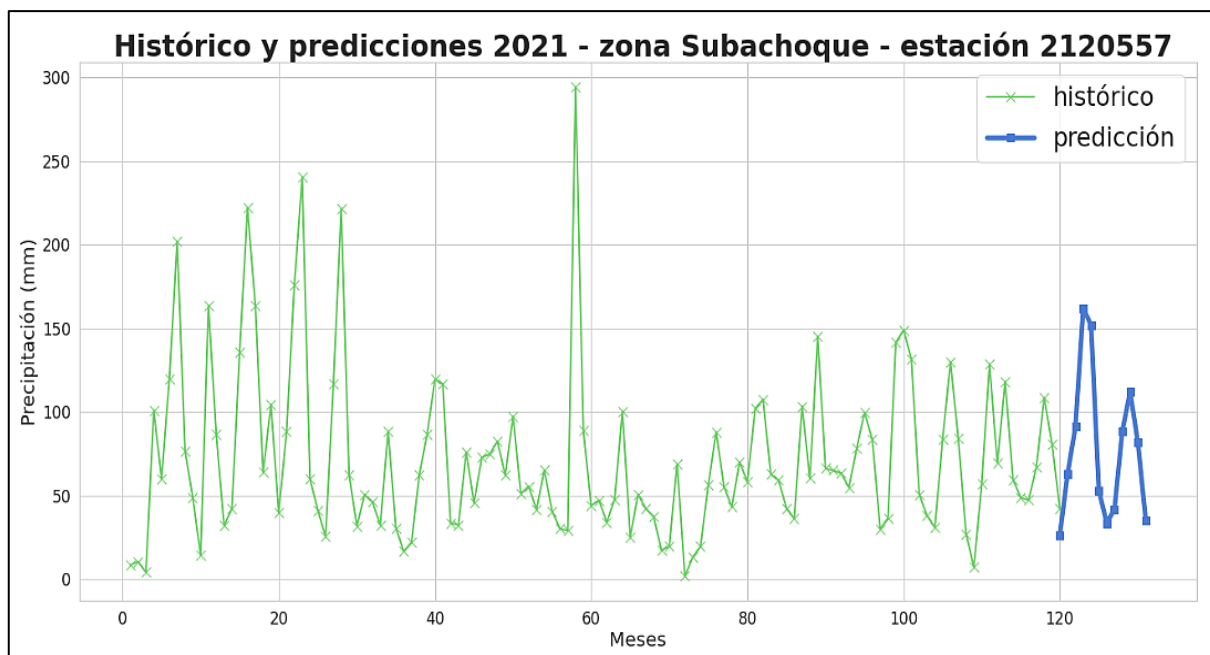


Figura 29. Histórico y predicciones de precipitación - estación 2120557 (elaboración propia)

Las predicciones de precipitación de las zonas analizadas y sus correspondientes gráficas se encuentran en el Anexo II del presente documento.

Analizando las tendencias de las predicciones generadas por el modelo para las diferentes estaciones hidrometeorológicas se observan valores mensuales de precipitación semejantes a las cifras registradas en el histórico de los últimos 10 años. A continuación, se realizará un análisis de las predicciones generadas, sus correspondientes valores y distribución. En la Tabla 8 se registra una estadística descriptiva de los valores generados en las predicciones de precipitación.

Tabla 8. Análisis valores predicciones precipitación

Característica Climática	Descripción
Precipitación	Unidad de Medida: milímetros de agua (mm) o litros de agua por unidad de superficie (m ²). Es decir, 1 milímetro de agua (mm) de lluvia equivale a 1 L de agua por m ² .
	Rango Valores: [4.40 - 340.63]
	Valor Promedio: 76.25
	Valores atípicos: Las predicciones no presenta valores negativos, tampoco registran valores extremos de precipitación (valores superiores a 700 mm).

Fuente: elaboración propia

En la Figura 30 se evidencia la distribución de los valores de niveles de precipitación generados por el modelo y en los cuales aproximadamente el 60% de las predicciones registran niveles de precipitación mensual en el rango de 0 a 100 mm de agua.

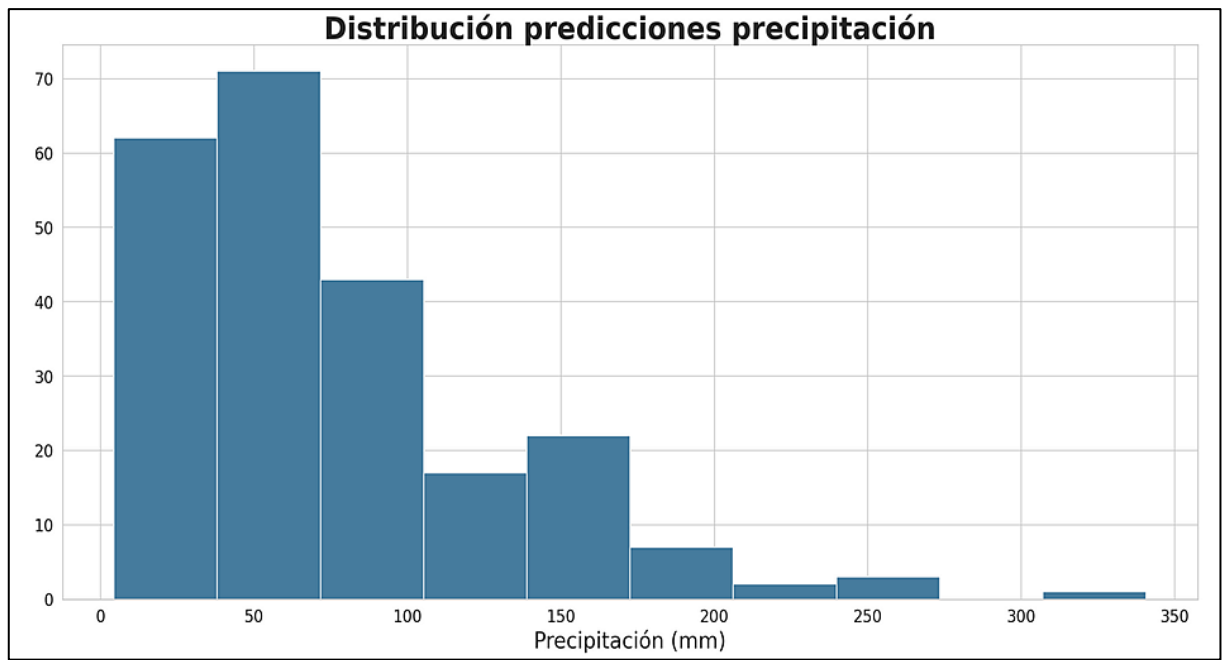


Figura 30. Histograma predicciones niveles precipitación (elaboración propia)

4.7 Identificación zonas agrícolas

Utilizando las predicciones de los niveles de precipitaciones generados en las 19 zonas examinadas se realizará la identificación de los territorios apropiados para actividades agrícolas en Cundinamarca. En la Figura 31 se puede observar el identificador y nombre de las zonas examinadas.

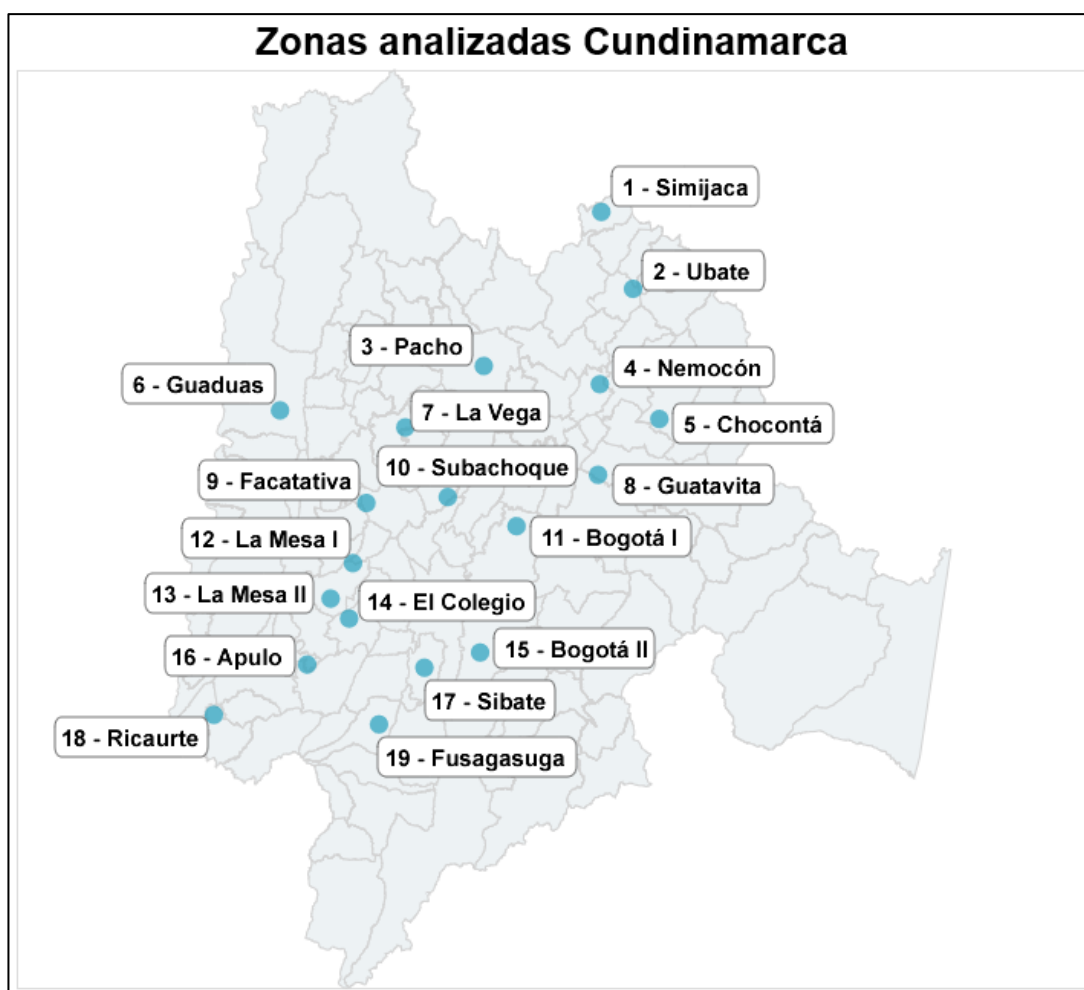


Figura 31. Zonas analizadas Cundinamarca

La identificación de zonas aptas para la agricultura se realizará utilizando el criterio de estimación de la demanda de agua realizada por la entidad Cenicafe (Centro Nacional de Investigaciones de Café) entidad encargada de estudiar los aspectos relacionados a la cosecha y producción de las fincas cafeteras. La entidad en sus estudios define que un cultivo en territorio colombiano necesita de 27 a 50 mm de agua lluvia mensual para su desarrollo (Builes et al., s. f.). En el desarrollo de este trabajo se definirá un valor de 40mm de agua lluvia mensual para considerar una zona apta para actividades agrícolas. Posterior a la definición del umbral de precipitación se realizará la construcción de los mapas en los cuales se

efectuará una georreferenciación de las estaciones hidrometeorológicas y su correspondiente predicción de nivel de precipitación para cada uno de los meses del 2021. En la Figura 32 se puede observar el mapa de Cundinamarca, la ubicación y nombre de las zonas analizadas, el código de la estación climática, el nivel de precipitación estimado y su correspondiente categorización para actividades agrícolas en la temporada enero 2021. La zona es clasificada como adecuada para actividades agrícolas si los niveles de precipitación son superiores a 40mm. El tamaño de los círculos indica el nivel de precipitación y el color indica su clasificación: verde para zonas adecuadas para la agricultura y rojo para zonas inadecuadas para actividades agrícolas. El mapa incluye un componente interactivo que permite identificar el nombre de la ciudad, código de la estación climática y su correspondiente nivel de precipitación ubicando el puntero del mouse sobre las estaciones analizadas.

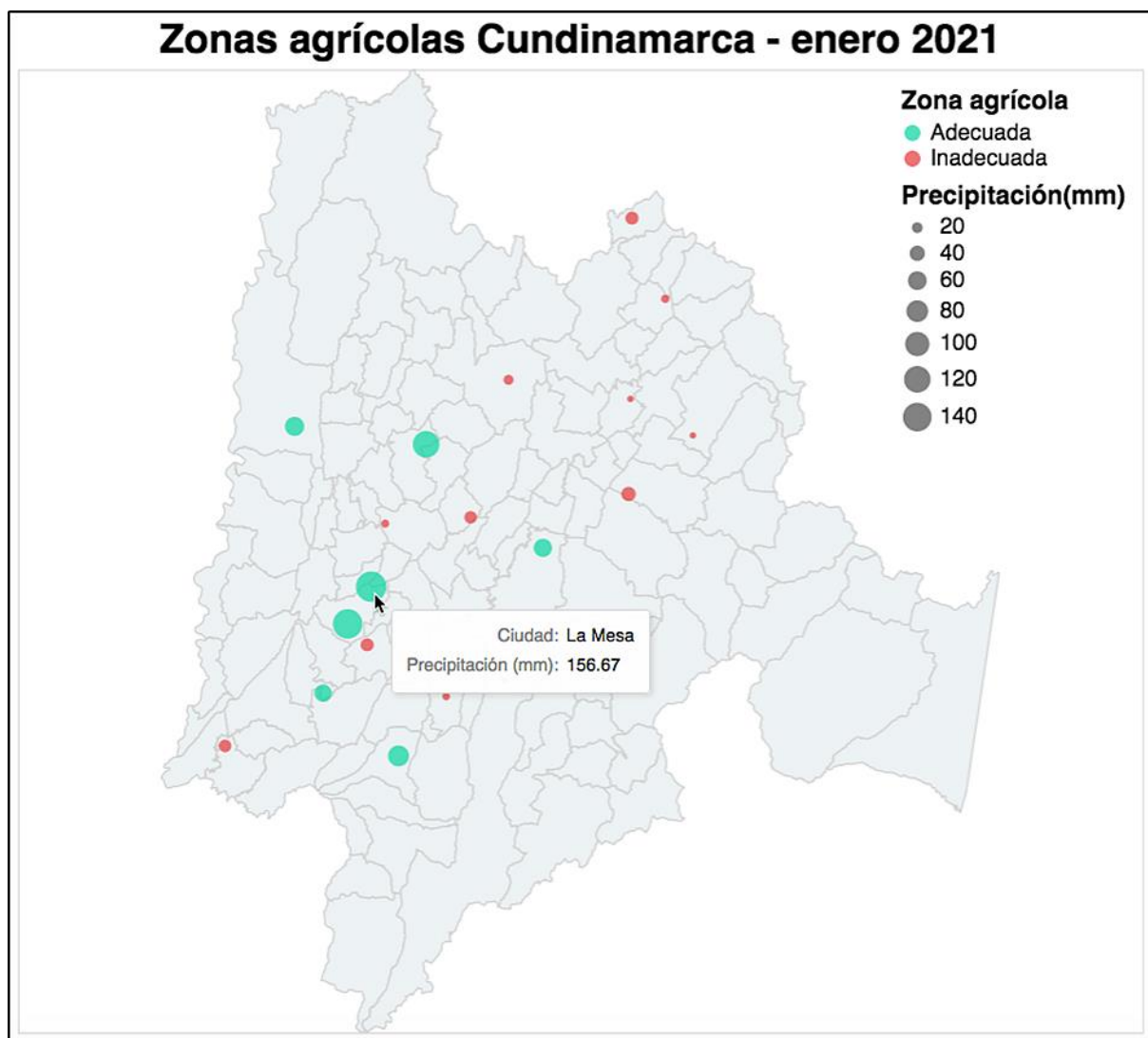


Figura 32. Identificación zonas agrícolas - enero 2021 (elaboración propia)

A continuación, en las Figuras 33-34 se exponen los resultados de la clasificación de zonas agrícolas en Cundinamarca de febrero y marzo respectivamente. En el Anexo III se exponen las gráficas de clasificación mensual de zonas agrícolas para el año 2021.

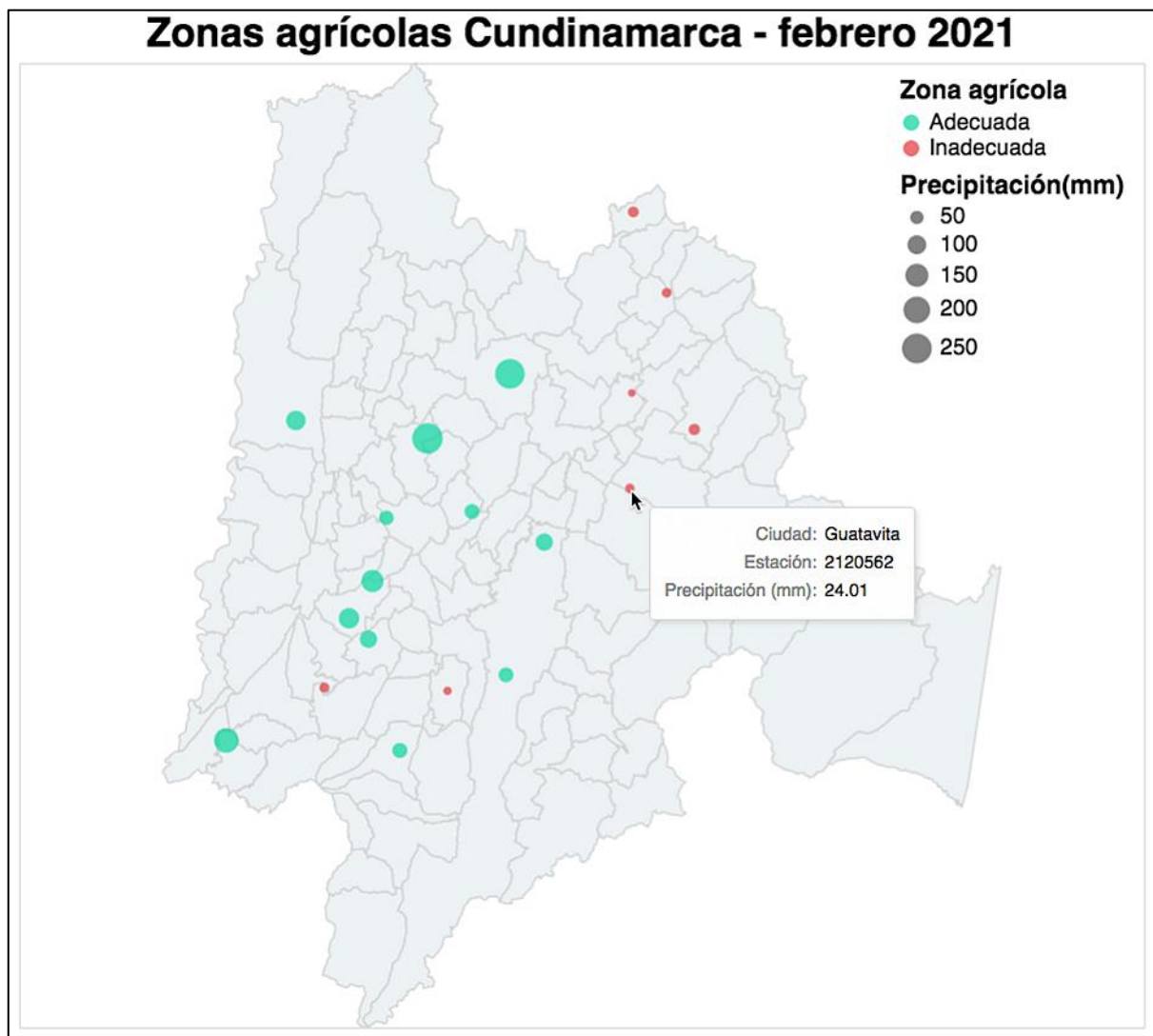


Figura 33. Identificación zonas agrícolas - febrero 2021 (elaboración propia)

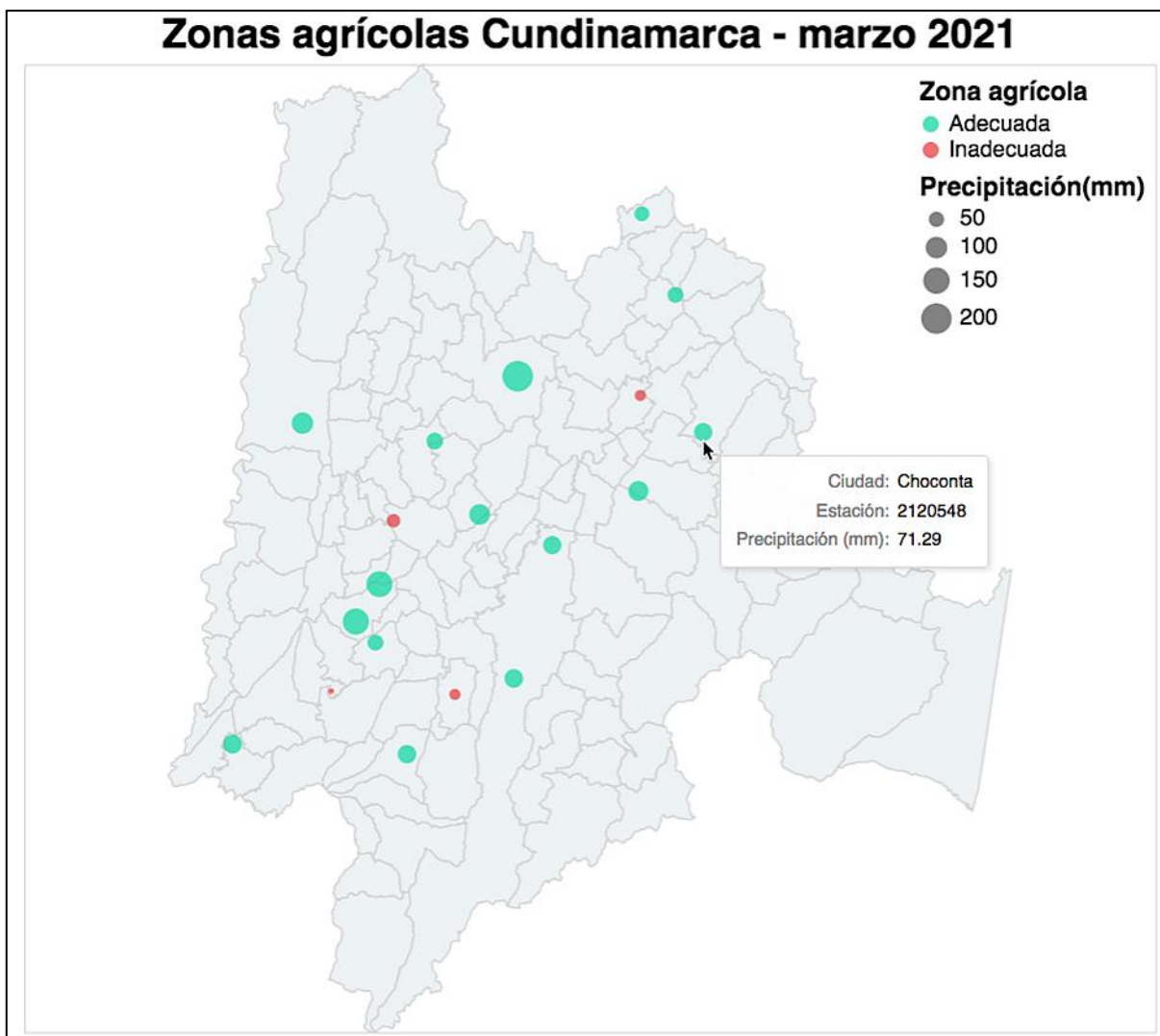


Figura 34. Identificación zonas agrícolas - marzo 2021 (elaboración propia)

La identificación de las zonas propicias para actividades agrícolas durante el año 2021 en el territorio de Cundinamarca se efectuará evaluando las predicciones de lluvias mensuales que satisfacen el criterio de selección de zonas (lluvias mayores a 40 mm). En cada zona se realizará el cálculo del porcentaje de meses que cumplen el criterio de precipitación. Las estaciones que presenten un resultado mayor a 70% serán catalogadas como zonas adecuadas para actividades agrícolas, igualmente las zonas que tengan un resultado inferior a 70% recibirán la clasificación de inadecuadas para la agricultura. La clasificación de todas las zonas agrícolas para el año 2021 se puede observar en la Tabla 9.

Tabla 9. Clasificación zonas agrícolas 2021.

Código estación	Zona	Porcentaje meses lluvias mayores a 40mm	Clasificación Zonas Agrícolas
2120559	Bogotá I	91%	Adecuada
2306516	La Vega	91%	Adecuada
2120646	El Colegio	83%	Adecuada
2120647	La Mesa I	83%	Adecuada
2306507	Pacho	83%	Adecuada
2306517	Guaduas	83%	Adecuada
2119514	Fusagasugá	75%	Adecuada
2120548	Chocontá	75%	Adecuada
2120557	Subachoque	75%	Adecuada
2120562	Guatavita	75%	Adecuada
2120630	Bogotá II	75%	Adecuada
2120639	La Mesa II	75%	Adecuada
2401519	Ubaté	75%	Adecuada
2120540	Nemocón	58%	Inadecuada
2120629	Facatativá	58%	Inadecuada
2120644	Ricaurte	58%	Inadecuada
2120561	Sibaté	50%	Inadecuada
2401513	Simijaca	50%	Inadecuada
2120641	Apulo	33%	Inadecuada

Fuente: elaboración propia

Acorde a los resultados de la clasificación de zonas, para el año 2021 el 68.4% de las zonas presentarán condiciones de lluvias favorables para el desarrollo de actividades agrícolas en el territorio de Cundinamarca, el restante 31.6% de las zonas serán clasificadas como inadecuadas para la agricultura como consecuencia de registrar la menor cantidad de meses con registros de precipitación mayores a 40mm. En la Figura 35 se puede observar la clasificación de zonas para actividades agrícolas en el territorio de Cundinamarca.



Figura 35. Clasificación zonas agrícolas Cundinamarca 2021 (elaboración propia)

Posterior a la clasificación de las zonas agrícolas para el año 2021, se realizará una comparación con las zonas agrícolas identificadas utilizando el histórico de los últimos diez años de precipitaciones. En la Figura 36 se exponen los resultados de la comparación realizada. Las zonas agrícolas identificadas como adecuadas en el análisis histórico representan el 57.9% de los territorios, mientras que en la clasificación del 2021 las zonas adecuadas registran el 68.4% de los terrenos, incrementando en 10.5% los territorios clasificados como adecuados para actividades agrícolas. Por último, se realizó la validación de la cantidad de zonas clasificadas como adecuadas en los dos grupos (histórico y 2021) detectando nueve zonas comunes en las clasificaciones, representando el 81.8% de similitud en la clasificación de zonas adecuadas para actividades agrícolas respecto a la totalidad de zonas apropiadas registradas en la clasificación histórica de los territorios.

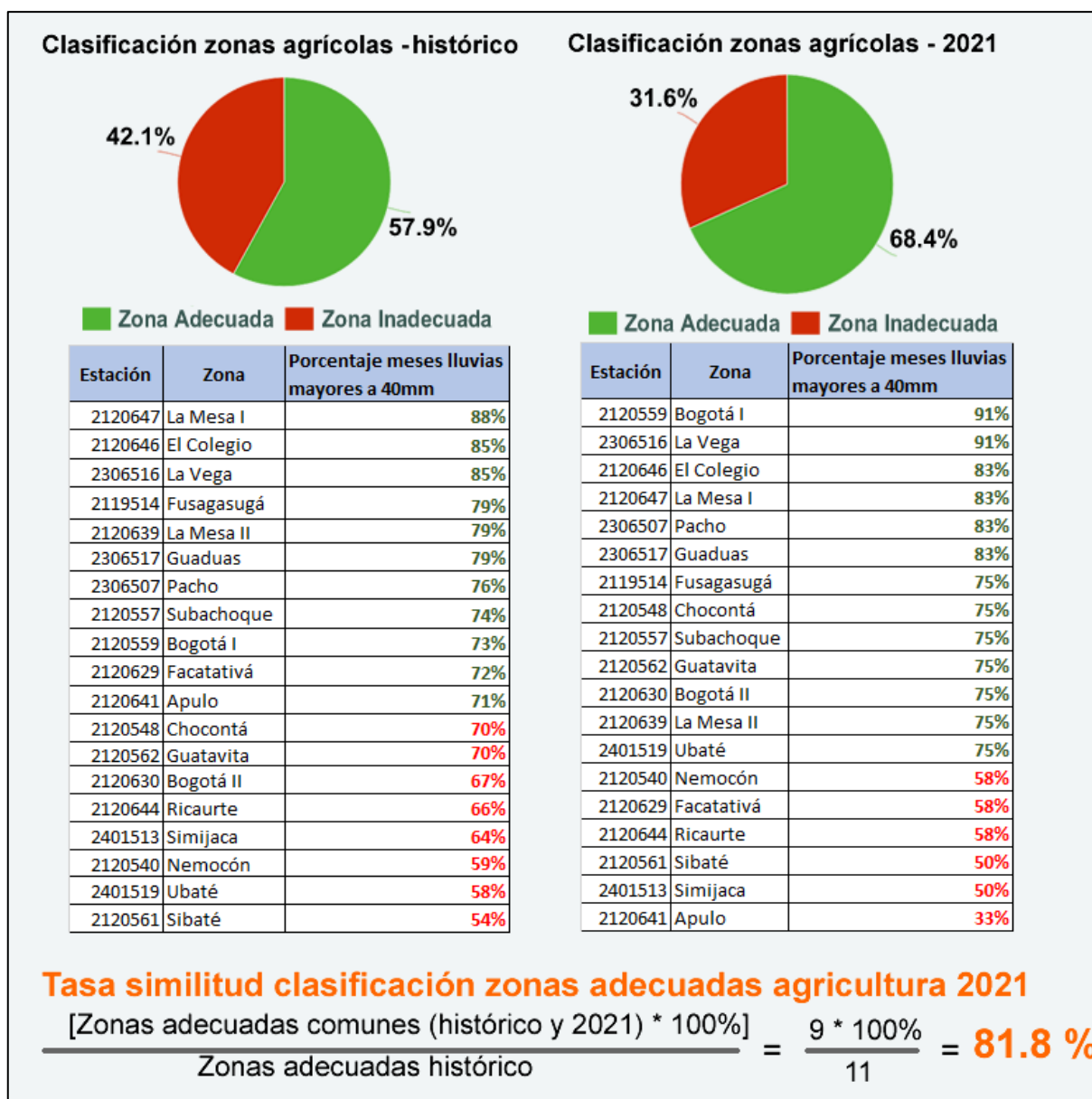


Figura 36. Comparación clasificación zonas agrícolas (elaboración propia)

El 81.8% obtenido en la tasa de similitud de clasificación de zonas adecuadas para actividades agrícolas en 2021, demuestra la precisión que tiene el modelo para realizar estimaciones de precipitaciones, adicionalmente se puede evidenciar como las predicciones del modelo siguen la tendencia general de precipitaciones de los últimos diez años en las zonas analizadas. La identificación de los meses apropiados para actividades agrícolas durante el año 2021 en el territorio de Cundinamarca se realizará evaluando las predicciones mensuales de lluvia que cumplen el criterio de selección (lluvias mayores a 40 mm). Para cada uno de los meses se realizará el cálculo del porcentaje de los registros que cumplen el criterio de precipitación. Los meses que presenten un resultado superior a 70% serán catalogados como adecuados para actividades agrícolas, igualmente los meses que registren un resultado inferior a 70% recibirán

la clasificación de inadecuados para labores agrícolas. La clasificación de los meses del año 2021 se puede observar en la Tabla 10.

Tabla 10. Clasificación meses actividades agrícolas 2021

Mes	Porcentaje registros con lluvias mayores a 40mm	Clasificación actividades agrícolas
Octubre	100%	Adecuado
Noviembre	100%	Adecuado
Mayo	89%	Adecuado
Abril	84%	Adecuado
Marzo	78%	Adecuado
Junio	78%	Adecuado
Septiembre	73%	Adecuado
Febrero	63%	Inadecuado
Julio	52%	Inadecuado
Diciembre	52%	Inadecuado
Agosto	42%	Inadecuado
Enero	36%	Inadecuado

Fuente: elaboración propia

Analizando los resultados generados en la clasificación, en el año 2021 el 58.3% de los meses presentarán condiciones favorables para la ejecución de actividades agrícolas en Cundinamarca, el restante 41.7% de los meses serán clasificados como inadecuados como consecuencia de los bajos niveles de lluvias requeridos para la agricultura. En la Figura 37 se puede observar la clasificación mensual para actividades agrícolas correspondiente al año 2021.

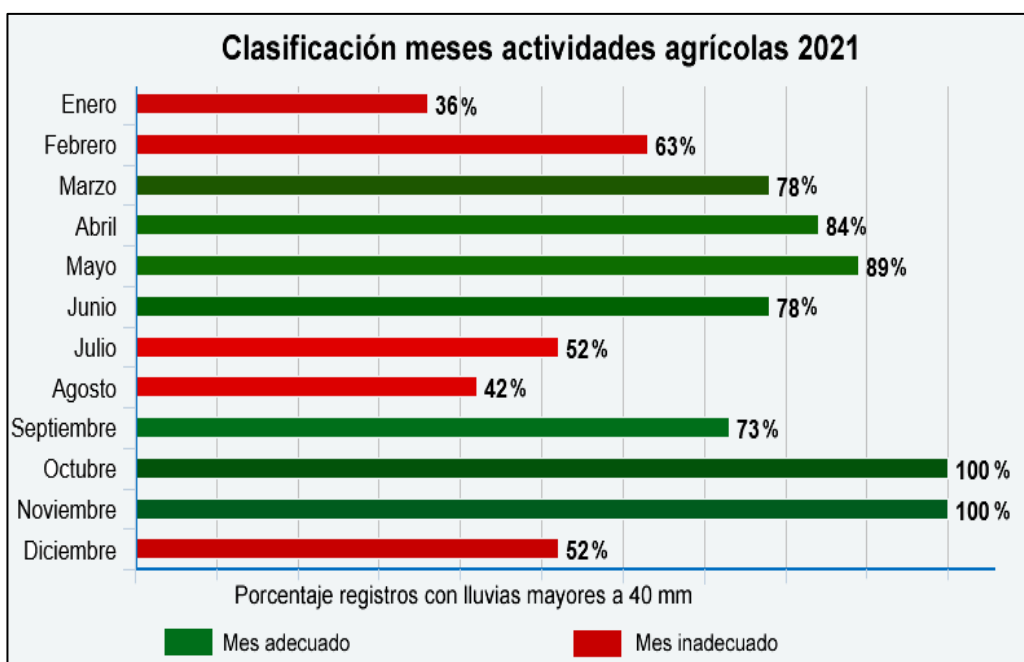


Figura 37. Clasificación mensual zonas agrícolas Cundinamarca 2021 (elaboración propia)

Posterior a la clasificación mensual de zonas agrícolas para el año 2021, se realizará una comparación con la categorización mensual desarrollada utilizando el histórico de los últimos diez años de precipitaciones. En la Figura 38 se registran los resultados de la comparación realizada. Los meses identificados como adecuados para actividades agrícolas en el análisis histórico representan el 41.7% de los meses, mientras que en la clasificación del 2021 los meses adecuados registran el 58.3% de los meses, aumentando en 16.6% los meses categorizados como adecuados para actividades agrícolas.

Finalmente se efectuó la verificación de la cantidad de meses clasificados como adecuados en los dos grupos (histórico y 2021) identificando que la totalidad de los meses catalogados como adecuados para la agricultura en el análisis histórico se encuentran también clasificados como adecuados en las estimaciones del 2021, representando un grado de similitud de 100% en la identificación de los meses propicios para el desarrollo de actividades agrícolas en el territorio de Cundinamarca. Analizando los resultados y la tasa de similitud obtenida de 100% en la clasificación de los meses favorables a la agricultura se demuestra el grado de precisión que tiene el modelo para generar las predicción de los niveles de lluvia.

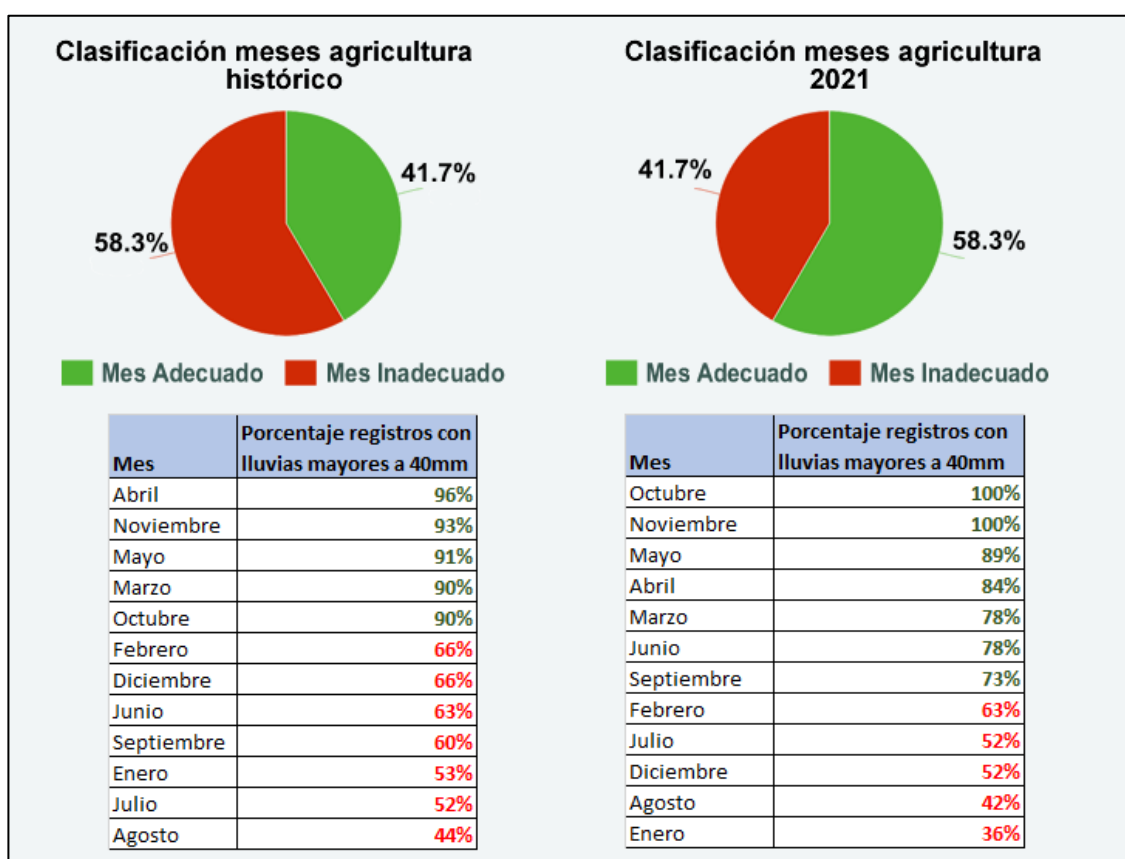


Figura 38. Comparación clasificación meses agricultura (elaboración propia)

Examinando las estimaciones de los niveles de precipitación generados por el modelo para el año 2021 en el territorio de Cundinamarca, se realizó la identificación de las zonas y meses

que registraran precipitaciones superiores a 200 mm, esta actividad se desarrolló con el propósito de referenciar las temporadas y zonas con probabilidades de presentar inundaciones como consecuencia de los elevados niveles de precipitación estimados. Acorde a la identificación de zonas registrada en la Tabla 11, el mes que presenta el mayor riesgo de inundaciones es mayo con precipitaciones estimadas de hasta 340 mm, adicionalmente en mayo se observa la mayor cantidad de zonas con precipitaciones superiores a 200 mm, particularmente las zonas identificadas son: El Colegio, La Vega y Guaduas.

Tabla 11. Identificación zonas precipitaciones elevadas - Cundinamarca 2021

Zona	Estación	Mes	Precipitación (mm)
Pacho	2306507	Febrero	245.75
La Vega	2306516	Febrero	260.33
Pacho	2306507	Marzo	200.48
La Mesa I	2120647	Abril	203.84
El Colegio	2120646	Mayo	217.16
La Vega	2306516	Mayo	240.01
Guaduas	2306517	Mayo	340.63
La Mesa I	2120647	Noviembre	205.49
Pacho	2306507	Noviembre	222.17

Fuente: elaboración propia

5. Conclusiones y trabajo futuro

5.1. Conclusiones

El presente trabajo pretendía abordar la problemática referente a la selección de terrenos adecuados para actividades agrícolas en la región de Cundinamarca, a fin de remediar la problemática descrita, se planteó la creación un modelo predictivo de redes neuronales LSTM orientado a la estimación de las precipitaciones mensuales. Posteriormente, analizando las predicciones generadas se definió un proceso de clasificación con el propósito de identificar en el año 2021 las zonas y meses apropiados para actividades agrícolas en Cundinamarca. Las predicciones de los valores de precipitación permitirán conocer la cantidad de lluvia mensual proyectada en las zonas analizadas en cada uno de los meses del año 2021.

La contribución generada en el desarrollo de este trabajo radica en la creación y verificación de un proceso capaz de estimar los niveles mensuales de precipitación y además identificar las zonas adecuadas para actividades agrícolas en el territorio de Cundinamarca, analizando los datos climáticos históricos de la región. Finalmente, con el propósito de facilitar la comprensión de los resultados generados se realizó la construcción de un componente visual de tipo mapa en el cual se integraron la ubicación geográfica de la zonas analizadas, la estimación de los niveles de lluvia y la clasificación de las zonas adecuadas para la agricultura en Cundinamarca.

Analizando los resultados generados en el presente trabajo se concluye que el proceso desarrollado presenta una gran estabilidad y precisión en la identificación de zonas adecuadas para actividades agrícolas en el territorio de Cundinamarca, la exactitud en la identificación de las zonas fue contrastada respecto a la clasificación de zonas utilizando los datos climáticos históricos de los territorios. Respecto a la precisión de los niveles de lluvias obtenidos por el modelo predictivo se concluye que las estimaciones generadas se encuentran en un rango de valores factibles y conservan la tendencia de lluvias presentada en los últimos diez años en las zonas analizadas. En relación con la identificación de los meses adecuados para la agricultura en Cundinamarca se determina una elevada tasa aciertos en la categorización de los meses respecto a la clasificación realizada utilizando los datos históricos. Finalmente es relevante considerar que aun cuando las predicciones presentan un grado de confiabilidad, los niveles de lluvia son susceptibles a cambios generados por alteraciones súbitas en las condiciones atmosféricas complejas de predecir a largo plazo.

5.2. Líneas de trabajo futuro

Respecto a las líneas de trabajo futuras que permitan ampliar el alcance del presente trabajo, se referencian las siguientes actividades:

- Incorporación de diferentes fuentes de datos climáticas que presenten múltiples frecuencias de medición (diaria, semanal, quincenal) con el objetivo de generar predicciones con distinta frecuencia.
- Exploración y evaluación de algoritmos diferentes a las redes neuronales artificiales enfocados a la generación de predicciones climáticas.
- Integración de nuevas características climáticas y propiedades como la calidad del suelo que permitan ampliar el número de factores analizados con el propósito de fortalecer el proceso de identificación de zonas agrícolas.
- Implementación de un componente de retroalimentación en el modelo predictivo, a fin de ajustar los cálculos y predicciones generadas analizando datos climáticos en tiempo real.
- Optimización del componente visual con el propósito de aportar más datos respecto a las zonas analizadas, datos referentes a la calidad del suelo de la zona aportarían nuevos criterios para la selección de territorios apropiados para actividades agrícolas.

6. Bibliografía

- Alpaydin, E. (2020). Introduction. En *Introduction to Machine Learning* (pp. 3-4). MIT Press.
- Ayzel, G., Scheffer, T., & Heistermann, M. (2020). RainNet v1. 0: A convolutional neural network for radar-based precipitation nowcasting. *Geoscientific Model Development*, 13(6), 2631–2644.
- Bjerknes, V. (1999). The Problem of Weather Forecasting as a Problem in Mechanics and Physics. En M. A. Shapiro & S. Grønås (Eds.), *The Life Cycles of Extratropical Cyclones* (pp. 1-4). American Meteorological Society. https://doi.org/10.1007/978-1-935704-09-6_1
- Builes, V. H. R., Robledo, Á. J., & Pulgarín, J. A. (s. f.). *RANGOS ADECUADOS DE LLUVIA PARA EL CULTIVO DE CAFÉ EN COLOMBIA*. 8.
- Cox, J. D. (2002). *Storm Watchers: The Turbulent History of Weather Prediction from Franklin's Kite to El Niño* (John Wiley&Sons, pp. 85-90). John Wiley & Sons.
- Easy-TensorFlow. (2018). *Easy TensorFlow—RNN* [2018]. Easy TensorFlow - RNN. <https://www.easy-tensorflow.com/tf-tutorials/recurrent-neural-networks/many-to-many>
- Erickson, J. (1991). *Las Tormentas: De las antiguas creencias a la moderna meteorología* (pp. 60-64). McGraw-Hill.
- Fandango, A. (2018). *Mastering TensorFlow 1.x: Advanced machine learning and deep learning concepts using TensorFlow 1.x and Keras*. Packt Publishing.
- FAO. (2019). FAO-ITU E-agriculture In Action: Big Data for Agriculture | E-Agriculture. *Unpacking the data driven digital revolution*, 25-28.
- Friedman, R. M. (1993). *Appropriating the Weather: Vilhelm Bjerknes and the Construction of a Modern Meteorology* (pp. 89-90). Cornell University Press.
- Fritz, W. B. (1996). The women of ENIAC. *IEEE Annals of the History of Computing*, 18(3), 13-28. <https://doi.org/10.1109/85.511940>
- Fuentes, J., & Fuentes, Á. (2015). NIMBUS nº 11-12. En *NIMBUS nº 11-12* (pp. 151-157). Universidad Almería.

- Goloubinoff, M., Katz, E., & Lammel, A. (1997). *Antropología del clima en el mundo hispanoamericano*. 242.
- Gulli, A., & Pal, S. (2017). *Deep Learning with Keras*. Packt Publishing Ltd.
- Hansen, J. W. (2002). Realizing the potential benefits of climate prediction to agriculture: Issues, approaches, challenges. *Agricultural Systems*, 74(3), 309-330. [https://doi.org/10.1016/S0308-521X\(02\)00043-4](https://doi.org/10.1016/S0308-521X(02)00043-4)
- Hartree, D. R. (1946). The Eniac, an Electronic Calculating Machine. *Nature*, 157(3990), 527-527. <https://doi.org/10.1038/157527a0>
- Hassoun, M. H., & Hassoun, A. P. of C. E. M. H. (1995). Preface. En *Fundamentals of Artificial Neural Networks* (p. XIII-XV). MIT Press.
- Hope, T., Resheff, Y. S., & Lieder, I. (2017). Introduction. En *Learning TensorFlow: A Guide to Building Deep Learning Systems* (pp. 5-6). O'Reilly Media, Inc.
- Jakkula, V. (2006). Tutorial on support vector machine (svm). *School of EECS, Washington State University*, 37.
- Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*. <http://arxiv.org/abs/1412.6980>
- Koshkarov, A. V. (2018). Machine learning methods in digital agriculture: Algorithms and cases. *International Journal of Advanced Studies*, 8(1), 11–26.
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, 22(11), 6005-6022. <https://doi.org/10.5194/hess-22-6005-2018>
- Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20(2), 130-141. [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2)
- Mayer-Schönberger, V., & Cukier, K. (2013). Value. En *Big Data: A Revolution that Will Transform how We Live, Work, and Think* (pp. 117-120). Houghton Mifflin Harcourt.
- Mehrotra, K., Mohan, C. K., & Ranka, S. (1997). Prediction Networks. En *Elements of Artificial Neural Networks* (p. 136). MIT Press.

- metrics, S., Perlis, A. J., Sayward, F., & Shaw, M. (1981). Software Project Forecasting. En *Software Metrics: An Analysis and Evaluation* (pp. 80-81). MIT Press.
- Michie, D. (1968). «Memo» Functions and Machine Learning. *Nature*, 218, 19-22.
- Nayak, M. A., & Ghosh, S. (2013). Prediction of extreme rainfall event using weather pattern recognition and support vector machine classifier. *Theoretical and Applied Climatology*, 114(3), 583-603. <https://doi.org/10.1007/s00704-013-0867-3>
- Olson, D. L., & Delen, D. (2008). SEMMA. En *Advanced Data Mining Techniques* (pp. 19-22). Springer Science & Business Media.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 6.
- Petterssen, S. (1968). *Introducción a la meteorología* (pp. 12-13). UABC.
- Platzman, G. W. (1979). The ENIAC computations of 1950—Gateway to numerical weather prediction. *Bulletin of the American Meteorological Society*, 60(4), 302–312.
- Poornima, S., & Pushpalatha, M. (2019). Prediction of rainfall using intensified lstm based recurrent neural network with weighted linear units. *Atmosphere*, 10(11), 668.
- Python Foundation. (2020). *Welcome to Python.org*. Python.Org. <https://www.python.org/>
- Richardson, L. F. (2007). *Weather Prediction by Numerical Process*. Cambridge University Press.
- Sahasrabudde, D. V. (2015). *Data Structure for Representation of Big Data of Weather Forecasting: A Review*. 3(6), 9.
- SAS Institute. (2017, agosto 30). *SAS Help Center: Introduction to SEMMA*. Introduction to SEMMA.
<https://documentation.sas.com/?docsetId=emref&docsetTarget=n061bzurmej4j3n1nj8bbj1a2.htm&docsetVersion=14.3&locale=en>

- Singh, M. S., Warren, R. A., & Jakob, C. (2019). A steady-state model for the relationship between humidity, instability, and precipitation in the tropics. *Journal of Advances in Modeling Earth Systems*, 11(12), 3973–3994.
- Sivakumar, M. (2006). Climate prediction and agriculture: Current status and future challenges. *Climate Research*, 33, 12-15. <https://doi.org/10.3354/cr033003>
- Sønderby, C. K., Espeholt, L., Heek, J., Dehghani, M., Oliver, A., Salimans, T., Agrawal, S., Hickey, J., & Kalchbrenner, N. (2020). MetNet: A Neural Weather Model for Precipitation Forecasting. *arXiv preprint arXiv:2003.12140*.
- Taub, L. (2004). *Ancient Meteorology* (p. 113). Routledge.
- Thorpe, A. J., Volkert, H., & Ziemiański, M. J. (2003). The Bjerknes' Circulation Theorem: A Historical PerspectiveA Historical Perspective. *Bulletin of the American Meteorological Society*, 84(4), 471-480. <https://doi.org/10.1175/BAMS-84-4-471>
- Vicuña, G. (2009). *Manual de meteorología popular* (pp. 5-6). Editorial MAXTOR.
- West, R. C. (2014). *Handbook of Middle American Indians, Volume 1: Natural Environment and Early Cultures*. University of Texas Press.
- World Meteorological Organization. (2010). *Guide to Agricultural Meteorological Practices*. 134, Chapter 5.
- Yegnanarayana, B. (2009). Feedback Neural Networks. En *Artificial Neural Networks* (pp. 142-143). PHI Learning Pvt. Ltd.

Anexos

Anexo I. Catálogo estaciones hidrometeorológicas

A continuación, se exponen las características de las estaciones hidrometeorológicas ubicadas en el territorio de Cundinamarca.

Código Estación	Nombre Estación	Latitud	Longitud	Altura (m.s.n.m.)	Municipio	Vereda	Regional
2119022	PAJAS BLANCAS	4° 21' 16,884" N	74° 33' 24,192" W	769	NILO	PAJAS BLANCAS	3 - ALTO MAGDALENA
2119046	EL BATAN	4° 15' 42,391" N	74° 18' 45,375" W	2692	PASCA	SANTA TERESITA	12 - SUMAPAZ
2119047	HACIENDA LA MESA	4° 18' 19,584" N	74° 16' 43,392" W	2504	PASCA	LA MESA	12 - SUMAPAZ
2119514	UNIVERSIDAD FUSAGASUGÁ UDEC	4° 20' 8,484" N	74° 22' 12,900" W	1725	FUSAGASUGÁ	CASCO URBANO	12 - SUMAPAZ
2120027	SAUCIO	5° 6' 31,783" N	73° 42' 22,371" W	2643	CHOCONTÁ	TILATA	2 - ALMEIDAS Y GUATAVITA
2120043	VILLAPINZÓN	5° 12' 38,015" N	73° 36' 0,000" W	2730	VILLAPINZÓN	CASCO URBANO	2 - ALMEIDAS Y GUATAVITA
2120051	SIBATÉ APOSTÓLICA	4° 30' 21,744" N	74° 14' 37,021" W	2609	SIBATÉ	SAN BENITO	11 - SOACHA
2120055	LA UNIÓN	4° 52' 28,992" N	74° 15' 5,004" W	2652	EL ROSAL	BUENAVISTA	10 - SABANA OCCIDENTE
2120069	TESORO	4° 48' 1,403" N	74° 19' 4,008" W	2592	FACATATIVÁ	MOYANO	10 - SABANA OCCIDENTE
2120071	LAS MARGARITAS	4° 49' 57,288" N	74° 15' 0,684" W	2602	EL ROSAL	CAMPO ALEGRE	10 - SABANA OCCIDENTE

2120074	ZIQUAIRÁ	5° 1' 26,687" N	73° 59' 56,688" W	2613	ZIQUAIRÁ	CASCO URBANO	9 - SABANA CENTRO
2120075	BOJACÁ	4° 43' 53,045" N	74° 20' 32,513" W	2601	BOJACÁ	CASCO URBANO	10 - SABANA OCCIDENTE
2120077	TORCA	4° 47' 27,097" N	74° 1' 41,088" W	2596	BOGOTÁ, D.C.	TORCA	9 - SABANA CENTRO
2120080	SAN JOSÉ	4° 50' 59,711" N	73° 54' 16,596" W	2660	GUASCA	SAN ISIDRO	1 - BOGOTÁ Y LA CALERA
2120085	EL BOSQUE - CISACA	4° 28' 24,86" N	74° 5'3,82" W	3175	BOGOTÁ, D.C.	EL BOSQUE SURORIEN TAL	1 - BOGOTÁ Y LA CALERA
2120088	EL SALITRE	5° 10' 46,200" N	74° 0' 14,292" W	3166	TAUSA	EL SALITRE	14 - UBATÉ
2120096	EL CONSUELO	4° 59' 26,880" N	73° 46' 51,312" W	2933	SESQUILÉ	TIERRA NEGRA	2 - ALMEIDAS Y GUATAVITA
2120103	SANTA TERESA	4° 44' 44,484" N	73° 55' 52,392" W	3002	LA CALERA	SANTA HELENA	1 - BOGOTÁ Y LA CALERA
2120112	LA CASITA	4° 38' 9,819" N	74° 1' 12,286" W	3100	LA CALERA	EL HATO	1 - BOGOTÁ Y LA CALERA
2120136	SANTA INÉS	4° 48' 23,400" N	74° 8' 29,400" W	2591	TENJO	CHACAL	9 - SABANA CENTRO
2120138	EL ENCANTO	5° 10' 39,288" N	73° 53' 30,192" W	3180	TAUSA	LA FLORIDA	14 - UBATÉ
2120141	ACANDY	5° 3' 50,795" N	73° 53' 56,796" W	2596	NEMOCÓN	LA PUERTA	9 - SABANA CENTRO
2120159	ALCO	4° 58' 4,189" N	74° 0' 28,188" W	2603	CAJICÁ	CHUNTAM E	9 - SABANA CENTRO
2120166	FUTE	4° 35' 51,504" N	74° 16' 54,696" W	2633	SOACHA	CANOAS	11 - SOACHA
2120168	ALTO DE AIRE	5° 11' 33,503" N	73° 49' 9,012" W	2796	CUCUNUBÁ	ALTO DE AIRE	14 - UBATÉ
2120169	LADERA GRANDE	5° 10' 21,792" N	73° 51' 6,084" W	2956	TAUSA	LADERA GRANDE	14 - UBATÉ

2120170	EL LLANO	5° 10' 21,217" N	73° 49' 40,908" W	2838	SUESCA	BARRANCAS	2 - ALMEIDAS Y GUATAVITA
2120173	CAMPOBELLO	4° 45' 9,108" N	74° 18' 39,888" W	2561	MADRID	EL CORZO	10 - SABANA OCCIDENTE
2120174	MANJUI	4° 48' 8,713" N	74° 23' 18,492" W	3220	FACATATIVÁ	PUEBLO VIEJO	10 - SABANA OCCIDENTE
2120176	SANTA ISABEL	4° 59' 43,296" N	74° 4' 4,692" W	2729	TABIO	RIO FRI OCCIDENTAL	9 - SABANA CENTRO
2120177	ANCLAJE 14	4° 35' 35,593" N	74° 25' 28,308" W	931	EL COLEGIO	LUCERNA	13 - TEQUENDAMA
2120179	ARGENTINA	4° 44' 54,205" N	74° 31' 9,192" W	1154	QUIPILE	LA ARGENTINA	13 - TEQUENDAMA
2120182	PENAS BLANCAS	4° 33' 47,700" N	74° 22' 41,592" W	2118	EL COLEGIO	ANTIOQUIA	13 - TEQUENDAMA
2120183	DARÍO VALENCIA	4° 35' 54,493" N	74° 26' 9,996" W	717	EL COLEGIO	LUCERNA	13 - TEQUENDAMA
2120184	LAGUNA DEL INDIO	4° 24' 39,385" N	74° 27' 29,412" W	1691	VIOTÁ	EL PALMAR	13 - TEQUENDAMA
2120185	LA TRIBUNA	4° 51' 58,104" N	74° 24' 38,412" W	2789	FACATATIVÁ	LA TRIBUNA	10 - SABANA OCCIDENTE
2120187	BOMBAS SESQUILÉ	5° 2' 12,912" N	73° 48' 39,384" W	2578	SESQUILÉ	CHALECHE	2 - ALMEIDAS Y GUATAVITA
2120190	COL FRANCISCO JOSÉ DE CALDAS	4° 26' 33,684" N	74° 31' 32,196" W	594	VIOTÁ	CASCO URBANO	13 - TEQUENDAMA
2120193	GUANQUICA	5° 11' 26,592" N	73° 55' 33,708" W	3010	TAUSA	PARAMO BAJO	14 - UBATÉ
2120195	MONTECILLOS	4° 54' 51,011" N	73° 49' 26,688" W	2805	GUATAVITA	MONTECILLO	2 - ALMEIDAS Y GUATAVITA

2120213	EL CEDRAL	5° 3' 15,984" N	73° 59' 40,416" W	2666	COGUA	RODAMON TAL	9 - SABANA CENTRO
2120214	PARAMO DE GUERRERO	5° 6' 54,504" N	74° 2' 42,108" W	3256	ZIPAQUIRÁ	PARAMO DE GUERRERO	9 - SABANA CENTRO
2120215	AMÉRICA	4° 26' 54,384" N	74° 27' 39,888" W	1372	VIOTÁ	BUENA VISTA	13 - TEQUENDAMA
2120516	LA RAMADA	4° 42' 11,592" N	74° 10' 37,992" W	2571	FUNZA	CASCO URBANO	10 - SABANA OCCIDENTE
2120540	CHECUA	5° 7' 0,696" N	73° 51' 41,184" W	2603	NEMOCÓN	SUSATA	9 - SABANA CENTRO
2120541	REPRESA DEL NEUSA	5° 8' 19,7" N	73° 58' 16,3" W	3028	COGUA	COGUA	9 - SABANA CENTRO
2120557	LA PRIMAVERA	4° 51' 28,296" N	74° 12' 42,984" W	2623	EL ROSAL	EL RODEO	10 - SABANA OCCIDENTE
2120562	GUATAVITA	4° 54' 31,897" N	73° 51' 56,196" W	2679	GUATAVITA	SANTAMARIA	2 - ALMEIDAS Y GUATAVITA
2120629	VENECIA	4° 50' 38,904" N	74° 23' 58,308" W	2686	FACATATIVÁ	LA TRIBUNA	10 - SABANA OCCIDENTE
2120630	DONA JUANA	4° 30' 3,097" N	74° 8' 14,496" W	2800	BOGOTÁ, D.C.	DISTRITO CAPITAL	1 - BOGOTÁ Y LA CALERA
2120632	FORTUNALA	5° 15' 38,9" N	73° 35' 14,7" W	2814	VILLAPINZON	VILLAPINZON	2 - ALMEIDAS Y GUATAVITA
2120633	BARRANCAS	5° 10' 3,6" N	73° 48' 50,2" W	2913	TAUSA	TAUSA	14 - UBATÉ
2120634	PARAÍSO PERDIDO	4° 29' 51,685" N	74° 17' 28,104" W	2756	SIBATÉ	ALTO CHARCO	11 - SOACHA
2120637	LAS VIOLETAS	4° 22' 42,816" N	74° 41' 15,504" W	432	AGUA DE DIOS	LAS LOMAS	3 - ALTO MAGDALENA
2120640	LA VICTORIA	4° 27' 57,422" N	74° 36' 31,194" W	394	TOCAIMA	CATARNICA	3 - ALTO MAGDALENA

2120641	ESC SAMPER MADRID	4° 28' 23,303" N	74° 32' 8,592" W	541	APULO	SAN ANTONIO	13 - TEQUENDAMA
2120644	ARGELIA	4° 21' 27,397" N	74° 45' 2,808" W	315	RICAURTE	MANUEL DEL SUR	3 - ALTO MAGDALENA
2120646	MESITAS	4° 34' 43,787" N	74° 26' 20,616" W	1057	EL COLEGIO	CASCO URBANO	13 - TEQUENDAMA
2120652	PALOQUE MAO	4° 43' 47,101" N	74° 24' 23,508" W	2181	ZIPACÓN	PALO QUEMAO - ZELANDIA	10 - SABANA OCCIDENTE
2120691	PLANADAS	5° 3' 34,116" N	74° 6' 34,422" W	3357	SUBACHOQ UE	EL GUAMAL	10 - SABANA OCCIDENTE
2123502	EL PLACER	4° 40' 26,432" N	74° 43' 2,211" W	1335	PULÍ	EL PALMAR LA HOYA	7 - MAGDALENA CENTRO
2306033	AGUA FRÍA	5° 6' 1,512" N	74° 28' 23,196" W	1309	QUEBRADA NEGRA	AGUA FRIA	6 - GUALIVÁ
2306034	SAN ISIDRO	5° 2' 56,796" N	74° 30' 13,716" W	1132	VILLETA	CUNE	6 - GUALIVÁ
2306039	TIESTOS LOS	5° 21' 21,384" N	74° 24' 29,412" W	1659	LA PALMA	LA CAÑADA	8 - RIONEGRO
2306040	ALTO DE LA RANA	5° 7' 14,411" N	74° 32' 24,900" W	1669	GUADUAS	CINTA Y FRIA	6 - GUALIVÁ
2306308	NEGRETE	5° 5' 48,299" N	74° 9' 14,508" W	2314	PACHO	EL PINAL	8 - RIONEGRO
2306507	INST. AGRIC. ESC. VOCACION AL	5° 9' 31,212" N	74° 7' 43,788" W	1940	PACHO	LLANO DE LA HACIENDA	8 - RIONEGRO
2306517	GUADUAS	5° 3' 24,299" N	74° 35' 53,808" W	1023	GUADUAS	CASCO URBANO	4 - BAJO MAGDALENA
2312019	LOS PINOS	5° 14' 27,492" N	74° 1' 39,000" W	3475	SAN CAYETANO	LAGUNA VERDE	6 - GUALIVÁ
2401002	CARUPA HOSPITAL	5° 20' 53,592" N	73° 53' 53,412" W	2970	CARMEN DE CARUPA	CASCO URBANO	14 - UBATÉ

2401030	EL HATILLO	5° 9' 21,887" N	73° 47' 15,504" W	2989	SUESCA	CUAYÁ	2 - ALMEIDAS Y GUATAVITA
2401033	EL ESPINO	5° 19' 57,288" N	73° 44' 0,816" W	2570	LENGUAZA QUE	SIATAMA	14 - UBATÉ
2401036	MONSERRATE	5° 24' 11,915" N	73° 47' 51,108" W	2896	FÚQUENE	CASCO URBANO	14 - UBATÉ
2401037	SOCOTÁ	5° 24' 29,016" N	73° 55' 9,300" W	2952	CARMEN DE CARUPA	NAZARETH	14 - UBATÉ
2401038	EL PUENTE	5° 22' 7,104" N	73° 40' 12,288" W	2587	GUACHETÁ	FALDA DE MOLINO	14 - UBATÉ
2401039	EL TRIANGULO	5° 18' 39,205" N	73° 37' 9,516" W	2854	LENGUAZA QUE	TIBITA CENTRO	14 - UBATÉ
2401044	TRES ESQUINAS	5° 23' 9,204" N	73° 51' 28,116" W	3164	SUSA	PAUNITA	14 - UBATÉ
2401049	EL PEDREGAL	5° 13' 10,092" N	73° 52' 38,712" W	2908	SUTATAUSA	SALITRE	14 - UBATÉ
2401051	EL ZARZAL	5° 26' 32,315" N	73° 39' 27,684" W	2923	RÁQUIRA	FARFAN	5 - CHIQUINQUIRÁ
2401052	HATO 1	5° 17' 16,800" N	73° 55' 27,012" W	3034	CARMEN DE CARUPA	LLANO GRANDE	14 - UBATÉ
2401053	HATO 2	5° 15' 49,717" N	73° 55' 31,692" W	3280	CARMEN DE CARUPA	LLANO GRANDE	14 - UBATÉ
2401057	HATO 6	5° 13' 26,400" N	73° 57' 2,412" W	3253	CARMEN DE CARUPA	SALINAS	14 - UBATÉ
2401110	ISLA DEL SANTUARIO	5° 28' 4,800" N	73° 44' 3,912" W	2560	FÚQUENE	LAGUNA DE FUQUENE	14 - UBATÉ
2401115	CAMPAMENTO UBATÉ	5° 17' 56,113" N	73° 48' 52,416" W	2587	VILLA DE SAN DIEGO DE UBATÉ	CASCO URBANO	14 - UBATÉ
2401116	LOS PINOS	5° 11' 32,784" N	73° 46' 49,512" W	2883	CUCUNUBÁ	CARRIZAL	2 - ALMEIDAS Y GUATAVITA
2401511	BOYERALA	5° 18' 13,789" N	73° 51' 6,300" W	2616	VILLA DE SAN DIEGO DE UBATE	SUCUNCHUQUE	14 - UBATÉ

2401515	CARRIZAL	5° 12' 14,184" N	73° 46' 3,216" W	2885	CUCUNUBÁ	CARRIZAL	9 - SABANA CENTRO
2401519	NOVILLER OS	5° 20' 8,412" N	73° 47' 5,712" W	2590	VILLA DE SAN DIEGO DE UBATE	PALOGOR DO	14 - UBATÉ
2401534	REPRESA EL HATO	5° 18' 13,284" N	73° 54' 18,504" W	2849	CARMEN DE CARUPA	LLANO GRANDE	14 - UBATÉ
2120548	IBERIA LA	5° 2' 13,4" N	73° 43' 27,5" W	2721	CHOCONTA	TILATA	2 - ALMEIDAS Y GUATAVITA
2120559	APTO GUAIMARA L-USTA	4° 47' 27,600" N	74° 3' 12,492" W	2603	BOGOTÁ, D.C.	DISTRITO CAPITAL	1 - BOGOTÁ Y LA CALERA
2120561	MUNAL EL	4° 33' 2,6" N	74° 1' 5,4" W	2565	SIBATÉ	MUNA EL	11 - SOACHA
2401513	SIMIJACA	5° 30' 43,597" N	73° 51' 28,908" W	2622	SIMIJACA	SALITRE	14 - UBATÉ
2120639	LA MESA	4° 37' 28,704" N	74° 28' 54,300" W	1215	LA MESA	CASCO URBANO	13 - TEQUENDAMA
2120647	LA ESPERANZ A	4° 42' 21,888" N	74° 25' 49,008" W	1391	ZIPACÓN	EL OCASO	10 - SABANA OCCIDENTE

Anexo II. Predicciones precipitaciones estaciones hidrometeorológicas Cundinamarca

Año	Mes	ID Zona	Estación	Ciudad	Precipitación (mm)
2021	1	1	2401513	Simijaca	28.94
2021	2	1	2401513	Simijaca	33.33
2021	3	1	2401513	Simijaca	47.83
2021	4	1	2401513	Simijaca	69.33
2021	5	1	2401513	Simijaca	86.51
2021	6	1	2401513	Simijaca	58.28
2021	7	1	2401513	Simijaca	39.58
2021	8	1	2401513	Simijaca	35.32
2021	9	1	2401513	Simijaca	35.94
2021	10	1	2401513	Simijaca	115.14
2021	11	1	2401513	Simijaca	115.5
2021	12	1	2401513	Simijaca	31.77
2021	1	2	2401519	Ubaté	12.13
2021	2	2	2401519	Ubaté	27.18
2021	3	2	2401519	Ubaté	53.86
2021	4	2	2401519	Ubaté	89.6
2021	5	2	2401519	Ubaté	43.85
2021	6	2	2401519	Ubaté	56.78
2021	7	2	2401519	Ubaté	47.03
2021	8	2	2401519	Ubaté	41.65
2021	9	2	2401519	Ubaté	47.97
2021	10	2	2401519	Ubaté	98.57
2021	11	2	2401519	Ubaté	85.58
2021	12	2	2401519	Ubaté	18.31
2021	1	3	2306507	Pacho	17.22
2021	2	3	2306507	Pacho	245.75
2021	3	3	2306507	Pacho	200.48
2021	4	3	2306507	Pacho	153.35
2021	5	3	2306507	Pacho	150.45
2021	6	3	2306507	Pacho	36.57
2021	7	3	2306507	Pacho	56.63
2021	8	3	2306507	Pacho	52.2
2021	9	3	2306507	Pacho	86.37
2021	10	3	2306507	Pacho	167.24
2021	11	3	2306507	Pacho	222.17
2021	12	3	2306507	Pacho	93.05
2021	1	4	2120540	Nemocón	7.36
2021	2	4	2120540	Nemocón	16.87
2021	3	4	2120540	Nemocón	27.02

2021	4	4	2120540	Nemocón	89.26
2021	5	4	2120540	Nemocón	60.33
2021	6	4	2120540	Nemocón	42.91
2021	7	4	2120540	Nemocón	45.55
2021	8	4	2120540	Nemocón	35.6
2021	9	4	2120540	Nemocón	42.23
2021	10	4	2120540	Nemocón	65.69
2021	11	4	2120540	Nemocón	66.23
2021	12	4	2120540	Nemocón	29.72
2021	1	5	2120548	Chocontá	7.12
2021	2	5	2120548	Chocontá	37.39
2021	3	5	2120548	Chocontá	71.29
2021	4	5	2120548	Chocontá	78.39
2021	5	5	2120548	Chocontá	123.24
2021	6	5	2120548	Chocontá	127.37
2021	7	5	2120548	Chocontá	111.54
2021	8	5	2120548	Chocontá	96.14
2021	9	5	2120548	Chocontá	67.35
2021	10	5	2120548	Chocontá	76.95
2021	11	5	2120548	Chocontá	49.72
2021	12	5	2120548	Chocontá	18.59
2021	1	6	2306517	Guaduas	62.01
2021	2	6	2306517	Guaduas	106.04
2021	3	6	2306517	Guaduas	98.34
2021	4	6	2306517	Guaduas	90.68
2021	5	6	2306517	Guaduas	340.63
2021	6	6	2306517	Guaduas	37.8
2021	7	6	2306517	Guaduas	48.08
2021	8	6	2306517	Guaduas	18.2
2021	9	6	2306517	Guaduas	77.18
2021	10	6	2306517	Guaduas	196.93
2021	11	6	2306517	Guaduas	108.87
2021	12	6	2306517	Guaduas	89.5
2021	1	7	2306516	La Vega	119.44
2021	2	7	2306516	La Vega	260.33
2021	3	7	2306516	La Vega	59.78
2021	4	7	2306516	La Vega	150.47
2021	5	7	2306516	La Vega	240.01
2021	6	7	2306516	La Vega	50.85
2021	7	7	2306516	La Vega	37.53
2021	8	7	2306516	La Vega	55.38
2021	9	7	2306516	La Vega	76.75
2021	10	7	2306516	La Vega	122.21
2021	11	7	2306516	La Vega	183.98
2021	12	7	2306516	La Vega	150.01

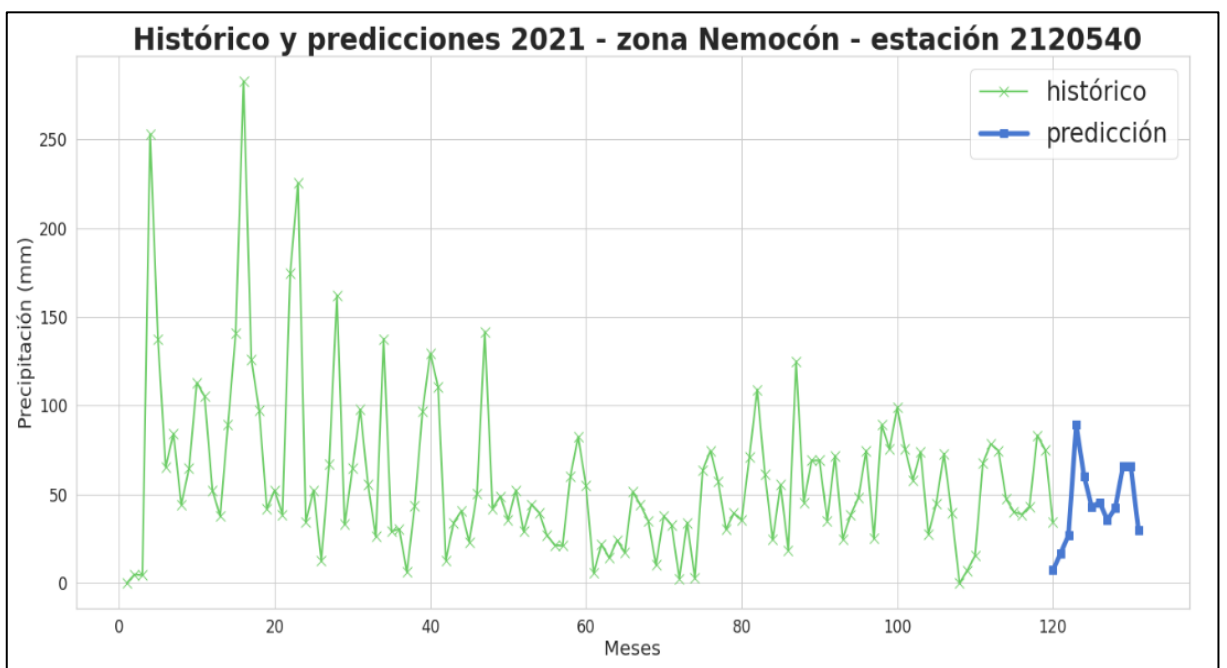
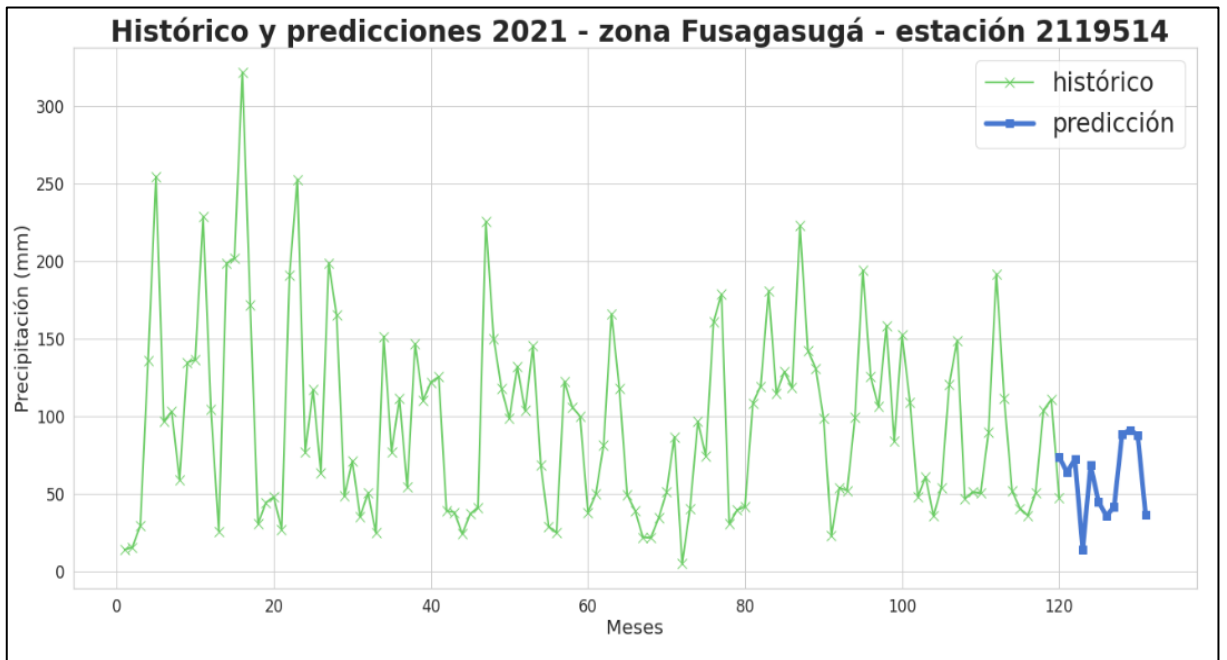
2021	1	8	2120562	Guatavita	34.95
2021	2	8	2120562	Guatavita	24.01
2021	3	8	2120562	Guatavita	85.82
2021	4	8	2120562	Guatavita	90.87
2021	5	8	2120562	Guatavita	67.47
2021	6	8	2120562	Guatavita	81.98
2021	7	8	2120562	Guatavita	93.29
2021	8	8	2120562	Guatavita	36.84
2021	9	8	2120562	Guatavita	54.06
2021	10	8	2120562	Guatavita	77.62
2021	11	8	2120562	Guatavita	49.08
2021	12	8	2120562	Guatavita	49.84
2021	1	9	2120629	Facatativá	11.32
2021	2	9	2120629	Facatativá	58.29
2021	3	9	2120629	Facatativá	39.26
2021	4	9	2120629	Facatativá	35.96
2021	5	9	2120629	Facatativá	144.98
2021	6	9	2120629	Facatativá	41.06
2021	7	9	2120629	Facatativá	18.56
2021	8	9	2120629	Facatativá	23.75
2021	9	9	2120629	Facatativá	59.86
2021	10	9	2120629	Facatativá	106.94
2021	11	9	2120629	Facatativá	101.31
2021	1	10	2120557	Subachoque	26.41
2021	2	10	2120557	Subachoque	63.07
2021	3	10	2120557	Subachoque	91.13
2021	4	10	2120557	Subachoque	161.92
2021	5	10	2120557	Subachoque	151.82
2021	6	10	2120557	Subachoque	53
2021	7	10	2120557	Subachoque	33.53
2021	8	10	2120557	Subachoque	41.88
2021	9	10	2120557	Subachoque	88.53
2021	10	10	2120557	Subachoque	111.97
2021	11	10	2120557	Subachoque	81.57
2021	12	10	2120557	Subachoque	35.26
2021	1	11	2120559	Bogotá	56.63
2021	2	11	2120559	Bogotá	84.1
2021	3	11	2120559	Bogotá	71.57
2021	4	11	2120559	Bogotá	77.1
2021	5	11	2120559	Bogotá	79.89
2021	6	11	2120559	Bogotá	55.08
2021	7	11	2120559	Bogotá	46.72
2021	8	11	2120559	Bogotá	68.17
2021	9	11	2120559	Bogotá	49.67
2021	10	11	2120559	Bogotá	78.27

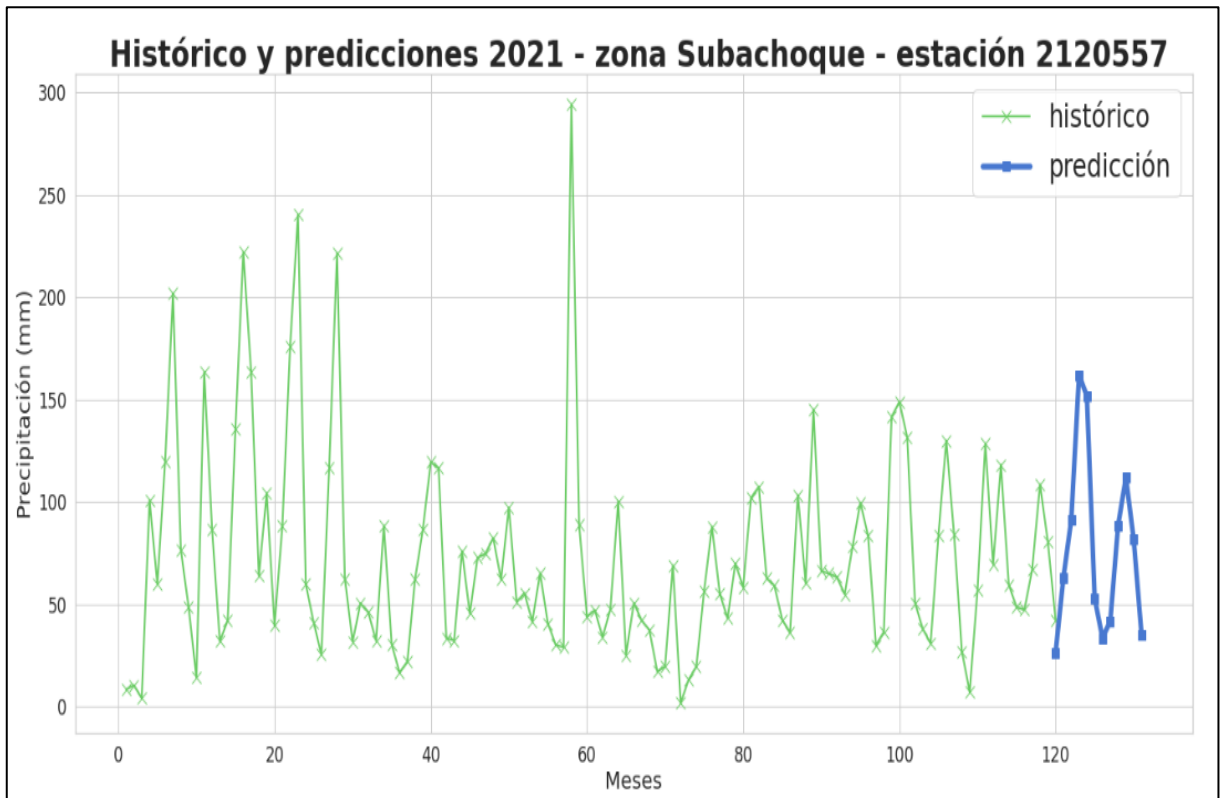
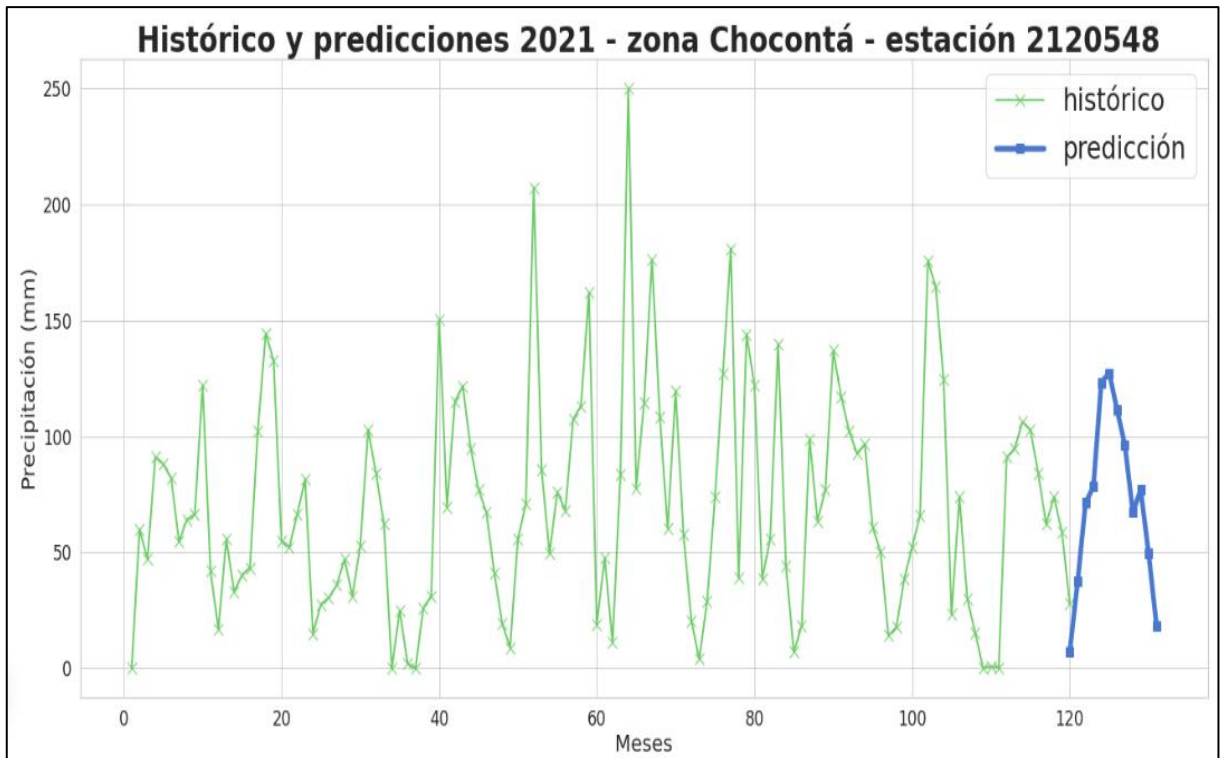
2021	11	11	2120559	Bogotá	100.27
2021	12	11	2120559	Bogotá	36.18
2021	1	12	2120647	La Mesa	156.67
2021	2	12	2120647	La Mesa	138.08
2021	3	12	2120647	La Mesa	144.23
2021	4	12	2120647	La Mesa	203.84
2021	5	12	2120647	La Mesa	194.89
2021	6	12	2120647	La Mesa	93.33
2021	7	12	2120647	La Mesa	25.29
2021	8	12	2120647	La Mesa	8.08
2021	9	12	2120647	La Mesa	94.22
2021	10	12	2120647	La Mesa	199.39
2021	11	12	2120647	La Mesa	205.49
2021	12	12	2120647	La Mesa	138.02
2021	1	13	2120639	La Mesa	147.07
2021	2	13	2120639	La Mesa	117.16
2021	3	13	2120639	La Mesa	147.05
2021	4	13	2120639	La Mesa	145.56
2021	5	13	2120639	La Mesa	76.41
2021	6	13	2120639	La Mesa	47.5
2021	7	13	2120639	La Mesa	41.34
2021	8	13	2120639	La Mesa	13.32
2021	9	13	2120639	La Mesa	28.56
2021	10	13	2120639	La Mesa	163.01
2021	11	13	2120639	La Mesa	165.66
2021	12	13	2120639	La Mesa	37.02
2021	1	14	2120646	El Colegio	29.05
2021	2	14	2120646	El Colegio	83.98
2021	3	14	2120646	El Colegio	54.34
2021	4	14	2120646	El Colegio	143.92
2021	5	14	2120646	El Colegio	217.16
2021	6	14	2120646	El Colegio	106.54
2021	7	14	2120646	El Colegio	40.9
2021	8	14	2120646	El Colegio	38.69
2021	9	14	2120646	El Colegio	97.28
2021	10	14	2120646	El Colegio	156.45
2021	11	14	2120646	El Colegio	168.37
2021	12	14	2120646	El Colegio	101.31
2021	1	15	2120630	Bogotá	6.06
2021	2	15	2120630	Bogotá	61.64
2021	3	15	2120630	Bogotá	74.7
2021	4	15	2120630	Bogotá	66.9
2021	5	15	2120630	Bogotá	33.54
2021	6	15	2120630	Bogotá	68.35
2021	7	15	2120630	Bogotá	68.78

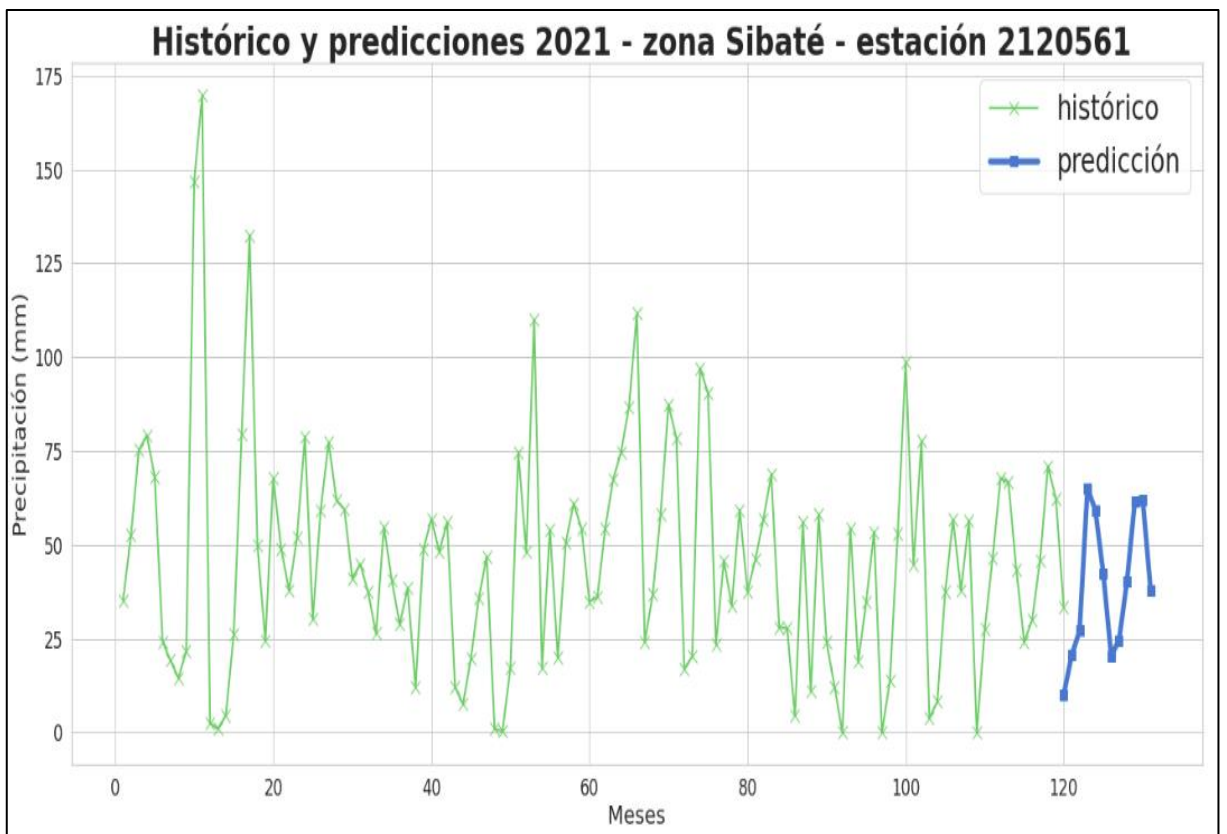
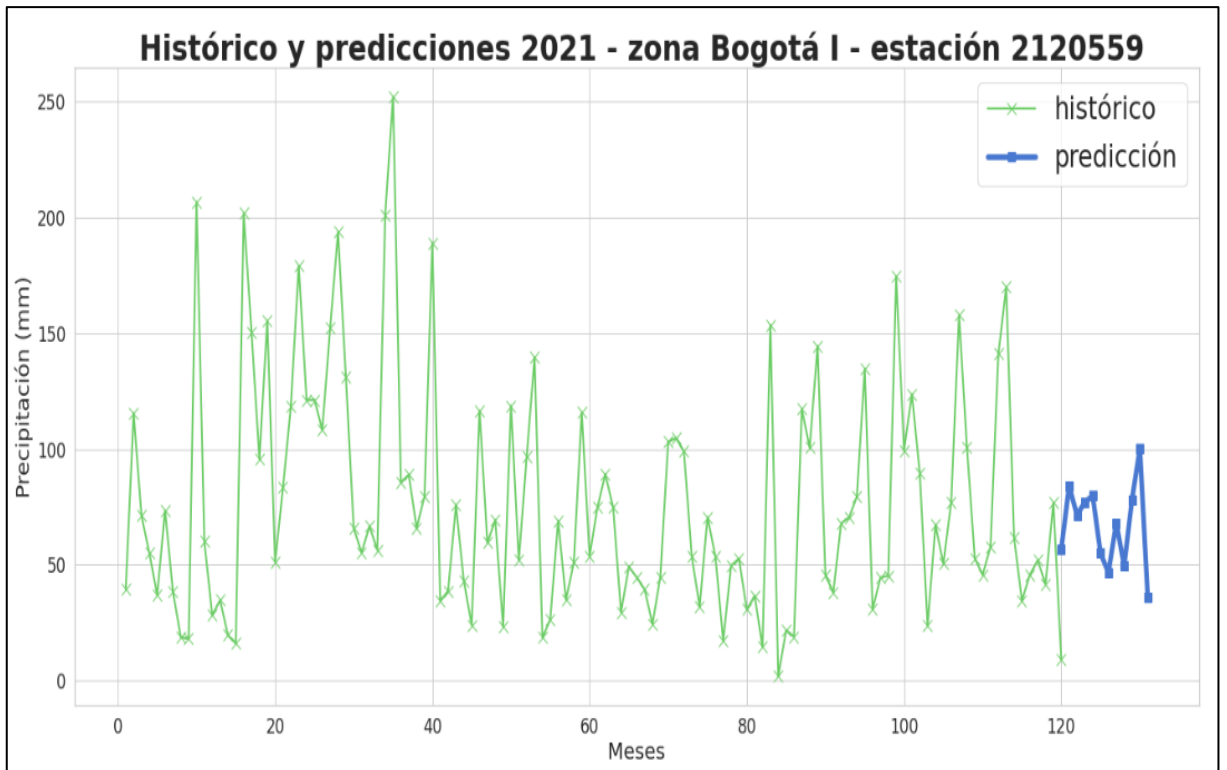
2021	8	15	2120630	Bogotá	57.6
2021	9	15	2120630	Bogotá	37.01
2021	10	15	2120630	Bogotá	61.22
2021	11	15	2120630	Bogotá	77.31
2021	12	15	2120630	Bogotá	40.4
2021	1	16	2120641	Apulo	50.54
2021	2	16	2120641	Apulo	28.17
2021	3	16	2120641	Apulo	6.91
2021	4	16	2120641	Apulo	12.05
2021	5	16	2120641	Apulo	26.46
2021	6	16	2120641	Apulo	4.41
2021	7	16	2120641	Apulo	15.64
2021	8	16	2120641	Apulo	29.3
2021	9	16	2120641	Apulo	39.33
2021	10	16	2120641	Apulo	144.3
2021	11	16	2120641	Apulo	108.25
2021	12	16	2120641	Apulo	47.17
2021	1	17	2120561	Sibaté	10.12
2021	2	17	2120561	Sibaté	20.56
2021	3	17	2120561	Sibaté	27.1
2021	4	17	2120561	Sibaté	65.03
2021	5	17	2120561	Sibaté	59.29
2021	6	17	2120561	Sibaté	42.47
2021	7	17	2120561	Sibaté	20.45
2021	8	17	2120561	Sibaté	24.66
2021	9	17	2120561	Sibaté	40.29
2021	10	17	2120561	Sibaté	61.73
2021	11	17	2120561	Sibaté	61.8
2021	12	17	2120561	Sibaté	37.84
2021	1	18	2120644	Ricaurte	26.68
2021	2	18	2120644	Ricaurte	168.37
2021	3	18	2120644	Ricaurte	76.28
2021	4	18	2120644	Ricaurte	169.99
2021	5	18	2120644	Ricaurte	142.42
2021	6	18	2120644	Ricaurte	28.51
2021	7	18	2120644	Ricaurte	20.6
2021	8	18	2120644	Ricaurte	31.16
2021	9	18	2120644	Ricaurte	36.68
2021	10	18	2120644	Ricaurte	109.5
2021	11	18	2120644	Ricaurte	44.08
2021	12	18	2120644	Ricaurte	43.19
2021	1	19	2119514	Fusagasugá	73.88
2021	2	19	2119514	Fusagasugá	63.8
2021	3	19	2119514	Fusagasugá	72.38
2021	4	19	2119514	Fusagasugá	13.95

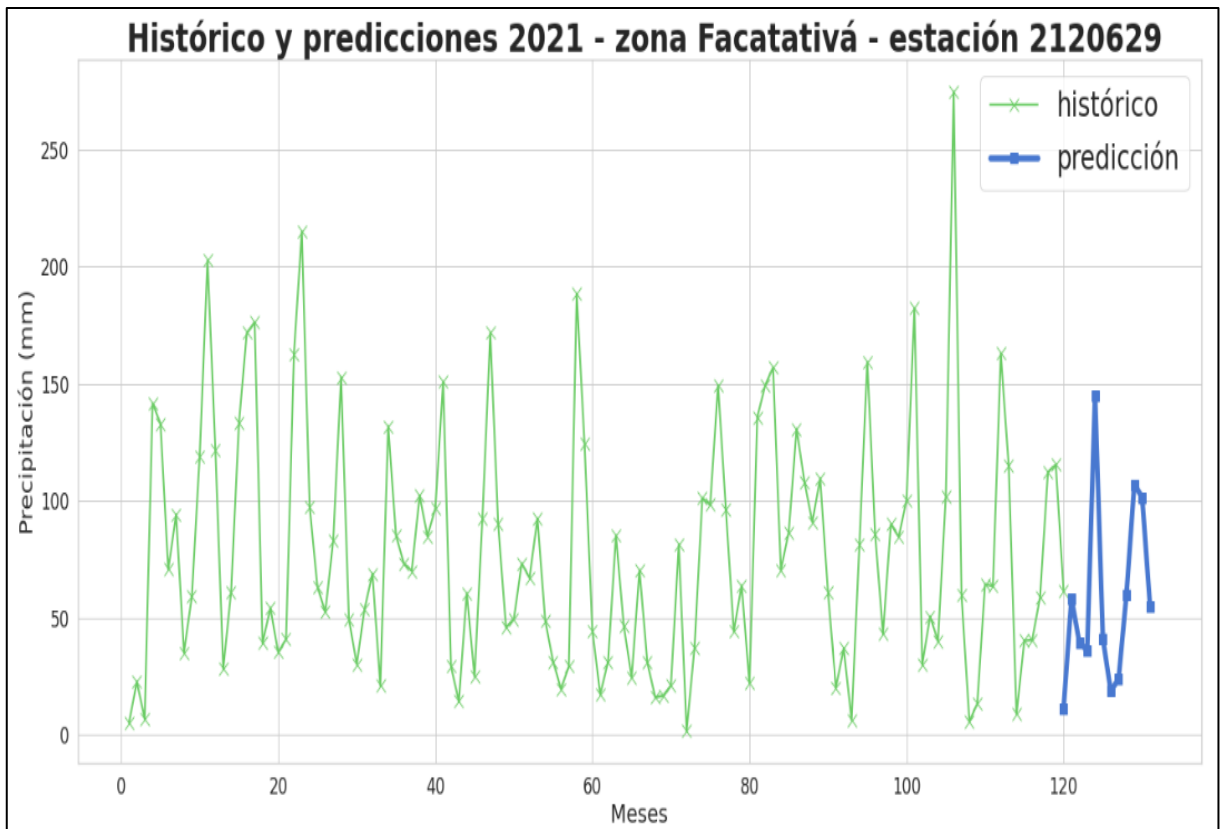
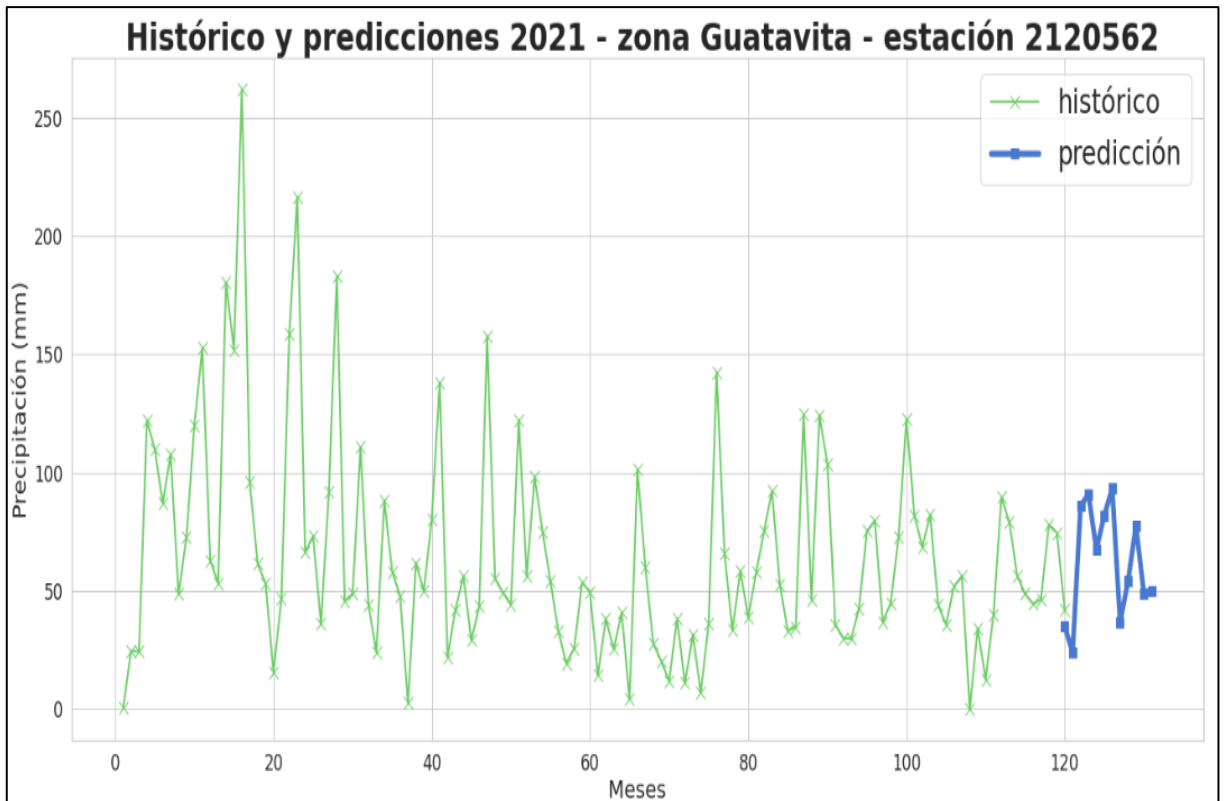
2021	5	19	2119514	Fusagasugá	68.57
2021	6	19	2119514	Fusagasugá	45.14
2021	7	19	2119514	Fusagasugá	35.75
2021	8	19	2119514	Fusagasugá	41.56
2021	9	19	2119514	Fusagasugá	88.22
2021	10	19	2119514	Fusagasugá	91.26
2021	11	19	2119514	Fusagasugá	87.97
2021	12	19	2119514	Fusagasugá	36.6

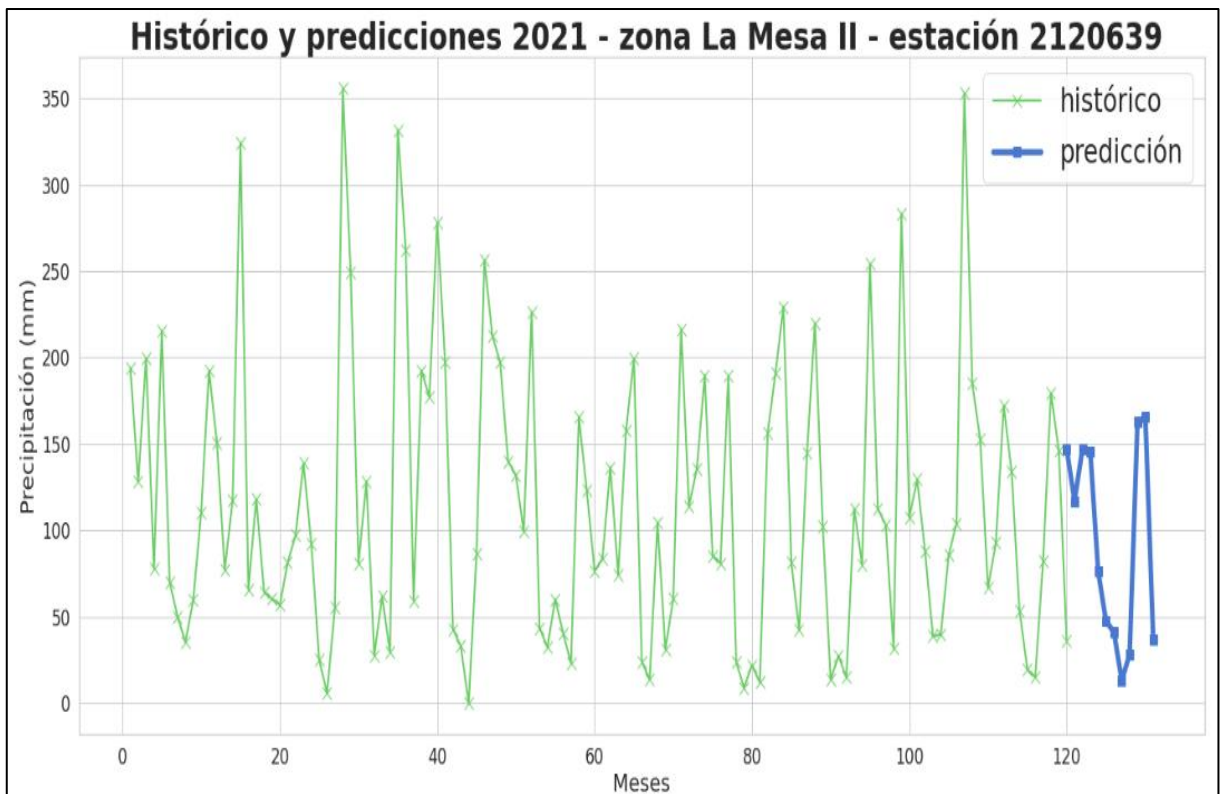
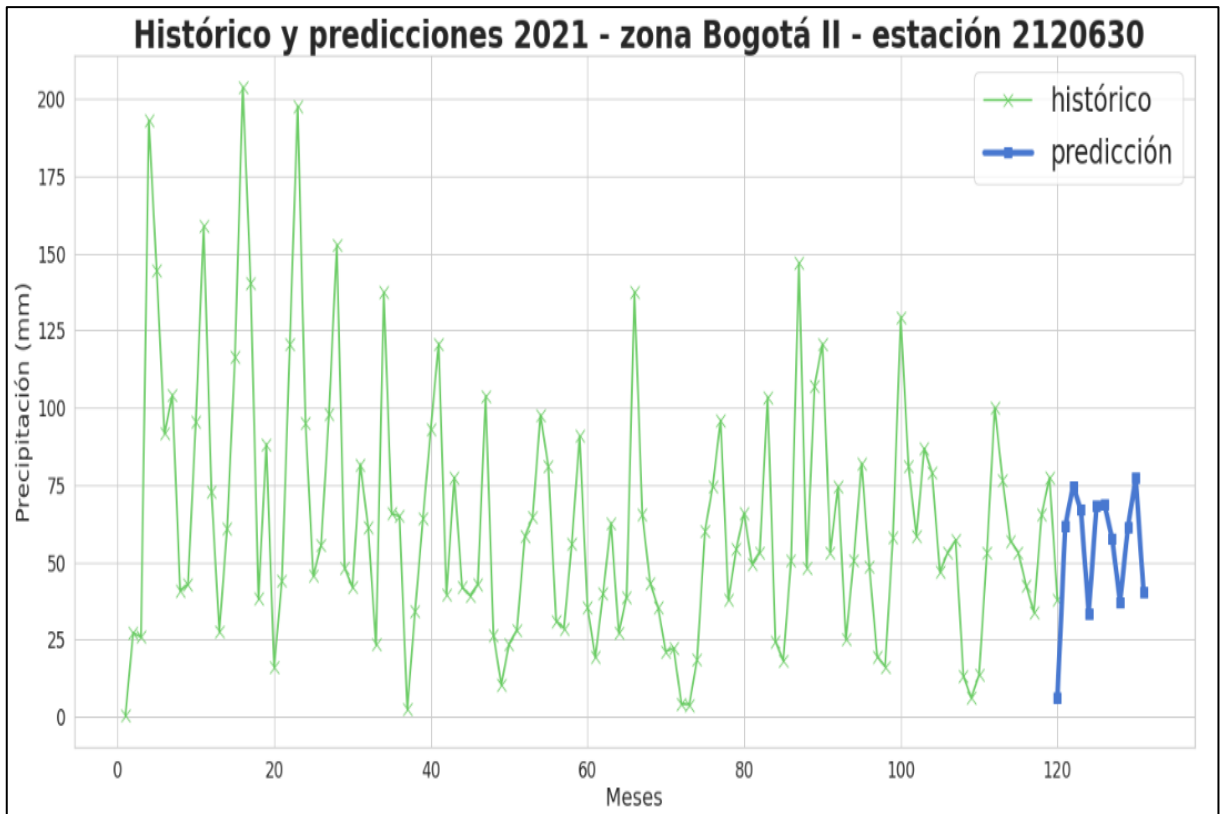
Gráficas predicciones niveles de precipitación

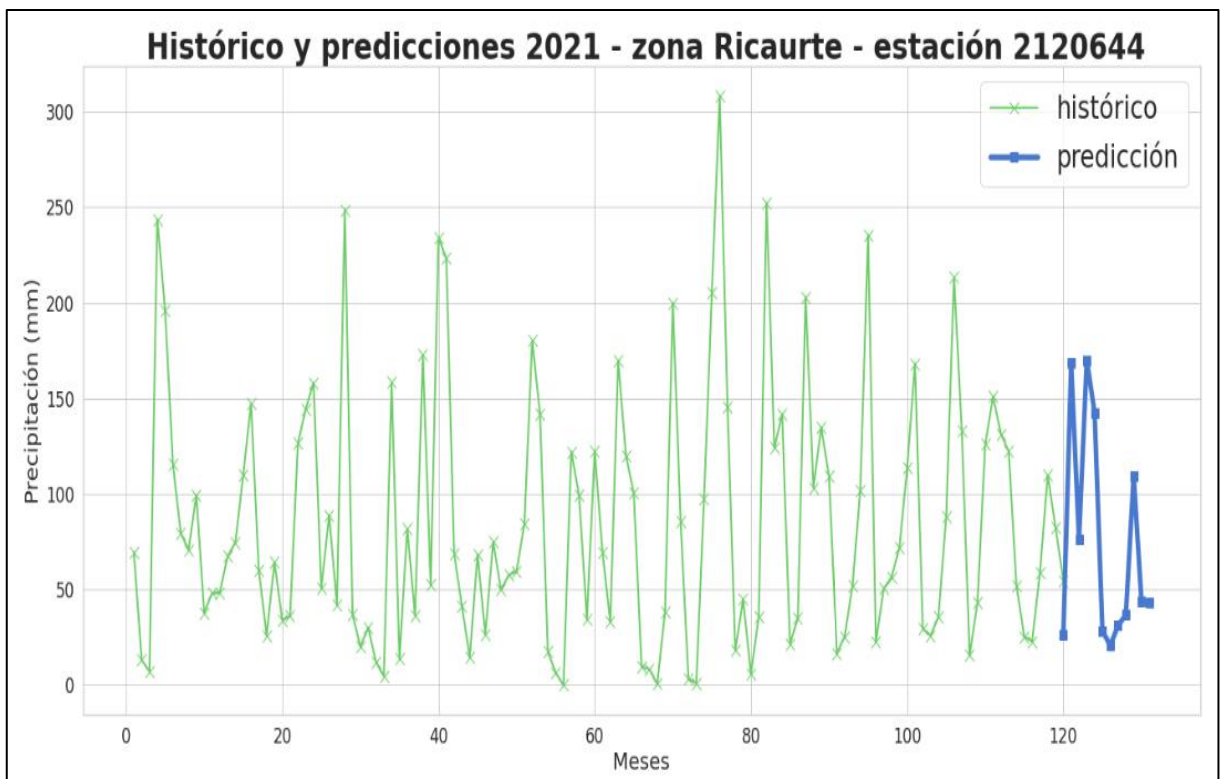
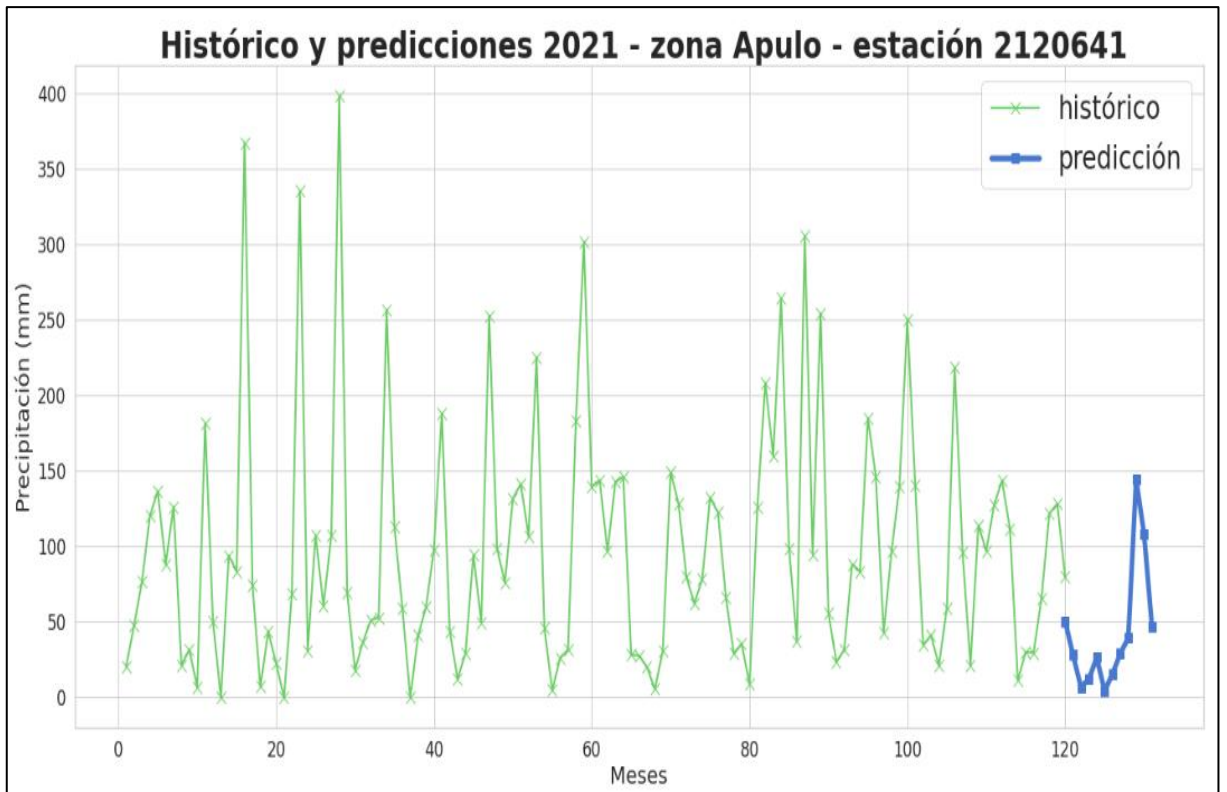


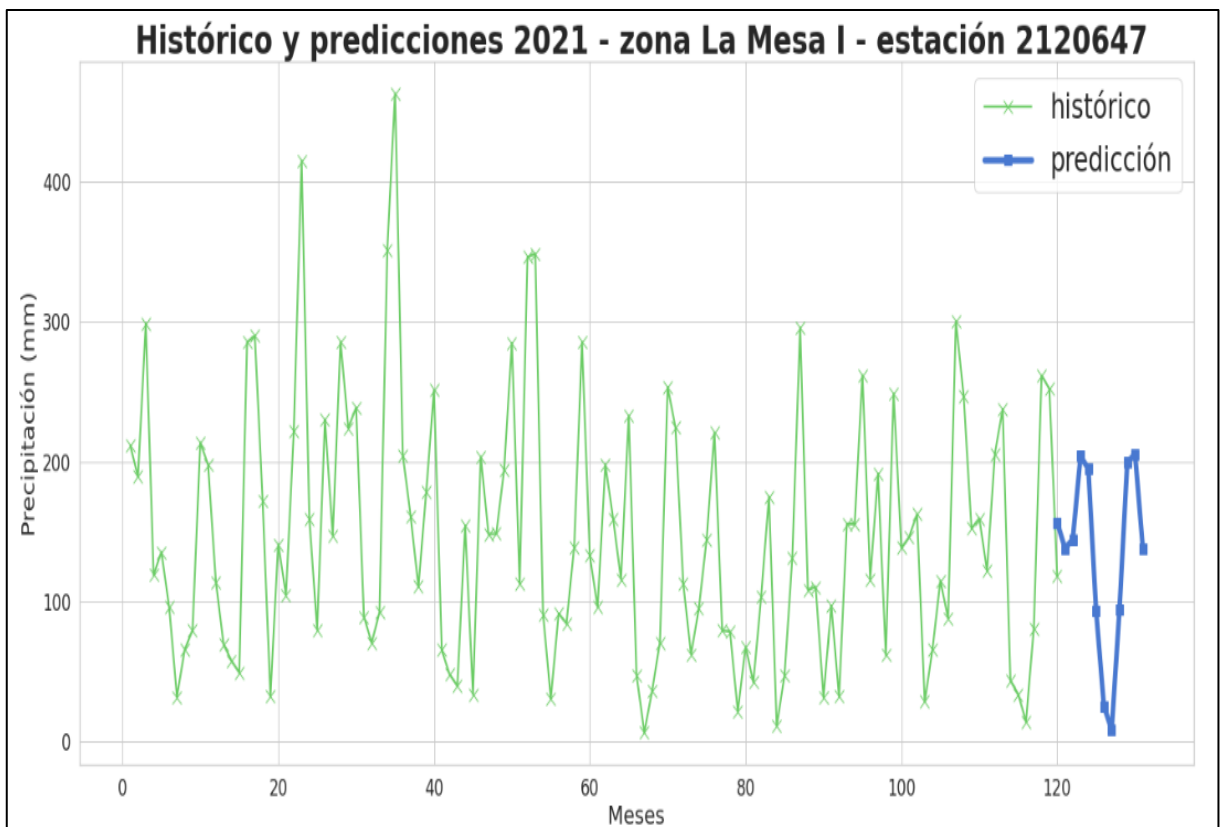
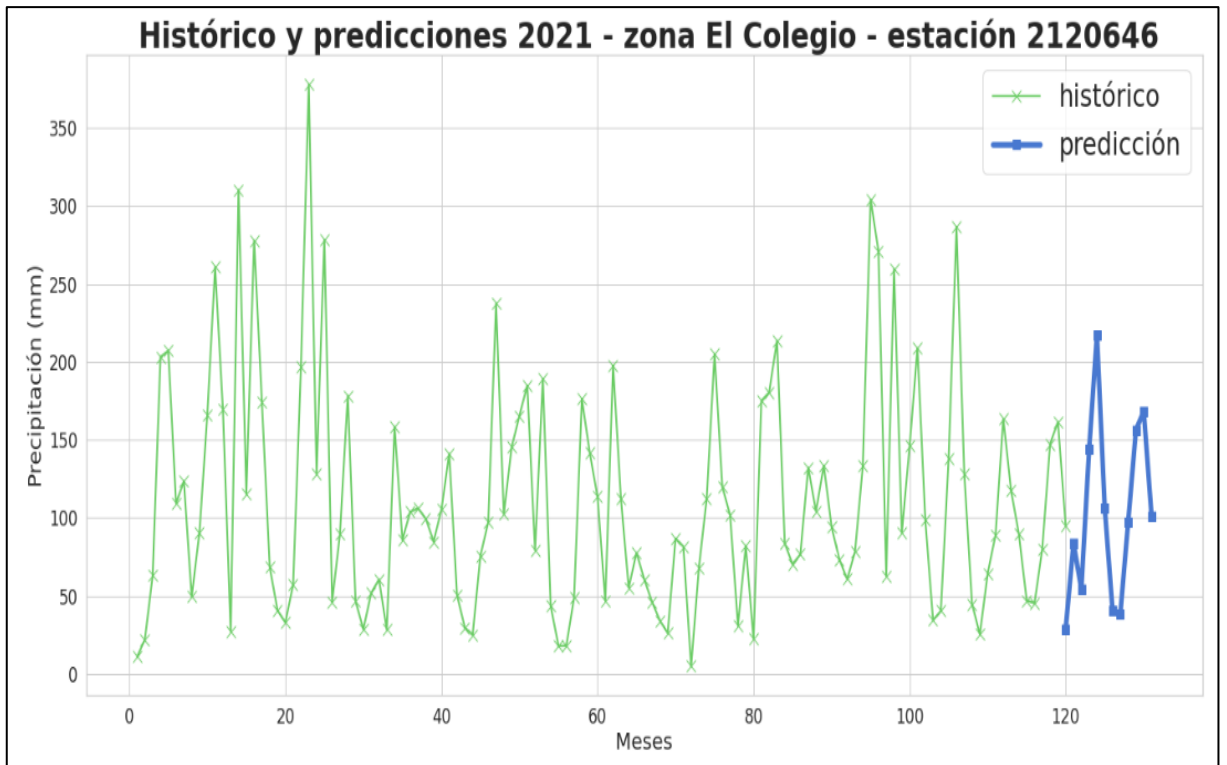


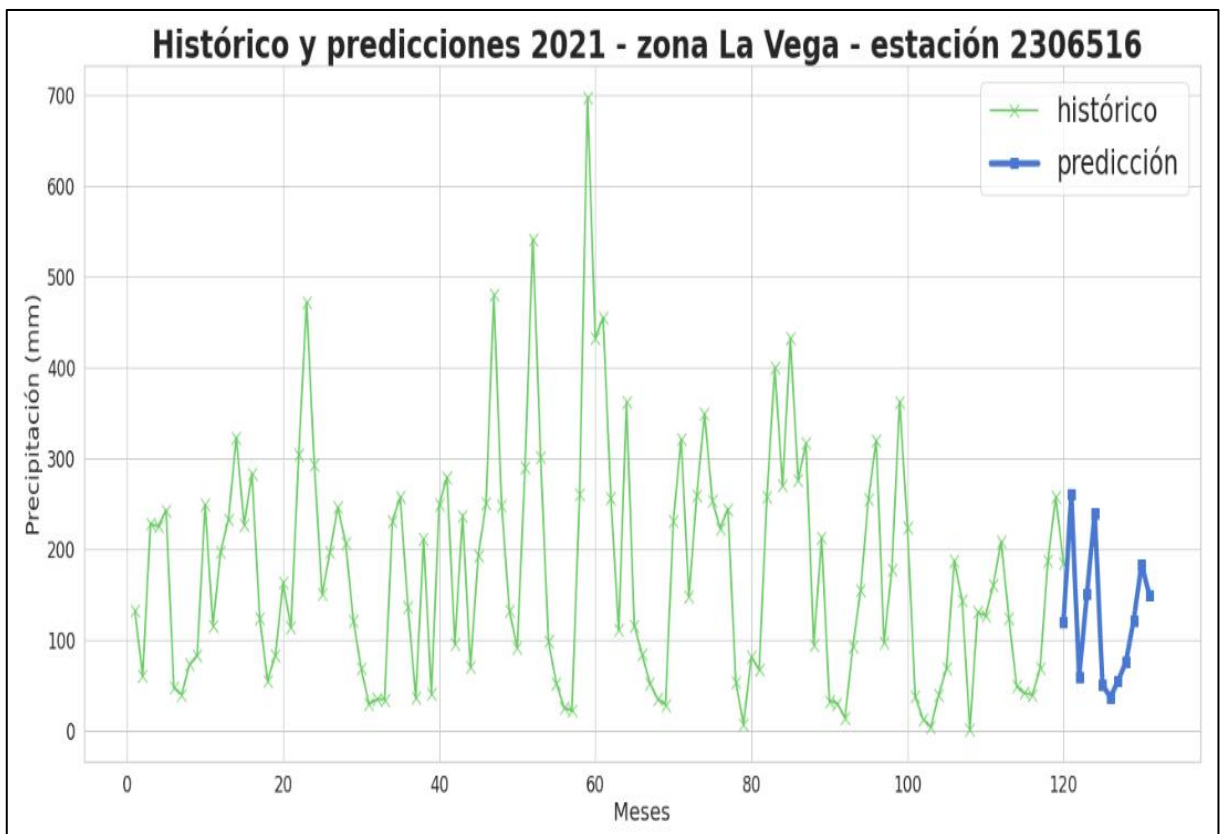
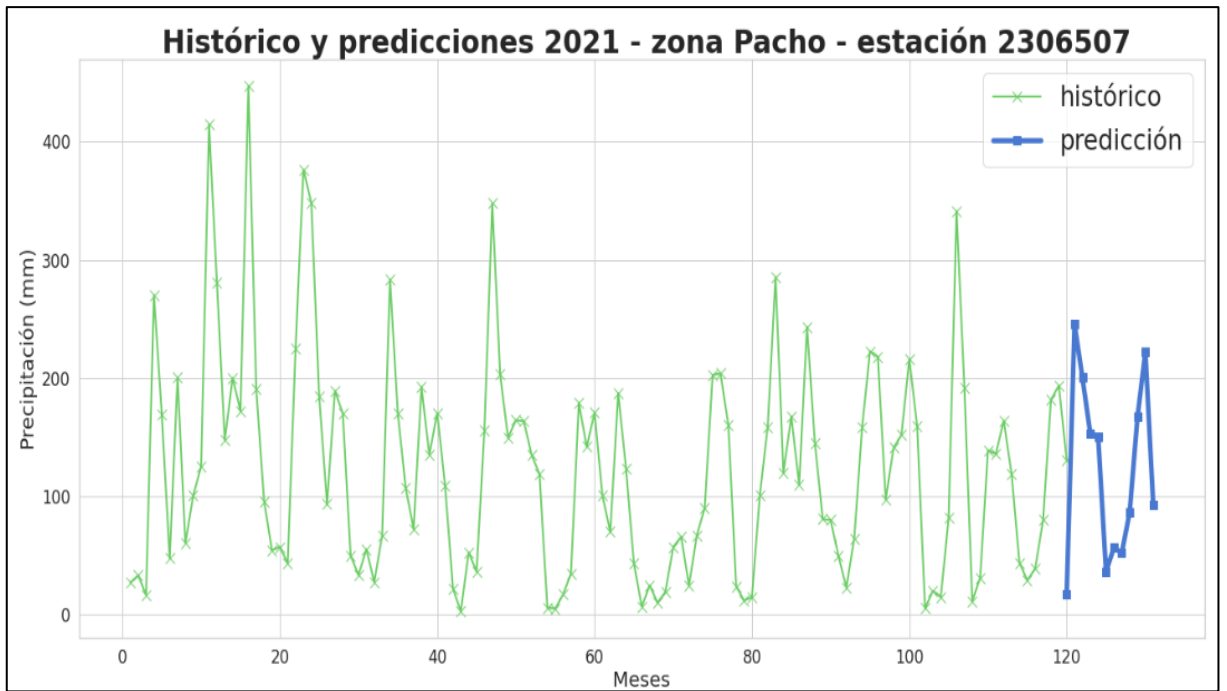


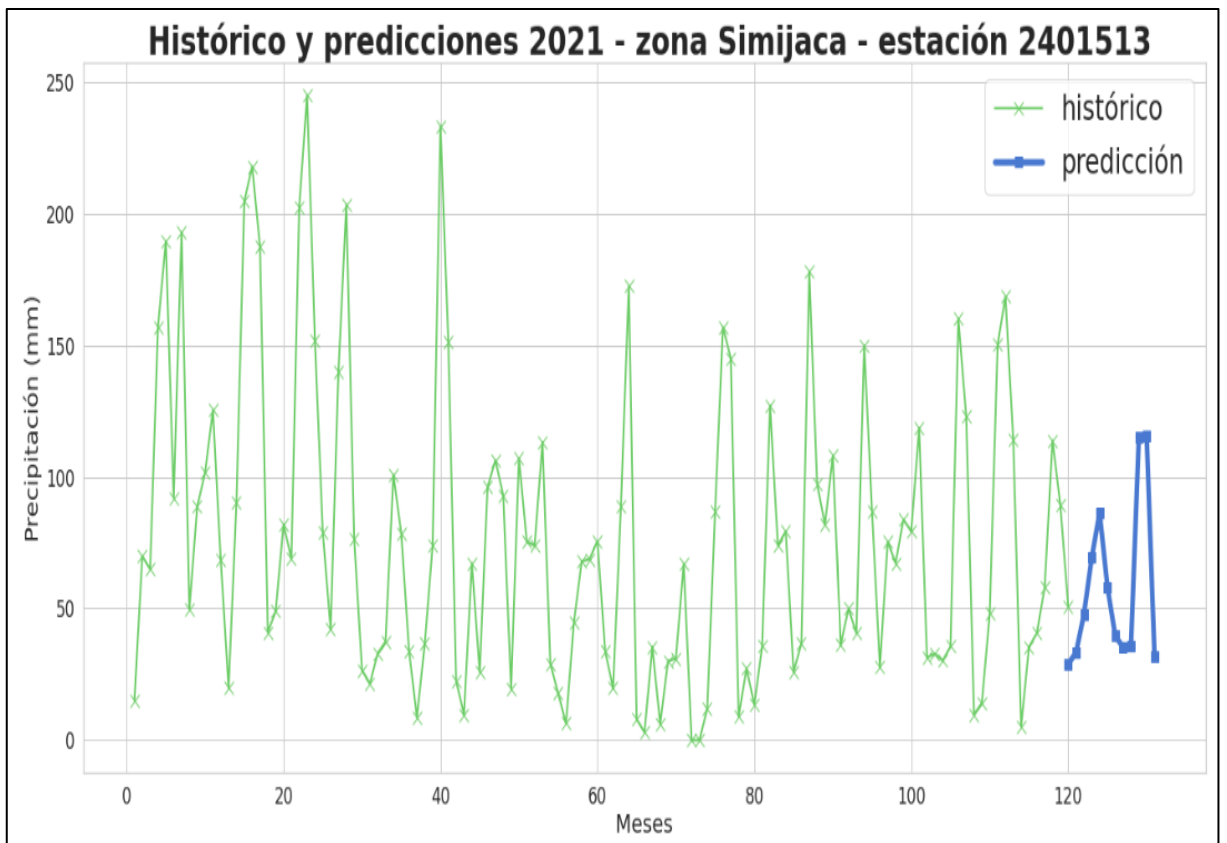
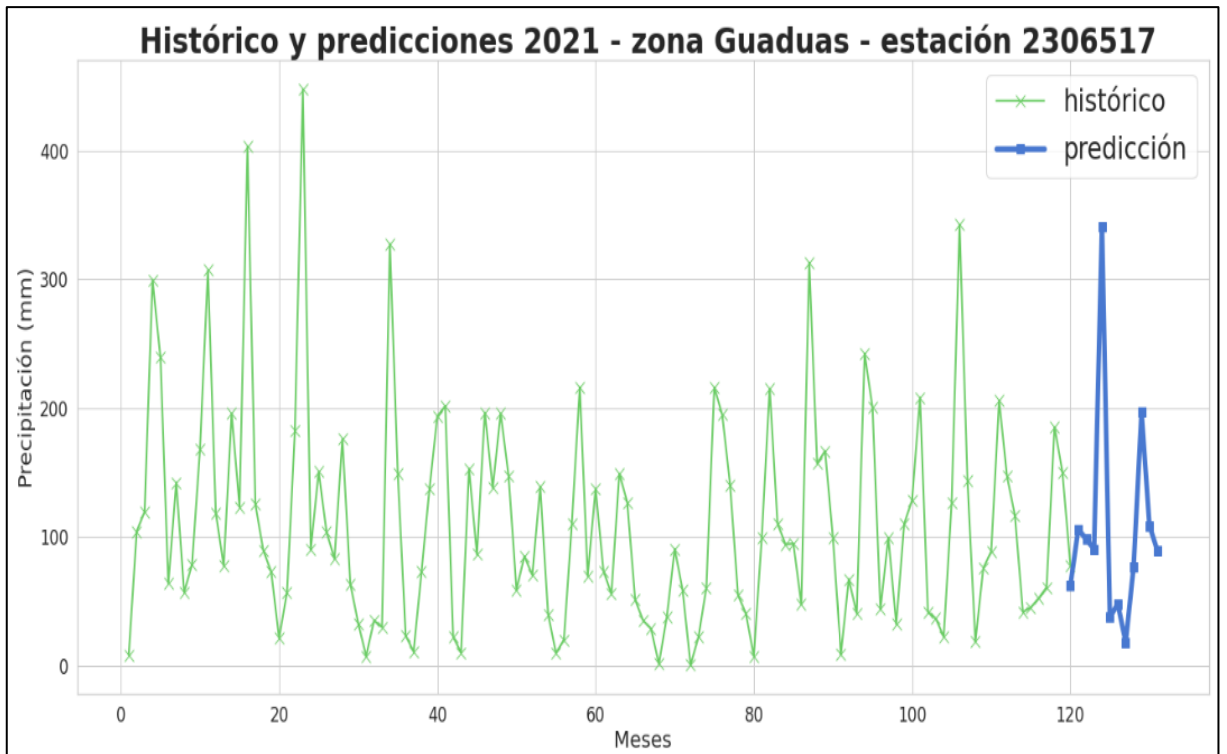


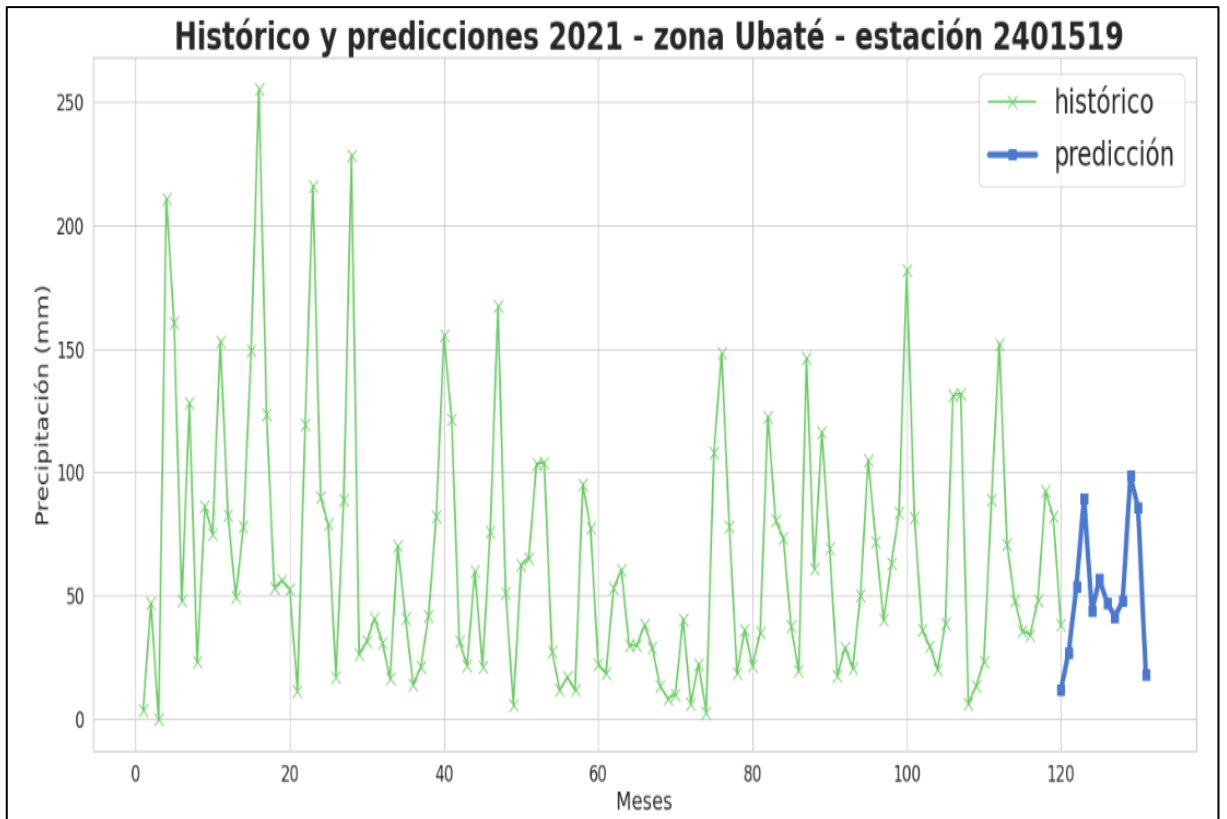












Anexo III. Clasificación zonas agrícolas Cundinamarca 2021

A continuación, se expone la clasificación mensual de zonas agrícolas en el territorio de Cundinamarca para el año 2021.

