



Pontificia Universidad  
**JAVERIANA**  
Cali

**RECONSTRUCCIÓN DE SERIES CLIMÁTICAS POR MEDIO DE LA  
COMBINACIÓN DE DATOS DE ESTACIONES CLIMÁTICAS EN TIERRA E  
IMÁGENES SATELITALES DENTRO DEL TERRITORIO COLOMBIANO  
APLICANDO MÉTODOS ESTADÍSTICOS Y DE APRENDIZAJE AUTOMÁTICO**

*Raúl Fernando Castro Suárez*

*Código 8975815*

*Edwin Alexander Ochoa Sánchez*

*Código 8975367*

*Juan Carlos Villalba Acevedo*

*Código 8975646*

*Proyecto Aplicado para optar al título de  
Magister en Ciencia de Datos*

Director

Profesor David Arango Londoño

FACULTAD DE INGENIERÍA Y CIENCIAS  
MAESTRÍA EN CIENCIA DE DATOS  
SANTIAGO DE CALI, 7 DE JUNIO DE 2024

## TABLA DE CONTENIDO

INTRODUCCIÓN .....	9
1. DEFINICIÓN DEL PROBLEMA .....	10
1.1 PLANTEAMIENTO DEL PROBLEMA .....	10
1.2 FORMULACIÓN DEL PROBLEMA.....	11
2. OBJETIVOS DEL PROYECTO.....	12
2.1 OBJETIVO GENERAL.....	12
2.2 OBJETIVOS ESPECÍFICOS.....	12
3. MARCO TEÓRICO.....	13
3.1 ANTECEDENTES .....	13
3.2 PREDICCIÓN DE LA PRECIPITACIÓN.....	18
3.2.1 Aprendizaje Automático.....	18
3.2.2 Aprendizaje Estadístico .....	19
3.3 ÁREA DE ESTUDIO .....	19
3.3.1 Climatología.....	20
3.3.2 Aspectos Fisiográficos .....	21
4. METODOLOGÍA.....	23
4.1 SELECCIÓN FUENTE DE DATOS.....	23
4.1.1 Recopilación de datos de estaciones meteorológicas en tierra.....	24
4.1.2 Extracción de datos satelitales .....	25
4.1.3 Lectura de archivos .....	27
4.2 ANÁLISIS EXPLORATORIO DE DATOS.....	28
4.3 ANÁLISIS ESPACIAL DE DATOS.....	28
4.4 TRANSFORMACIÓN DE DATOS.....	29
4.5 APLICACIÓN TÉCNICAS DE IMPUTING .....	29
4.6 SELECCIÓN DE MODELOS .....	29
4.6.1 Modelos implementados .....	30
4.7 APLICACIÓN MODELOS DE REGRESIÓN .....	37
4.7.1 GridSearch .....	38
4.8 APLICACIÓN DE MODELOS SERIES DE TIEMPO .....	39
4.9 APLICACIÓN DE MODELOS DE CLASIFICACIÓN .....	39

4.9.1	Clasificación por intensidad de lluvia .....	39
4.10	APLICACIÓN DE REDES NEURONALES .....	40
4.11	EVALUACIÓN DE RESULTADOS .....	42
4.11.1	Métricas Regresión.....	43
4.11.2	Métricas Clasificación.....	43
5.	ANÁLISIS DE DATOS.....	45
5.1	ANÁLISIS EXPLORATORIO DE DATOS.....	45
5.1.1	Conjunto de datos .....	45
5.1.2	Periodo Analizado .....	45
5.1.3	Estadística Descriptiva.....	45
5.1.4	Visualización Datos.....	47
5.1.5	Correlación de Datos .....	48
5.1.6	Análisis de puntos con baja correlación .....	49
5.2	AGREGACIONES TEMPORALES .....	52
5.2.1	Comportamiento temporal de la lluvia con diferentes agregaciones.....	53
5.2.2	Correlación de la lluvia con diferentes agregaciones temporales .....	57
6.	RESULTADOS MODELOS .....	60
6.1	APRENDIZAJE AUTOMÁTICO .....	60
6.1.1	Modelos de regresión .....	60
6.1.2	Modelos de clasificación .....	65
6.2	APRENDIZAJE ESTADÍSTICO .....	71
6.2.1	Modelos series de tiempo.....	71
6.3	COMPARACIÓN DE RESULTADOS.....	81
6.3.1	Modelos de regresión .....	81
6.3.2	Modelos series de tiempo.....	85
6.3.3	Modelos de clasificación .....	86
6.3.4	Otros modelos .....	89
7.	CONCLUSIONES Y OBSERVACIONES .....	91
7.1	CONCLUSIONES .....	91
7.2	OBSERVACIONES Y POSIBLES TRABAJOS FUTUROS .....	92
	REFERENCIAS BIBLIOGRÁFICAS .....	94

## LISTA DE FIGURAS

Fig. 1. Mapa de localización del área de estudio .....	20
Fig. 2. Relieve / orografía Valle del Cauca.....	22
Fig. 3. Esquema comprensión del proyecto. ....	23
Fig. 4. Totalidad estaciones IDEAM en el departamento del Valle del Cauca.....	24
Fig. 5. Estaciones hidrometereológicas seleccionadas para el proyecto. ....	25
Fig. 6. Precipitaciones mundiales.....	26
Fig. 7. Precipitaciones anuales promedio. ....	26
Fig. 8. Precipitación anual promedio para Colombia. ....	27
Fig. 9. Dataframe estaciones en tierra.....	28
Fig. 10. Dataframe datos satelitales.....	28
Fig. 11. Dataframe unificado.....	29
Fig. 12. Modelos implementados en el proyecto.....	30
Fig. 13. Ejemplo obtención de parámetros de un estimador / modelo.....	38
Fig. 14. Ejemplo aplicación GridSearchCV para ajuste de hiperparámetros de un estimador / modelo. .....	38
Fig. 15. Intensidad de lluvia diaria [32] .....	40
Fig. 16. Intensidad de lluvia por horaria, diaria y mensual [33].....	40
Fig. 17. Clasificación intensidad de lluvia diaria NCHM [34] .....	40
Fig. 18. Arquitectura de la red neuronal completamente conectada.....	41
Fig. 19. Resumen modelo FCNN.....	42
Fig. 20. Periodo analizado conjunto de datos.....	45
Fig. 21. Datos faltantes por Estación.....	46
Fig. 22. Mapa de calor Datos faltantes en el periodo analizado.....	47
Fig. 23. Mapa de calor de correlación para las Estaciones – datos diarios.....	48
Fig. 24. Mapa de calor de correlación para Satélite – datos diarios.....	49
Fig. 25. Localización de estaciones con correlación débil.....	50
Fig. 26. Correlación y porcentaje de días sin lluvia (ceros) - datos diarios .....	51
Fig. 27. Mapa de calor lluvia acumulada mensual .....	52
Fig. 28. Lluvia acumulada diaria - Estación 26070760.....	53
Fig. 29. Lluvia acumulada semanal - Estación 26070760 .....	53
Fig. 30. Lluvia acumulada mensual - Estación 26070760.....	54
Fig. 31. Lluvia acumulada anual - Estación 26070760.....	54
Fig. 32. Matriz de dispersión lluvia acumulada diaria - Estación 26070760 .....	55
Fig. 33. Matriz de dispersión lluvia acumulada semanal- Estación 26070760.....	55
Fig. 34. Matriz de dispersión lluvia acumulada mensual- Estación 26070760.....	56
Fig. 35. Matriz de dispersión lluvia acumulada anual- Estación 26070760 .....	56
Fig. 36. Correlación y porcentaje de días sin lluvia (ceros) - datos semanales .....	57
Fig. 37. Correlación y porcentaje de días sin lluvia (ceros) - datos mensuales.....	58

Fig. 38. Correlación y porcentaje de días sin lluvia (ceros) - datos anuales.....	58
Fig. 39. Mapa de calor de correlación para datos en Satélite y Estaciones Tierra – datos mensuales. .....	59
Fig. 40. Resultados R2 predicciones diarias OLS .....	60
Fig. 41. Resultados R2 predicciones semanales OLS .....	61
Fig. 42. Resultados R2 predicciones mensuales OLS.....	61
Fig. 43. Resultados R2 predicciones diarias SVR .....	62
Fig. 44. Resultados R2 predicciones semanales SVR.....	62
Fig. 45. Resultados R2 predicciones mensuales SVR.....	63
Fig. 46. Resultados R2 predicciones diarias DTR.....	63
Fig. 47. Resultados R2 predicciones semanales DTR.....	64
Fig. 48. Resultados R2 predicciones mensuales DTR .....	64
Fig. 49. Resultados Accuracy predicciones diarias Logit .....	65
Fig. 50. Resultados Accuracy predicciones semanales Logit.....	66
Fig. 51. Resultados Accuracy predicciones mensuales Logit.....	66
Fig. 52. Resultados Accuracy predicciones diarias SVC.....	67
Fig. 53. Resultados Accuracy predicciones semanales SVC.....	68
Fig. 54. Resultados Accuracy predicciones mensuales SVC .....	68
Fig. 55. Resultados Accuracy predicciones diarias DTC.....	69
Fig. 56. Resultados Accuracy predicciones semanales DTC .....	69
Fig. 57. Resultados Accuracy predicciones mensuales DTC.....	70
Fig. 58. Comparación resultados Accuracy Logit por Intensidad de Lluvia .....	70
Fig. 59. Mayor cantidad de valores consecutivos no nulos por estaciones .....	71
Fig. 60. Comportamiento temporal de la estación con mejor convergencia entre la correlación y mayor cantidad de valores no nulos consecutivos - [26070760].....	72
Fig. 61. Descomposición de la serie temporal en sus componentes principales: tendencia, estacionalidad y residuos - [26070760] .....	72
Fig. 62. Resultados de la prueba estadística Dickey Fuller - [26070760] .....	73
Fig. 63. Estimación de los parámetros (p, d, q) del modelo ARIMA a aplicar - [26070760].....	73
Fig. 64. Summary del modelo SARIMA implementado - [26070760] .....	74
Fig. 65. Métricas del desempeño del modelo SARIMA - [26070760] .....	74
Fig. 66. Gráfico de los errores residuales de las predicciones del modelo SARIMA - [26070760]....	75
Fig. 67. Resultados de las predicciones del modelo SARIMA - [26070760] .....	75
Fig. 68. Resultado predicción / pronóstico diario Prophet – Estación 26070760.....	76
Fig. 69. Resultado predicción / pronóstico semanal Prophet – Estación 26070760.....	77
Fig. 70. Resultado predicción / pronóstico mensual Prophet – Estación 26070760 .....	77
Fig. 71. Resultado predicción / pronóstico anual Prophet – Estación 26070760 .....	78
Fig. 72. Resultados R2 predicciones diarias Prophet .....	79
Fig. 73. Resultados R2 predicciones semanales Prophet .....	79
Fig. 74. Resultados R2 predicciones mensuales Prophet.....	80
Fig. 75. Resultados R2 predicciones mensuales Sktime .....	80
Fig. 76. Comparación resultados R2 OLS / regresión lineal .....	81
Fig. 77. Comparación resultados R2 SVR / máquinas de soporte vectorial .....	82

Fig. 78. Comparación resultados R2 DTR / árboles de decisión.....	82
Fig. 79. Comparación resultados R2 modelos de regresión.....	83
Fig. 80. Comparación de resultados R <sup>2</sup> por agrupación de tiempo y modelo.....	83
Fig. 81. Comparación de resultados MAE por agrupación de tiempo y modelo .....	84
Fig. 82. Comparación de resultados MSE por agrupación de tiempo y modelo .....	84
Fig. 83. Comparación de resultados error máximo por agrupación de tiempo y modelo .....	85
Fig. 84. Comparación resultados R2 Prophet / forecasting series de tiempo.....	86
Fig. 85. Comparación resultados Accuracy Logit / regresión logística.....	86
Fig. 86. Comparación resultados Accuracy SVC / Máquinas de soporte vectorial.....	87
Fig. 87. Comparación resultados Accuracy DTC / árboles de decisión .....	87
Fig. 88. Comparación resultados Accuracy modelos de clasificación .....	88
Fig. 89. Comparación de resultados por agrupación de tiempo, modelo y métricas de clasificación .....	88
Fig. 90. Comparación resultados R2 simple imputing.....	89
Fig. 91. Comportamiento de R2 para cada una de las épocas de la red neuronal en los conjuntos de entrenamiento y pruebas.....	90

## LISTA DE TABLAS

TABLA I. TOP 10 ESTACIONES CON MÁS DATOS FALTANTES.....	46
---	----

## LISTA DE ANEXOS

<b>ANEXO A</b> .....	96
<b>ANEXO B</b> .....	106
<b>ANEXO C</b> .....	116
<b>ANEXO D</b> .....	140

## INTRODUCCIÓN

Comprender y predecir los patrones climáticos es fundamental para tomar decisiones informadas en diversas áreas, como la agricultura, la gestión de recursos naturales y la planificación urbana. Para lograr una comprensión precisa y detallada del clima, es necesario contar con datos climáticos confiables y completos [1]. Las estaciones climáticas en tierra proporcionan mediciones precisas en puntos específicos, pero pueden ser escasas en muchas regiones, lo que resulta en una cobertura espacial limitada. Por otro lado, las imágenes de satélite ofrecen una cobertura global, lo que permite a los investigadores observar y monitorear los patrones climáticos a gran escala, pero pueden tener limitaciones en términos de resolución temporal y espacial [2].

En el proyecto se recopilieron datos de más 30 años de datos históricos para 58 estaciones climáticas en tierra del Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia (IDEAM) [3] [4] instaladas en el departamento del Valle del Cauca, de la misma manera se obtuvieron los datos históricos satelitales del *Climate Hazards Group InfraRed Precipitation with Station* (CHIRPS) [5] [6] para los mismos puntos. Se realizó el análisis exploratorio de los datos, las tendencias de las series de tiempo estudiadas, la dispersión y correlación de los datos.

La comprensión de los datos se realizó utilizando tanto algoritmos en R, como notebooks en Python para el cargue, descripción y visualización y análisis de los datos.

El análisis de los datos del proyecto permite evidenciar que como era de esperar todas las 58 estaciones en tierra cuentan con datos faltantes (entre el 0.02% al 27.25% de los datos), así como observar que en el área de estudio se presenta una tendencia bimodal de lluvia (épocas de lluvia en abril-mayo y octubre-noviembre).

Los diferentes modelos utilizados tanto de aprendizaje automático, como estadístico permiten evidenciar que el problema abordado se desempeña de mejor manera a nivel de clasificación que a nivel de regresión y dentro de este último, los valores predichos presentan mejores métricas para agregaciones mensuales que con predicciones diarias o semanales de lluvia. A la hora de hacer pronósticos de las series de tiempo, se evidencia que algunos modelos aprenden bien la tendencia de los datos históricos de los datos, pero no predicen muy bien los picos altos o extremos de lluvia.

# 1. DEFINICIÓN DEL PROBLEMA

## 1.1 PLANTEAMIENTO DEL PROBLEMA

El cambio climático es uno de los desafíos más apremiantes que enfrenta la humanidad en la actualidad. Comprender y predecir los patrones climáticos es fundamental para tomar decisiones informadas en diversas áreas, como la agricultura, la gestión de recursos naturales y la planificación urbana. Para lograr una comprensión precisa y detallada del clima, es necesario contar con datos climáticos confiables y completos [1].

Las estaciones climáticas en tierra y las imágenes de satélite son fuentes importantes de información climática. Sin embargo, cada una de estas fuentes tiene limitaciones inherentes. Las estaciones climáticas en tierra proporcionan mediciones precisas en puntos específicos, pero pueden ser escasas en muchas regiones, lo que resulta en una cobertura espacial limitada. Por otro lado, las imágenes de satélite ofrecen una cobertura global, lo que permite a los investigadores observar y monitorear los patrones climáticos a gran escala, pero pueden tener limitaciones en términos de resolución temporal y espacial [2].

Los satélites climatológicos proporcionan información continua y global de la tierra, permitiendo observar los sistemas climáticos en diferentes regiones y escalas temporales. También, pueden detectar y rastrear fenómenos climáticos importantes, como tormentas tropicales, huracanes, ciclones, frentes fríos, sequías y otros eventos extremos. Estas observaciones son vitales para predecir la trayectoria y la intensidad de los eventos climáticos y emitir advertencias tempranas a las poblaciones afectadas. Las imágenes de satélite también se utilizan para estudiar los cambios en la cubierta terrestre, como la deforestación, la expansión urbana, la pérdida de glaciares y otros impactos humanos en el medio ambiente. Estos cambios tienen un impacto directo en el clima de una región determinada. Otro aporte importante es en el análisis de la temperatura de la superficie, mediante la medición en diferentes momentos del día y en diferentes regiones, aportando datos que ayudan a comprender las variaciones de temperatura en el tiempo y el espacio. Finalmente, para la estimación de variables climáticas, las imágenes satelitales son un insumo primordial, permitiendo estimar variables climáticas importantes, como la radiación solar, la evaporación, la humedad del suelo y la temperatura del agua superficial del océano. Estos datos son fundamentales para comprender los procesos climáticos y mejorar los modelos climáticos utilizados en la predicción del tiempo y del clima [7].

El proyecto mantuvo un enfoque que combina los datos de estaciones climáticas en tierra e imágenes de satélite para reconstruir series con un alto nivel de completitud y detalle. Para lograr esto, se utilizaron métodos estadísticos y de aprendizaje automático para analizar y procesar los datos, aprovechando las fortalezas de cada fuente de información. Este enfoque permitió superar

las limitaciones individuales de las estaciones climáticas en tierra y las imágenes de satélite, y proporcione una visión más completa y precisa del clima [8].

A pesar de lo anterior, existen desafíos técnicos y metodológicos que se abordaron en este proyecto. Algunos de los cuales incluyen la integración y normalización de los datos de diferentes fuentes, la selección adecuada de variables climáticas relevantes, el diseño de modelos de aprendizaje automático que sean capaces de capturar las complejidades de los datos climáticos, y la validación de los resultados obtenidos [9].

## **1.2 FORMULACIÓN DEL PROBLEMA**

El presente trabajo, nos lleva a responder la siguiente pregunta principal de investigación:

¿De qué manera beneficia aplicar el análisis estadístico y las herramientas de aprendizaje automático para la reconstrucción y mejora de la integridad de las series climáticas haciendo uso de estaciones en tierra e imágenes satelitales para el territorio colombiano?

Y las siguientes preguntas complementarias al estudio:

¿Qué tipo de imágenes de satélite son las más apropiadas para generar los datos faltantes dentro de una serie climática?

¿Existen bases de datos especializadas con datos históricos del comportamiento del clima y que brinden información validada de modo tal que se puedan comparar los resultados obtenidos en el desarrollo del proyecto?

¿Se logra un impacto positivo y significativo al combinar los datos de dos fuentes de información meteorológica, que son las estaciones en tierra y las imágenes satelitales, con el propósito de reconstruir los datos faltantes en lo que respecta a las series climáticas, utilizando métodos estadísticos y modelos de aprendizaje automático?

## **2. OBJETIVOS DEL PROYECTO**

### **2.1 OBJETIVO GENERAL**

Desarrollar un modelo basado en datos de estaciones climáticas en tierra e imágenes satelitales para la reconstrucción de datos de series climáticas en Colombia por medio del análisis estadístico y herramientas de aprendizaje automático.

### **2.2 OBJETIVOS ESPECÍFICOS**

- Seleccionar las fuentes de datos de estaciones climáticas en tierra e imágenes de satélite más apropiadas para analizar y reconstruir las series climáticas.
- Identificar patrones y tendencias mediante el análisis de las series climáticas estudiadas de cara a realizar una adecuada imputación de valores faltantes.
- Evaluar el desempeño de los modelos seleccionados a través de las métricas definidas, las cuales son el  $R^2$  (coeficiente de determinación), el MSE (error cuadrático medio) y el MAE (error absoluto medio). Lo anterior entorno al resultado de la imputación de valores faltantes en los conjuntos de datos de las estaciones en tierra, así como el ajuste de los modelos al comportamiento de los datos combinados.

### **3. MARCO TEÓRICO**

#### **3.1 ANTECEDENTES**

Dentro de la gran variedad de definiciones que se han planteado con relación al clima, destacan dos por su aceptación y conveniencia en diferentes contextos. Por un lado, está de definición de Hann la cual establece lo siguiente: “el clima es el conjunto de fenómenos meteorológicos que caracteriza el estado medio de la atmósfera en un punto cualquiera de la Tierra”; mientras que Sorre define el clima como: “la serie de los estados de la atmósfera por encima de un lugar en su sucesión habitual” [10]. Sin embargo, es posible establecer un factor común entre todas las definiciones, las cuales concuerdan con que, al referirse al clima, se está hablando de condiciones atmosféricas promedio, habituales o que caracterizan, mientras que el concepto de tiempo se asocia a condiciones meteorológicas instantáneas.

El estudio del clima parte o encuentra un importante sustento en el registro de información captado por medio de las estaciones meteorológicas, las cuales capturan información de diferentes variables meteorológicas tales como la velocidad del viento, estado del tiempo, espesor de nieve, visibilidad, presión, humedad, precipitaciones, temperatura, entre otras; estas dos últimas serán el objeto de análisis del presente proyecto. Dicho consolidado de datos hacen referencia a las denominadas series climáticas por medio de las que es posible estimar diferentes estados o características meteorológicas de una zona haciendo uso de cálculos estadísticos. Adicionalmente, la comparación entre series climáticas ha permitido un importante avance en el estudio del evolutivo del clima, los cambios del clima y el saber de los factores que intervienen en este.

En síntesis, las series climáticas son un conjunto de datos cuyos registros representan estados consecutivos de tiempo con relación a una zona o lugar determinado, registros medidos durante cierto periodo de tiempo, pudiendo ser de manera mensual, anual o diaria.

Los registros meteorológicos al ser observaciones tomadas en momentos específicos del día son un recurso vital para investigadores y meteorológicos, pudiendo ser utilizados en múltiples disciplinas. Como es el caso, por ejemplo, de la investigación llevada a cabo para esclarecer el accidente del avión de pasajeros L-1011 de la empresa Delta Airlines en el aeropuerto internacional de Dallas (EEUU) en agosto de 1985, en donde se consultaron registros de este tipo en la reconstrucción del mapa meteorológico buscando identificar condiciones climáticas que hubiesen podido influir en la seguridad del vuelo [11].

De igual manera, los satélites meteorológicos prestan una utilidad primordial en el entendimiento del funcionamiento del clima, puesto que permiten analizar las características de la atmósfera y demás factores que están presentes en el comportamiento del clima y los fenómenos atmosféricos. El concepto y aplicación de imágenes satelitales meteorológicas surgió en 1946, pero fue hasta 1960 que la NASA puso en órbita el primer satélite de manera exitosa bajo el nombre de TIROS. Seguido a eso, ESSA y NOAA pusieron sus propios satélites meteorológicos en órbita a finales de ese mismo

año. En 1969 el satélite Nimbus 3 mejoró significativamente los datos meteorológicos pudiendo medir la información de temperatura en la troposfera [12].

Partiendo del hecho de que los satélites meteorológicos son lanzados con la finalidad de recopilar datos del clima, estos deben equipar un conjunto de herramientas que les permitan capturar información básica tal como la temperatura del aire, las precipitaciones y el viento, como mediciones más complejas referentes a la capa de nieve y hielo de la tierra, incendios de grandes magnitudes, erupciones volcánicas o fenómenos del clima espacial como tormentas geomagnéticas.

Los satélites meteorológicos se clasifican en dos grandes categorías, los geoestacionarios y los de órbita polar. Los primeros orbitan la Tierra a una altitud de 35880 km, lo que les permite tener una posición estacionaria respecto a la rotación de la Tierra, haciendo posible la transmisión de imágenes y datos de manera continua utilizando sensores de luz visibles e infrarrojos, para su posterior aprovechamiento por los servicios meteorológicos, destacando el hecho de que este tipo de satélites pueden tomar fotografías de un mismo lugar de la Tierra cada 30 minutos, o de áreas más pequeñas entre 5 a 15 minutos. La segunda categoría corresponde a los satélites meteorológicos de órbita polar, cuya altitud es notablemente inferior a los anteriores (menor a 850 km) conservando una dirección norte – sur y sur – norte. Sin embargo, a diferencia de los satélites fijos, pueden observar cualquier lugar de la tierra, generando imágenes con resoluciones mucho mejores que los primeros debido a su cercanía con la tierra, pudiendo observar con un nivel de detalle más alto fenómenos/sucesos como tormentas, volcanes, deslizamientos de tierra, incendios forestales, etc.

Si bien es cierto que la mayoría de los satélites meteorológicos son operados y de propiedad de gobiernos nacionales, existen ciertos operadores comerciales que venden sus imágenes a determinadas organizaciones como cadenas de televisión interesadas en mostrar mapas meteorológicos. Junto a las estaciones meteorológicas, radares, boyas y globos, los satélites representan los principales instrumentos meteorológicos de medición.

A saber, se destacan ciertas desventajas de los satélites asociados al deficiente funcionamiento cerca del océano o de la costa, así como la pérdida de confianza de los datos capturados en condiciones tormentosas en las que el viento coincide con precipitaciones como la lluvia [13]. La reconstrucción de las series climáticas vistas propiamente como conjuntos de datos puede ser abordada desde dos perspectivas principales; haciendo una imputación de valores analizando la relación entre variables existentes, o, efectuando un ejercicio de pronóstico basado en métodos estadísticos. En este último, se han implementado métodos de postprocesamiento que van desde métodos simples de corrección de sesgo hasta estadística de salida de modelos basados en regresión lineal múltiple más complicados, con lo que se busca mejorar los pronósticos brutos del modelo [1]. Lo anterior es especialmente útil para pronósticos meteorológicos a corto plazo. Sin embargo, cuando de pronósticos a largo plazo se trata, sale a colación las limitaciones de los modelos lineales, puesto que es sabido que no siempre es posible obtener una buena relación de este tipo entre los predictores y los pronósticos. En ese caso, surge la oportunidad de utilizar técnicas de redes neuronales y aprendizaje automático, encontrando múltiples ventajas dentro de las cuales está el explorar características o relaciones no lineales para un conjunto de datos, además de extraer patrones más sofisticados y covariabilidades ocultas dentro de los predictores y

pronósticos multidimensionales, generando de esa manera, información estadística de alto nivel que podrá ser utilizada para corregir los pronósticos iniciales, enfocados principalmente en los considerados de largo plazo (entre la semana 3 y 4), obteniendo una eficiencia considerablemente superior en comparación a los modelos tradicionales de regresión lineal múltiple.

La reconstrucción de las series climáticas vistas propiamente como conjuntos de datos puede ser abordada desde dos perspectivas principales; haciendo una imputación de valores analizando la relación entre variables existentes, o, efectuando un ejercicio de pronóstico basado en métodos estadísticos. En este último, se han implementado métodos de postprocesamiento que van desde métodos simples de corrección de sesgo hasta estadística de salida de modelos basados en regresión lineal múltiple más complicados, con lo que se busca mejorar los pronósticos brutos del modelo [14]. Lo anterior es especialmente útil para pronósticos meteorológicos a corto plazo. Sin embargo, cuando de pronósticos a largo plazo se trata, se evidencian las limitaciones de los modelos lineales, puesto que es conocido que no siempre es posible obtener una buena relación de este tipo entre los predictores y los pronósticos. En ese caso, surge la oportunidad de utilizar técnicas de redes neuronales y aprendizaje automático, encontrando múltiples ventajas dentro de las cuales está el explorar características o relaciones no lineales para un conjunto de datos, además de extraer patrones más sofisticados y covariabilidades ocultas dentro de los predictores y pronósticos, generando de esa manera, información estadística de alto nivel que podrá ser utilizada para corregir los pronósticos iniciales, enfocados principalmente en los considerados de largo plazo (entre la semana 3 y 4), obteniendo una eficiencia considerablemente superior en comparación a los modelos tradicionales de regresión lineal múltiple.

Como parte de la revisión de la literatura para el presente proyecto, se encontró la aplicación de modelos lineales multinivel de efectos mixtos para analizar y predecir datos anidados y jerárquicos de forma múltiple, propiamente aplicado para la predicción de pronósticos climáticos en las variables de temperatura y precipitaciones basados en observaciones en la zona de Dar Chaoui al norte de Marruecos entre 1979 y 2014 [15]. A diferencia de la utilización de modelos de regresión lineal simple, en los modelos lineales mixtos de efectos de múltiples niveles, los parámetros se tratan por medio de una estimación conjunta de los efectos fijos, los efectos aleatorios y la estimación de parámetros de varianza. Dicho ejercicio demostró que los valores estimados estaban muy próximos al valor real.

Más aún, se ha innovado en múltiples alternativas concernientes a realizar predicciones en tiempo real, nuevamente, de las variables climáticas de temperatura y precipitación. Lo anterior es el caso de la Oficina de Recuperación y Servicio Nacional de Meteorología y Oceanografía (EEUU), la cual creó una subestación climática llamada Forecast Rodeo, con el objetivo de predecir acertadamente estas variables en el oeste de Estados Unidos con una anticipación enmarcada en dos periodos de tiempo, de 2 - 4 y de 4 - 6 semanas [16]. Como parte del desarrollo, se utilizaron dos modelos de regresión no lineal. El primero integra las mediciones meteorológicas y pronósticos realizados por medio de modelos dinámicos aplicados a los conjuntos de datos obtenidos de la subestación Rodeo, para posteriormente, eliminar predictores irrelevantes haciendo uso de procedimientos personalizados de selección de modelos multitarea. El segundo establece la utilización restringida de mediciones históricas de las variables objetivo (temperatura y precipitación) complementados con características multitarea de estaciones cercanas en una regresión lineal local ponderada.

Para finalizar el estudio de antecedentes, se realizó la búsqueda de trabajos o investigaciones previas relacionados con la combinación de datos provenientes de observaciones de estaciones climáticas en tierra y de satélites que orbitan la tierra. Algunos de esos estudios se expondrán a continuación:

El artículo titulado "Spatiotemporal geostatistical analysis of precipitation combining ground and satellite observations" por E. A. Varouchakis, D. T. Hristopulos, G. P. Karatzas, G. A. Corzo Perez y V. Diaz, publicado en la revista Hydrological Research en marzo de 2021, se centra en el análisis geoespacial y geostadístico de la precipitación al combinar observaciones de estaciones terrestres y observaciones satelitales. En este estudio, los autores exploran cómo integrar y analizar conjuntamente datos de precipitación obtenidos de diferentes fuentes para comprender mejor la variabilidad espacial y temporal de la precipitación. Utilizan métodos geoestadísticos para modelar la dependencia espacial de los datos de precipitación y realizar estimaciones en ubicaciones no muestreadas. El objetivo principal del estudio es mejorar la precisión y la resolución espacial de las estimaciones de precipitación al combinar datos de estaciones terrestres y observaciones satelitales. Esto tiene implicaciones importantes para el monitoreo y la gestión de los recursos hídricos, la planificación del uso del agua y la predicción de eventos extremos relacionados con la precipitación [17].

El artículo titulado "Combining the Perspective of Satellite- and Ground-Based Observations to Analyze Cloud Frontal Systems" por A. Hünerbein et al., publicado en el Journal of Applied Meteorology and Climatology en noviembre de 2014, se centra en el análisis de sistemas frontales de nubes utilizando observaciones tanto de satélites como de estaciones terrestres. En este estudio, los autores investigan cómo combinar y utilizar de manera conjunta las perspectivas proporcionadas por las observaciones satelitales y las mediciones en estaciones climáticas en tierra para analizar y comprender los sistemas frontales de nubes. Los sistemas frontales son áreas de transición entre diferentes masas de aire y son de gran importancia para el clima y las condiciones meteorológicas. El artículo describe los métodos utilizados para integrar y analizar datos de satélite y estaciones terrestres, así como las técnicas de procesamiento y análisis aplicadas. También presenta los resultados obtenidos en términos de la caracterización de los sistemas frontales de nubes y la identificación de patrones y características relevantes [18].

El artículo titulado "Estimating snow depth by combining satellite data and ground-based observations over Alaska: A deep learning approach" por J. Wang et al., publicado en el Journal of Hydrology en junio de 2020, se centra en la estimación de la profundidad de la nieve al combinar datos satelitales y observaciones en tierra utilizando un enfoque de aprendizaje profundo (deep learning). En este estudio, los autores proponen un método para estimar la profundidad de la nieve en Alaska al combinar datos provenientes de satélites y mediciones realizadas en estaciones terrestres. Utilizan un enfoque basado en técnicas de aprendizaje profundo, que es una rama de la inteligencia artificial que permite a los modelos analizar y extraer características complejas de los datos. El artículo describe la metodología empleada para entrenar y utilizar el modelo de aprendizaje profundo, así como los conjuntos de datos utilizados. También presenta los resultados de las estimaciones de la profundidad de la nieve obtenidas a partir de la combinación de los datos satelitales y las observaciones terrestres. El enfoque propuesto tiene como objetivo mejorar la

precisión y la resolución espacial de las estimaciones de la profundidad de la nieve, lo que es de gran importancia para la monitorización y la gestión de los recursos hídricos, el pronóstico de inundaciones y otras aplicaciones relacionadas con el clima y el agua [19].

El artículo titulado "Agreements between ground-based and satellite-based observations" por S. I. Akasofu, D. et al., publicado en *Planetary and Space Science* en diciembre de 1990, se centra en la comparación y evaluación de las observaciones realizadas tanto en tierra como desde satélites. En este estudio, los autores examinan las concordancias y discrepancias entre las observaciones realizadas desde estaciones terrestres y las mediciones obtenidas mediante observaciones satelitales. El objetivo principal del artículo es evaluar y comprender la consistencia y la concordancia entre estas dos fuentes de datos. El artículo describe los métodos utilizados para realizar las comparaciones y las evaluaciones, así como los conjuntos de datos utilizados en el estudio. También presenta los resultados obtenidos según las similitudes y diferencias entre las observaciones terrestres y las observaciones satelitales. El estudio de las concordancias entre las observaciones en tierra y desde satélites es importante para validar y mejorar la calidad de los datos obtenidos mediante ambos métodos. Esto tiene implicaciones para diversas áreas de estudio, como la meteorología, la climatología y la ciencia espacial, ya que permite una mejor comprensión de los fenómenos naturales y la mejora de los modelos y pronósticos [20].

El artículo titulado "Combining satellite precipitation and long-term ground observations for hydrological monitoring in China" por X. Zhang y Q. Tang, publicado en el *Journal of Geophysical Research: Atmospheres* en julio de 2015, se centra en la combinación de datos de precipitación provenientes de observaciones satelitales y mediciones de larga duración en estaciones terrestres para el monitoreo hidrológico en China. En este estudio, los autores investigan cómo integrar y utilizar conjuntamente los datos de precipitación obtenidos tanto desde satélites como desde estaciones terrestres para mejorar el monitoreo hidrológico en China. La precipitación es un factor clave en los estudios hidrológicos, ya que influye en la disponibilidad de agua y en los procesos de flujo y almacenamiento en los sistemas hídricos. El artículo describe los métodos utilizados para combinar los datos de precipitación de diferentes fuentes y las técnicas de análisis aplicadas. También presenta los resultados obtenidos en términos de la mejora del monitoreo hidrológico, la evaluación de la calidad de los datos y la validación de los resultados obtenidos mediante la combinación de datos satelitales y terrestres. La combinación de datos de precipitación satelital y observaciones terrestres proporciona una cobertura espacial más amplia y una mayor resolución temporal, lo que permite una mejor comprensión de los patrones de precipitación y su variabilidad en el tiempo y el espacio. Esto tiene implicaciones importantes para la gestión de recursos hídricos, la predicción de inundaciones y sequías, y la planificación de la infraestructura hidrológica en China [21].

Desde siempre, el interés por el entendimiento de la precipitación en cuanto a su comportamiento e incidencia como elemento climatológico clave en el desarrollo de una gran variedad de actividades económicas del país, ha ocupado un espacio importante en los proyectos de investigación nacientes en la academia. Es de saber que tanto las inundaciones como las sequías son un problema serio para la economía, el medio ambiente y la sociedad en general. Por tal razón, es imperativo poder predecir dicha variabilidad, anticipando la ocurrencia de un evento de esta índole, y así, mitigar el impacto adverso que puedan llegar a tener.

En este contexto, surge la idea de utilizar mediciones tanto de las estaciones terrestres como de los satélites como método de calibración y/o validación para construir un modelo funcional aditivo que pueda predecir los datos de precipitación de las estaciones terrestres de manera más precisa. Se menciona un modelo funcional debido a que los datos, aunque se evalúan de manera discreta en el tiempo, en realidad tienen una naturaleza continua. En lugar de utilizar datos discretos por estación e imágenes satelitales para cada año, se propone emplear un enfoque funcional que caracterice el comportamiento de manera anual. Además, dado que la modelación de la precipitación diaria resultaría complicada debido a su gran variabilidad, se sugiere trabajar con la precipitación diaria acumulada anual. Esto permitiría obtener curvas más suaves para cada año y, finalmente, construir diferentes modelos incorporando factores y covariables para mejorar la capacidad predictiva.

Este método que fusiona datos tanto terrestres como satelitales, y emplea técnicas de modelado funcional, presenta una estrategia completa para mejorar la predicción de la lluvia y, en consecuencia, mitigar el impacto de sucesos climáticos extremos en Colombia.

Desde un punto de vista estadístico, el análisis de un gran volumen de datos espaciotemporales, como los utilizados en el actual proyecto, presenta desafíos significativos. Anteriormente, se han propuesto diversas técnicas de modelación de la precipitación, que abarcan modelos lineales generalizados, modelos lineales mixtos, series temporales, cadenas de Markov, regresión no paramétrica, modelación de la precipitación pentadal, modelación Integrando imágenes satelitales, geoestadística, entre otros. Sin embargo, la elección de la herramienta estadística depende de los objetivos específicos perseguidos, como la descripción espaciotemporal, la estimación de la variación media, la predicción en el tiempo y la predicción espacial.

En resumen, se destaca la importancia de comprender y predecir la precipitación en el Valle del Cauca, dado su impacto en la economía regional. Se propone el uso de herramientas estadísticas, en particular el análisis de datos funcionales, junto con modelos de *machine learning*, para abordar los desafíos relacionados con la variabilidad espacial y temporal de la precipitación y su modelación en una región con una economía sensible a las condiciones climáticas.

## **3.2 PREDICCIÓN DE LA PRECIPITACIÓN**

En el estudio, para la predicción de la precipitación se aplican técnicas de aprendizaje automático y estadístico. Estos métodos / herramientas se definen a continuación.

### **3.2.1 Aprendizaje Automático**

Acorde a Géron [22], el *machine learning* (ML) o aprendizaje automático es la ciencia (y arte) de programar computadores para que ellos puedan aprender de los datos. En el presente estudio se

aborda el problema con aprendizaje supervisado, modelos de regresión y clasificación como se definen a continuación.

Los sistemas de ML pueden clasificarse según la cantidad y el tipo de supervisión que reciben durante el entrenamiento. Existen varias categorías, pero en el proyecto se utilizó aprendizaje supervisado [22].

### 3.2.1.1 *Aprendizaje Supervisado*

En el aprendizaje supervisado, el conjunto de entrenamiento con el que se alimenta al algoritmo incluye las soluciones deseadas, denominadas etiquetas [22]. Hay dos tipos principales de tarea en el aprendizaje supervisado:

#### 3.2.1.1.1 Regresión

En los problemas de regresión, la etiqueta de salida es una cantidad o valor continuo. El objetivo es predecir un valor numérico (milímetros o pulgadas de lluvia).

#### 3.2.1.1.2 Clasificación

En los problemas de clasificación la etiqueta es una categoría o clase. El objetivo es predecir correctamente a qué clase pertenece un nuevo dato u observación (llueve o no llueve).

## 3.2.2 Aprendizaje Estadístico

El aprendizaje estadístico se refiere a un conjunto de herramientas y métodos utilizados para dar sentido a conjuntos de datos complejos [23]. Estos métodos provienen de la estadística y el aprendizaje automático, y su objetivo principal es hacer inferencias o predicciones a partir de datos. El aprendizaje estadístico se centra en el desarrollo y la aplicación de modelos que pueden aprender de los datos, identificar patrones y tomar decisiones con mínima intervención humana.

## 3.3 **ÁREA DE ESTUDIO**

Las estaciones meteorológicas analizadas se encuentran localizadas en el departamento del Valle del Cauca, localizado en la parte suroccidental de Colombia entre las coordenadas geográficas 3° y 5° latitud Norte; 75°45' y 77°35' longitud Oeste, la máxima altura es 4080 m.s.n.m. y se encuentra en los farallones de Cali. Limita por el Norte con el departamento del Chocó y Risaralda, por el Este

con Tolima y Quindío, por el Sur con el Cauca, y por el Occidente con el Océano Pacífico. Tiene un área de 21140 km<sup>2</sup>, lo que representa el 1.95% del territorio nacional. En la Fig. 1 se puede apreciar un mapa con la localización del área de estudio descrita.

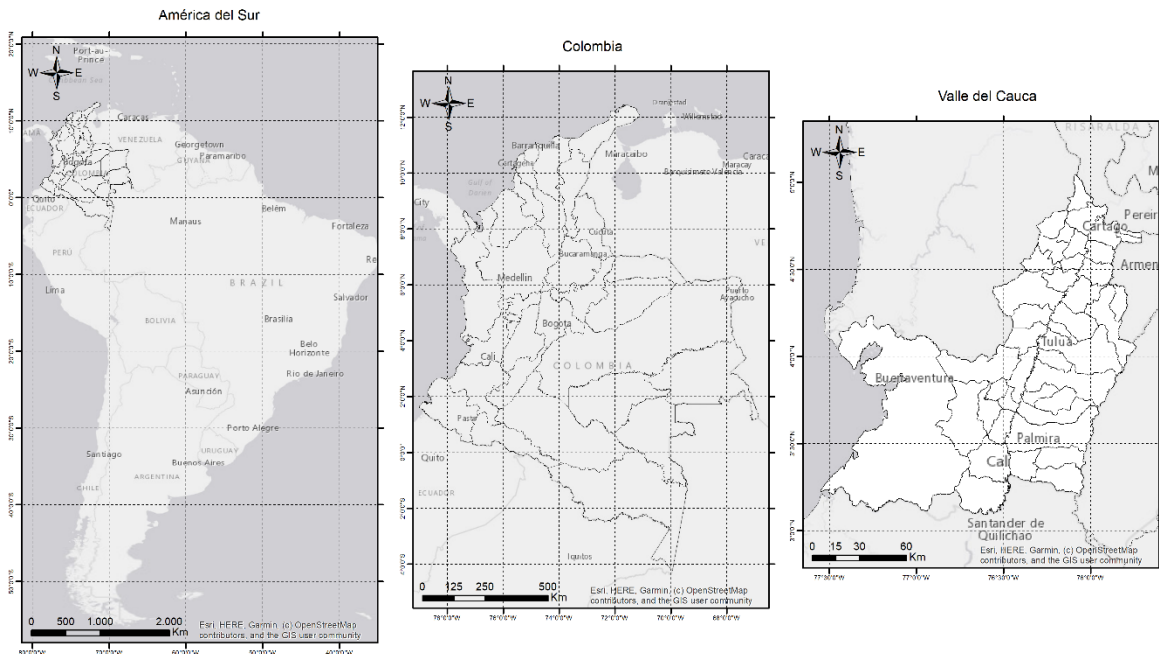


Fig. 1. Mapa de localización del área de estudio  
Nota: elaboración propia.

### 3.3.1 Climatología

Este departamento es conocido por tener un clima variado debido a su topografía diversa que incluye áreas costeras, valles y montañas. Su posición geográfica en la zona intertropical, muy cerca al Ecuador, determina un régimen de altas temperaturas durante todo el año; sin embargo, el factor altitudinal, presente en las cordilleras, ocasiona una disminución de la temperatura del casi 0.5°C por cada 100 m de ascenso.

Las masas nubosas procedentes del Océano Pacífico chocan con la vertiente occidental y cima de la cordillera occidental, al igual que con la vertiente occidental expuesta de la Cordillera Central, causando la ocurrencia de lluvias por lo que aparecen ambientes húmedos a muy húmedos. La ladera o vertiente oriental de la Cordillera Occidental, partes media y baja, toda el área plana del valle geográfico del río Cauca y la ladera baja de la cordillera central por estar en una posición oculta, con relación a los vientos cargados de humedad del Océano Pacífico, reciben menor precipitación lo que determina ambientes secos a muy secos.

Los máximos valores de precipitación se presentan en los periodos abril – junio y octubre – noviembre, aunque para las áreas próximas a Buenaventura el comportamiento de lluvias es prácticamente igual durante todo el año, por la humedad imperante en el océano.

El clima del Valle del Cauca presenta un régimen bimodal, caracterizado por dos trimestres secos y dos trimestres lluviosos. La mayor parte de la precipitación anual ocurre en los meses lluviosos, representando aproximadamente el 70% de la precipitación total registrada en la región. Dado que la economía del departamento se basa en gran medida en la agricultura, los cambios bruscos en las condiciones climáticas, especialmente en los patrones de precipitación, afectan la disponibilidad de agua y, en última instancia, repercuten en la producción agrícola.

### 3.3.2 Aspectos Fisiográficos

Los aspectos geomorfológicos, climáticos, altitudinales y latitudinales se combinan para crear cuatro zonas fisiográficas muy contrastadas: vertiente occidental de la Cordillera Central, valle geográfico del río Cauca, Cordillera Occidental y llanura marina del Pacífico. La orografía del Valle del Cauca se puede apreciar en el mapa de la Fig. 2 en la página siguiente.

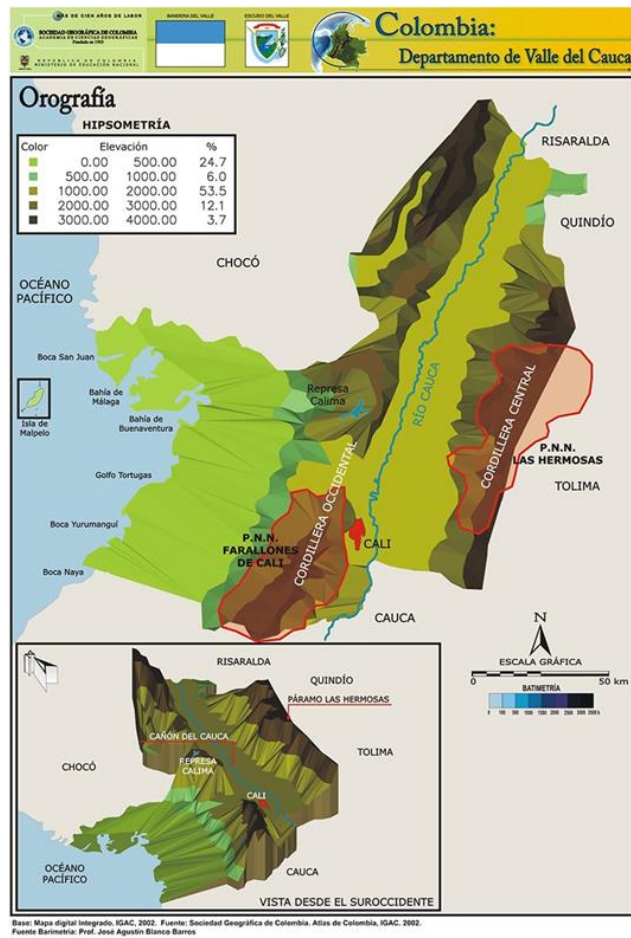


Fig. 2. Relieve / orografía Valle del Cauca.  
 Nota: tomado de [24].

## 4. METODOLOGÍA

El esquema general del proyecto se puede observar en la figura siguiente (Fig. 3).

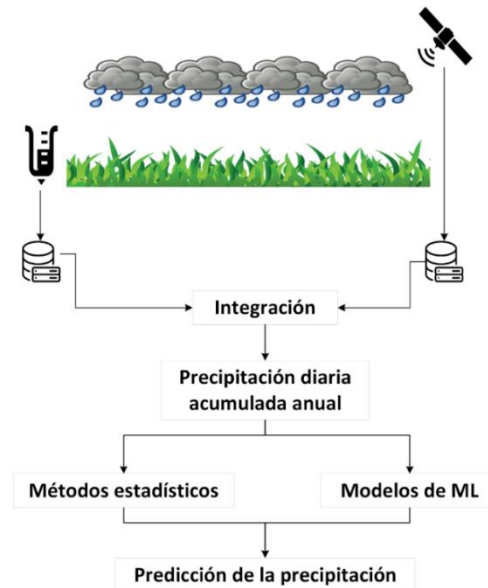


Fig. 3. Esquema comprensión del proyecto.  
Nota: elaboración propia.

Con el esquema anterior en mente, a continuación, se describe de manera detallada cada uno de los pasos de la metodología aplicada en el estudio.

### 4.1 SELECCIÓN FUENTE DE DATOS

Para el estudio se utilizaron dos fuentes de datos, la primera son datos capturados por estaciones meteorológicas en tierra y la segunda son datos proporcionados por imágenes capturadas por satélites climatológicos que orbitan alrededor de la Tierra. Ambas fuentes proporcionan información de una variedad de datos hidrometeorológicos que son esenciales para el monitoreo y el estudio de las condiciones atmosféricas. Sin embargo, para el caso de este proyecto, la variable objeto de estudio es la precipitación. Considerando que los datos de las estaciones en tierra presentan pérdida de datos por su ubicación geográfica, fallas en los instrumentos o falta de mantenimiento, los datos suministrados por los satélites permiten reconstruir las series climáticas en un periodo de tiempo. De esta manera, para este estudio la variable dependiente ( $X$ ) está representada por los datos provenientes de las imágenes satelitales y la variable independiente ( $Y$ ) por los datos de las estaciones en tierra. A continuación, se detallan las características de cada fuente de datos.

#### 4.1.1 Recopilación de datos de estaciones meteorológicas en tierra

En el presente estudio se analizaron los datos de 58 estaciones hidrometeorológicas en tierra en el departamento del Valle del Cauca del Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia (IDEAM), abarcando 32 años de registros desde 1983 hasta 2014, como base para el análisis exploratorio de los datos. Los *datasets* se descargaron como CSV comprimidos (.zip) de la base de datos DHIME en línea del IDEAM. Cada archivo cuenta tanto con los de las estaciones, así como la fecha (año-mes-día), así como los registros respectivos de precipitación en milímetros de lluvia en cada caso.

En la Fig. 4 se puede observar la distribución geográfica de la totalidad de estaciones históricas del IDEAM disponibles en el área de estudio.

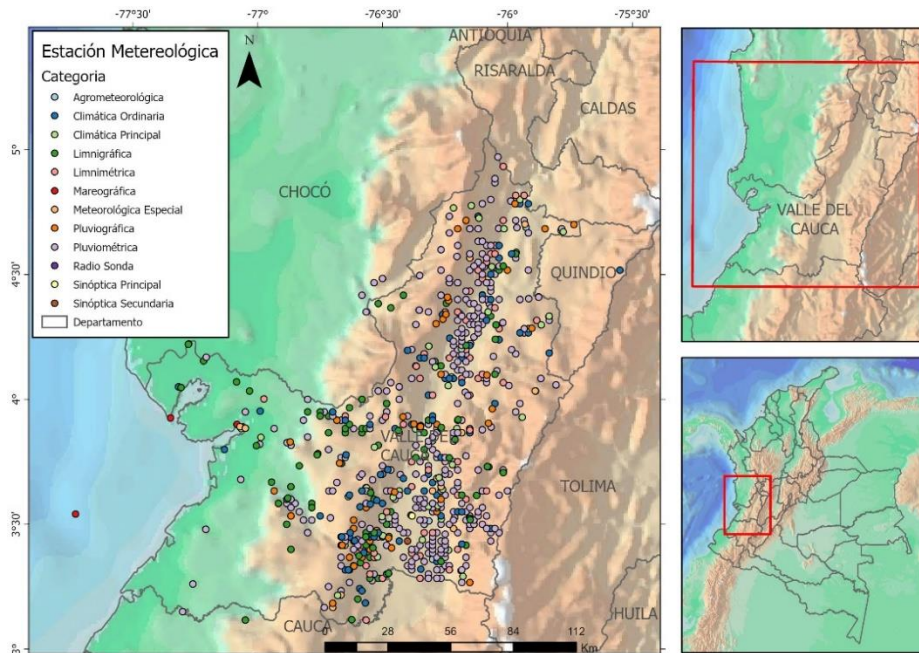


Fig. 4. Totalidad estaciones IDEAM en el departamento del Valle del Cauca.  
Nota: elaboración propia.

De las estaciones disponibles en el departamento del Valle del Cauca, se realizó la selección de 58 estaciones meteorológicas con datos representativos de precipitación que se analizaron en este estudio. La distribución y ubicación de las estaciones seleccionadas se puede observar en el mapa de la Fig. 5.

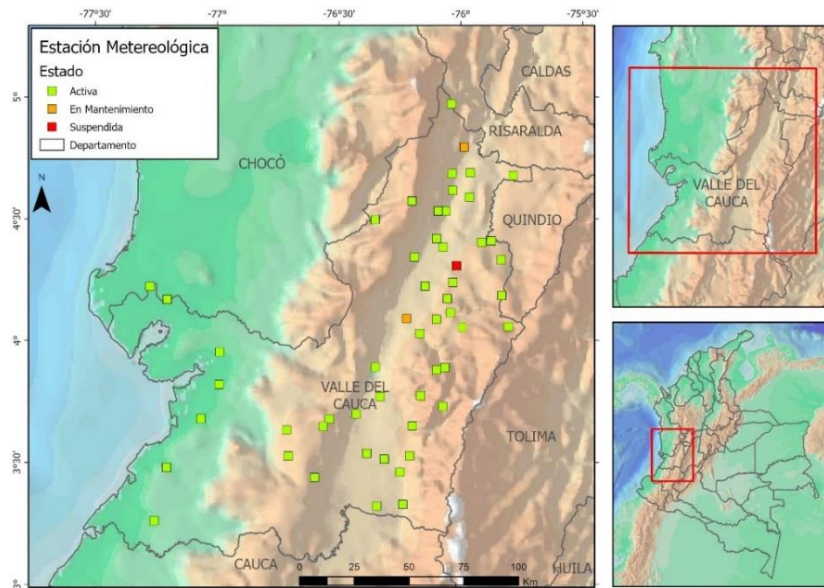


Fig. 5. Estaciones hidrometeorológicas seleccionadas para el proyecto.  
Nota: elaboración propia.

#### 4.1.2 Extracción de datos satelitales

Los datos satelitales se obtuvieron del *dataset* disponible en *Climate Hazards Group InfraRed Precipitation with Station (CHIRPS)*, quienes han creado una base de datos de precipitación que abarca desde 1981 hasta la actualidad. Estos datos se presentan en formato ráster con una resolución espacial de 5 kilómetros a lo largo del Ecuador. Es posible acceder de manera gratuita a esta información en el sitio web del grupo CHIRPS. Los datos se encuentran recopilados en un archivo zip que contiene 12 archivos GeoTiff (.tif), uno para cada mes del año (enero es el 1; diciembre es el 12) [25].

Dentro de la información disponible se cuenta con los valores de precipitación mundial mensual para cada año, lo cual se puede apreciar en la Fig. 6.

En las figuras siguientes se puede visualizar los datos climáticos satelitales obtenidos de WordClim, fuente de datos climáticos históricos mensuales de la variable precipitación, variable objeto de estudio del presente proyecto. Las gráficas fueron procesadas con *R*, haciendo uso de las librerías: *Raster*, *Ggplot2*, y *Leaflet*.

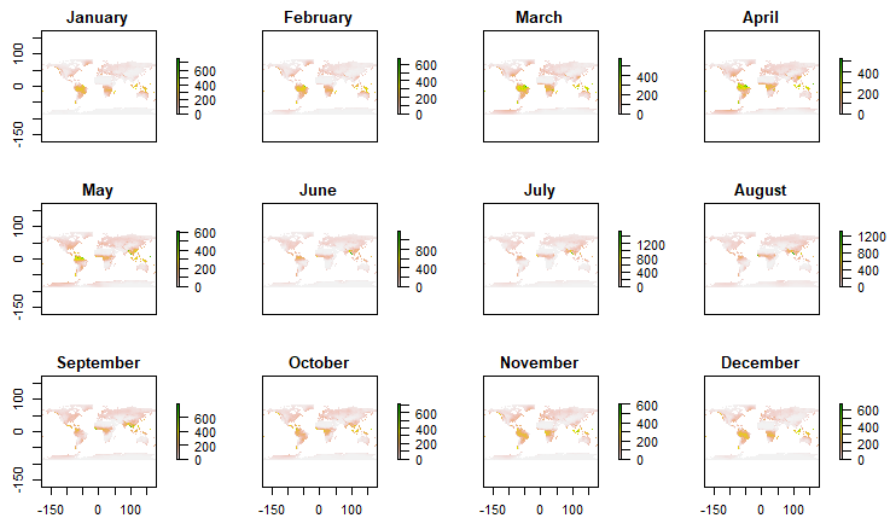


Fig. 6. Precipitaciones mundiales.  
Nota: elaboración propia.

Se obtuvo el promedio mensual de precipitación mundial con el objetivo de recopilar toda la información disponible en una sola imagen (ver Fig. 7).

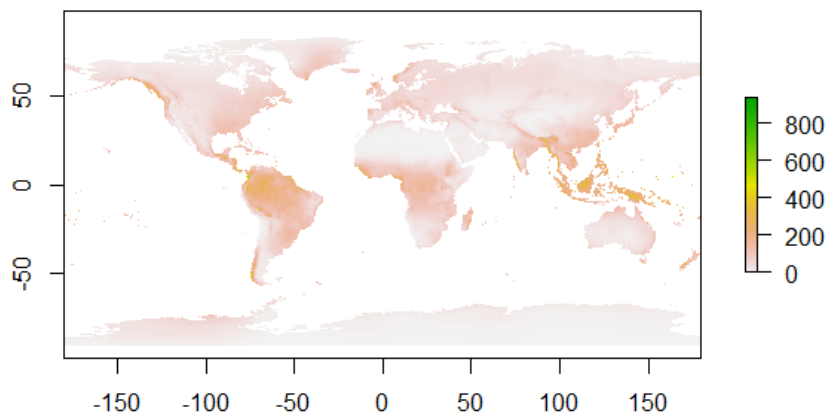


Fig. 7. Precipitaciones anuales promedio.  
Nota: elaboración propia.

Luego, se procedió a obtener los valores de precipitación promedio para el territorio colombiano (tal como se puede apreciar en la Fig. 8). Se observa la distribución anual de precipitaciones para Colombia con una escala en metros, destacando zonas con altas precipitaciones en regiones como el Pacífico y la Amazonia con valores superiores a 8500 mm, seguido por la zona central del país, donde las precipitaciones se sitúan en valores menores a 8000 mm. La zona de La Guajira presenta precipitaciones más bajas que el resto del país, con valores por debajo de 4000 mm anuales.

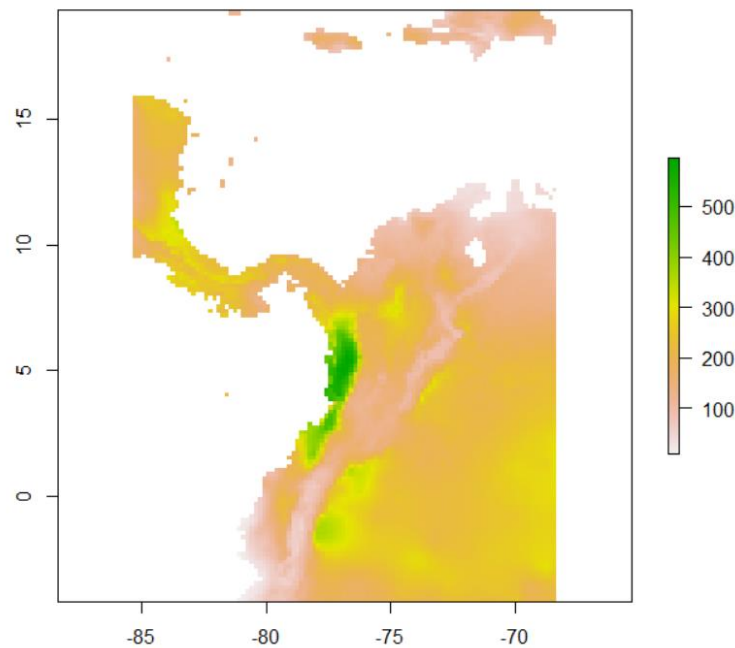


Fig. 8. Precipitación anual promedio para Colombia.  
Nota: elaboración propia.

Con esta información de precipitación para Colombia, y considerando las coordenadas geográficas de las estaciones en tierra del IDEAM, se extrajo el valor diario de la lluvia para cada ubicación, obteniendo la precipitación de los datos satelitales de cada estación en el período de tiempo (1983-2014). Los criterios para delimitar este periodo de tiempo fueron la cantidad de datos faltantes y la disponibilidad de estaciones en tierra. Los datos se consolidaron en una hoja de cálculo (.xlsx) que contiene para cada día (*day - month - year*) los registros de precipitación en milímetros de lluvia de los 58 puntos satelitales.

#### 4.1.3 Lectura de archivos

Como se comentó anteriormente, los datos de precipitación de cada una de las estaciones en tierra se descargaron por grupos en formato CSV y comprimidos ZIP para cada una de las estaciones seleccionadas del IDEAM y luego se consolidaron en un solo archivo de texto CSV con todos los registros de lluvia de las estaciones para día del periodo histórico analizado. De la misma manera, los datos satelitales extraídos se consolidaron a una hoja de cálculo XLSX para los puntos satelitales / coordenadas equivalentes a las estaciones en tierra para cada día del periodo analizado.

Con los datos en formatos mencionados, se procedió a su lectura y la creación con Python respectiva de los *dataframes* tanto a nivel estaciones y satélite. A partir de las columnas de año, mes y día se crea un índice *datetime* para todos los registros. En las figuras siguientes (Fig. 9 y Fig. 10) se puede apreciar el encabezado para cada *dataframe*.

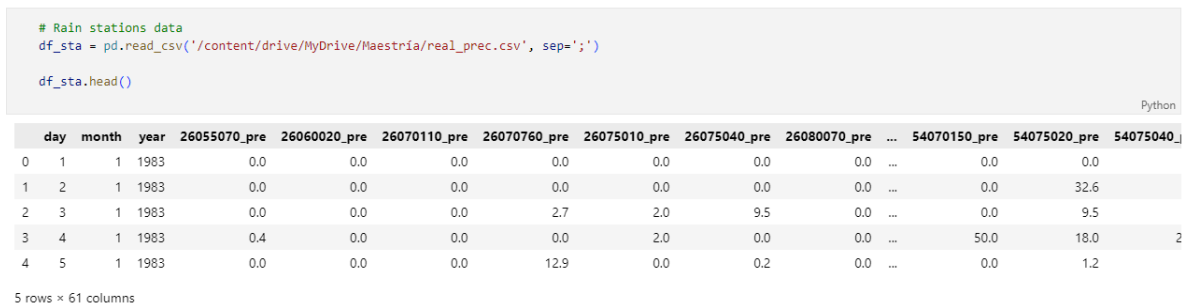


Fig. 9. Dataframe estaciones en tierra.  
Nota: elaboración propia.

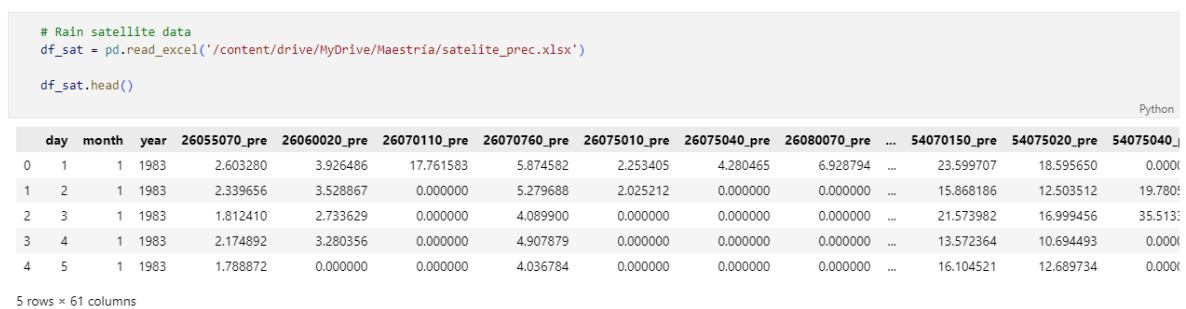


Fig. 10. Dataframe datos satelitales.  
Nota: elaboración propia.

## 4.2 ANÁLISIS EXPLORATORIO DE DATOS

Para el análisis descriptivo de los datos se utilizaron las librerías Pandas, Numpy y Statmodels. Para la visualización de datos se generan diferentes gráficas de interés (series de tiempo, histogramas, *boxplots*, etc.) con las librerías Matplotlib, Seaborn y Plotly. El resultado del análisis exploratorio / EDA se puede consultar en el capítulo 5 y en el ANEXO A. Dicho análisis incluye el estudio de las tendencias de las lluvias en las diferentes agregaciones temporales de interés, la correlación entre los puntos estudiados, así como la identificación de los datos faltantes de las estaciones en tierra del IDEAM.

## 4.3 ANÁLISIS ESPACIAL DE DATOS

Cuando fue necesario, se realizó un análisis espacial de las estaciones y su efecto en los registros obtenidos (p.ej. estaciones lluviosas en el Pacífico), adicional a esto la ubicación de los puntos de las estaciones y la generación de los mapas requeridos para el proyecto se realizaron con las librerías Folium, Leaflet y el software ArcGIS.

#### 4.4 TRANSFORMACIÓN DE DATOS

Los *dataframes* de datos satelitales y de estaciones en tierra se unieron (*merge*) en uno solo conjunto de datos usando como criterio el índice de fechas (*datetime*). Las columnas de los datos satelitales se renombran como el sufijo “\_x” y las de las estaciones en tierra con el sufijo “\_y” (ver encabezado en la Fig. 11). Para crear las agregaciones temporales estudiadas se pasa de *datetimes* diarios a semanales, mensuales y anuales aplicando con Pandas los *resamples* correspondientes a cada periodo de interés. Para convertir los registros de precipitación en categorías o intensidades de interés se aplican funciones teniendo en cuenta los rangos de lluvia.

```
# Dataframe merge
df = pd.merge(left=df_sat, right=df_sta, left_index=True, right_index=True)
df.head()
```

	day_x	month_x	year_x	26055070_x	26060020_x	26070110_x	26070760_x	26075010_x	26075040_x	26080070_x	...	54070150_y	54075020_y	54075040_y	26070
1983-01-01	1	1	1983	2.603280	3.926486	17.761583	5.874582	2.253405	4.280465	6.928794	...	0.0	0.0	1.0	
1983-01-02	2	1	1983	2.339656	3.528867	0.000000	5.279688	2.025212	0.000000	0.000000	...	0.0	32.6	1.5	
1983-01-03	3	1	1983	1.812410	2.733629	0.000000	4.089900	0.000000	0.000000	0.000000	...	0.0	9.5	2.0	
1983-01-04	4	1	1983	2.174892	3.280356	0.000000	4.907879	0.000000	0.000000	0.000000	...	50.0	18.0	24.8	
1983-01-05	5	1	1983	1.788872	0.000000	0.000000	4.036784	0.000000	0.000000	0.000000	...	0.0	1.2	0.0	

5 rows × 122 columns

Fig. 11. Dataframe unificado.  
Nota: elaboración propia.

#### 4.5 APLICACIÓN TÉCNICAS DE IMPUTING

Previo a la aplicación de modelos como tal de regresión y clasificación, se aplicaron técnicas de *imputing* univariado, multivariado y de vecinos más cercanos (KNN) con Scikit-Learn para completar los datos faltantes. Para evaluar los resultados con dichos estimadores se calcularon para cada estación y agregación temporal las métricas seleccionadas en el proyecto (ver numeral 4.11).

#### 4.6 SELECCIÓN DE MODELOS

En el presente estudio se implementaron técnicas de aprendizaje automático (*machine learning* – ML), estadístico y profundo (*Deep learning*) a las series climáticas de precipitación seleccionadas, a continuación, en la Fig. 12 se presenta un diagrama de los modelos implementados en el proyecto.

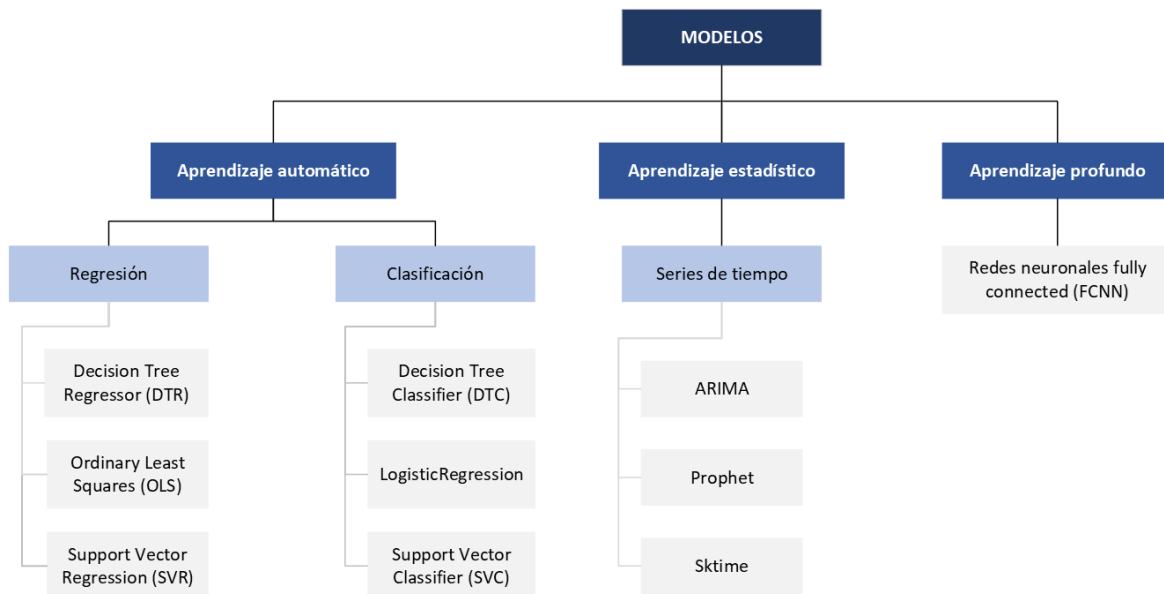


Fig. 12. Modelos implementados en el proyecto  
Nota: elaboración propia.

Los modelos seleccionados e implementados en la figura anterior se describen con mayor detalle en los numerales siguientes.

#### 4.6.1 Modelos implementados

En el proyecto se utilizaron las librerías de Python: scikit-learn [26], sktime [27], statmodels [28] y Prophet [29]. Los principales estimadores / modelos implementados en el proyecto de dichas librerías se describen a continuación con base a la documentación de cada una.

##### 4.6.1.1 Regresión

###### 4.6.1.1.1 LinearRegression

El modelo de Regresión Lineal de *Scikit-Learn* es probablemente el más básico y conocido. Asume una relación lineal entre las variables independientes y la variable dependiente. Calcula los coeficientes para las variables independientes que minimizan el error cuadrático total entre las predicciones y los valores reales.

Este modelo se puede representar matemáticamente de la siguiente manera:

Dado un conjunto de datos con  $n$  muestras y  $m$  características, y denotando  $X$  como la matriz de características (con  $n$  filas y  $m$  columnas) e  $y$  como el vector de etiquetas/*targets*, el objetivo del

modelo de regresión lineal es encontrar los coeficientes  $\Theta$  que minimizan la suma de los cuadrados de las diferencias entre las predicciones y los valores reales.

La fórmula para el modelo de regresión lineal simple (una sola característica) es:

$$y = \theta_0 + \theta_1 \cdot x$$

Donde:

- $y$  es la variable dependiente (*target*/etiqueta).
- $x$  es la variable independiente (característica).
- $\theta_0$  es la intersección en el eje  $y$  (ordenada al origen).
- $\theta_1$  es la pendiente de la línea (la tasa de cambio de  $y$  con respecto a  $x$ ).

Para encontrar estos coeficientes, *scikit-learn* utiliza el método de mínimos cuadrados ordinarios (OLS) o algún método de optimización más avanzado dependiendo de los parámetros específicos configurados.

#### 4.6.1.1.2 SVR

La Regresión de Vectores de Soporte, parte de la librería *scikit-learn*, es una versión de las Máquinas de Soporte Vectorial (SVM) adaptada para regresión. En lugar de tratar de ajustar un hiperplano en un espacio de características para separar dos clases, el SVR intenta ajustar un hiperplano de tal manera que se maximice el margen y se minimice el error de predicción dentro de un límite determinado.

La representación matemática del SVR es la siguiente:

Dado un conjunto de datos de entrenamiento  $(X, y)$ , donde  $X$  es una matriz de características e  $y$  es el vector de etiquetas, el objetivo es encontrar la función  $f(X)$  que se ajuste lo mejor posible a los datos.

La formulación básica del problema de SVR es la siguiente:

$$\min_{w, b, \xi, \xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*)$$

sujeto a las restricciones:

$$\begin{cases} y_i - \langle w, x_i \rangle - b \leq \epsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \epsilon + \xi_i^* \end{cases}$$

Y

$$\xi_i, \xi_i^* \geq 0$$

Donde:

- $w$  es el vector de pesos que define el hiperplano.

- $b$  es el sesgo.
- $\xi$  y  $\xi^*$  son variables de holgura que representan errores de margen.
- $\varepsilon$  es el parámetro de holgura.
- $C$  es un parámetro de regularización que controla el *trade-off* entre la complejidad del modelo y la penalización por errores.
- $\phi(x)$  es una función de mapeo no lineal que proyecta los datos de entrada a un espacio de características de mayor dimensión.

El objetivo es minimizar la suma de los errores de margen y la complejidad del modelo, sujeto a las restricciones impuestas por las etiquetas de los datos de entrenamiento.

Esta formulación define un problema de optimización cuadrático convexo que se puede resolver utilizando técnicas de optimización convexa. En la práctica, se utilizan algoritmos eficientes de optimización numérica para encontrar los valores óptimos de  $w$  y  $b$ .

#### 4.6.1.1.3 DecisionTreeRegressor

Es un modelo de regresión basado en árboles de decisión. Funciona dividiendo el espacio de los datos en regiones más pequeñas y manejables. Para cada región, el modelo calcula el valor medio de la variable objetivo, que se utiliza para hacer predicciones. Los árboles de decisión son útiles por su simplicidad y capacidad para modelar relaciones no lineales.

Matemáticamente se podría describir este modelo de la siguiente manera: Se tiene un conjunto de datos de entrenamiento que consiste en  $n$  observaciones  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , donde  $x_i$  representa las características e  $y_i$  representa la variable objetivo (la etiqueta correspondiente o el valor que queremos predecir).

El objetivo del algoritmo *DecisionTreeRegressor* es encontrar la mejor forma de dividir los datos en cada nodo del árbol para minimizar la varianza de las etiquetas de destino dentro de cada subconjunto.

Supongamos que, en un nodo dado  $m$ , tenemos un conjunto de datos  $D_m$  que consiste en  $n_m$  observaciones. Se desea dividir este conjunto de datos en dos subconjuntos  $D_{m,izquierda}$  y  $D_{m,derecha}$  mediante una regla de división  $\theta = (j, t)$ , donde  $j$  es el índice de la característica y  $t$  es el umbral.

Para cada nodo  $m$ , queremos minimizar la función de costo:

$$\text{Costo}(m) = \frac{n_m}{n} * \text{Var}(D_m)$$

Donde  $\text{Var}(D_m)$  es la varianza de las etiquetas de destino en el conjunto de datos  $D_m$ .

El algoritmo busca recursivamente las divisiones que minimizan este costo hasta que se alcanza un cierto criterio de detención, como la profundidad máxima del árbol o el número mínimo de muestras requeridas para dividir un nodo.

#### 4.6.1.2 Clasificación

##### 4.6.1.2.1 LogisticRegression

Es un modelo de clasificación de *scikit-learn* que estima la probabilidad de que una instancia pertenezca a una clase particular. A pesar de su nombre, es un modelo de clasificación y no de regresión. Es particularmente útil para problemas de clasificación binaria. El modelo utiliza la función logística para modelar una función de decisión binaria y puede extenderse a la clasificación multiclase (p.ej., usando la técnica "one-vs-rest").

La función logística está dada por:

$$P(y = 1 | x) = \frac{1}{1 + e^{-(wx+b)}}$$

Donde:

- $P(y = 1|x)$  es la probabilidad de que la clase sea 1 dada la entrada  $x$ .
- $w$  es el vector de pesos (coeficientes).
- $b$  es el término de sesgo.
- $e$  es la base del logaritmo natural ( $\sim 2.71828$ ).

La predicción final se realiza aplicando un umbral a la probabilidad estimada. Por ejemplo, si la probabilidad estimada es mayor que 0.5, se clasifica como la clase positiva; de lo contrario, se clasifica como la clase negativa.

El modelo de regresión logística se entrena ajustando los parámetros  $w$  y  $b$  para maximizar la probabilidad conjunta de las observaciones en el conjunto de entrenamiento. Esto se hace típicamente maximizando la función de verosimilitud logarítmica o minimizando la función de pérdida logística, que está dada por:

$$L(w, b) = -\frac{1}{n} \sum_{i=1}^n [y^i \log(P(y = 1|x_i)) + (1 - y_i) \log(1 - P(y = 1|x_i))]$$

Donde:

- $n$  es el número de observaciones.
- $y_i$  es la etiqueta de clase real de la observación  $x_i$ .

El modelo se puede regularizar agregando términos de penalización a los coeficientes  $w$  para evitar el sobreajuste.

La regresión logística es un modelo lineal generalizado y es ampliamente utilizado debido a su simplicidad, eficiencia computacional y capacidad para proporcionar probabilidades bien calibradas como salida.

#### 4.6.1.2.2 SVC

El SVC de *scikit-learn* es un clasificador que utiliza la técnica de Máquinas de Soporte Vectorial para realizar la clasificación de datos. Es efectivo en espacios de alta dimensión y en situaciones donde el número de dimensiones es mayor que el número de muestras. Este modelo es útil en problemas de clasificación binaria y multiclase y funciona bien con márgenes de separación claros. La matemática detrás de SVR es la siguiente:

Dado un conjunto de datos de entrenamiento  $(X, y)$ , donde  $X$  es una matriz de características e  $y$  es el vector de etiquetas, el objetivo es encontrar la función  $f(X)$  que se ajuste lo mejor posible a los datos.

El objetivo del SVR es encontrar la función que se ajuste dentro de un margen de tolerancia especificado, que se define mediante dos parámetros:  $\varepsilon$  y  $C$ .

La formulación matemática del problema de optimización para SVR es la siguiente:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*)$$

Sujeto a las restricciones:

$$y_i(w, x_i + b) \geq 1 + \xi_i$$

$$\xi_i \geq 0$$

donde  $C$  es un parámetro de regularización que controla el equilibrio entre maximizar el margen y minimizar el error de clasificación. El parámetro  $C$  controla la penalización por errores de clasificación: un  $C$  mayor penaliza más los errores, lo que lleva a un hiperplano de separación más ajustado a los datos de entrenamiento. Por otro lado, un  $C$  menor permite un margen más amplio, lo que puede llevar a un mayor error de clasificación en el conjunto de entrenamiento, pero posiblemente mejor generalización en datos no vistos.

#### 4.6.1.2.3 DecisionTreeClassifier

Es un modelo de clasificación de *scikit-learn* basado en árboles de decisión. Al igual que su contraparte de regresión, divide repetidamente el espacio de los datos, creando un árbol donde cada nodo representa una decisión basada en una característica.

Matemáticamente, un árbol de decisión clasificador puede representarse como una función  $f(x)$  que asigna una etiqueta de clase a un vector de características  $x$ . Esta función se define por las reglas de decisión almacenadas en el árbol.

En un árbol de decisión,  $f(x)$  se puede expresar como una secuencia de reglas *if - else* anidadas. Por ejemplo, podríamos tener algo como:

*Si  $x_1 \leq \text{corte}_1$ , entonces*

*Si  $x_2 \leq \text{corte}_2$ , entonces*

$Clase = C_1$

*Sino*

$Clase = C_2$

*Sino*

*Si  $x_3 \leq corte_3$ , entonces*

$Clase = C_3$

*Sino*

$Clase = C_4$

Donde  $x_1, x_2, x_3$  son las características,  $corte_1, corte_2, corte_3$  son los valores de corte seleccionados durante el entrenamiento y  $C_1, C_2, C_3, C_4$  son las etiquetas de clase asignadas a cada región.

El proceso de entrenamiento de un árbol de decisión implica encontrar la mejor manera de dividir el espacio de características para maximizar la pureza de las hojas. Esto se hace de manera recursiva hasta que se alcanza algún criterio de parada, como la profundidad máxima del árbol o el número mínimo de muestras requeridas para realizar una división.

En resumen, aunque la formulación matemática exacta de un árbol de decisión no se expresa típicamente como una ecuación única, se representa mediante un conjunto de reglas de decisión que definen cómo se clasifican las observaciones basadas en sus características.

#### 4.6.1.3 Series de tiempo

##### 4.6.1.3.1 ARIMA

Con relación al proyecto abordado, se contempla el uso del modelo estadístico ARIMA, ya que ha demostrado ser eficaz para capturar tendencias y estacionalidades en los datos climáticos. Puede modelar patrones cíclicos y comportamientos a largo plazo presentes en las series temporales, permitiendo así la generación de proyecciones más precisas.

Por otra parte, la versatilidad del modelo ARIMA facilita la incorporación de múltiples fuentes de datos, como en este caso, las mediciones de estaciones en tierra y la información recopilada a partir de imágenes satelitales. Esto proporciona una base sólida para la reconstrucción de series climáticas con una amplia variedad de datos.

Es de saberse que las series climáticas pueden sufrir de irregularidades en las mediciones debido a diversos factores, como problemas con los dispositivos de medición o falta de datos en ciertos períodos. El modelo ARIMA por lo demás, se presenta inicialmente como una buena opción capaz de manejar datos irregulares pudiendo imputar valores faltantes, garantizando la integridad y consistencia de las series reconstruidas.

ARIMA, es propiamente un modelo de regresión el cual está compuesto por los siguientes tres términos:

- AR (auto-regresión): modelo que muestra una variable cambiante que retrocede sobre sus propios valores atrasados/anteriores.
- I (integrado): diferenciación de observaciones sin procesar para permitir que la serie temporal se vuelva estacionario.
- MA (promedio móvil): dependencia entre una observación y un error residual de un modelo de promedio móvil.

Ahora bien, los parámetros  $p$ ,  $d$  y  $q$  establecen el modelo ARIMA a implementar, parámetros descritos a continuación:

- $p$ : número de observaciones de retraso en el modelo; también denominado como el orden de retardo.
- $d$ : número de veces que se diferencian las observaciones sin procesar; también denominado como el grado de diferenciación.
- $q$ : tamaño de la ventana del promedio móvil; también denominado como el orden de la media móvil.

El modelo SARIMA (*Seasonal Autoregressive Integrated Moving Average*) es una extensión del modelo ARIMA que incorpora componentes estacionales para capturar patrones repetitivos en la serie temporal. Es útil cuando los datos exhiben variaciones estacionales que no pueden ser capturadas adecuadamente por un modelo ARIMA estándar. Al incluir términos estacionales en el modelo SARIMA, se puede mejorar la capacidad predictiva y la capacidad de capturar patrones estacionales en los datos.

#### 4.6.1.3.2 Prophet

Prophet es un algoritmo de Meta / Facebook para la predicción / pronóstico de datos de series de tiempo basado en un modelo aditivo en el que las tendencias no lineales se ajustan con estacionalidad anual, semanal, diaria y por festivos. Funciona mejor con series de tiempo que tienen fuertes efectos estacionales y varias estaciones de datos históricos. Prophet es robusto a los datos faltantes, a los cambios en la tendencia, y maneja bien los valores atípicos.

#### 4.6.1.3.3 NaiveForecaster

El *NaiveForecaster* es un modelo de predicción de series temporales dentro de *Sktime* que utiliza estrategias simples pero efectivas para predecir futuros valores basándose en los datos históricos disponibles. Es parte de los *forecasters* estadísticos y se usa comúnmente como un modelo de referencia debido a su simplicidad y eficacia en ciertos tipos de datos.

Para calcular los cuantiles de predicción, *sktime* estima el error estándar de los residuos de predicción bajo la suposición de residuos no correlacionados. La varianza del pronóstico se calcula entonces multiplicando la varianza residual por una constante. Esta constante es un ajuste de sesgo

de muestra pequeña y cada método (media, última, deriva) tiene diferentes fórmulas para calcular la constante (las fórmulas se pueden consultar en [30]). Por último, bajo la suposición de que los residuos siguen una distribución normal, *NaiveForecaster* usa la varianza del pronóstico y las puntuaciones  $z$  de una distribución normal para estimar los cuantiles de predicción.

#### 4.6.1.4 Redes neuronales

##### 4.6.1.4.1 SequentialRegressor

Esta es una adaptación de modelos de redes neuronales de *Keras* [31] para tareas de regresión. Permite construir modelos de aprendizaje profundo personalizados y flexibles, adecuados para datos complejos y no lineales.

Este modelo secuencial sigue los siguientes pasos en el contexto de la regresión de series temporales:

- **Arquitectura del modelo:** Los modelos secuenciales pueden tomar varias formas, como redes neuronales recurrentes (RNN), redes neuronales convolucionales (CNN), redes neuronales *transformer*, entre otros. Cada uno de estos modelos tiene su propia arquitectura y forma de procesar las secuencias de entrada para realizar la regresión.
- **Entrada y salida:** En el contexto de la regresión de series temporales, la entrada al modelo suele ser una secuencia de puntos de datos temporales, y la salida es una predicción para un punto de datos futuro (o varios puntos de datos futuros, dependiendo del diseño del modelo).
- **Función de pérdida:** Durante el entrenamiento del modelo, se utiliza una función de pérdida para comparar las predicciones del modelo con los valores reales. La función de pérdida típicamente utilizada para la regresión incluye el error cuadrático medio (MSE, *Mean Squared Error*) o alguna variante que tenga en cuenta la naturaleza secuencial de los datos.
- **Optimización:** Se utiliza un algoritmo de optimización para ajustar los parámetros del modelo de manera que la función de pérdida se minimice. Esto implica el cálculo de gradientes y la actualización iterativa de los parámetros del modelo.
- **Entrenamiento y predicción:** Una vez entrenado el modelo, se puede utilizar para hacer predicciones sobre nuevas secuencias de datos temporales.

## 4.7 APLICACIÓN MODELOS DE REGRESIÓN

Para cada estimador / modelo de regresión descrito anteriormente se procedió a crear un bucle con un *pipeline* de trabajo para cada uno de los modelos / estimadores para cada estación / punto estudiado en el proyecto (58 en total). El pipeline incluye el respectivo train / split de los datos, asignando 70% a los datos de entrenamiento (*train*) y 30% o 0.3 a los datos de prueba (*test*). El

resultado final del bucle es un *dataframe* con las métricas de evaluación cada uno de los puntos / modelos (ver ANEXO C).

#### 4.7.1 GridSearch

Para el ajuste de los hiperparámetros de los estimadores utilizados, se obtuvieron en primer lugar la lista de parámetros de cada modelo con el método *get\_params* (ver ejemplo en la Fig. 13). Con los parámetros identificados para cada estimador, fue posible aplicar dentro de un bucle / pipeline un diccionario / grilla de valores / opciones a evaluar con la función *GridSearchCV* de Scikit-Learn que arroja los mejores parámetros y las métricas correspondiente para cada uno de ellos en cada punto / estación analizada en el estudio ().

```
# SVR Parameters
SVR().get_params()

{'C': 1.0,
 'cache_size': 200,
 'coef0': 0.0,
 'degree': 3,
 'epsilon': 0.1,
 'gamma': 'scale',
 'kernel': 'rbf',
 'max_iter': -1,
 'shrinking': True,
 'tol': 0.001,
 'verbose': False}
```

Fig. 13. Ejemplo obtención de parámetros de un estimador / modelo.

Nota: elaboración propia.

```
# Parameters grid
param_grid = {
    'svr__C': [0.1, 1, 10, 100],
    'svr__kernel': ['linear', 'rbf', 'poly'],
    'svr__gamma': ['scale', 'auto']
}

# GridSearchCV to pipeline with grid parameters
reg_search = GridSearchCV(reg, param_grid, cv=5,
                          scoring=['r2', 'neg_mean_absolute_error', 'neg_mean_squared_error', 'max_error'],
                          refit='r2', n_jobs=-1)

# GridSearchCV fit
reg_search.fit(X_train, y_train)

# GridSearchCV best parameters
print('Best parameters: ', reg_search.best_params_)

# Best reg GridSearchCV model
reg_best = reg_search.best_estimator_
```

Fig. 14. Ejemplo aplicación GridSearchCV para ajuste de hiperparámetros de un estimador / modelo.

Nota: elaboración propia.

## 4.8 APLICACIÓN DE MODELOS SERIES DE TIEMPO

En el análisis exploratorio de datos, se evidenció que varias estaciones en tierra cuentan con periodos continuos (semanas, meses e incluso años) de datos faltantes (ver numeral 5.1.3.1). Dicho suceso, se debe a que al presentar fallas la estación, el hecho de reparar o reemplazar los sensores o todo el sistema, requiere un proceso contractual público que como es de esperar no es inmediato por las etapas que deben surtirse. Por lo anterior, en el proyecto se identificó que aplicar modelos de series de tiempo para realizar pronósticos futuros con los datos históricos previos a la pérdida de datos, es una opción viable para reconstruir las series climáticas como la precipitación.

Como se describió en el numeral 4.6.1.3, en el proyecto se aplicaron las librerías Pmdarima / ARIMA, Prophet y Sktime para realizar pronósticos / *forecasting* de las series de tiempo de lluvia y sus diferentes agregaciones temporales. Para cada modelo fue necesario realizar la transformación de los datetime a periodos y hacer un *temporal train / train split* del 70% y 30% basado en la serie de tiempo (lo que implica que no se usó un *suffle* aleatorio) y luego con un bucle / pipeline de trabajo se encuentra para cada estimador / agregación las métricas de cada punto / estación evaluada.

## 4.9 APLICACIÓN DE MODELOS DE CLASIFICACIÓN

En vista de los resultados iniciales de las métricas de regresión para las precipitaciones diarias, en el proyecto se planteó que adicional a predecir los valores de precipitación de lluvia, sería de utilidad saber con base a los datos históricos si va a llover o no el futuro y era necesario validar las métricas en dicho ejercicio de aprendizaje supervisado. Para lo anterior, fue necesario clasificar de manera binaria los registros diarios de precipitación, en días sin lluvia ("0" = 0 mm) o días lluviosos o con registros de precipitación ("1" = > 0 mm).

Para los datos categorizados y como se describió en el numeral 4.6.1.2 se aplicaron los algoritmos de regresión logística (*LogisticRegression*), máquinas de soporte vectorial (SVC) y árboles de decisión (*DecisionTreeClassifier*) de Scikit-Learn.

En cada caso y tal como con los estimadores de regresión, se crearon bucles con pipelines de trabajo para obtener las métricas para cada modelo, agregación y punto estudiado en el proyecto.

### 4.9.1 Clasificación por intensidad de lluvia

Más allá de si en un día o un mes llueve o no, la precipitación se puede clasificar por su intensidad en diferentes categorías. A continuación se presentan en las figuras siguientes (Fig. 15, Fig. 16 y Fig. 17), diferentes clasificaciones de intensidad para la lluvia diaria y mensual acorde a Wang [32], Latupapua et al. [33] y el Centro Nacional de Hidrología y Meteorología (NCHM) de Bután [34].

Rainfall Intensity	Rain Accumulation in a Day (mm)
Light rain	$R < 5$
Moderate rain	$5 \leq R < 20$
Heavy rain	$20 \leq R < 50$
Violent rain	$R \geq 50$

Fig. 15. Intensidad de lluvia diaria [32]

Nota: tomado de [32].

No.	Rainfall intensity					Class
	Per hour (Mm)	Per day (Mm)	Per month			
			Rainy day (day)	Estimated Total Rainfall (Mm)	Cumulative Rainfall Month (Mm)	
1.	<1	<5	5-6	10-15	10-15	very mild
2.	1-5	5-20	6-7	60-70	70-85	mild
3.	5-10	21-50	6-7	180-210	250-295	moderate
4.	10-20	51-100	2-4	150-250	400-545	heavy
5.	> 20	> 100	1-2	110-300	510-845	very heavy

Fig. 16. Intensidad de lluvia por horaria, diaria y mensual [33]

Nota: tomado de [33].

#### Rainfall Classification: Intensity of Rainfall in 24 Hours

Term	Definitions
No rain	Rainfall amount realised in a day is 0.0 mm
Very Light Rain	Rainfall amount realised in a day is between 0.1 to 0.9 mm
Light Rain	Rainfall amount realised in a day is between 1.0 mm to 10 mm
Moderate Rain	Rainfall amount realised in a day is between 11 to 30 mm
Heavy Rain	Rainfall amount realised in a day is between 31.0 to 70.0 mm
Very Heavy Rain	Rainfall amount realised in a day is between 71.0 to 150 mm
Extremely Heavy Rain	Rainfall amount realised in a day is equal or more than 151 mm
Rainy Day	Rainfall amount realised in a day is 1.0 mm or more.

Fig. 17. Clasificación intensidad de lluvia diaria NCHM [34]

Nota: tomado de [34]

Con las intensidades anteriores, en el estudio se crearon las categorías respectivas para los registros de lluvia acumulada diaria y mensual, y luego se aplicaron las clases acorde a los milímetros de precipitación del periodo, para evaluar el impacto de dichas categorías en la evaluación de los modelos de clasificación seleccionados.

#### 4.10 APLICACIÓN DE REDES NEURONALES

En presente estudio, no podía dejarse de lado la aplicación de redes neuronales en busca de un panorama prometedor o, al menos, algo más optimista que los resultados obtenidos hasta el momento. Es importante reconocer que las redes neuronales tienen su mayor aplicación en el

campo de la visión por computador y el reconocimiento de imágenes. Sin embargo, no se quería descartar su implementación sin haber obtenido resultados sistemáticos que soportaran esta decisión, aplicados al proyecto actual. De esta manera, se llegó al punto de adaptar el ejercicio de predicción de la precipitación de la lluvia a una red neuronal completamente conectada, cuya arquitectura puede verse en el siguiente diagrama de bloques de la Fig. 18.

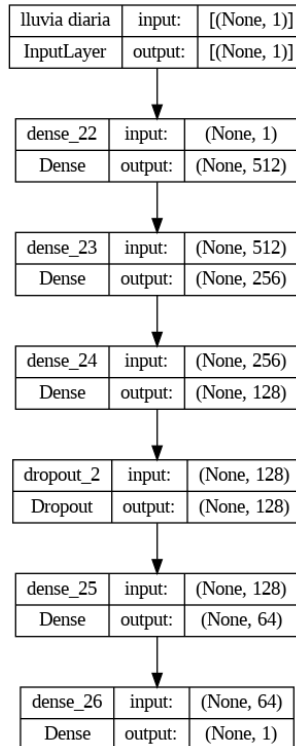


Fig. 18. Arquitectura de la red neuronal completamente conectada  
Nota: elaboración propia.

Como capa de entrada, se tiene un vector unidimensional con los valores de lluvia captados por las imágenes satelitales. El parámetro establecido como "None" indica que no se ha especificado el tamaño del lote (*batch size*).

La primera capa densa (oculta) establece una dimensión de entrada y 512 dimensiones de salida, interpretadas como 512 neuronas. Este comportamiento se replica en las siguientes cuatro capas ocultas, donde estas se conectan de manera secuencial, teniendo una disminución paulatina de la cantidad de unidades en cada capa en términos de potencias de 2, hasta llegar a 64 unidades previo a la capa de salida. La capa de salida establece una única unidad de salida, que será la predicción de la lluvia para las estaciones en tierra.

Aunque no se especifica en el diagrama, todas las capas ocultas tienen una función de activación ReLU (*Rectified Linear Unit*), mientras que en la capa de salida se implementó una función de activación lineal (*linear*), adecuada en problemas de regresión donde se predice un valor numérico. Como función de pérdida, se utilizó el error cuadrático medio / MSE para calcular la diferencia entre

las predicciones y los valores reales. Además, como métrica de evaluación, y siguiendo la misma línea de los modelos anteriores, se optó por el coeficiente de determinación  $R^2$ .

Para evitar un posible sobreajuste (*overfitting*) del modelo debido a la alta presencia de ceros en el vector de entrada del modelo, se decidió incluir una capa de *dropout* como técnica de regularización, con una fracción de desactivación de neuronas del 50% en cada paso del entrenamiento.

En la Fig. 91, se presenta el resumen de la red *Fully Connected Neural Network* (FCNN) implementada y se aprecia capa por capa la cantidad de parámetros que el modelo debe ser capaz de estimar.

Model: "Prediccion\_Lluvia\_Diaria"

Layer (type)	Output Shape	Param #
lluvia_diaria (InputLayer)	[(None, 1)]	0
dense_32 (Dense)	(None, 512)	1024
dense_33 (Dense)	(None, 256)	131328
dense_34 (Dense)	(None, 128)	32896
dropout_4 (Dropout)	(None, 128)	0
dense_35 (Dense)	(None, 64)	8256
dense_36 (Dense)	(None, 1)	65

=====  
 Total params: 173569 (678.00 KB)  
 Trainable params: 173569 (678.00 KB)  
 Non-trainable params: 0 (0.00 Byte)

Fig. 19. Resumen modelo FCNN  
 Nota: elaboración propia.

El algoritmo de optimización utilizado fue Adam, con una tasa de aprendizaje adaptativa de 0.01. Nuevamente, el establecimiento tanto de la arquitectura de la red como de los hiperparámetros del modelo se diseñó con el propósito de llevar esta herramienta a un ejercicio de predicción de un valor numérico continuo.

#### 4.11 EVALUACIÓN DE RESULTADOS

A continuación, se presentan las métricas utilizadas para evaluar los modelos seleccionados en el proyecto, las definiciones y fórmulas corresponden a las encontradas en *scikit-learn* [26].

### 4.11.1 Métricas Regresión

A continuación, se definen las métricas utilizadas para los modelos de regresión.

#### 4.11.1.1 $R^2$ (Coeficiente de Determinación)

El  $R^2$  (coeficiente de determinación) mide cuánta variación en los datos de respuesta puede ser explicada por el modelo. La mejor puntuación posible es 1, la peor 0, aunque puede ser negativa (porque el modelo puede ser arbitrariamente peor y en realidad no tiene por qué ser el cuadrado de una cantidad  $R$ ).

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

#### 4.11.1.2 MAE (Error Absoluto Medio)

El MAE (*Mean Absolute Error*) mide el promedio de los errores absolutos entre las predicciones y los valores reales, sin considerar la dirección.

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i|$$

#### 4.11.1.3 MSE (Error Cuadrático Medio)

El MSE (*Mean Squared Error*) mide el promedio de los cuadrados de los errores entre las predicciones y los valores reales. Es sensible a los datos atípicos ya que eleva los errores al cuadrado.

$$MSE = \frac{1}{n} \sum (y_i - \hat{y}_i)^2$$

#### 4.11.1.4 Max Error

El *Max Error* mide el error máximo entre las predicciones y los valores reales. Es la diferencia absoluta más grande entre un valor real y su predicción correspondiente.

$$Max Error = \max |y_i - \hat{y}_i|$$

### 4.11.2 Métricas Clasificación

Ahora se definen, las métricas utilizadas con los modelos de clasificación. En el estudio se analizan las métricas tanto si se convierten los registros de lluvia en días con lluvia o no (llueve o no llueve) o en la clasificación de la intensidad de lluvia en diferentes categorías.

#### 4.11.2.1 Exactitud (*Accuracy*)

La Exactitud o *Accuracy* mide la proporción de predicciones correctas (tanto positivas como negativas) entre el total de casos.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

#### 4.11.2.2 Sensibilidad (*Recall*)

La Sensibilidad o *Recall* mide la proporción de positivos reales que fueron identificados correctamente.

$$Recall = \frac{TP}{TP + FN}$$

#### 4.11.2.3 Precisión (*Precision*)

La Precisión mide la proporción de predicciones positivas que fueron correctas.

$$Precision = \frac{TP}{TP + FP}$$

#### 4.11.2.4 F1-Score

El F1-Score es la media armónica de la Precisión y la Sensibilidad / *Recall*, proporcionando un balance entre estas dos métricas.

$$F1 = 2 \frac{Precision * Recall}{Precision + Recall}$$

Como se ha mencionado en la aplicación de los modelos de regresión, clasificación y series de tiempo, las métricas para cada modelo, estación y agregación se realizó por medio de bucles / pipelines de trabajo que al final arrojaban un *dataframe* con las métricas consolidadas por punto.

## 5. ANÁLISIS DE DATOS

### 5.1 ANÁLISIS EXPLORATORIO DE DATOS

#### 5.1.1 Conjunto de datos

Como se comentó en el capítulo anterior, a partir de los datos recopilados, se crearon conjunto de datos / *datasets* / *dataframe*: uno para las estaciones IDEAM (en adelante referido en tablas y gráficas como “Estación o Estaciones”) y otro para los datos satelitales CHIRPS (referido indistintamente como “Satélite o Satélites”). Cada columna / *variable* / *feature* de dichos *datasets* corresponde a una estación / punto estudiado en el proyecto.

#### 5.1.2 Periodo Analizado

Como se muestra en la figura siguiente, para ambos conjuntos de datos, se recopilaron y analizaron 32 años de datos (11688 días / registros).

```
Datos Estaciones lluvia / precipitación
Inicio = 1983-01-01 00:00:00
Fin = 2014-12-31 00:00:00
Período = relativedelta(years=+31, months=+11, days=+30)

Datos Satelites lluvia / precipitación
Inicio = 1983-01-01 00:00:00
Fin = 2014-12-31 00:00:00
Período = 31.99791919067469
```

Fig. 20. Periodo analizado conjunto de datos.

Nota: elaboración propia.

#### 5.1.3 Estadística Descriptiva

En el ANEXO A se presentan los estadísticos descriptivos principales para los datos de Estaciones y Satélites los cuales proporcionan una comprensión general de la distribución, dispersión y forma de los datos e insumo para la generación y análisis de los siguientes gráficos.

### 5.1.3.1 Datos faltantes

Como era de esperarse de equipos y sensores electrónicos debido a fallas o falta de mantenimiento, en el *dataset* de las Estaciones se evidenciaron datos faltantes (NaNs / Nulls) para los puntos analizados.

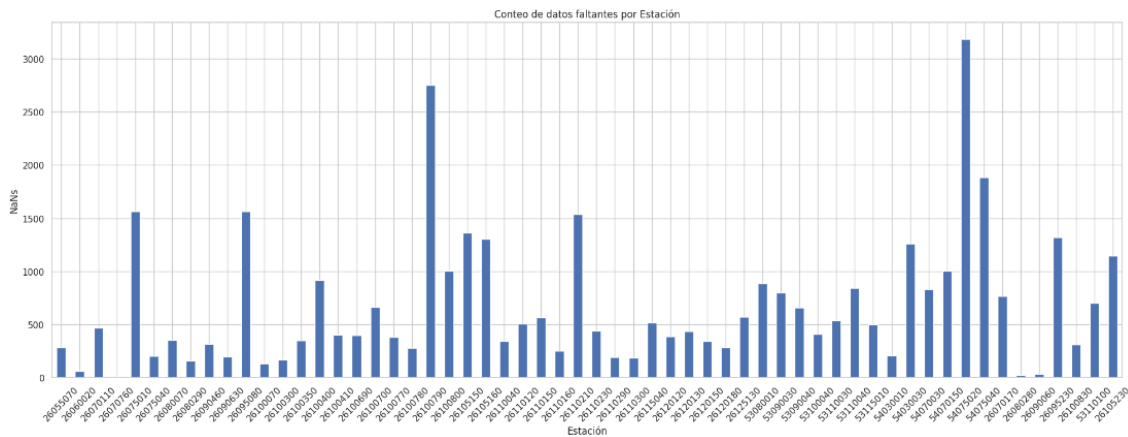


Fig. 21. Datos faltantes por Estación.  
Nota: elaboración propia.

Siendo las estaciones<sup>1</sup> con datos faltantes mayores al 10% las presentadas en la tabla siguiente:

TABLA I. TOP 10 ESTACIONES CON MÁS DATOS FALTANTES.

#	Estación	% Datos Faltantes
1	54075020_y	27.25
2	26100790_y	23.52
3	54075040_y	16.09
4	26075010_y	13.36
5	26095080_y	13.35
6	26110210_y	13.12
7	26105150_y	11.64
8	26095230_y	11.29
9	26105160_y	11.14
10	54030030_y	10.75

Nota: elaboración propia.

La pérdida de datos en las estaciones analizadas es significativa, lo que puede tener diversas implicaciones en el contexto del análisis, la toma de decisiones y la comprensión de fenómenos naturales en la zona de estudio. Es importante garantizar la integridad y disponibilidad de la

<sup>1</sup> Como se ha mencionado anteriormente, en el presente estudio, en los *dataframes* y gráficas desarrolladas, las columnas / variables de las estaciones en tierra llevan el sufijo “\_y” y las de los puntos satelitales “\_x”.

información de las estaciones en tierra, sin embargo, cuando las áreas de estudio presentan condiciones de acceso difíciles o los recursos no son suficientes para revertir la pérdida de información, los datos de satélite se convierten en una fuente de datos alterna sin importar la ubicación geográfica.

Con la ayuda del mapa de calor de la Fig. 22 se puede identificar en cuales de las estaciones en tierra hay una mayor ausencia de datos. En general, las estaciones en tierra tienen periodos de tiempo en los cuales se presenta pérdida de datos, lo que interrumpe las series de tiempo. Por esta razón, se pretende obtener información a partir de los datos satelitales que puedan completar de manera consistente los periodos de tiempo sin información en las estaciones en tierra. Aunque se escapa del alcance del presente estudio, es bueno conocer siempre si la gran cantidad de datos faltantes en algunas estaciones se debe a su ubicación geográfica, condiciones climatológicas, errores de comunicación o por fallas en los instrumentos o sensores.

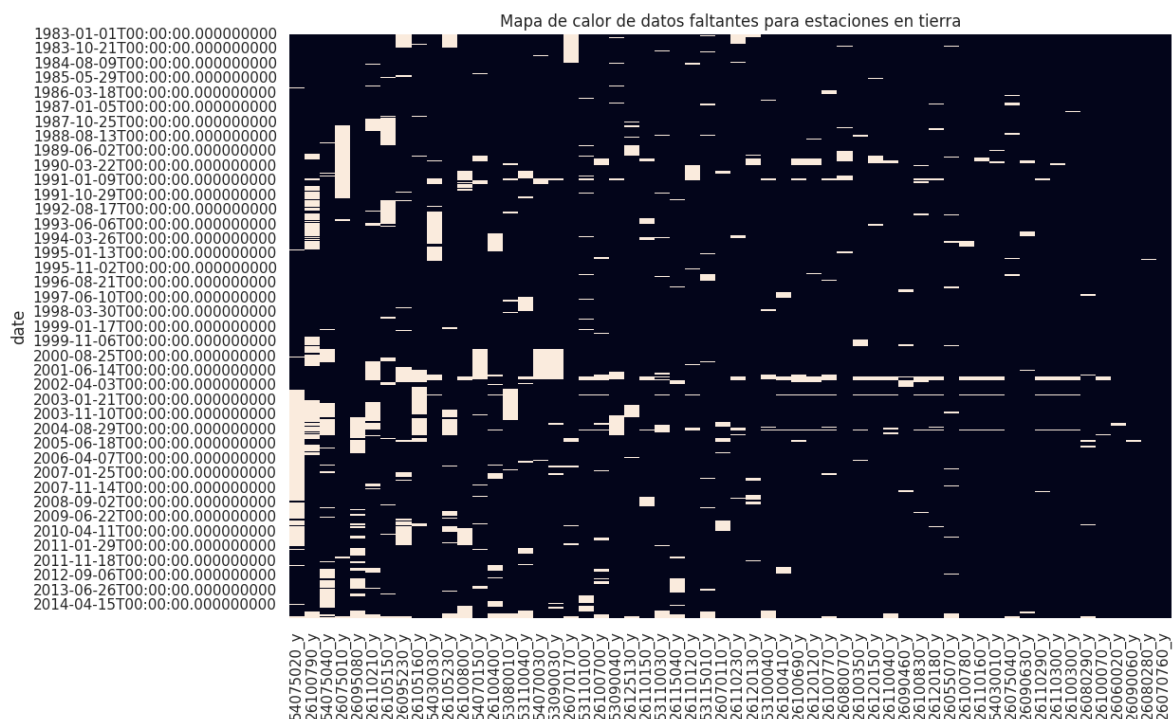


Fig. 22. Mapa de calor Datos faltantes en el periodo analizado.

Nota: elaboración propia.

### 5.1.4 Visualización Datos

En el ANEXO A se pueden consultar las principales visualizaciones realizadas para los datos de precipitación estudiados tanto de las estaciones en tierra, como de satélite.

### 5.1.5 Correlación de Datos

En la figura siguiente (Fig. 23) se presenta el mapa de calor de correlación para las Estaciones en tierra, esta visualización permite analizar la relación entre las estaciones en forma de una matriz de colores.

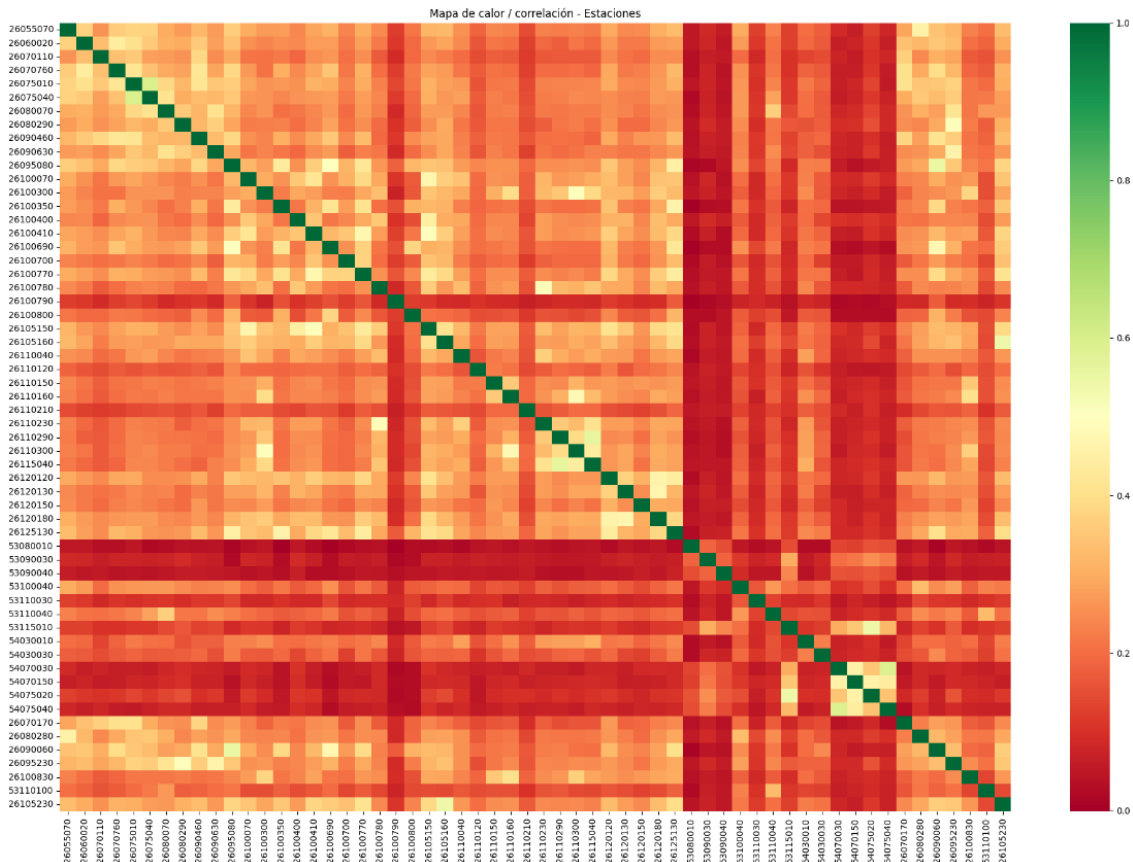


Fig. 23. Mapa de calor de correlación para las Estaciones – datos diarios.

Nota: elaboración propia.

Se observa que la correlación entre estaciones en tierra se encuentra entre moderada y débil, con magnitudes por debajo de los 0.5. Estos valores de correlación pueden deberse al distanciamiento entre las estaciones, lo que permite que los datos capturados por cada una de las estaciones sean valores significativos en la zona de estudio. Con este mapa de calor también se corrobora que hay un grupo de estaciones con correlación muy baja con las otras (celdas con color rojo valores con cercanos a cero de la correlación).

Ahora, en la Fig. 24 en la página siguiente se puede apreciar el mapa de calor de correlación para los datos de satélite de precipitación.



Como se observa en la Fig. 25 las estaciones rojas tienen una magnitud de correlación baja, su localización está distante de la mayoría de las estaciones y, además, se encuentran en una zona denominada geomorfológicamente como la llanura marina del Pacífico, región con un clima tropical húmedo con altas precipitaciones y temperaturas cálidas todo el año. Esta combinación de clima y condiciones oceánicas pueden darnos una idea del por qué estas estaciones presentan datos con una distribución mayor de su mediana en el rango intercuartil (50% central de los datos), con más cantidad de valores atípicos y con una correlación más baja comparadas frente al resto de estaciones terrestres localizadas en las regiones de la Cordillera Central, el valle geográfico del río Cauca y la Cordillera Occidental.

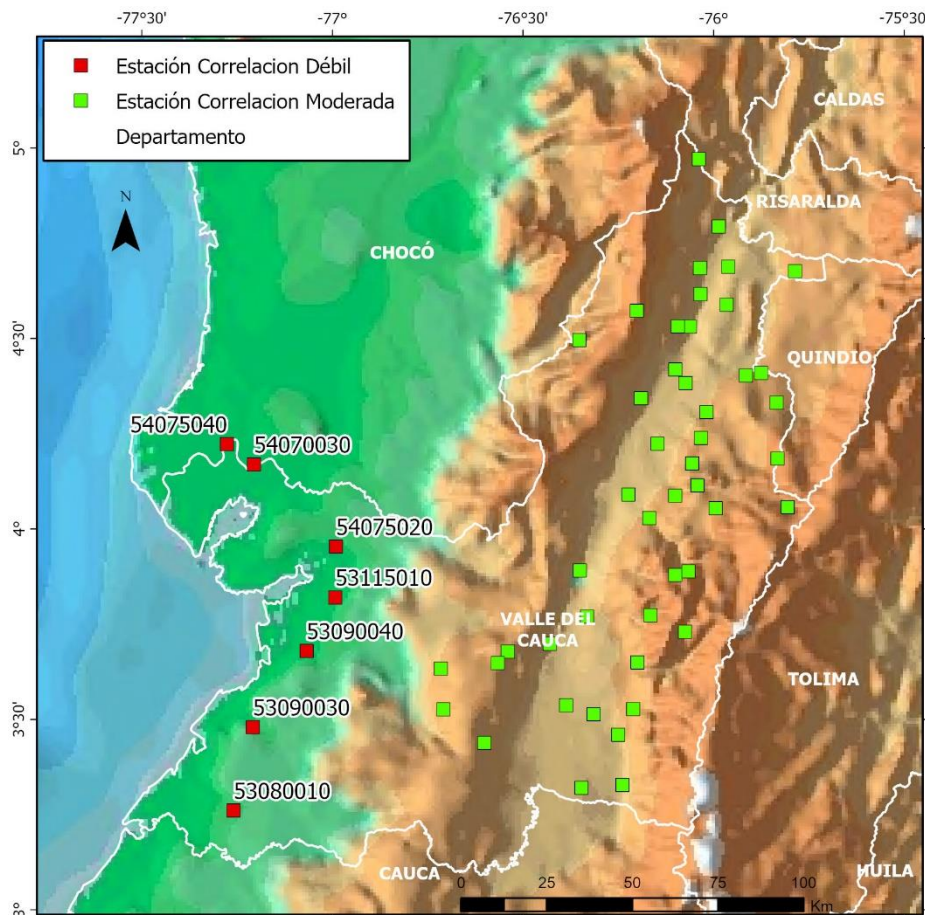


Fig. 25. Localización de estaciones con correlación débil  
Nota: elaboración propia.

Lo anterior se evidencia a nivel de correlaciones, en el siguiente gráfico (Fig. 26) que combina el resultado de la correlación entre las estaciones (tres medidas comunes: Pearson, Spearman y Kendall) y el valor del porcentaje de días sin lluvia con una temporalidad diaria. Las tres medidas de correlación muestran un comportamiento muy similar (se encuentran alrededor de 0.2), esto sugiere que las variables están relacionadas de manera consistente, dicho de otra manera, la

relación entre variables es robusta y no está influenciada significativamente por supuesto específicos de cada método.

De igual forma, se observa como factor común, la alta presencia de valores ceros, interpretado como días sin lluvia. Esto, claramente representaría la clase dominante en un problema de clasificación. Y, por la naturaleza del problema, esa desbalanceada distribución es la representación precisa de la realidad. Realidad en donde la clase minoritaria (días con lluvia) representa eventos inusuales, por lo que mantener dicho desbalance refleja más fielmente la frecuencia real de estos eventos en el mundo real.

Es así como, y para este caso en particular, el balancear el conjunto de datos mediante diferentes técnicas puede llevar a la generación de datos artificiales o a la pérdida de información valiosa. Esto puede causar que el modelo se sobreajuste de especial manera a los datos desbalanceados, en lugar de llegar a una buena generalización de los datos reales.

Por otra parte, se considera que la evaluación de los modelos de clasificación en conjuntos con datos desbalanceados puede ser más realista para entender su rendimiento en escenarios prácticos.

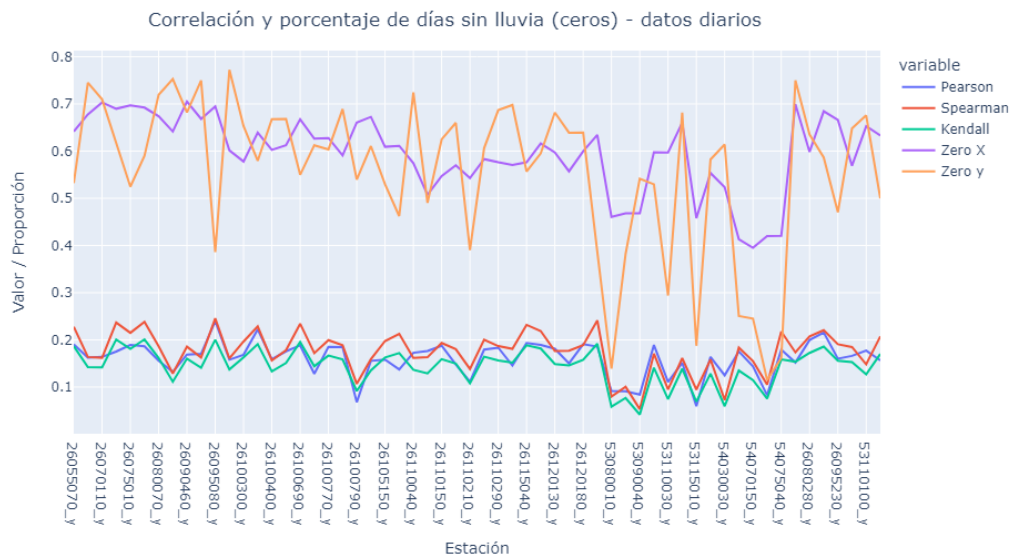


Fig. 26. Correlación y porcentaje de días sin lluvia (ceros) - datos diarios  
Nota: elaboración propia.

El siguiente mapa de calor (Fig. 27) nos muestra el comportamiento de los datos tanto de las estaciones en tierra como de los datos provenientes de las imágenes satelitales, permitiéndonos ver unos valores de precipitación acumulada mensual altos en las mismas estaciones que referenciamos anteriormente con una correlación débil.

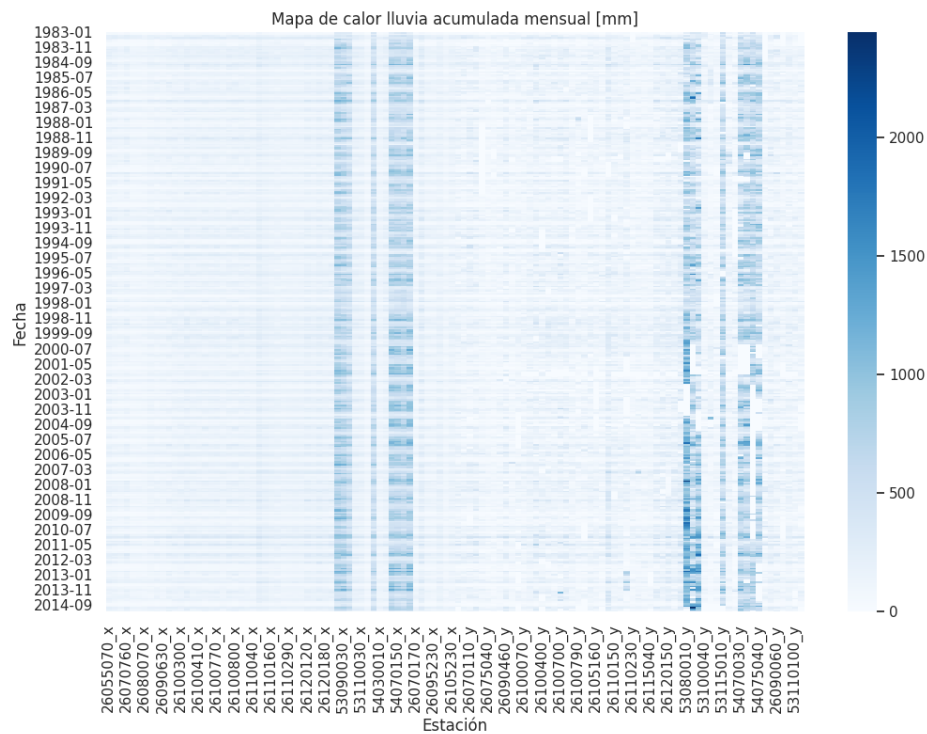


Fig. 27. Mapa de calor lluvia acumulada mensual  
Nota: elaboración propia.

Lo anterior es clave tenerlo presente, ya que como se mostrará en los capítulos siguientes los modelos utilizados no se desempeñan bien en las estaciones / puntos con alta precipitación.

## 5.2 AGREGACIONES TEMPORALES

Como se comentó en el numeral 5.1.2, se cuenta con un periodo de 11688 días de datos de lluvia. Es importante precisar también, que la precipitación se suele analizar a nivel horario, diario, semanal, mensual o anual dependiendo del tipo de estudio que se quiera realizar. Como se presentará en el capítulo siguiente en el proyecto se probaron los modelos seleccionados con agregaciones (lluvia acumulada) diarias, semanales, mensuales y anuales, aunque para la mayoría de los modelos de regresión los resultados mejoraron a nivel semanal (1671) y mensual (384 meses) respecto a los datos diarios, no pasó lo mismo con datos anualizados (32 años de estudio).

### 5.2.1 Comportamiento temporal de la lluvia con diferentes agregaciones

En las figuras siguientes (Fig. 28, Fig. 29, Fig. 30 y Fig. 31) se presenta como cambian las series de tiempo para la Estación 26070760 en las agregaciones utilizadas en el estudio (diaria, semanal, mensual y anual), tanto para datos de las estaciones en tierra, como para los satelitales.

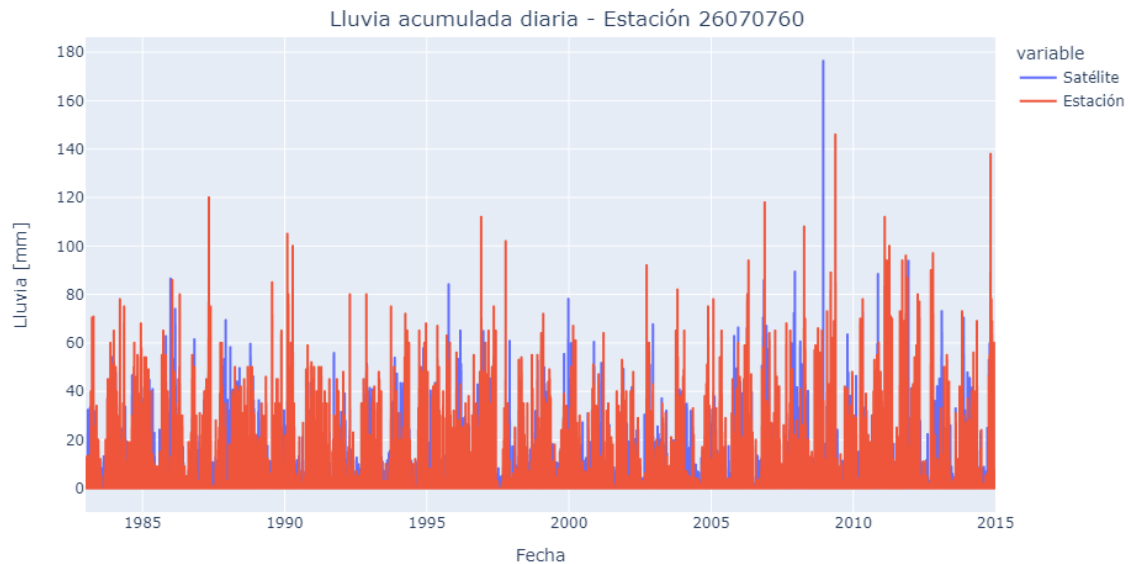


Fig. 28. Lluvia acumulada diaria - Estación 26070760

Nota: elaboración propia.

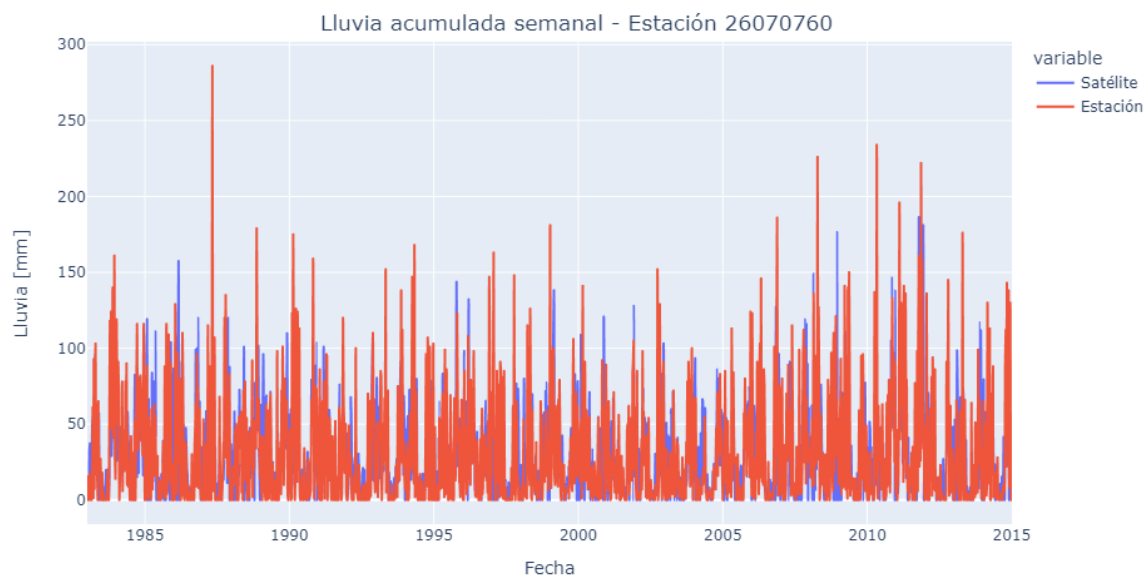


Fig. 29. Lluvia acumulada semanal - Estación 26070760

Nota: elaboración propia.

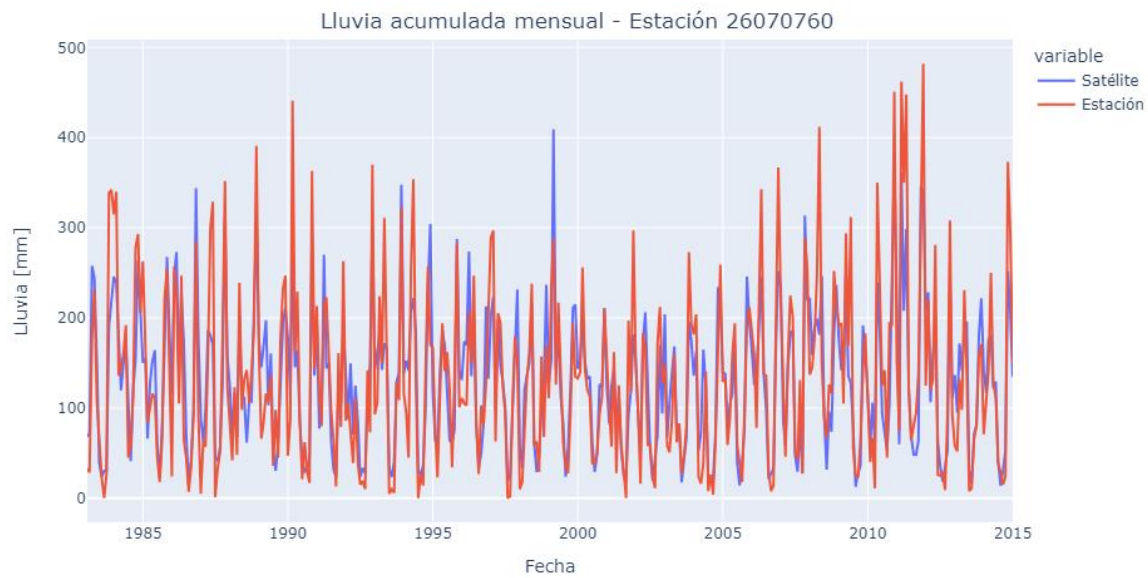


Fig. 30. Lluvia acumulada mensual - Estación 26070760  
Nota: elaboración propia.

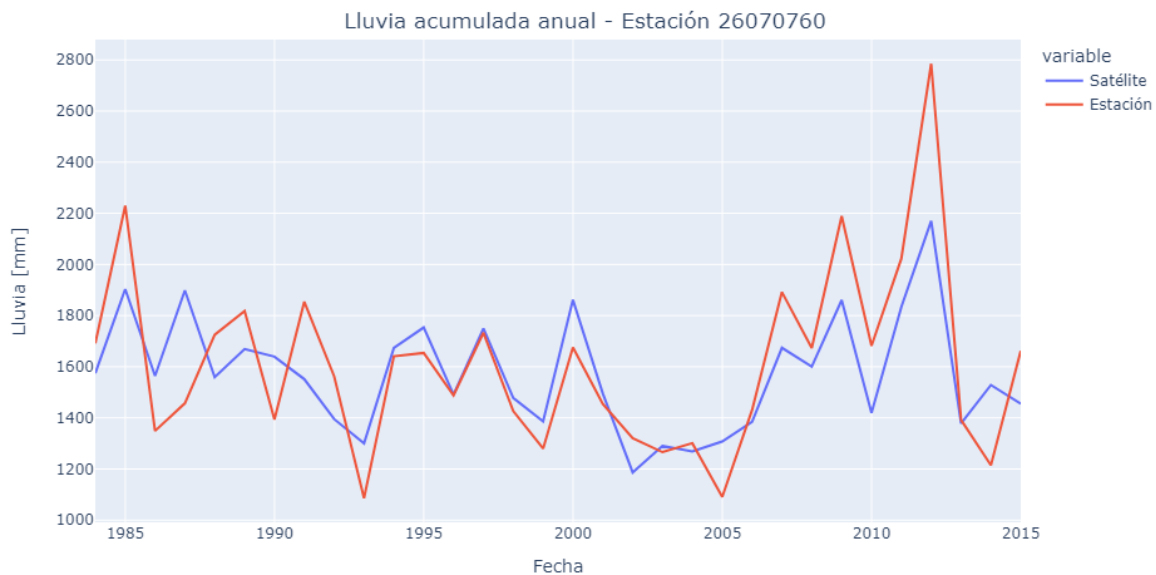


Fig. 31. Lluvia acumulada anual - Estación 26070760  
Nota: elaboración propia.

La estación seleccionada para las gráficas anteriores es la que menos datos faltantes cuenta del *dataset* y en general se pueda observar que los datos satelitales y de las estaciones en tierra presentan tendencias más afines a agregaciones mayores.

Para evaluar el impacto de las agregaciones estudiadas en la correlación entre la estación en tierra y satelital para el punto referido (26070760), a continuación, se presentan en las figuras siguientes

(Fig. 32, Fig. 33, Fig. 34 y Fig. 35), las matrices o gráficas de dispersión para la lluvia acumulada diaria, semanal, mensual y anual.

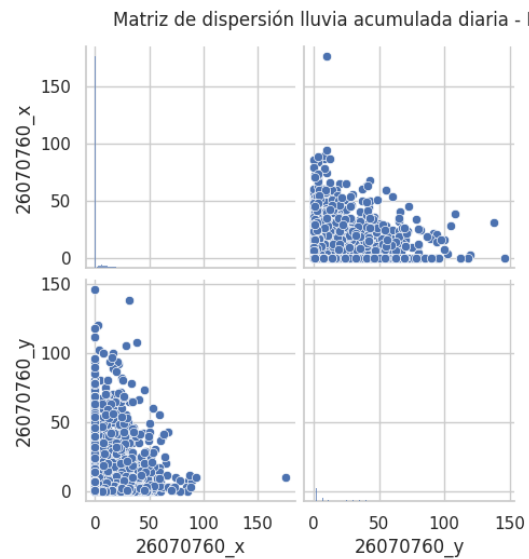


Fig. 32. Matriz de dispersión lluvia acumulada diaria - Estación 26070760  
Nota: elaboración propia.

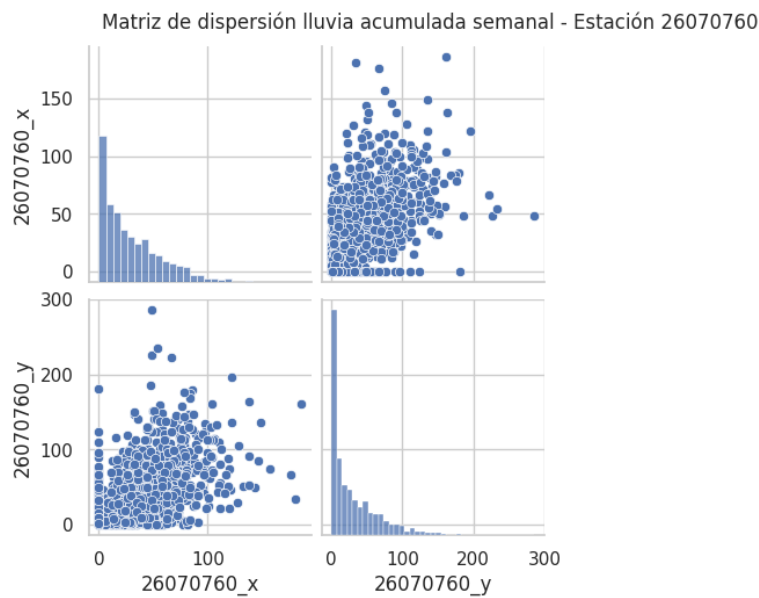


Fig. 33. Matriz de dispersión lluvia acumulada semanal- Estación 26070760  
Nota: elaboración propia.

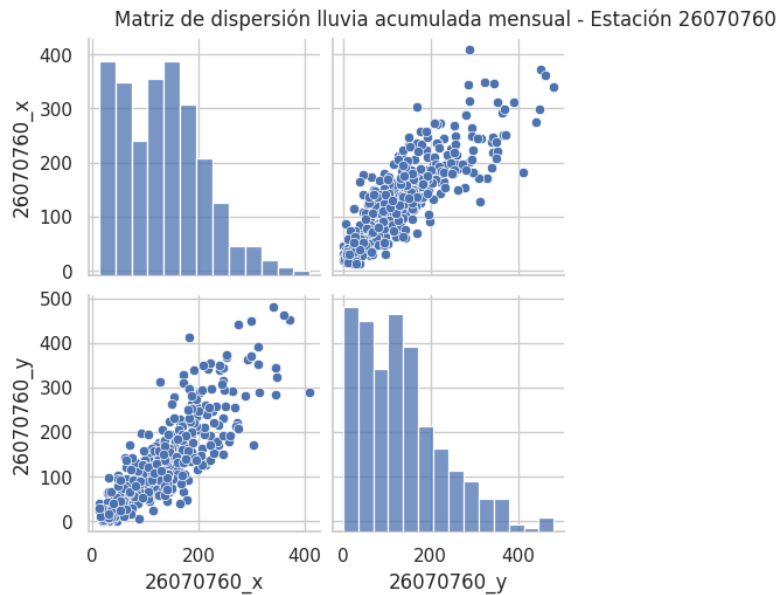


Fig. 34. Matriz de dispersión lluvia acumulada mensual- Estación 26070760  
Nota: elaboración propia.

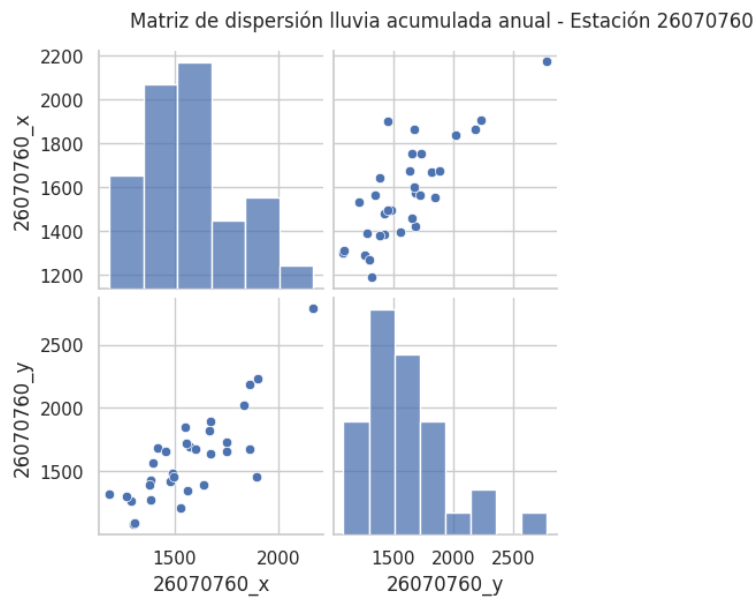


Fig. 35. Matriz de dispersión lluvia acumulada anual- Estación 26070760  
Nota: elaboración propia.

Como se puede observar en las figuras anteriores, con datos diarios no se encuentra ninguna tendencia o correlación clara entre los datos en tierra y satelitales, mientras que con datos semanales empieza a formarse una tendencia, la cual es clara para los datos mensuales y empieza perderse en las lluvias acumuladas anuales debido a la poca cantidad de datos.

## 5.2.2 Correlación de la lluvia con diferentes agregaciones temporales

Tal como se presentó en la Fig. 26 para los datos diarios, para los datos de lluvias semanales, mensuales y anuales también se calcularon las correlaciones de Pearson, Spearman y Kendall para todas las estaciones y la existencia de días sin lluvias (porcentaje de ceros) en cada una de las series, los resultados se pueden apreciar en las figuras siguientes (Fig. 36, Fig. 37 y Fig. 38).

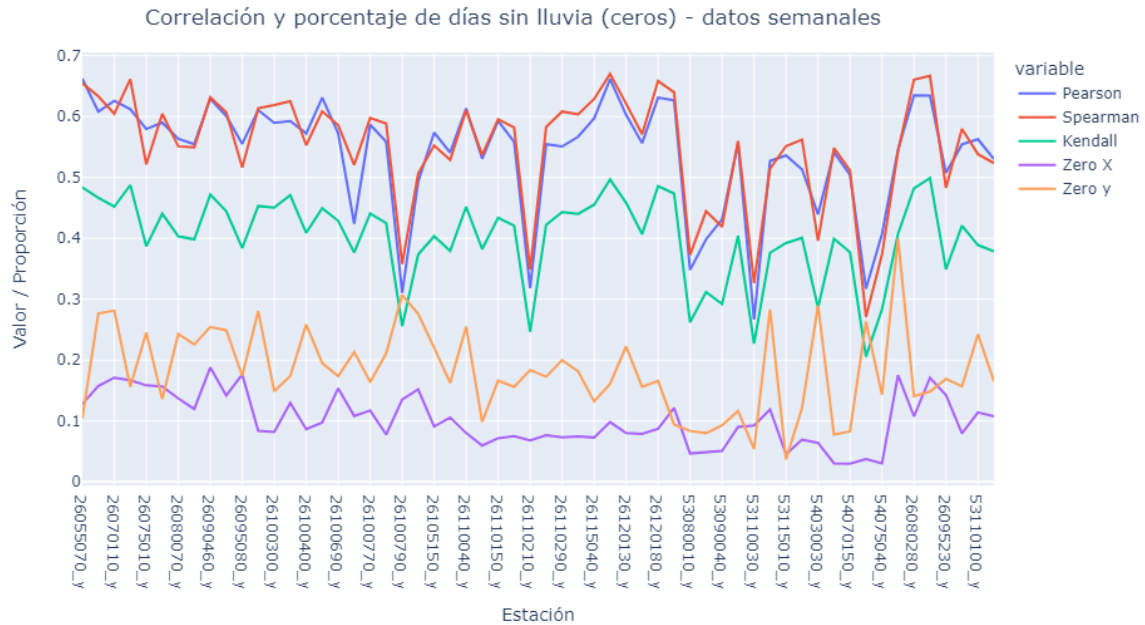


Fig. 36. Correlación y porcentaje de días sin lluvia (ceros) - datos semanales

Nota: elaboración propia.

Correlación y porcentaje de días sin lluvia (ceros) - datos mensuales

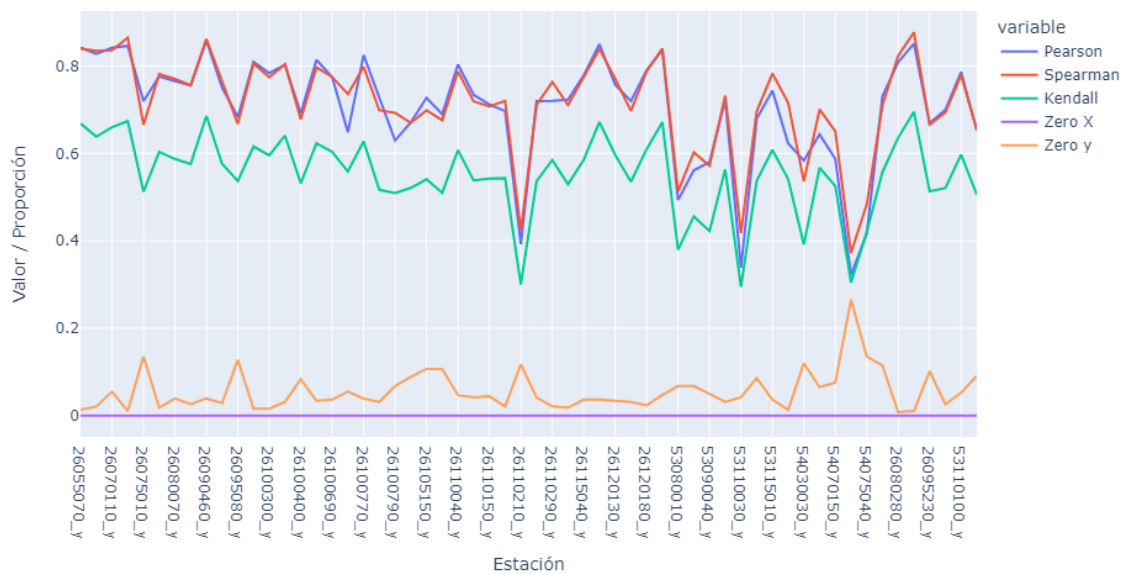


Fig. 37. Correlación y porcentaje de días sin lluvia (ceros) - datos mensuales  
Nota: elaboración propia.

Correlación y porcentaje de días sin lluvia (ceros) - datos anuales



Fig. 38. Correlación y porcentaje de días sin lluvia (ceros) - datos anuales  
Nota: elaboración propia.

Tal como se apreció con las dispersiones de la estación 26070760, en las figuras anteriores se puede observar que los valores de correlación aumentan a nivel semanal respecto a los diarios (pasan de 0.2 a 0.5), teniendo mejores resultados Pearson y Spearman que Kendall; para los datos mensuales

se dan las mejores correlaciones (entre 0.6 y 0.8) y para datos anuales es interesante notar que algunas estaciones presentan correlaciones bajas e incluso negativas. Los ceros (periodos sin lluvia) como es de esperar disminuyen a medida que aumenta el periodo / agregación.

Para los datos de lluvia mensual que presentaron las mejores correlaciones, se presenta en la página siguiente en la Fig. 39 el mapa de calor de las correlaciones entre las estaciones en tierra y satelitales.

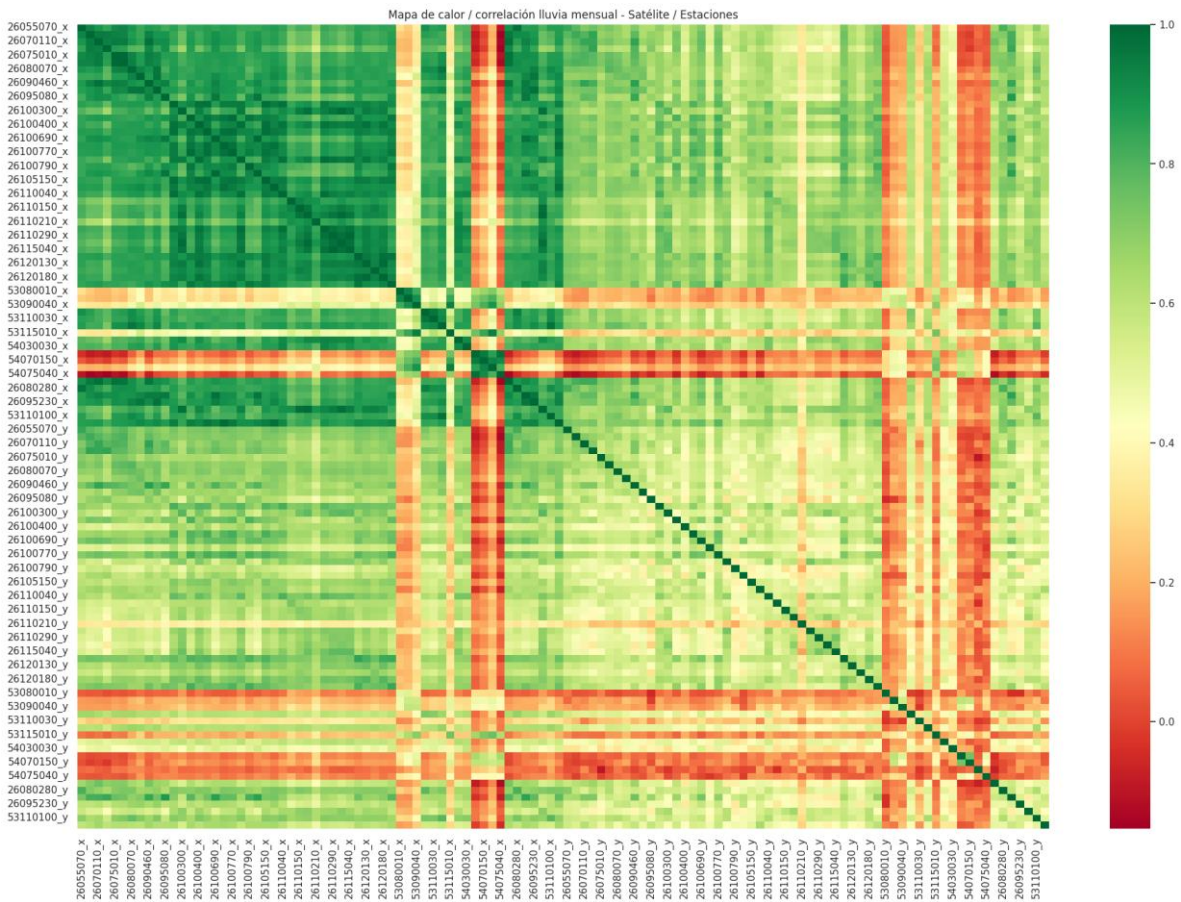


Fig. 39. Mapa de calor de correlación para datos en Satélite y Estaciones Tierra – datos mensuales.  
Nota: elaboración propia.

La Fig. 39 muestra la correlación entre los datos mensuales para Satélite / Estaciones (en tierra), donde se observa en general que la fuerza de la correlación es alta entre los datos, solo en unos pocos casos se presentan correlaciones débiles. Con esto se puede inferir que la mayoría de los datos de las estaciones y de los satélites tienden a aumentar y disminuir juntos. Pese a que los datos tienen una periodicidad mensual sigue observándose una correlación débil en las mismas ocho estaciones que fueron identificadas en el mapa de calor de correlación para los datos diarios (ver numeral 5.1.5).

## 6. RESULTADOS MODELOS

### 6.1 APRENDIZAJE AUTOMÁTICO

A continuación, se presentan los resultados de las métricas principales obtenidos tanto a nivel de regresión, como de clasificación con modelos de aprendizaje automático seleccionados en el proyecto.

#### 6.1.1 Modelos de regresión

Para los modelos de regresión se presentan las gráficas siguientes con los valores obtenidos de  $R^2$ , para cada modelo referenciado y el valor medio o promedio para todas las estaciones (línea punteada de color rojo). Las gráficas de resultados de  $R^2$  para cada modelo de regresión, se pueden consultar en el ANEXO B. Las tablas de todas las métricas obtenidas para los modelos de regresión incluyendo MAE, MSE y MaxError se pueden consultar en el ANEXO C.

##### 6.1.1.1 LinearRegression

En las figuras siguientes (Fig. 40, Fig. 41 y Fig. 42) se presentan los resultados del  $R^2$  obtenidos mediante regresión lineal para las predicciones diarias, semanales y mensuales.

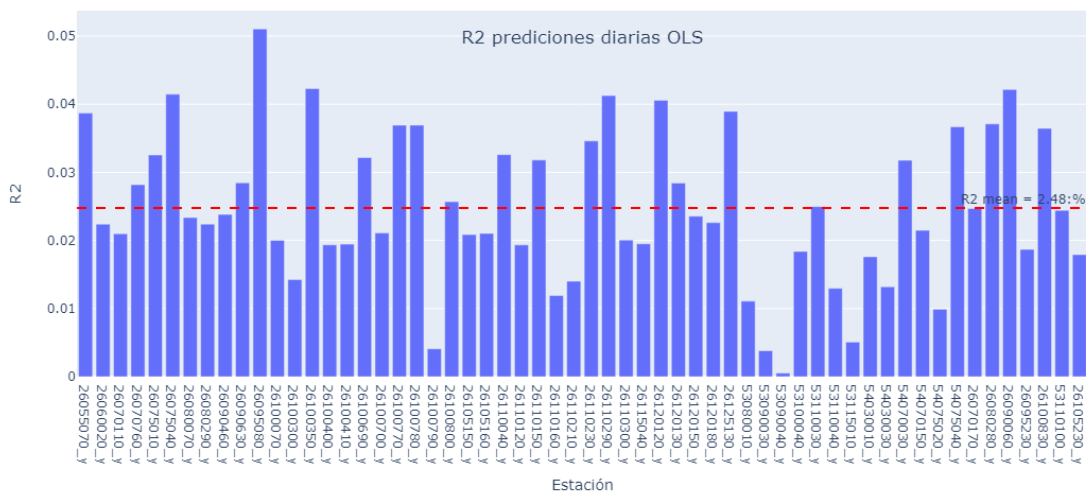


Fig. 40. Resultados  $R^2$  predicciones diarias OLS

Nota: elaboración propia.

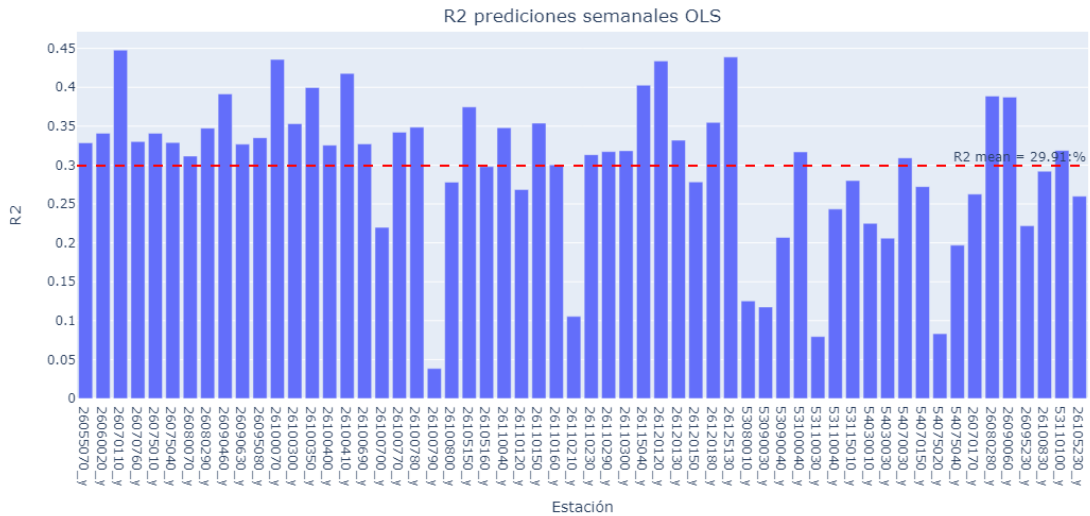


Fig. 41. Resultados R2 predicciones semanales OLS  
Nota: elaboración propia.

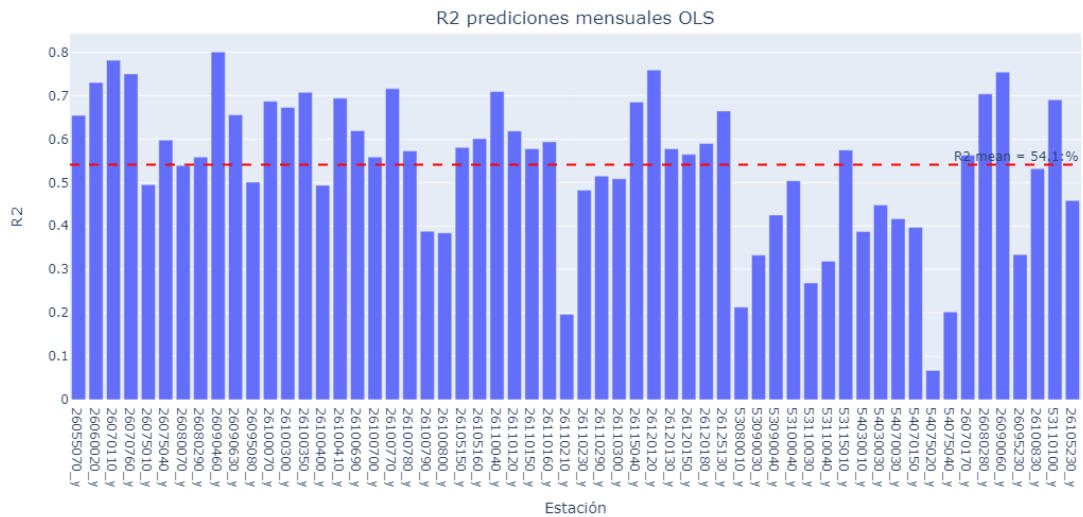


Fig. 42. Resultados R2 predicciones mensuales OLS  
Nota: elaboración propia.

Como se puede observar en las figuras anteriores, los modelos de regresión lineal / OLS no presentan buenos resultados de  $R^2$ , para agregaciones diarias, pero el valor promedio de dicha métrica pasa de cerca de 2.5% a casi 30% con datos semanales y al 54% para las agregaciones mensuales. En el proyecto se aplicaron otros estimadores / modelos lineales como *Ridge*, *Lasso*, *ElasticNet* de *scikit-learn* obteniendo en esencia los mismos resultados anteriores con *LinearRegression*.

### 6.1.1.2 SVR

En las figuras siguientes (Fig. 43, Fig. 44 y Fig. 45) se presentan los resultados del  $R^2$  obtenidos mediante máquinas de soporte vectorial (SVR) para las predicciones diarias, semanales y mensuales.

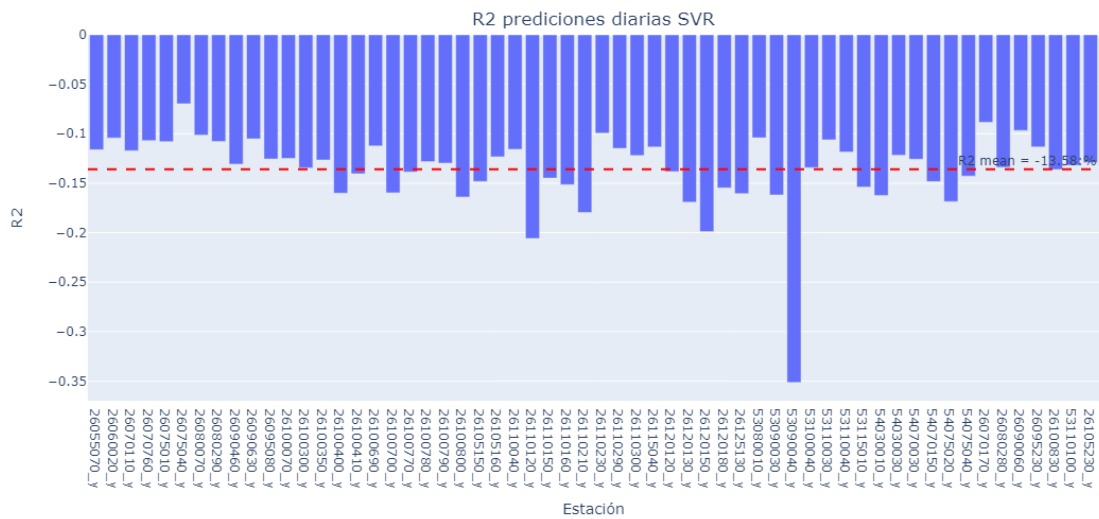


Fig. 43. Resultados  $R^2$  predicciones diarias SVR

Nota: elaboración propia.

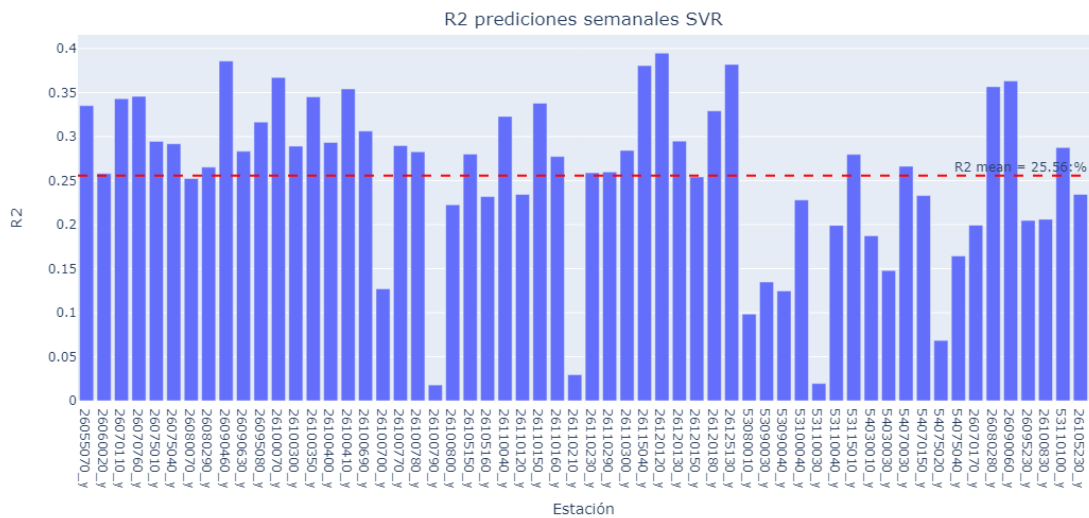


Fig. 44. Resultados  $R^2$  predicciones semanales SVR

Nota: elaboración propia.



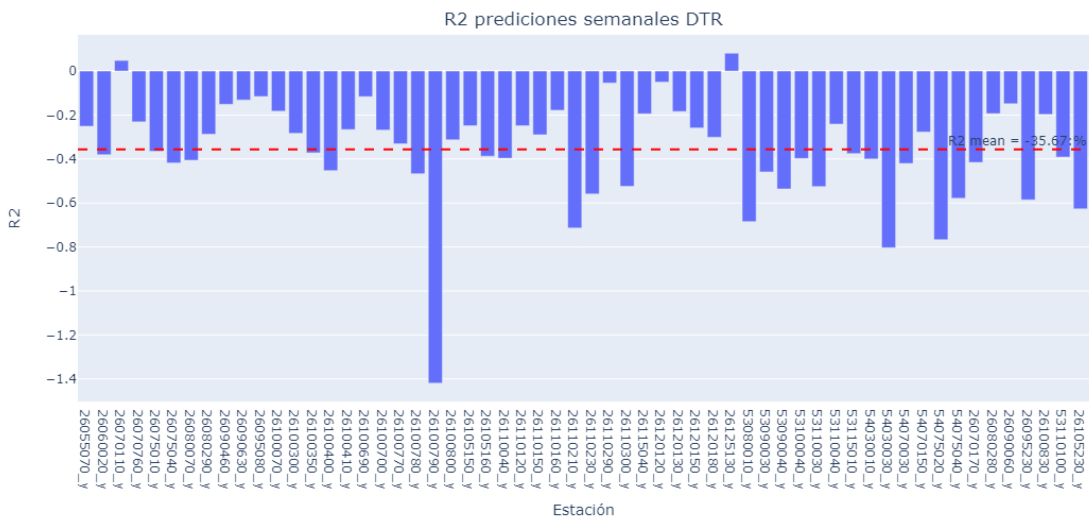


Fig. 47. Resultados R2 predicciones semanales DTR  
Nota: elaboración propia.

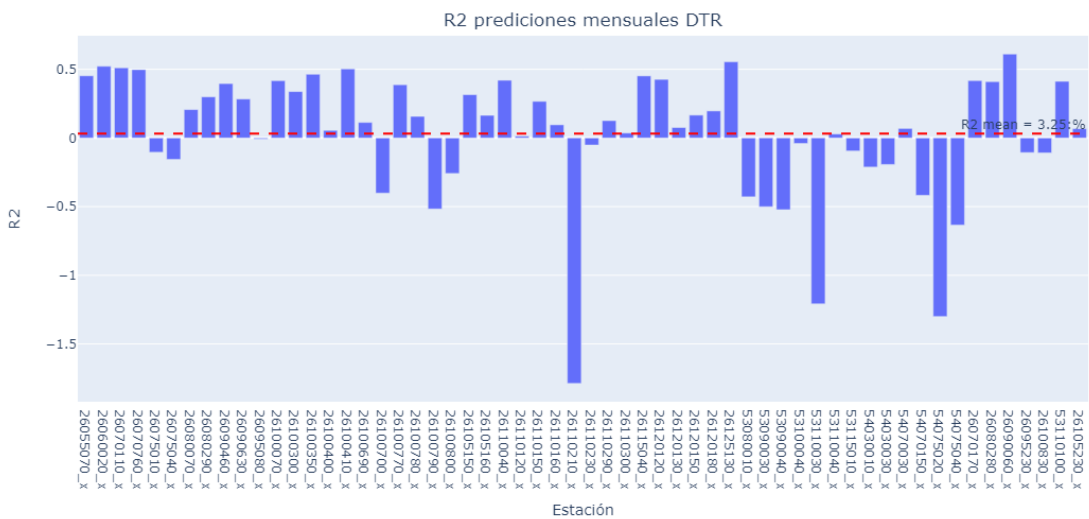


Fig. 48. Resultados R2 predicciones mensuales DTR  
Nota: elaboración propia.

Como se puede observar en las tres figuras anteriores, los árboles de decisión no dan métricas adecuadas de  $R^2$  para los registros diarios de lluvia, no mejoran mucho al agregarse semanalmente y en general permanecen bajos para las lluvias mensuales. Es de notar, que en cada agregación hay una estación diferente que da un desempeño muy malo y arrastra el promedio de la métrica para todas las estaciones.

## 6.1.2 Modelos de clasificación

A continuación, se presentan las gráficas con los resultados de la métrica *Accuracy* (exactitud) para los modelos de clasificación utilizados en el estudio, así como el valor promedio para todas las estaciones analizadas. Las tablas de todas las métricas obtenidas para los modelos de clasificación incluyendo *recall*, *precisión* y *F1 score* se pueden consultar en el ANEXO C.

### 6.1.2.1 LogisticRegression

En las figuras siguientes (Fig. 49, Fig. 50 y Fig. 51) se presentan los resultados de *Accuracy* obtenidos mediante regresión logística (*logit*) para las predicciones diarias, semanales y mensuales. En el eje y se visualiza la precisión obtenida y en el X la estación objeto de la predicción.

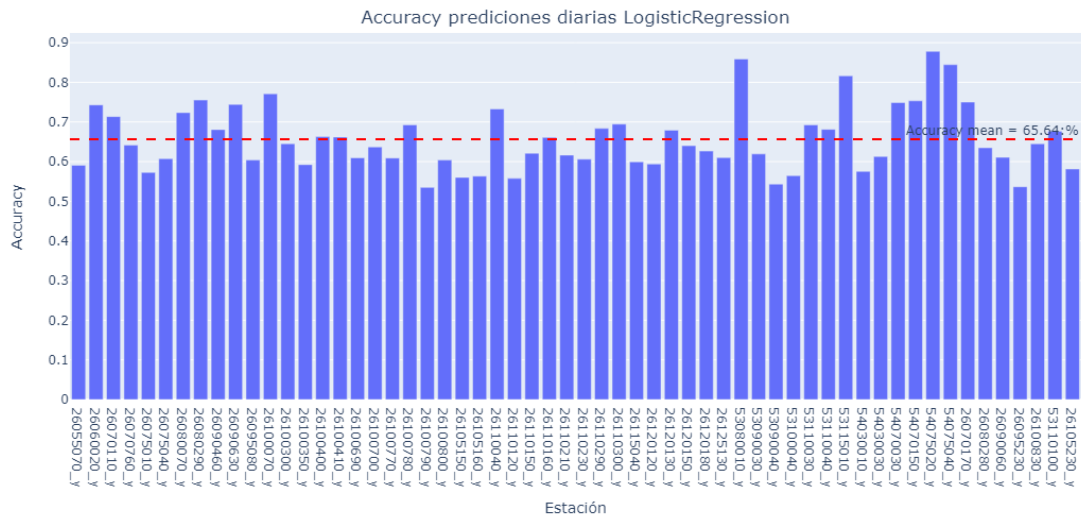


Fig. 49. Resultados Accuracy predicciones diarias Logit  
Nota: elaboración propia.

Como se puede observar en la Fig. 57 la precisión del modelo para predicciones diarias es de cerca 66%, lo que significa que el modelo predijo correctamente un poco más de la mitad de los casos correctamente.



Aunque son métricas diferentes, es claro que el desempeño de los modelos de clasificación para predecir si lluevo o no mejora respecto a cuando se quiere predecir el valor de milímetros de precipitación de un día usando modelos de regresión.

### 6.1.2.2 SVC

A continuación, se relacionan las figuras Fig. 52, Fig. 53 y Fig. 54, las cuales presentan los resultados de *Accuracy* obtenidos mediante máquinas de soporte vectorial (SVC) para las predicciones / clasificaciones diarias, semanales y mensuales. Aunque no se presentan en este numeral, en el estudio se probaron otros estimadores basados en máquinas de soporte vectorial adicional a SVC, como *NuSVC* y *LinearSVC* de *scikit-learn* obteniendo en esencia los mismos resultados que se muestran en las figuras siguientes.

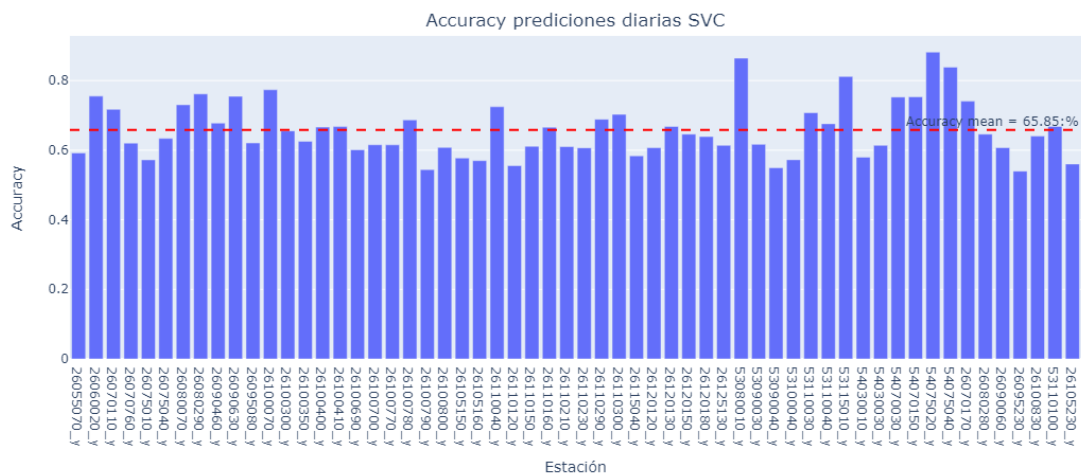


Fig. 52. Resultados Accuracy predicciones diarias SVC  
Nota: elaboración propia.

En la Fig. 60 se observa una precisión moderada para el modelo SVC con predicciones diarias, similar a la exactitud del 66% arrojada por el modelo de regresión logística / logit para los registros diarios.

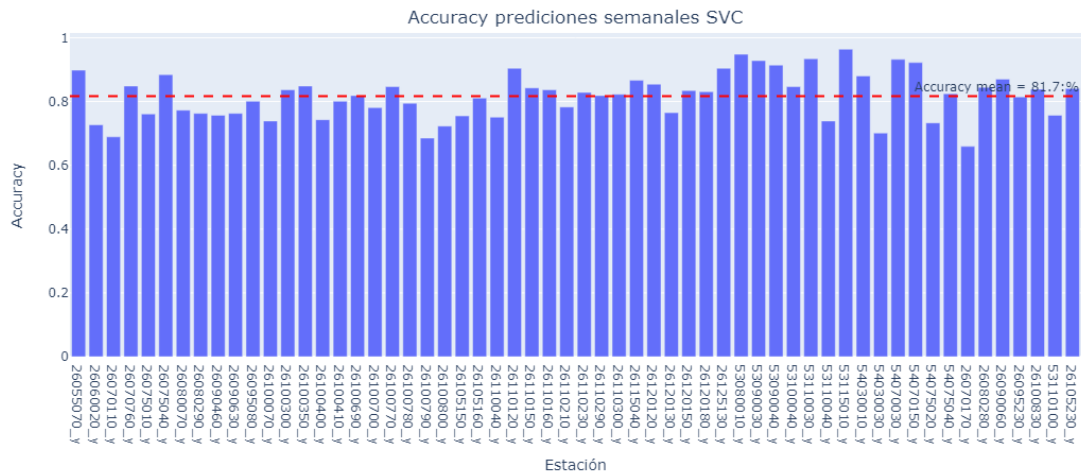


Fig. 53. Resultados Accuracy predicciones semanales SVC  
Nota: elaboración propia.

Para el modelo SVC con predicciones semanales (Fig. 61) se observa un incremento en la exactitud del 17% aproximadamente y llega a un 82%, respecto a las clasificaciones diarias. El resultado obtenido con las máquinas de soporte vectorial permite confirmar que para el caso de las clasificaciones de lluvia a mayores agregaciones (sin contar la anual) se obtienen mejores *accuracies*.

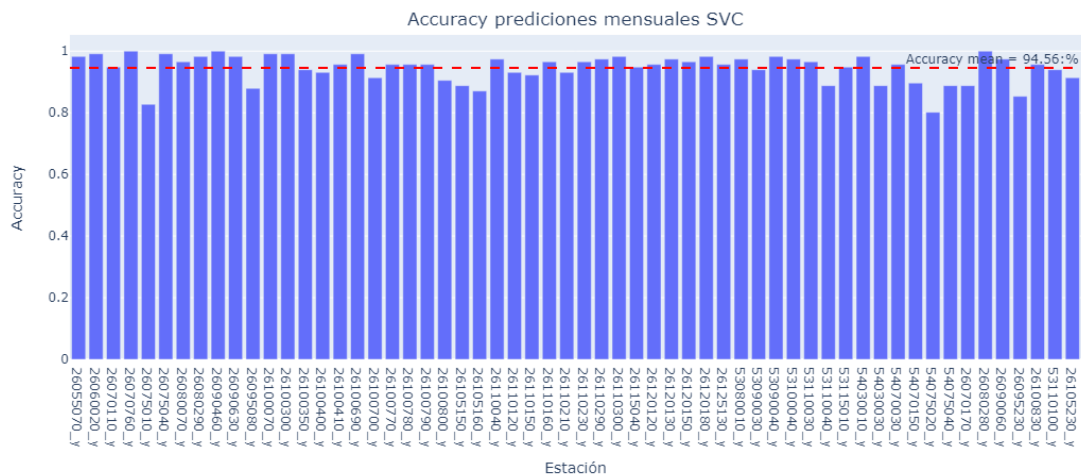


Fig. 54. Resultados Accuracy predicciones mensuales SVC  
Nota: elaboración propia.

La precisión observada en la Fig. 62 correspondiente al modelo SVC para las predicciones mensuales, refuerza lo afirmado con las predicciones semanales: el valor del *accuracy* aumentará de manera significativa (cerca de 30%) al clasificar la precipitación desde agregaciones diarias a agregaciones mensuales.

### 6.1.2.3 DecisionTreeClassifier

En las figuras siguientes (Fig. 55, Fig. 56 y Fig. 57) se presentan los resultados de *Accuracy* obtenidos mediante árboles de decisión (DTC) para las clasificaciones diarias, semanales y mensuales.

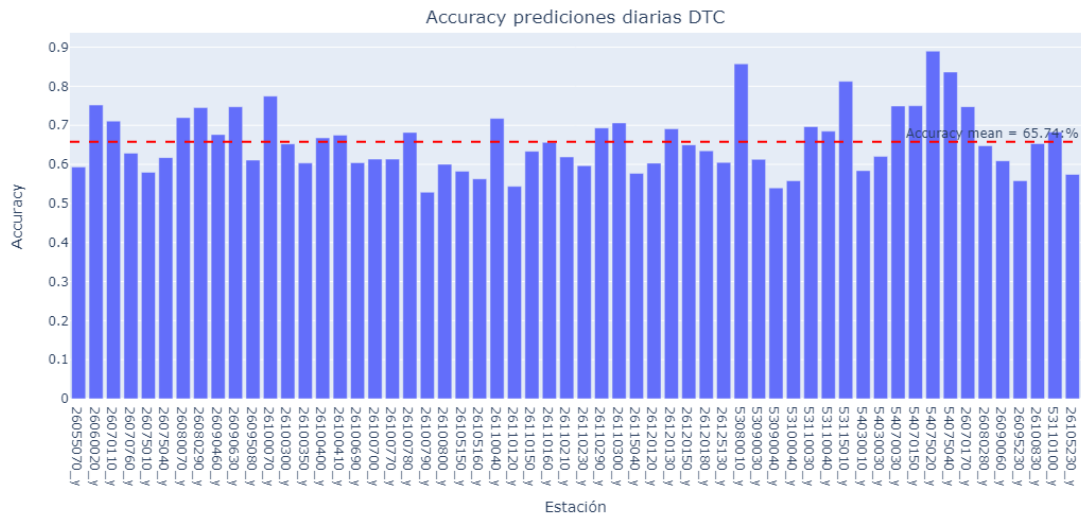


Fig. 55. Resultados Accuracy predicciones diarias DTC  
Nota: elaboración propia.

En la Fig. 63 se observa una exactitud alrededor de los 0.66 para las predicciones diarias con DTC tal como se observó usando los modelos logit y SVC.

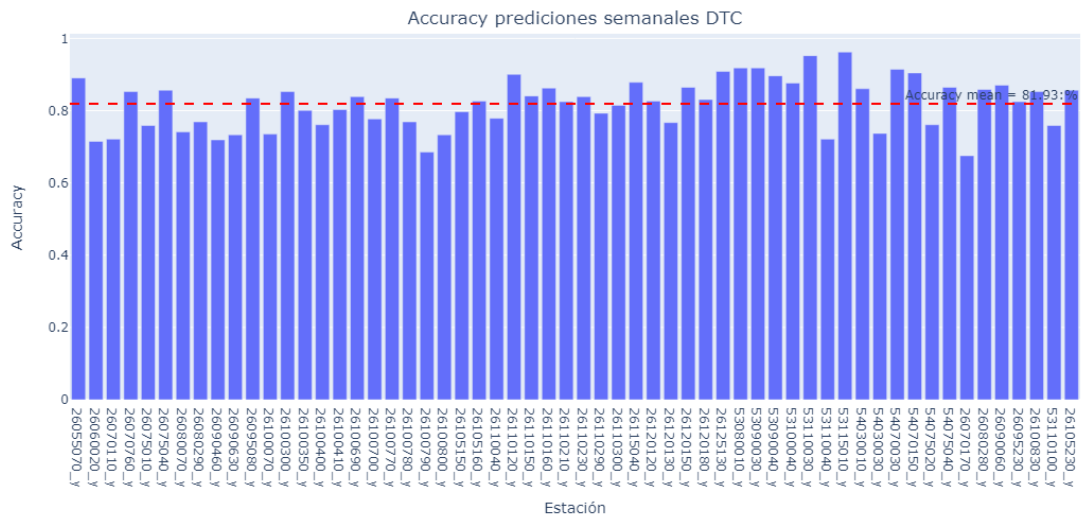


Fig. 56. Resultados Accuracy predicciones semanales DTC  
Nota: elaboración propia.

Nuevamente, como los modelos de clasificación anteriores, se observa un incremento al 82% (Fig. 64) al agregar de manera semanal los datos.

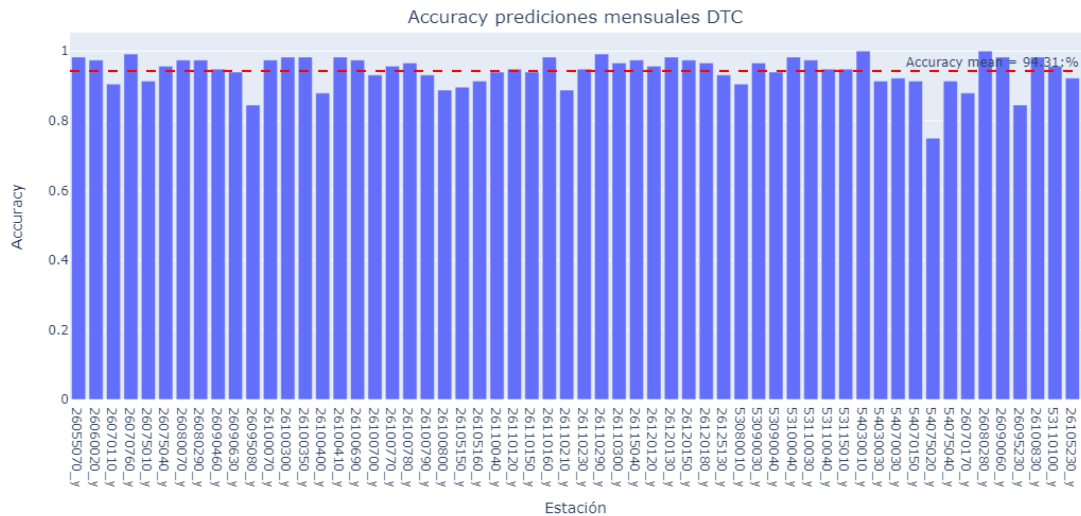


Fig. 57. Resultados Accuracy predicciones mensuales DTC  
Nota: elaboración propia.

Al comparar los resultados de *accuracy* obtenidos para los tres modelos / estimadores utilizados (*LogisticRegression*, *SCV* y *DecisionTreeClassifier*) se observan resultados similares al aumentar la agregación. Esto sugiere que el periodo de tiempo más adecuado para lograr una alta exactitud en la clasificación es mensualmente. Los modelos ajustados a este periodo de tiempo pueden ser mucho más simples, lo que les permite centrarse en patrones a largo plazo, que suelen ser más estables y fáciles de predecir.

Ahora, al aplicar las diferentes intensidades de lluvia descritas anteriormente (ver numeral 4.9.1), se generaron *dataframes* con las categorías de cada intensidad y se utilizó el modelo de clasificación de regresión logística (logit) para evaluar el cambio en las métricas seleccionadas en el proyecto. Los resultados de *accuracy* obtenidos se presentan en la figura siguiente (Fig. 58).

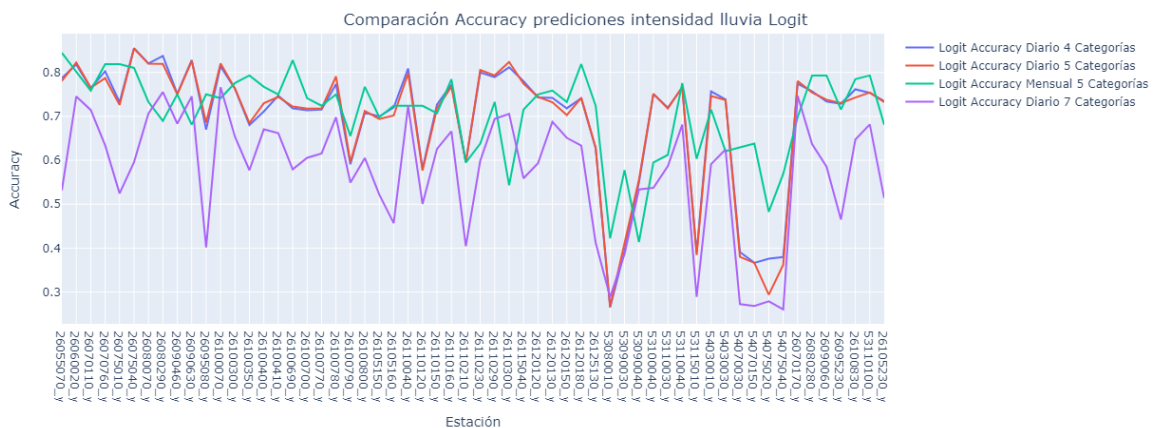


Fig. 58. Comparación resultados Accuracy Logit por Intensidad de Lluvia  
Nota: elaboración propia.

De la figura anterior, se aprecia claramente que el aumento del número de categorías disminuye el valor de la exactitud obtenida en cada estación por el modelo de regresión logística, tanto que los valores obtenidos para 5 categorías / intensidades diarias presentan mejores *accuracies* en algunos puntos que las predicciones mensuales para el mismo número de categorías.

## 6.2 APRENDIZAJE ESTADÍSTICO

A continuación, se presentan los resultados obtenidos para los modelos estadísticos de series de tiempo seleccionados en el proyecto.

### 6.2.1 Modelos series de tiempo

#### 6.2.1.1 ARIMA

Es importante señalar que la estrategia empleada se fundamentó en identificar la estación con la mejor convergencia entre la correlación más fuerte entre las mediciones en tierra y los valores satelitales, y la secuencia más extensa de valores no nulos, para de ese modo, ejemplificar la aplicación del modelo sobre dicha estación; salvaguardando así el precepto de temporalidad de los datos que rige al modelo ARIMA (ver Fig. 59 y Fig. 60).

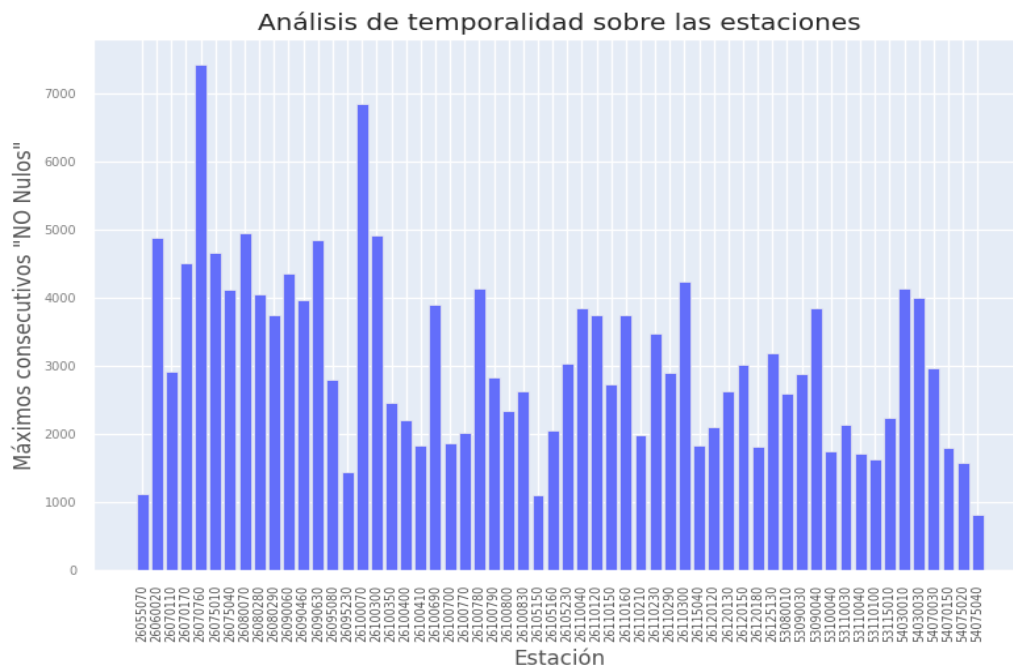


Fig. 59. Mayor cantidad de valores consecutivos no nulos por estaciones

Nota: elaboración propia.

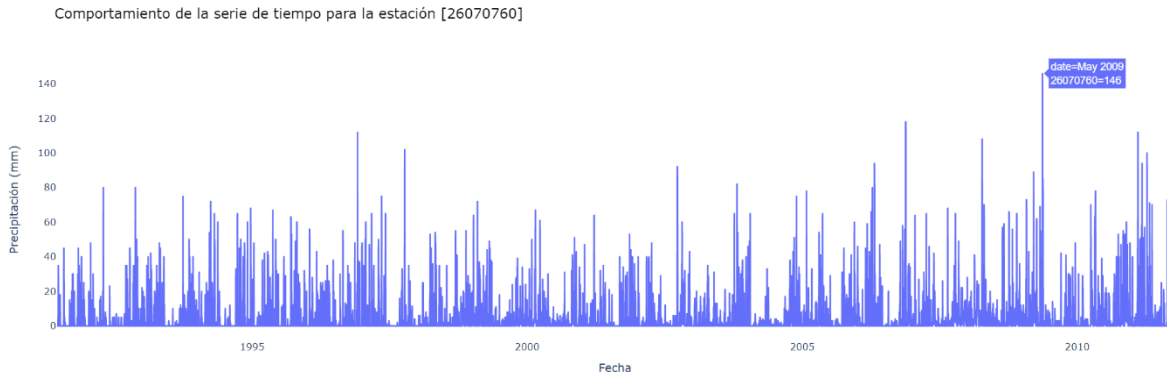


Fig. 60. Comportamiento temporal de la estación con mejor convergencia entre la correlación y mayor cantidad de valores no nulos consecutivos - [26070760]  
Nota: elaboración propia.

Ahora bien, la naturaleza estocástica del modelo ARIMA lo hace adecuado para adaptarse a las variaciones geográficas presentes en el área establecida en el proyecto. Esto es especialmente relevante al combinar datos de estaciones terrestres ubicadas en áreas específicas con información satelital que cubre regiones más extensas.

La Fig. 61 muestra claramente que los datos de la estación 26070760 en cuestión son estacionales, estacionalidad que se analiza a continuación.

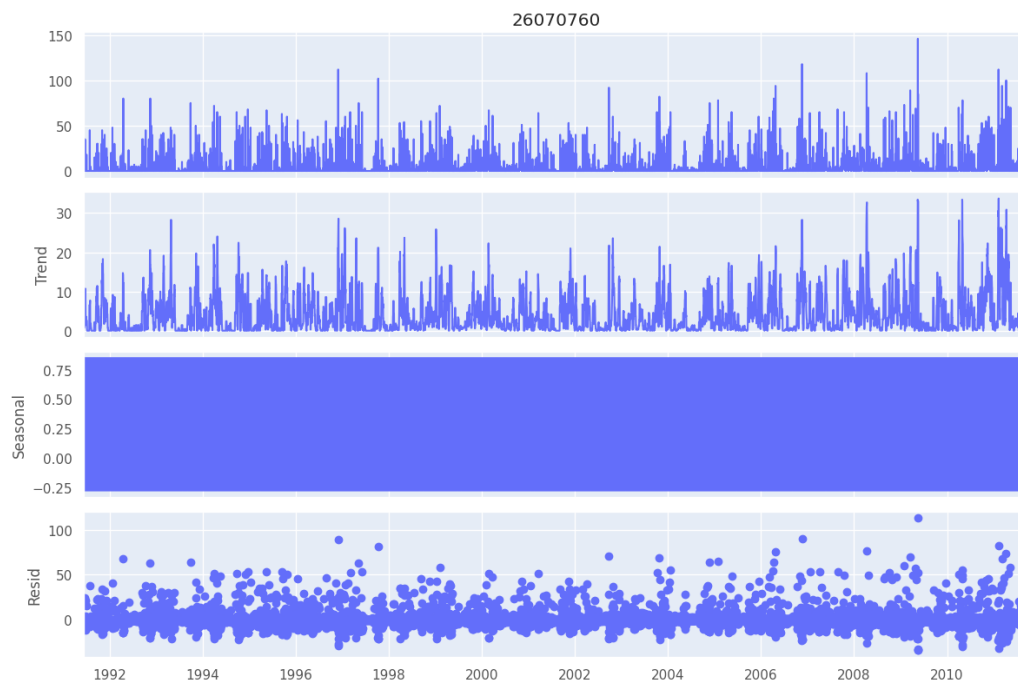


Fig. 61. Descomposición de la serie temporal en sus componentes principales: tendencia, estacionalidad y residuos - [26070760]  
Nota: elaboración propia.

A continuación, se presentan los resultados de prueba estadística Dickey Fuller (ver Fig. 62), buscando determinar de manera teórica la estacionalidad de los datos.

```
Resultados de la prueba de Dickey-Fuller para la estación: [26070760]
Test Statistic          -1.246590e+01
p-value                 3.326200e-23
No Lags Used            2.500000e+01
Número de observaciones utilizadas  7.387000e+03
Critical Value (1%)     -3.431236e+00
Critical Value (5%)     -2.861931e+00
Critical Value (10%)    -2.566978e+00
dtype: float64
Conclusion:====>
Rechazar la hipótesis nula
Los datos son estacionarios
```

Fig. 62. Resultados de la prueba estadística Dickey Fuller - [26070760]

Nota: elaboración propia.

Por el resultado del p-value ( $<0.05$ ) es posible decir que la serie temporal es estacionaria, lo cual se establece al rechazar la hipótesis nula, aceptando la evidencia de que los datos de la serie de tiempo son generados por un proceso estacionario, con lo que se determina qué tan fuertemente se ven afectados por una tendencia.

Para la estimación de los parámetros del modelo ARIMA a aplicar, se utilizó la librería *auto\_arima* del módulo *pmdarima* en Python, lo cual arrojó los siguientes resultados (Fig. 63):

```
Performing stepwise search to minimize aic
ARIMA(0,1,0)(0,1,0)[12] : AIC=65404.561, Time=0.25 sec
ARIMA(1,1,0)(1,1,0)[12] : AIC=61228.306, Time=1.65 sec
ARIMA(0,1,1)(0,1,1)[12] : AIC=inf, Time=11.69 sec
ARIMA(1,1,0)(0,1,0)[12] : AIC=63373.346, Time=0.64 sec
ARIMA(1,1,0)(2,1,0)[12] : AIC=60408.050, Time=6.26 sec
ARIMA(1,1,0)(2,1,1)[12] : AIC=inf, Time=67.93 sec
ARIMA(1,1,0)(1,1,1)[12] : AIC=inf, Time=25.84 sec
ARIMA(0,1,0)(2,1,0)[12] : AIC=62481.172, Time=5.20 sec
ARIMA(2,1,0)(2,1,0)[12] : AIC=59731.568, Time=8.24 sec
ARIMA(2,1,0)(1,1,0)[12] : AIC=60553.299, Time=5.79 sec
ARIMA(2,1,0)(2,1,1)[12] : AIC=inf, Time=88.95 sec
ARIMA(2,1,0)(1,1,1)[12] : AIC=inf, Time=39.74 sec
ARIMA(3,1,0)(2,1,0)[12] : AIC=59304.733, Time=9.59 sec
ARIMA(3,1,0)(1,1,0)[12] : AIC=60120.095, Time=4.79 sec
ARIMA(3,1,0)(2,1,1)[12] : AIC=inf, Time=123.32 sec
ARIMA(3,1,0)(1,1,1)[12] : AIC=inf, Time=47.01 sec
ARIMA(4,1,0)(2,1,0)[12] : AIC=59076.071, Time=16.07 sec
ARIMA(4,1,0)(1,1,0)[12] : AIC=59877.820, Time=6.05 sec
ARIMA(4,1,0)(2,1,1)[12] : AIC=inf, Time=551.07 sec
ARIMA(4,1,0)(1,1,1)[12] : AIC=inf, Time=56.05 sec
ARIMA(4,1,1)(2,1,0)[12] : AIC=inf, Time=438.00 sec
ARIMA(3,1,1)(2,1,0)[12] : AIC=inf, Time=94.38 sec
ARIMA(4,1,0)(2,1,0)[12] intercept : AIC=59078.071, Time=257.36 sec

Best model: ARIMA(4,1,0)(2,1,0)[12]
Total fit time: 1865.890 seconds
ARIMA(4,1,0)(2,1,0)[12]
```

Fig. 63. Estimación de los parámetros (p, d, q) del modelo ARIMA a aplicar - [26070760]

Nota: elaboración propia.

Por lo anterior, se establecen los siguientes parámetros óptimos a utilizar:

**p: 4 - d: 1 - q: 0 - P: 2 - D: 1 - Q: 0**

Componente estacional con un periodo de 12 unidades de tiempo, indicando que, al tener fragmentados los datos de manera diaria, se indica al modelo que debe considerar una estacionalidad mensual.

En este momento, se crea y entrena el modelo utilizando la librería SARIMA. A continuación, se presenta el resumen del modelo en la figura siguiente (Fig. 64).

```

SARIMAX Results
Dep. Variable: 26070760          No. Observations: 7413
Model: SARIMAX(4, 1, 0)x(2, 1, 0, 12)  Log Likelihood -29531.036
Date: Wed, 20 Mar 2024          AIC 59076.071
Time: 02:54:43                 BIC 59124.436
Sample: 06-15-1991             HQIC 59092.691
- 09-30-2011

Covariance Type: opg
      coef  std err  z  P>|z|  [0.025  0.975]
ar.L1 -0.7527  0.008  -95.737  0.000  -0.768  -0.737
ar.L2 -0.5266  0.010  -54.585  0.000  -0.545  -0.508
ar.L3 -0.3620  0.010  -37.635  0.000  -0.381  -0.343
ar.L4 -0.1752  0.008  -21.739  0.000  -0.191  -0.159
ar.S.L12 -0.6689  0.007  -97.845  0.000  -0.682  -0.656
ar.S.L24 -0.3218  0.007  -47.439  0.000  -0.335  -0.309
sigma2 171.1508  1.082  158.227  0.000  169.031  173.271
Ljung-Box (L1) (Q): 3.69 Jarque-Bera (JB): 46620.97
Prob(Q): 0.05 Prob(JB): 0.00
Heteroskedasticity (H): 1.38 Skew: 1.83
Prob(H) (two-sided): 0.00 Kurtosis: 14.74

```

Fig. 64. Summary del modelo SARIMA implementado - [26070760]

Nota: elaboración propia.

En las siguientes figuras (Fig. 65, Fig. 66 y Fig. 67) se presentan las métricas de aplicar ARIMA, los errores residuales y las predicciones obtenidas.

Evaluación de los resultados del modelos SARIMA:

```

MSE: 239.11946340093002
MAE: 9.322276303777588
R2: 0

```

Fig. 65. Métricas del desempeño del modelo SARIMA - [26070760]

Nota: elaboración propia.

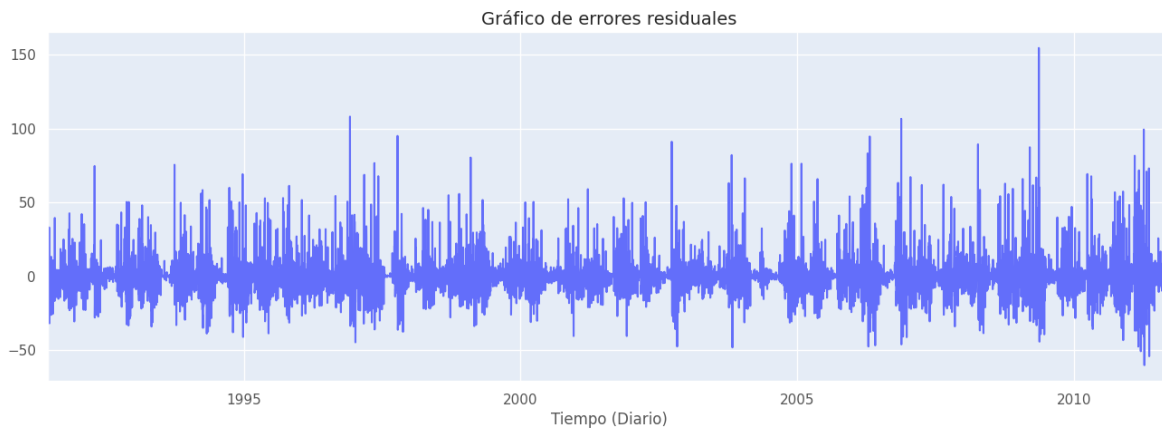


Fig. 66. Gráfico de los errores residuales de las predicciones del modelo SARIMA - [26070760]  
Nota: elaboración propia.

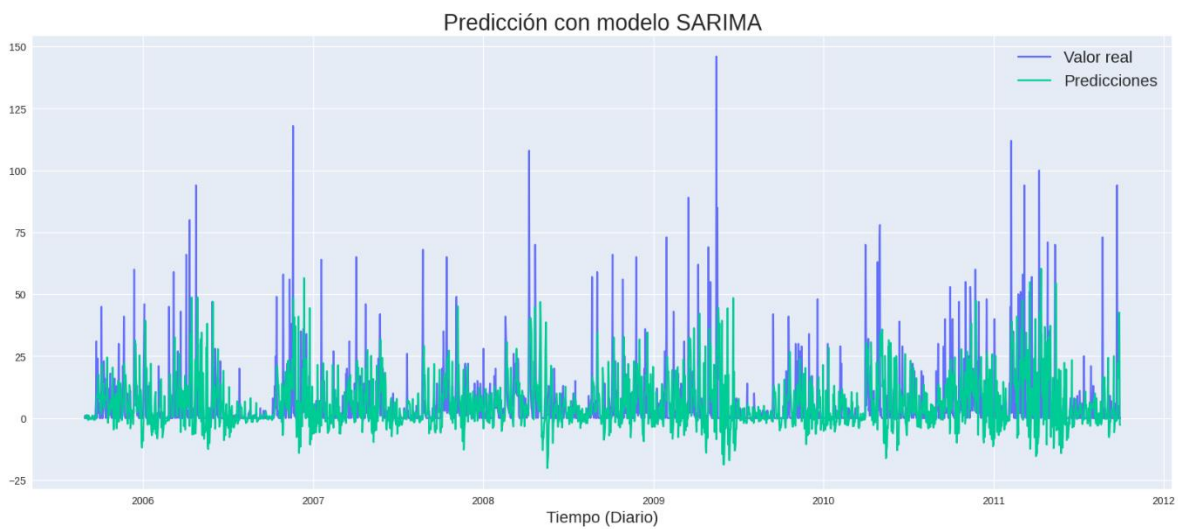


Fig. 67. Resultados de las predicciones del modelo SARIMA - [26070760]  
Nota: elaboración propia.

En ese orden de ideas, se procedió a aplicar el mismo modelo en todas las estaciones, obteniendo los resultados que se detallan en el ANEXO C (métricas modelo ARIMA, que incluyen  $R^2$ , MSE y MAE). Adicionalmente, se generaron las siguientes gráficas de resumen (Fig. 65).

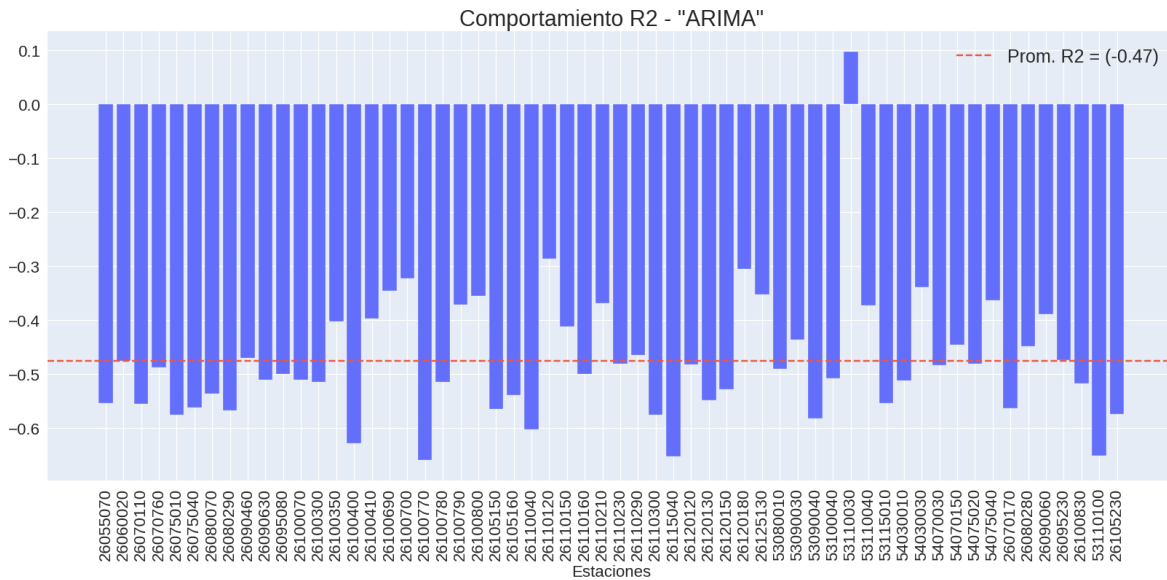


Fig. 72. Resultados R2 predicciones diaria – ARIMA  
Nota: elaboración propia.

### 6.2.1.2 Prophet

A continuación, se presentan las predicciones / pronósticos obtenidos con el modelo Prophet para valores diarios y mensuales de lluvia, en cada caso se representan (y distinguen por color) los datos usados para entrenamiento (*train* – azul), los valores de prueba (*test* – rojo) y las predicciones (pred - verde) del modelo. La Fig. 68 presenta las predicciones diarias, la Fig. 69 el pronóstico semanal, la Fig. 70 las mensuales y la Fig. 71 la predicción anual para la Estación 26070760 (la que menos datos faltantes tiene en el *dataset* original de las estaciones en tierra).

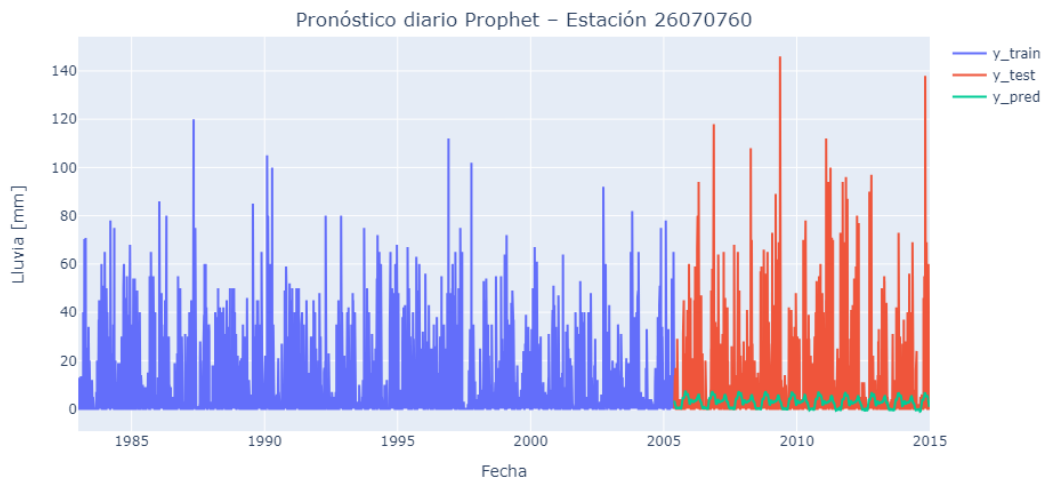


Fig. 68. Resultado predicción / pronóstico diario Prophet – Estación 26070760  
Nota: elaboración propia.

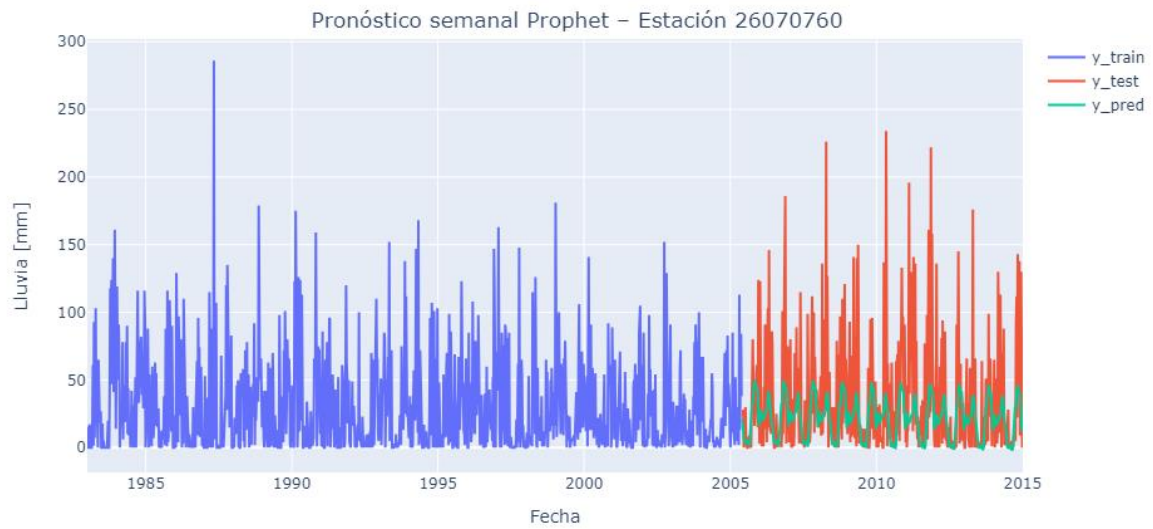


Fig. 69. Resultado predicción / pronóstico semanal Prophet – Estación 26070760  
Nota: elaboración propia.

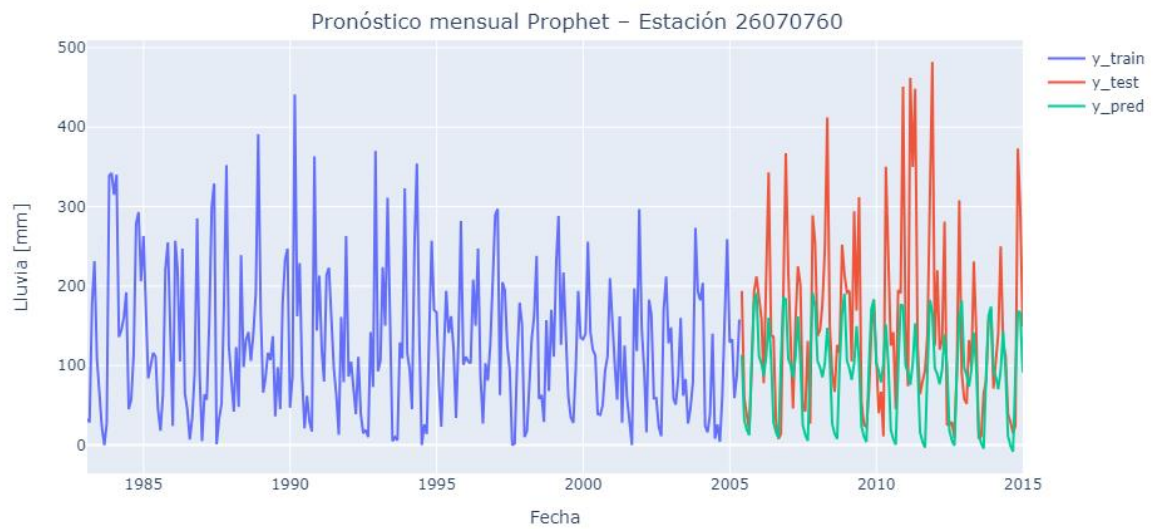


Fig. 70. Resultado predicción / pronóstico mensual Prophet – Estación 26070760  
Nota: elaboración propia.

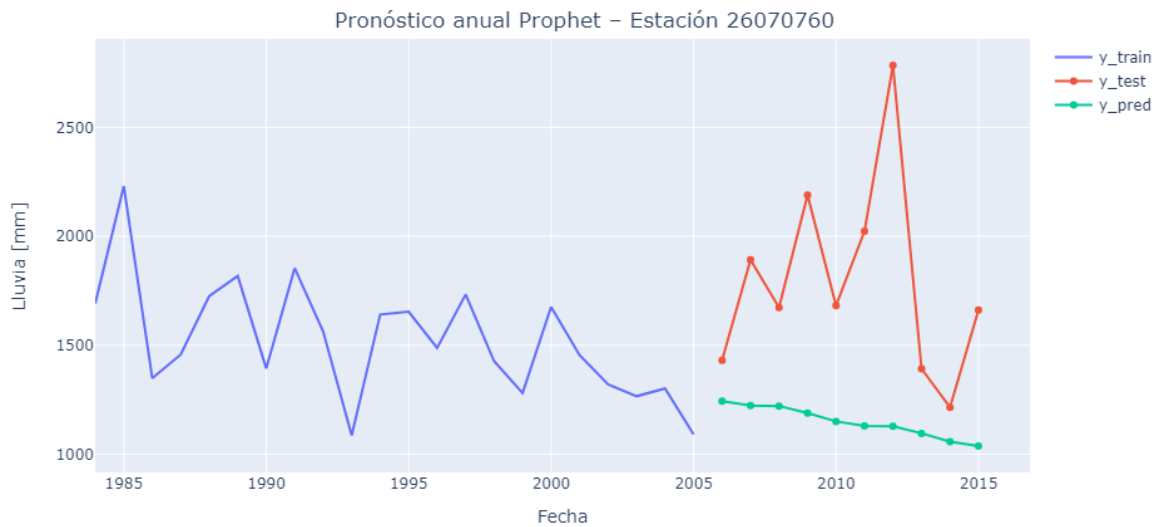


Fig. 71. Resultado predicción / pronóstico anual Prophet – Estación 26070760  
Nota: elaboración propia.

Se puede apreciar, especialmente en las predicciones mensuales, que el modelo predice bien las tendencias, aunque no capta o pronóstica bien los valores / picos extremos en cada uno de los años. Igual es importante reconocer que el modelo genera un patrón similar para cada uno de los casi 10 años pronosticados (0.3 o 30% son seleccionados para la prueba). Es evidente también que, aunque la tendencia del pronóstico mejora para los datos semanales respecto a los diarios y a su vez las predicciones de lluvia mensual respecto a la semanal, no pasa lo mismo para las predicciones anuales de lluvia que es claro que difieren significativamente tanto en la tendencia de la Fig. 71 y los valores obtenidos.

Con lo anterior en mente, se procede en las figuras siguientes a calcular las métricas para todas las estaciones / puntos analizados en el presente estudio. A continuación, se presenta los valores de la métrica  $R^2$  para cada uno de los puntos, en la Fig. 72 se presentan para los resultados diarios, la Fig. 73 presenta los resultados semanales y en la Fig. 74 para los mensuales.

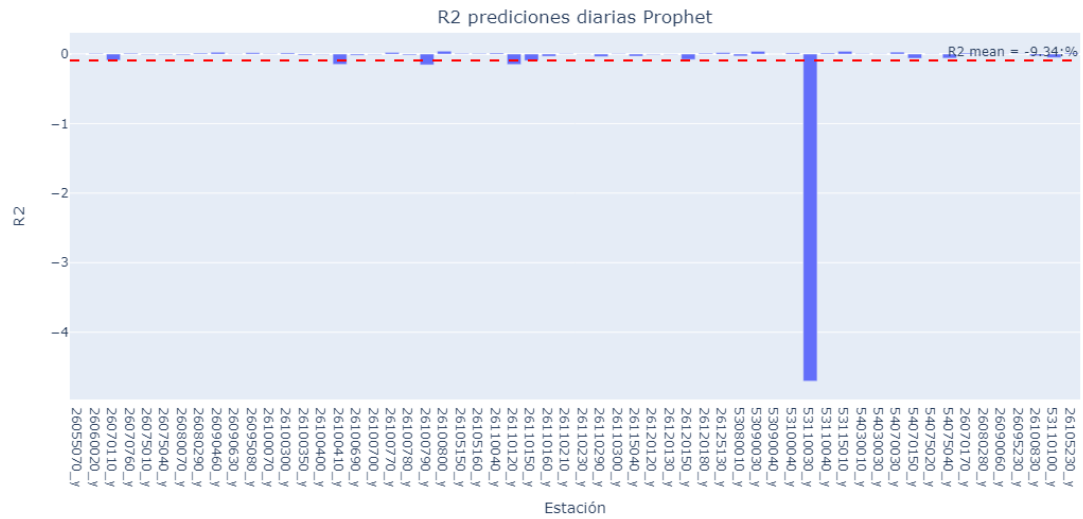


Fig. 72. Resultados R2 predicciones diarias Prophet  
Nota: elaboración propia.

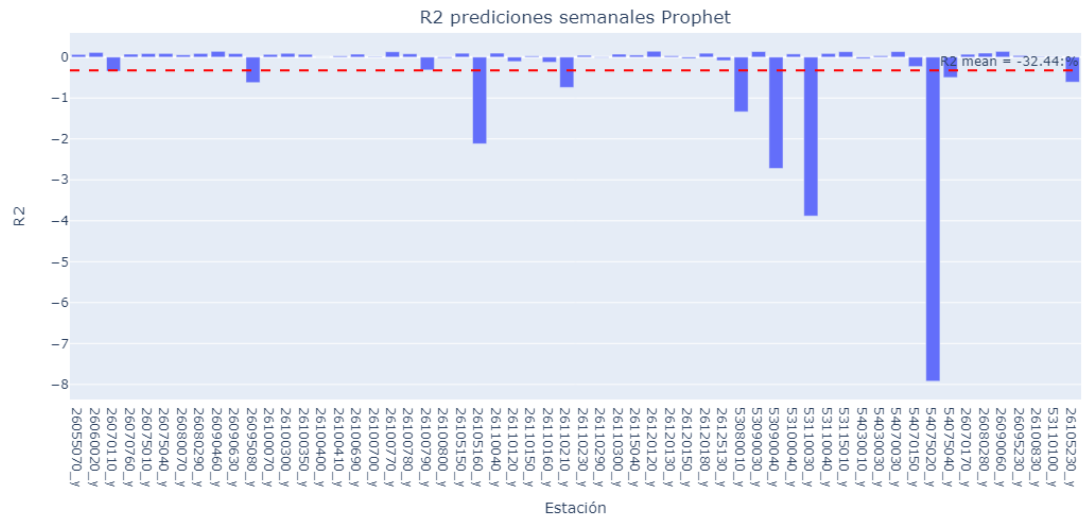


Fig. 73. Resultados R2 predicciones semanales Prophet  
Nota: elaboración propia.

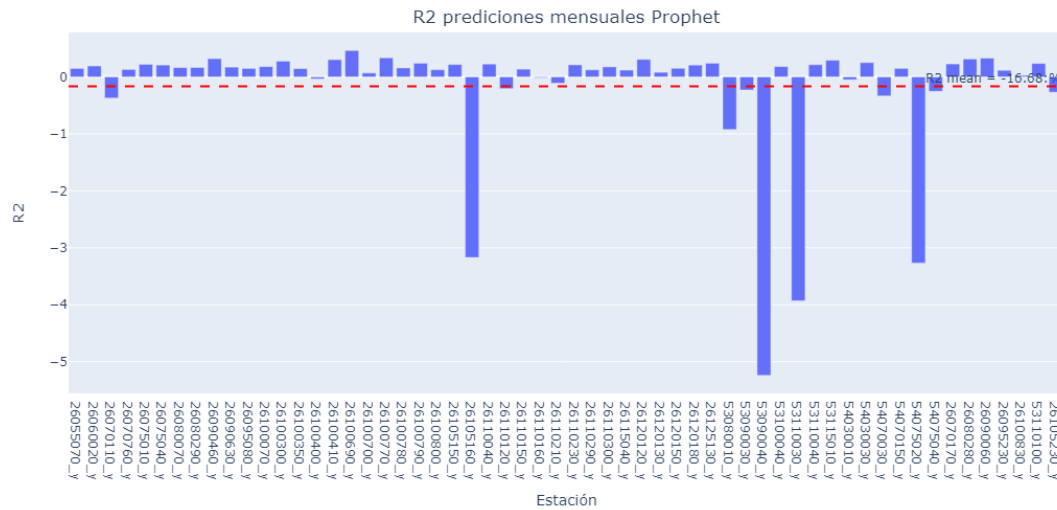


Fig. 74. Resultados R2 predicciones mensuales Prophet  
Nota: elaboración propia.

Como se puede apreciar de las tres figuras anteriores, los resultados de  $R^2$  usando Prophet para pronosticar las diferentes agregaciones de las series de tiempo no son buenas, siendo notorio una caída en del desempeño al pasar de la agregación diaria a la semanal y luego una leve mejoría al agregar mensualmente los datos.

### 6.2.1.3 Sktime

En la gráfica siguiente se presentan la métrica  $R^2$  para las predicciones mensuales obtenidas con los pronósticos con el *naiveforecaster* de Sktime. Es importante, tener presente que con el modelo en cuestión no fue posible generar predicciones diarias y semanales.

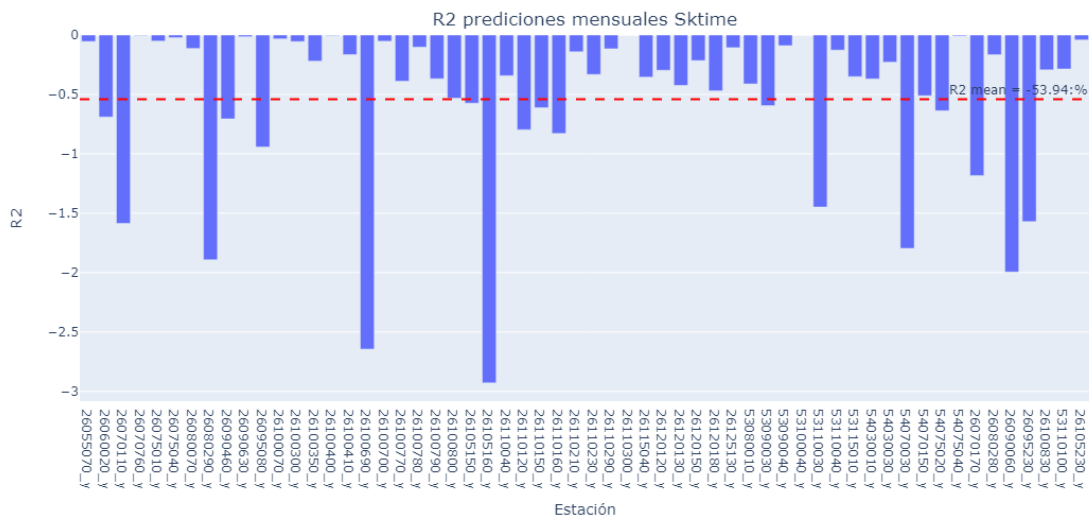


Fig. 75. Resultados R2 predicciones mensuales Sktime  
Nota: elaboración propia.

Como se puede apreciar en la Fig. 75, los resultados de  $R^2$  para los pronósticos mensuales no son buenos, en especial para siete estaciones que presentan métricas muy por debajo de las otras.

### 6.3 COMPARACIÓN DE RESULTADOS

En el presente numeral se realiza la comparación de los resultados obtenidos con las diferentes agregaciones y modelos utilizados en el estudio.

#### 6.3.1 Modelos de regresión

En las siguientes figuras (Fig. 76, Fig. 77 y Fig. 78) se comparan los resultados de  $R^2$  para cada uno de los modelos de regresión lineal utilizados en el estudio para las diferentes agregaciones temporales evaluadas (lluvia acumulada diaria, semanal y mensual).

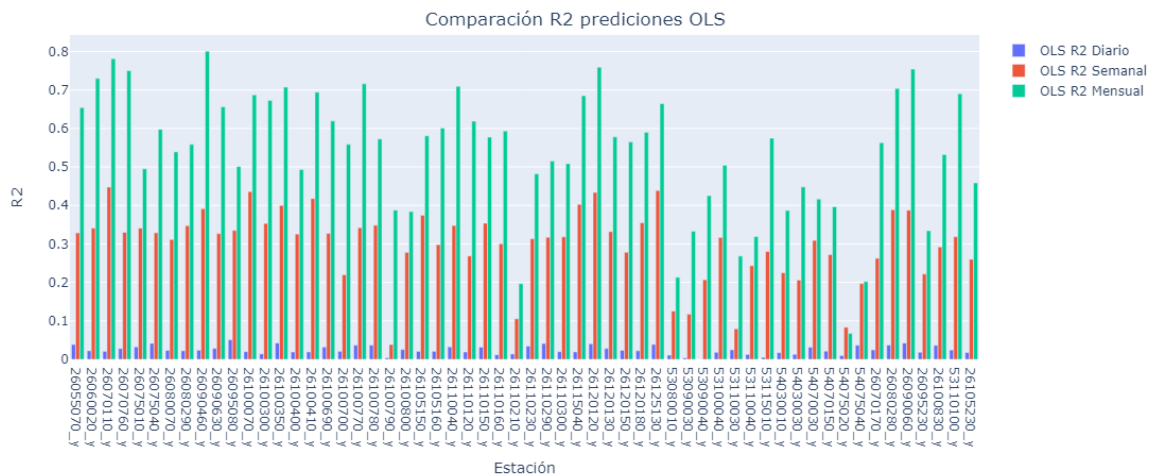


Fig. 76. Comparación resultados R2 OLS / regresión lineal  
Nota: elaboración propia.

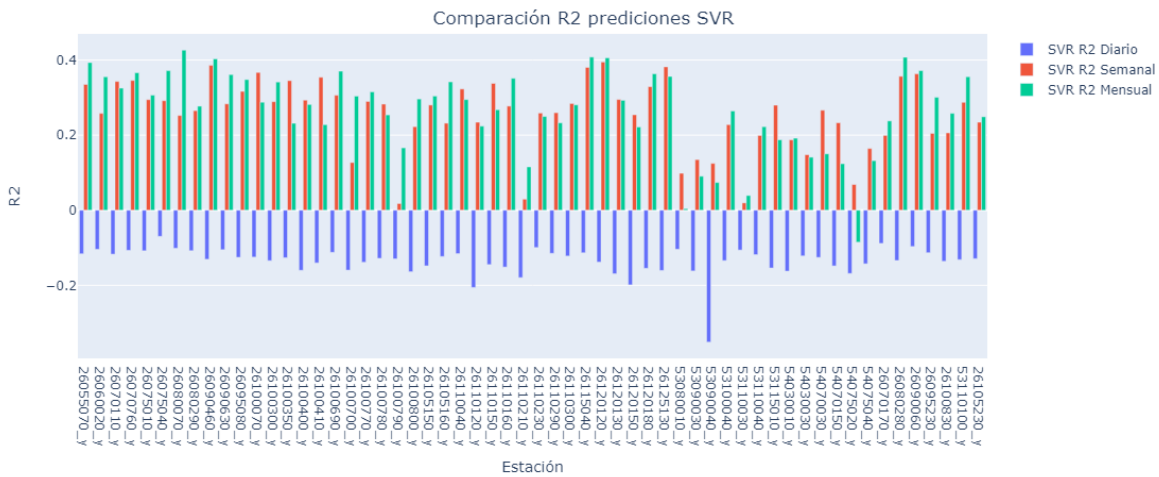


Fig. 77. Comparación resultados R2 SVR / máquinas de soporte vectorial  
Nota: elaboración propia.

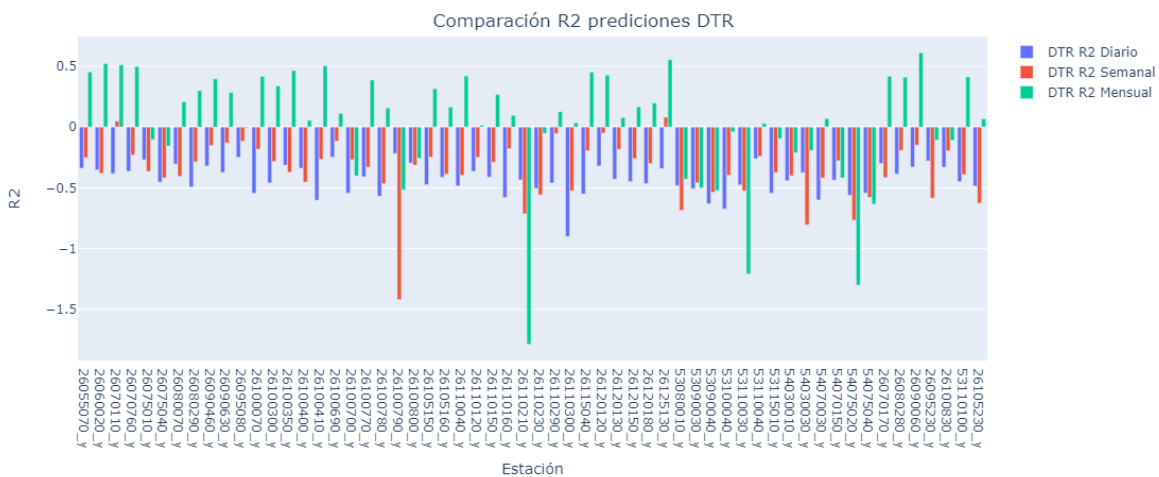


Fig. 78. Comparación resultados R2 DTR / árboles de decisión  
Nota: elaboración propia.

En los tres casos anteriores, es evidente que los mejores resultados se presentan con las agregaciones mensuales, respecto a las agregaciones semanales y diarias. Aunque se puede deducir por los valores obtenidos en las gráficas, en general de los tres modelos usados (regresión lineal, máquinas de soporte vectorial y árboles de decisión) los de regresión lineal presentan las mejores métricas en cada agregación, tal como se puede apreciar en la figura siguiente (Fig. 79). Para mayor claridad en el ANEXO D se presentan las gráficas de comparación de los  $R^2$  de los modelos de regresión para cada agregación por separado (diario, semanal y mensual). En dicho anexo, se observa fácilmente en las figuras que la regresión lineal / OLS presenta las mejores métricas y los árboles de decisión / DTR las más bajas, solamente en las predicciones semanales, el desempeño de OLS son muy similares a los de las máquinas de soporte vectorial / SVR.

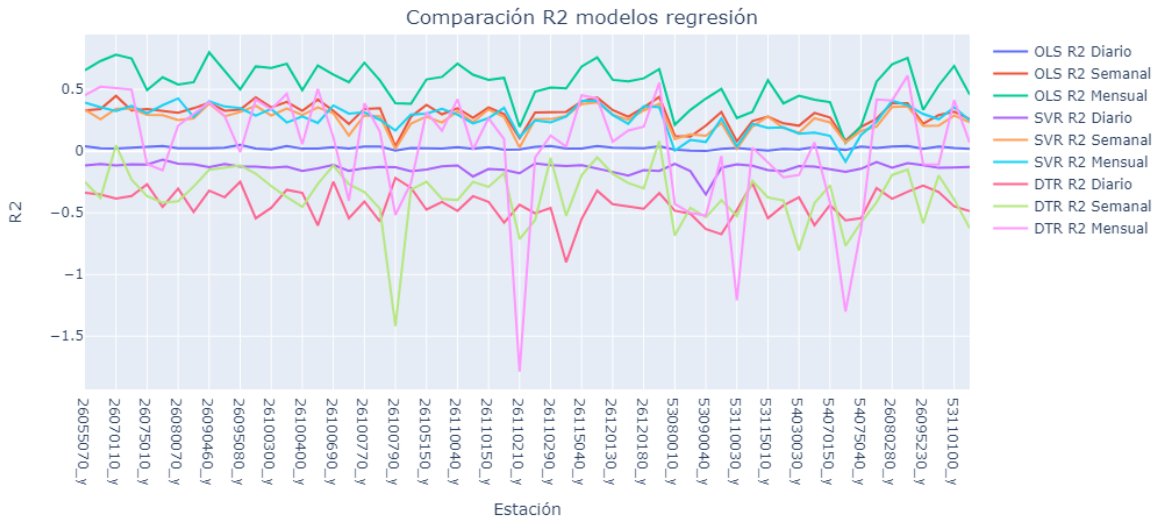


Fig. 79. Comparación resultados R2 modelos de regresión  
Nota: elaboración propia.

En la Fig. 80, se puede observar el mejor coeficiente de determinación ( $R^2$ ) promedio obtenido, medido sobre el  $R^2$  de la predicción en cada una de las estaciones, se da claramente en el modelo de regresión lineal para una agrupación mensual, alcanzando un valor máximo del 50%. Esto es significativamente mejor en términos de precisión al predecir el valor de lluvia mensual en comparación con las predicciones diarias. Este resultado sugiere que los modelos utilizados son más eficaces cuando se utilizan datos agregados mensualmente, lo que puede deberse a una reducción en la variabilidad y ruido presentes en los datos diarios, permitiendo así una mejor captura de las tendencias subyacentes.

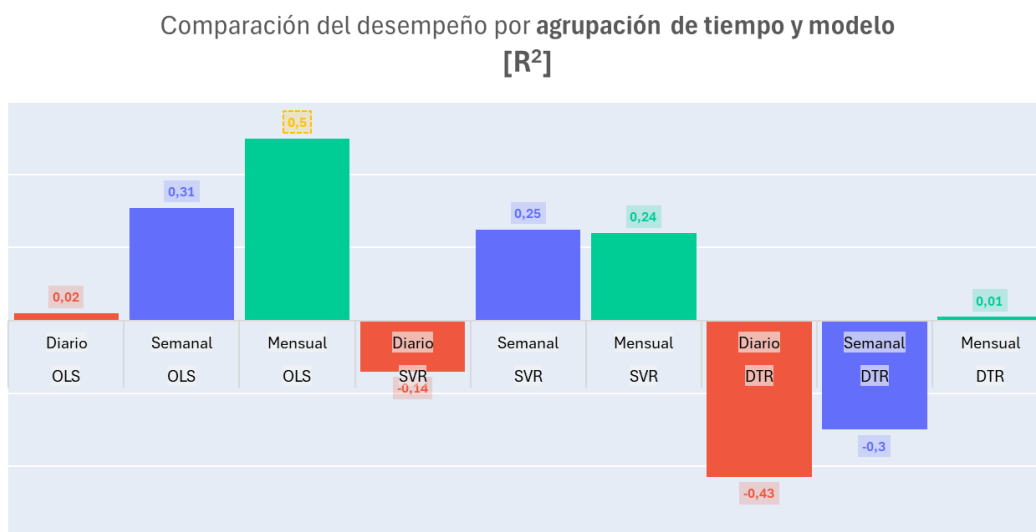


Fig. 80. Comparación de resultados R<sup>2</sup> por agrupación de tiempo y modelo  
Nota: elaboración propia.

En las figuras siguientes (Fig. 81, Fig. 82 y Fig. 83) se presentan la comparación del desempeño por agregación temporal y modelo de las métricas error absoluto medio (MAE), error cuadrático medio (MSE) y error máximo residual (MaxError) para los modelos de regresión implementados.

Comparación del desempeño por **agrupación de tiempo y modelo**  
[MAE]

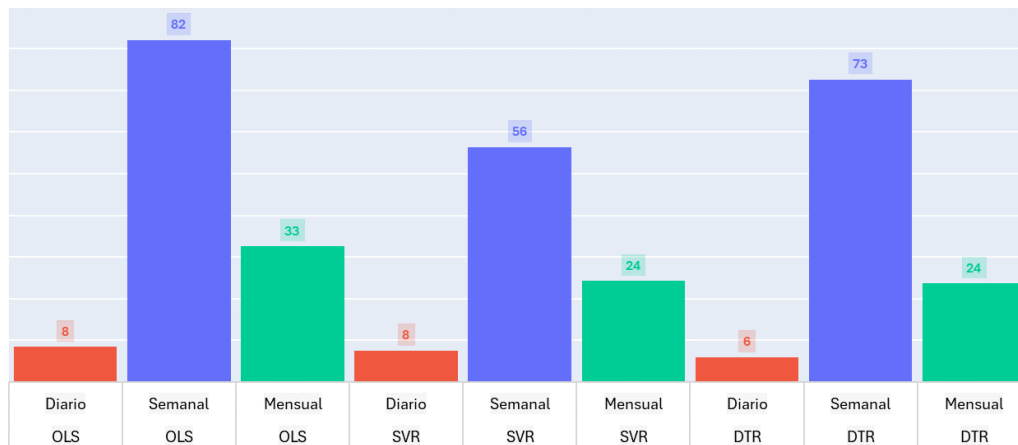


Fig. 81. Comparación de resultados MAE por agrupación de tiempo y modelo  
Nota: elaboración propia.

Los altos valores obtenidos para el MAE en las combinaciones de modelo-agrupación de tiempo, sugieren implicaciones en el desempeño de estos en cuanto a la capacidad de hacer predicciones precisas, lo que se interpreta como que, en promedio, las predicciones están lejos de los valores reales. Esto significa que los modelos no están capturando bien la relación subyacente entre la variable independiente y la variable dependiente.

Comparación del desempeño por **agrupación de tiempo y modelo**  
[MSE]

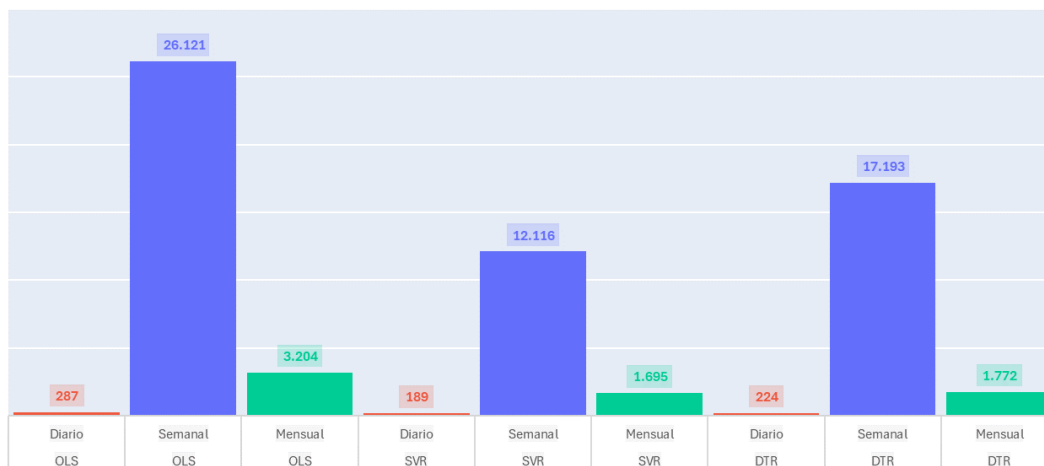


Fig. 82. Comparación de resultados MSE por agrupación de tiempo y modelo  
Nota: elaboración propia.

Como es conocido, al elevar al cuadrado las diferencias entre los valores reales y los valores predichos, en el MSE, se penalizan de mayor manera los errores grandes con relación al MAE. Sin embargo, es claro que un alto MSE indica que el modelo (sea cual quiera que se seleccione) tiene, en promedio, grandes errores de predicción. Lo que lleva a la posibilidad de la existencia de *outliers* que los modelos no logran captar bien.

Comparación del desempeño por agrupación de tiempo y modelo  
[MaxError]

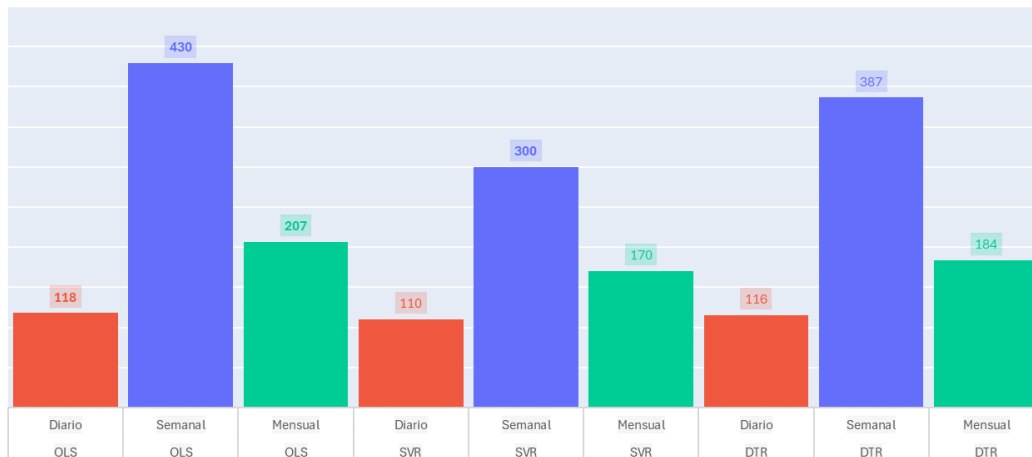


Fig. 83. Comparación de resultados error máximo por agrupación de tiempo y modelo  
Nota: elaboración propia.

Estos errores máximos obtenidos dan pie a entender, al menos en una primera instancia, el bajo desempeño de los indicadores analizados anteriormente. Es claro que en todos los casos hay al menos una observación para la que el modelo falla considerablemente en la predicción. Esto tiene implicaciones directas en los promedios de los indicadores aquí representados. Nuevamente, los casos atípicos son una opción de especial análisis, pero con la idea de analizarlos para determinar si son genuinamente atípicos, lo que podría indicar la posibilidad de tener datos mal etiquetados.

### 6.3.2 Modelos series de tiempo

Respecto a ARIMA, en el ANEXO B (Resultados  $R^2$  de las predicciones del modelo ARIMA para todas las estaciones), es posible observar el comportamiento de la predicción efectuada para cada una de las estaciones, tal como se explicó en el apartado 6.2.1.1. En las gráficas, es evidente la gran distorsión entre el valor real (azul) y el valor de la predicción (verde). Además, se puede notar que para ningún día en ninguna estación se logra llegar al valor real, lo que explica implícitamente los resultados del  $R^2$  de tal magnitud.

En la Fig. 84 se presentan se comparan los resultados de  $R^2$  obtenidos por medio de los pronósticos del modelo de series de tiempo Prophet para las diferentes agregaciones temporales evaluadas

(lluvia acumulada diaria, semanal y mensual). Como se puede apreciar en la figura de la página siguiente y ya se había establecido en el numeral 6.2.1.2, los resultados de Prophet a la hora de pronosticar las diferentes agregaciones de las series de tiempo no son buenos y se descartaría a la hora de pronosticar valores de precipitación.

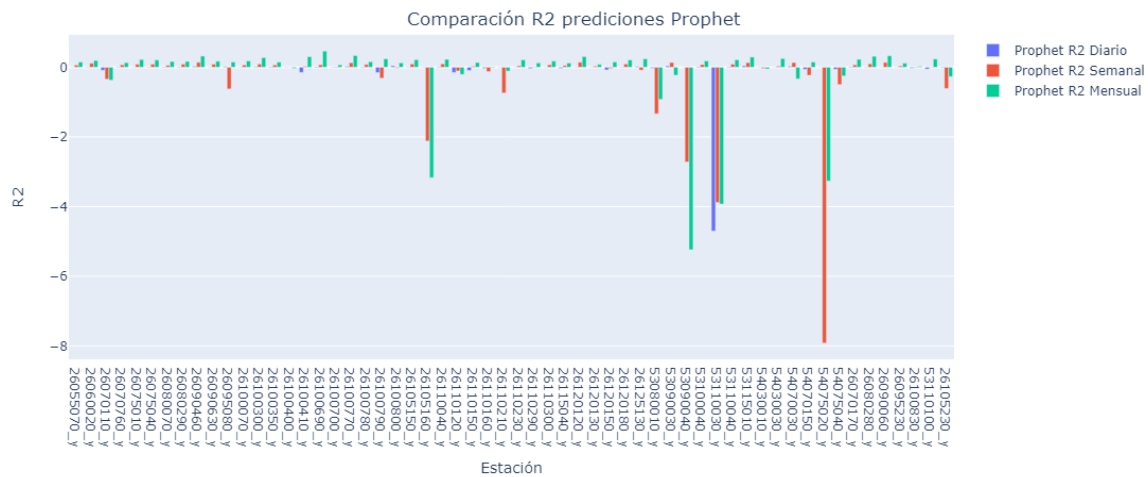


Fig. 84. Comparación resultados R2 Prophet / forecasting series de tiempo  
Nota: elaboración propia.

### 6.3.3 Modelos de clasificación

En las siguientes figuras (Fig. 85, Fig. 86 y Fig. 87) se comparan los resultados de exactitud (*accuracy*) para cada uno de los modelos de clasificación utilizados (*logit*, *SVC* y *DTC*) en el estudio para las diferentes agregaciones temporales evaluadas al clasificar la lluvia de manera binaria (existencia o no de lluvia diaria, semanal y mensual).

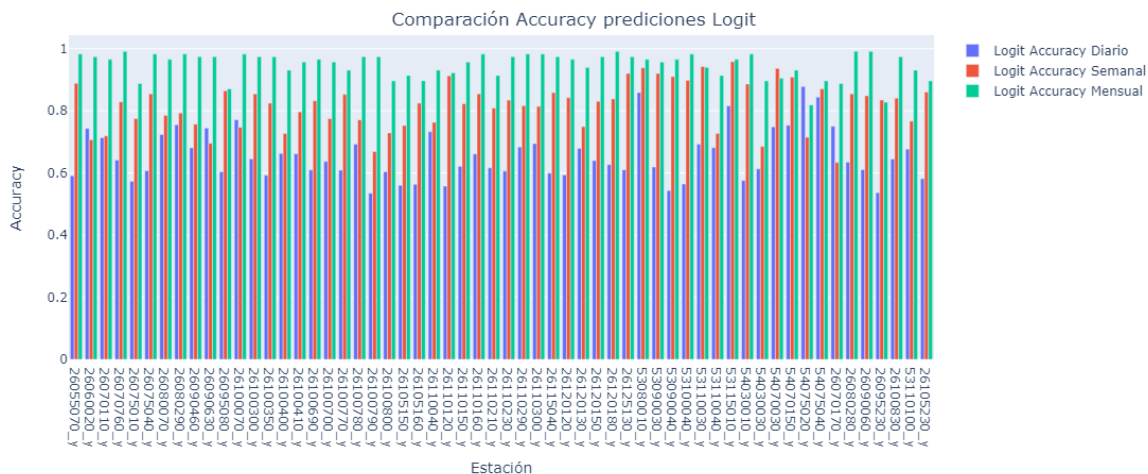


Fig. 85. Comparación resultados Accuracy Logit / regresión logística  
Nota: elaboración propia.

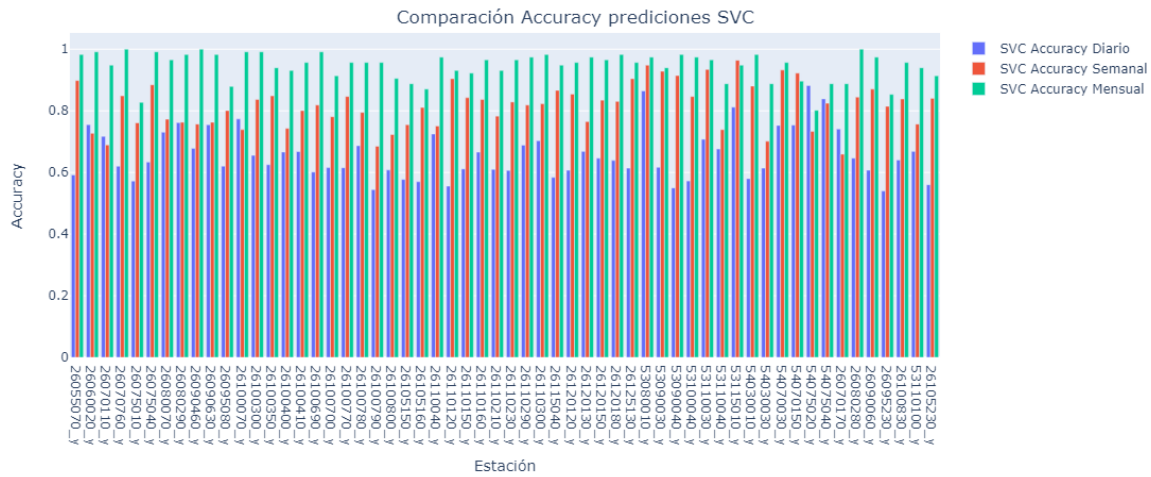


Fig. 86. Comparación resultados Accuracy SVC / Máquinas de soporte vectorial  
Nota: elaboración propia.

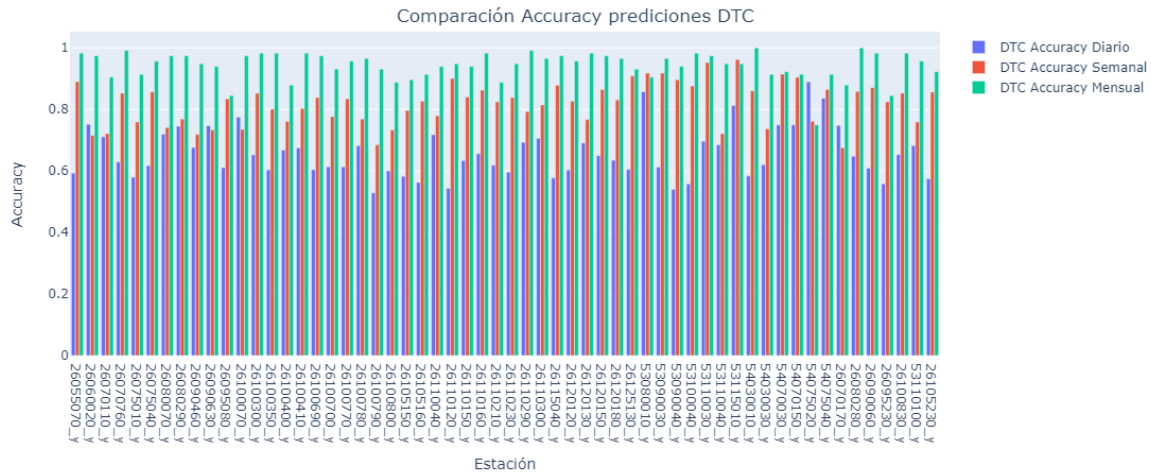


Fig. 87. Comparación resultados Accuracy DTC / árboles de decisión  
Nota: elaboración propia.

En las tres gráficas anteriores, es claro que el comportamiento y rango de valores para los tres diferentes modelos de clasificación son similares y que en general las agregaciones mensuales presentan los valores más alto de exactitud. A continuación, en la Fig. 88 se evidencia la similitud del *accuracy* / exactitud para los modelos *logit*, *SVC* y *DTC*. En el ANEXO D se pueden consultar las gráficas de comparación de los *accuracy* de los modelos de clasificación para cada agregación temporal por separado.

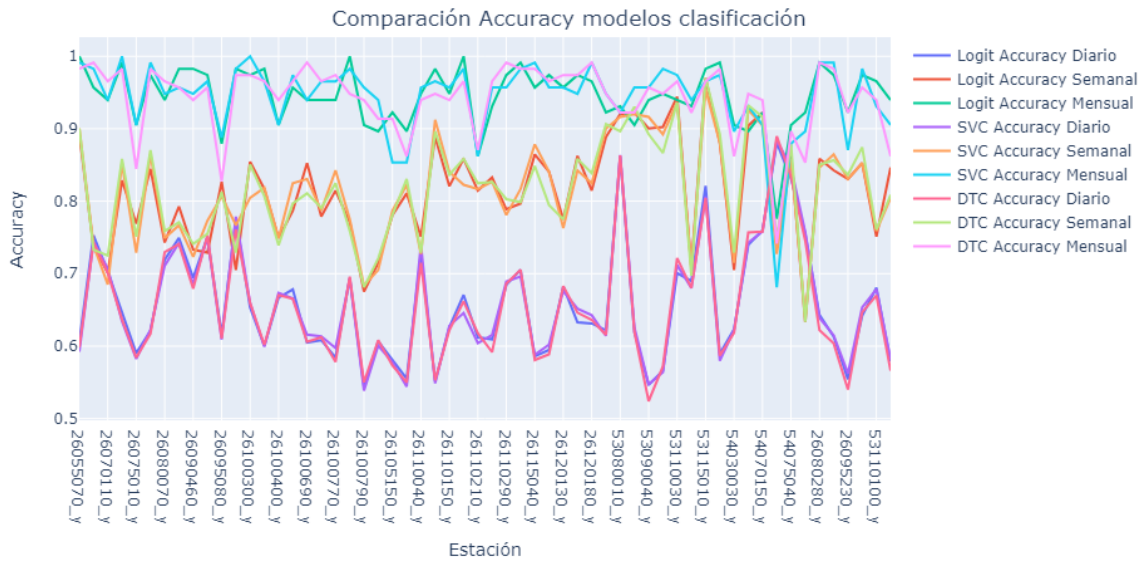


Fig. 88. Comparación resultados Accuracy modelos de clasificación  
Nota: elaboración propia.

Finalmente, en la Fig. 89 se evidencia una alta consistencia de las diferentes métricas de clasificación evaluadas al aplicar los diferentes modelos y las agrupaciones realizadas, donde destaca el comportamiento de la agrupación mensual por encima de las otras. El alto *accuracy* muestra que predicen correctamente en la mayoría de los casos. El *recall* perfecto indica que no se pierde ninguna instancia positiva, lo cual es crucial en aplicaciones sensibles. La alta precisión asegura que la mayoría de las predicciones positivas son correctas, minimizando los falsos positivos.

### Modelos de Clasificación

"Comparación del desempeño por agrupación de tiempo y modelo"

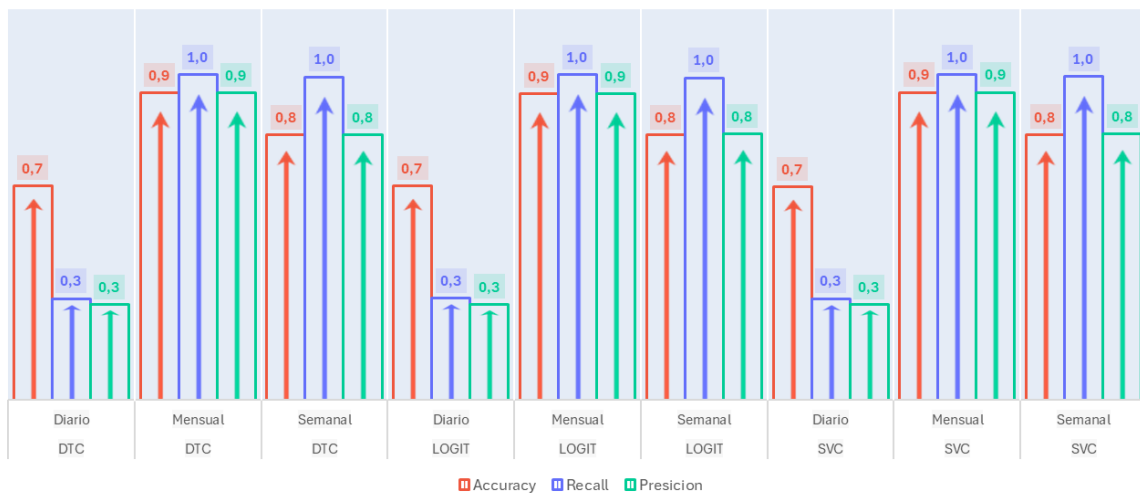


Fig. 89. Comparación de resultados por agrupación de tiempo, modelo y métricas de clasificación  
Nota: elaboración propia.

En resumen, los modelos de clasificación indican un desempeño robusto y bien ajustados al problema, tanto de manera mensual como semanal, con la capacidad de identificar todas las instancias positivas y mantener un alto grado de precisión en las predicciones, lo que a su vez se interpreta como la alta capacidad de predecir aquellas semanas o meses en los que habrá presencia de lluvia.

### 6.3.4 Otros modelos

En el proyecto se probaron otros modelos / estimadores, más allá de los mencionados anteriormente, que, aunque no dieron métricas aceptables (p.ej.  $R^2$ ) se resumen a continuación por el interés investigativo del proceso de prueba y error de diferentes algoritmos para la reconstrucción de series climáticas como la precipitación.

#### 6.3.4.1 Simple Imputing

A continuación, en la Fig. 90 se presentan los resultados de  $R^2$  al aplicar el estimador *SimpleImputer* de *Scikit-Learn* a las agregaciones diarias, semanales y mensuales de lluvia. Como se puede observar en dicha gráfica, los valores de la métrica no son buenos, pero se mantiene el hallazgo que en general las agregaciones mensuales presentan valores más altos que la métrica, respecto a los datos semanales y estos a su vez respecto a los diarios. En el estudio se aplicaron otros estimadores de *imputing* de *scikit-learn* como *IterativeImputer* que aplica un análisis multivariado y tiene en cuenta otras características y *KNNImputer* que aplica el método de k vecinos más cercanos, pero se obtuvieron prácticamente los mismos resultados que con *SimpleImputer*.

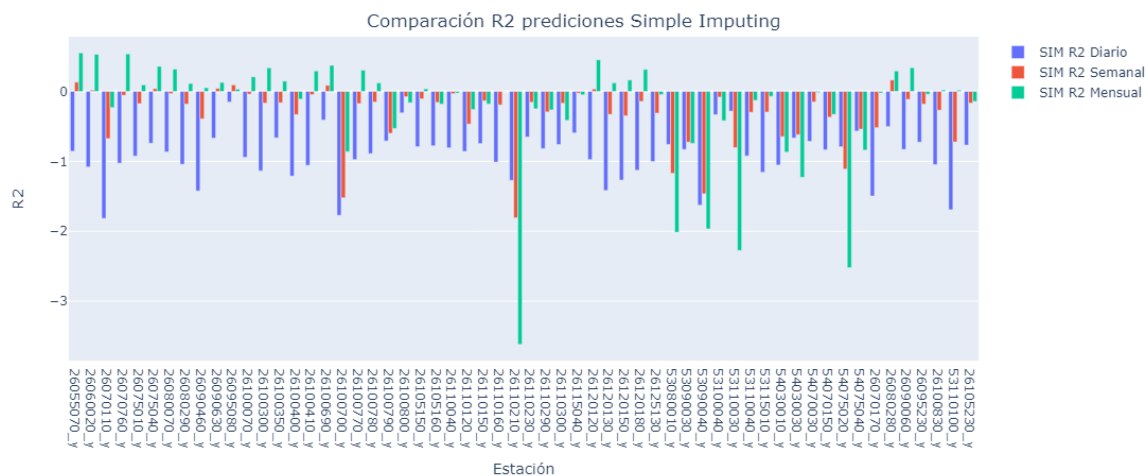


Fig. 90. Comparación resultados R2 simple imputing

Nota: elaboración propia.

### 6.3.4.2 Redes Neuronales

Una vez compilado el modelo de redes neuronales, se procedió a entrenarlo. Posteriormente, se graficaron los resultados del comportamiento de la métrica de evaluación definida previamente, el  $R^2$ , en cada una de las 20 épocas definidas para dicho proceso, tal como se observa en la figura Fig. 91 en la página siguiente.

Como en los análisis y modelos anteriores, la estación 26070760 se seleccionó para probar la red neuronal. Sin embargo, los resultados no fueron los mejores, ya que los valores obtenidos son incluso peores que los arrojados por los otros modelos de aprendizaje automático utilizados en el proyecto, con un comportamiento completamente desfasado de lo esperado en ambos conjuntos, tanto en el de entrenamiento como en el de pruebas. Eso sí, hubo una leve mejoría en términos de estabilidad, aunque no en la magnitud del número, para este segundo.

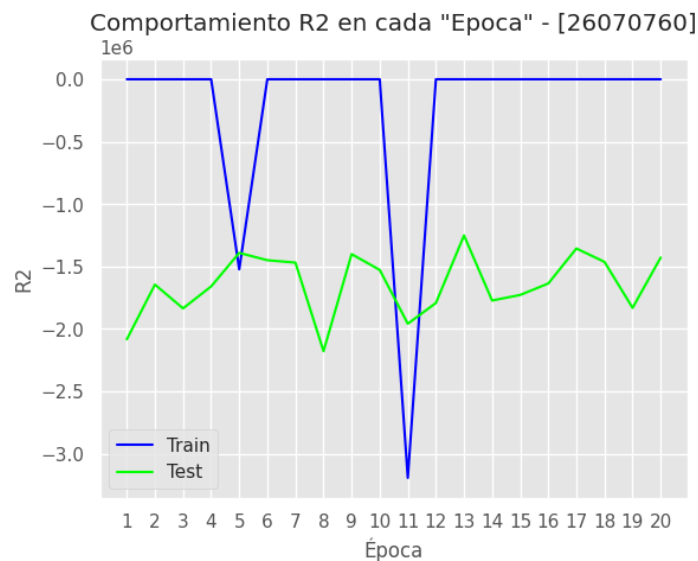


Fig. 91. Comportamiento de  $R^2$  para cada una de las épocas de la red neuronal en los conjuntos de entrenamiento y pruebas

Nota: elaboración propia.

## 7. CONCLUSIONES Y OBSERVACIONES

### 7.1 CONCLUSIONES

Se seleccionaron, procesaron y transformaron de manera exitosa los datos de precipitación de 58 estaciones climáticas en tierra del IDEAM y los datos satelitales CHIRPS asociados a cada punto en la zona de estudio, los cuales son la base del análisis exploratorio y la entrada para todos los modelos utilizados en el proyecto.

Luego de realizar el análisis exploratorio / EDA de los 32 años de datos comprendidos entre 1983 y 2014, se identificaron patrones y tendencias en los registros de precipitación, tal como la evolución temporal de las series, la tendencia bimodal de lluvia en todas las estaciones (épocas de lluvia en abril-mayo y octubre-noviembre) y la como era de esperar la existencia de datos faltantes (entre el 0.02% al 27.25%) en todas las estaciones analizadas.

Se evidenció en el estudio la baja correlación de los datos diarios de lluvia entre las estaciones, debido al alto porcentaje de ceros y que a medida que se aumenta la agregación de los registros a semanal y mensual, aumentan las correlaciones y disminuyen como era de esperar la proporción de ceros en los datos. Se identificaron ocho estaciones (53080010, 53090030, 53090040, 54070030, 54075020, 54075040 y 53115010) ubicadas en la llanura marina del Pacífico, región con un clima tropical húmedo con altas precipitaciones todo el año que presentan correlaciones débiles y en general los modelos no presentan buenos resultados en dichos puntos. La mayor correlación entre los diferentes puntos analizados se encuentra en el *dataset* de los Satélites.

Se probaron diferentes estimadores / modelos de aprendizaje automático, estadístico y profundo, los cuales fueron evaluados acorde a las métricas respectivas de regresión (*R2*, *MSE*, *MAE* y *MaxError*) y clasificación (*Accuracy*, *Recall*, *Precision* y *F1-Score*). Los resultados mostraron que:

- Regresión: aunque los resultados de  $R^2$  del modelo de regresión lineal / OLS para las predicciones diarias fueron bajos (alrededor del 2.5% en promedio para todas las estaciones), mejoraron significativamente para agregaciones semanales (cerca del 30%) y mensuales (alrededor del 54%). El estimador *LinearRegression* presenta el mejor desempeño en métricas en las diferentes agregaciones entre los modelos de regresión evaluados, seguidos por las máquinas de soporte vectorial / SVR especialmente en las predicciones semanales, y por último los árboles de decisión / DTR presentan las métricas más bajas y por lo tanto el desempeño más bajo.
- Clasificación: los modelos de clasificación evaluados (regresión logística / *logit*, máquinas de soporte vectorial / SVC y árboles de decisión / DTC) mostraron en general valores similares de exactitud / *accuracy*. Como en los modelos de regresión, la métrica aumentó al pasar de predicciones diarias a semanales y a mensuales, alcanzando un *accuracy* de aproximadamente 95% para predicciones mensuales. Al aumentar las categorías en los modelos de clasificación, como los resultantes de diferentes intensidades de lluvia en general disminuye el *accuracy* obtenido con los estimadores de clasificación como *LogisticRegression*.

- Series de tiempo: los modelos de series de tiempo evaluados (ARIMA, Prophet y Sktime) en general no lograron un buen desempeño en  $R^2$ , evidenciándose en el estudio que esto es especialmente cierto en la predicción de valores extremos o picos de precipitación. Sin embargo, para algunas estaciones los pronósticos captan de manera general la tendencia de la lluvia acumulada semanal y mensual.
- Redes neuronales: se implementó una red FCNN, pero los valores de  $R^2$  obtenidos presentan un desempeño más bajo que los anteriores modelos descritos.
- Otros modelos: se evaluaron diferentes estimadores de *imputing*, pero ninguna de los estimadores (univariados, multivariados y de vecinos más cercanos) presentaron buen desempeño de  $R^2$ .

De todo lo anterior se desprende, que las métricas de los modelos de regresión y clasificación para datos diarios mejoran siempre al agregar los datos de manera semanal y mensual, y aumentan significativamente las métricas al transformar el problema como clasificación (llueve o no llueve o por categorías de intensidad de lluvia).

Ciertamente, el ejercicio estuvo supeditado a prácticamente una sola característica: los datos satelitales. La temporalidad de estos datos se rompió debido a la presencia de valores nulos en las etiquetas: los datos de las estaciones en tierra, lo que sin duda generó una significativa penalización en los modelos de series de tiempo. Además, el comportamiento del fenómeno analizado en la región de estudio, con una alta presencia de ceros en la gran mayoría de las estaciones, dificultó el diagnóstico del no aprendizaje de los modelos.

Así las cosas, lo anterior abre la puerta a que futuras investigaciones aborden el problema seleccionando zonas geográficas donde las precipitaciones sean más frecuentes. De este modo, se podrían llegar a garantizar las condiciones de valor inicial que se persiguieron a lo largo del proyecto.

En todo caso, es claro que la estrategia correcta consiste en identificar el modelo adecuado para cada estación, en lugar de replicar un único modelo.

Más aún, como parte de la búsqueda de alternativas para alcanzar el objetivo principal de este proyecto, que era predecir el valor real de la lluvia en la zona de estudio, se consideró la hipótesis de que, aunque no se lograra obtener el valor exacto, se podría analizar el comportamiento de los residuales. El objetivo era identificar la tendencia que estos siguen para, junto con el valor de la predicción, componer una curva cercana a la curva de los valores reales de lluvia. No obstante, la curva obtenida en las predicciones no recreó la tendencia de los datos reales, lo cual se evidenció en el comportamiento de los errores cuadráticos medios (MAE).

## 7.2 OBSERVACIONES Y POSIBLES TRABAJOS FUTUROS

Para los modelos de regresión se puede apreciar que dichos algoritmos no funcionan para puntos / estaciones con alta precipitación, lo cual representa una futura área de investigación / trabajos de grado de maestría para profundizar en el tema y seleccionar / desarrollar para dichas condiciones. Lo anterior requiere profundizar en el estudio de modelos de predicción y pronóstico para zonas

altamente lluviosas como las encontradas en la zona de estudio en inmediaciones del Océano Pacífico y cerca al sur de Chocó (el departamento más lluvioso del país).

De la misma manera, es importante profundizar en el análisis de series de tiempo con alto porcentaje de datos con valor cero, como es el caso de la precipitación, especialmente en zonas secas o desérticas con pocos días de lluvia.

Por último, se ha identificado la necesidad de realizar estudios con modelos multivariados para la predicción de la precipitación, que incluya otras variables relacionadas como temperatura ambiente, humedad relativa, presión atmosférica, radiación solar y nubosidad (si se cuenta con datos disponibles en la zona de estudio).

## REFERENCIAS BIBLIOGRÁFICAS

- [1] J. Hansen, R. Ruedy, M. Sato, y K. Lo, “Global Surface Temperature Change”, *Rev. Geophys.*, vol. 48, núm. 4, 2010, doi: 10.1029/2010RG000345.
- [2] H. Hersbach *et al.*, “The ERA5 global reanalysis”, *Q. J. R. Meteorol. Soc.*, vol. 146, núm. 730, pp. 1999–2049, 2020, doi: 10.1002/qj.3803.
- [3] IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales, “Acerca de la Entidad”. Consultado: el 17 de abril de 2024. [En línea]. Disponible en: <http://www.ideam.gov.co/web/entidad/acerca-entidad>
- [4] IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales, “Banco de Datos del IDEAM - DHIME”. Consultado: el 17 de abril de 2024. [En línea]. Disponible en: <http://dhime.ideam.gov.co/atencionciudadano/>
- [5] C. Funk *et al.*, “CHIRPS: Rainfall Estimates from Rain Gauge and Satellite Observations | Climate Hazards Center - UC Santa Barbara”. Consultado: el 2 de mayo de 2023. [En línea]. Disponible en: <https://www.chc.ucsb.edu/data/chirps>
- [6] C. Funk *et al.*, “The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes”, *Sci. Data*, vol. 2, núm. 1, Art. núm. 1, dic. 2015, doi: 10.1038/sdata.2015.66.
- [7] M. Gillet, “Satellite meteorology, an introduction - Par Stanley Q. Kidder et Thomas H. Vonder Haar”, *La Météorologie*, vol. 1998, núm. 21, pp. 94–95, doi: 10.4267/2042/54503.
- [8] D. Maraun *et al.*, “Towards process-informed bias correction of climate change simulations”, *Nat. Clim. Change*, vol. 7, núm. 11, Art. núm. 11, nov. 2017, doi: 10.1038/nclimate3418.
- [9] Z. Hao, A. AghaKouchak, y T. J. Phillips, “Changes in concurrent monthly precipitation and temperature extremes”, *Environ. Res. Lett.*, vol. 8, núm. 3, p. 034014, ago. 2013, doi: 10.1088/1748-9326/8/3/034014.
- [10] A. Linés Escardó, “Contribución al concepto de clima”, *Territ. Rev. Dep. Ciènc. Terra*, núm. 1, pp. 203–214, 1998.
- [11] T. Vasquez, *International Weather Watchers Observer Handbook*. Weather Graphics Technologies, 1999. [En línea]. Disponible en: <http://www.weathergraphics.com/dl/obsman.pdf>
- [12] Axxess Networks, “Clima satelital y los satélites meteorológicos”. Consultado: el 9 de junio de 2023. [En línea]. Disponible en: <https://axessnet.com/clima-satelital-y-los-satelites-meteorologicos/>
- [13] I. Kuznetsov, “What are weather satellites and how do they help us get weather data”, WINDY.APP. Consultado: el 9 de junio de 2023. [En línea]. Disponible en: <https://windy.app/blog/what-are-weather-satellites.html>
- [14] Y. Fan, V. Krasnopolsky, H. van den Dool, C.-Y. Wu, y J. Gottschalck, “Using Artificial Neural Networks to Improve CFS Week-3–4 Precipitation and 2-m Air Temperature Forecasts”, *Weather Forecast.*, vol. 38, núm. 5, pp. 637–654, may 2023, doi: 10.1175/WAF-D-20-0014.1.
- [15] M. Beroho *et al.*, “Analysis and prediction of climate forecasts in Northern Morocco: application of multilevel linear mixed effects models using R software”, *Heliyon*, vol. 6, núm. 10, p. e05094, oct. 2020, doi: 10.1016/j.heliyon.2020.e05094.
- [16] A. Chattopadhyay, E. Nabizadeh, y P. Hassanzadeh, “Analog Forecasting of Extreme-Causing Weather Patterns Using Deep Learning”, *J. Adv. Model. Earth Syst.*, vol. 12, núm. 2, p. e2019MS001958, 2020, doi: 10.1029/2019MS001958.
- [17] E. A. Varouchakis, D. T. Hristopulos, G. P. Karatzas, G. A. Corzo Perez, y V. Diaz, “Spatiotemporal geostatistical analysis of precipitation combining ground and satellite observations”, *Hydrol. Res.*, vol. 52, núm. 3, pp. 804–820, mar. 2021, doi: 10.2166/nh.2021.160.
- [18] A. Hünerbein, H. Deneke, A. Macke, K. Ebell, y U. Görsdorf, “Combining the Perspective of Satellite- and Ground-Based Observations to Analyze Cloud Frontal Systems”, *J. Appl. Meteorol. Climatol.*, vol. 53, núm. 11, pp. 2538–2552, nov. 2014, doi: 10.1175/JAMC-D-13-0274.1.
- [19] J. Wang *et al.*, “Estimating snow depth by combining satellite data and ground-based observations over Alaska: A deep learning approach”, *J. Hydrol.*, vol. 585, p. 124828, jun. 2020, doi: 10.1016/j.jhydrol.2020.124828.

- [20] S.-I. Akasofu, D. Weimer, T. Iijima, B.-H. Ahn, y Y. Kamide, “Agreements between ground-based and satellite-based observations”, *Planet. Space Sci.*, vol. 38, núm. 12, pp. 1533–1540, dic. 1990, doi: 10.1016/0032-0633(90)90158-M.
- [21] X. Zhang y Q. Tang, “Combining satellite precipitation and long-term ground observations for hydrological monitoring in China”, *J. Geophys. Res. Atmospheres*, vol. 120, núm. 13, pp. 6426–6443, 2015, doi: 10.1002/2015JD023400.
- [22] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd ed. Sebastopol: O’Reilly, 2022. Consultado: el 16 de octubre de 2023. [En línea]. Disponible en: <https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/>
- [23] G. James, D. Witten, T. Hastie, R. Tibshirani, y J. Taylor, *An Introduction to Statistical Learning: with Applications in Python*. en Springer Texts in Statistics. Cham: Springer International Publishing, 2023. doi: 10.1007/978-3-031-38747-0.
- [24] “Relieve del Valle del Cauca: Relieve Departamento del Valle del Cauca Colombia”, [todacolombia.com](https://www.todacolombia.com). Consultado: el 22 de octubre de 2023. [En línea]. Disponible en: <https://www.todacolombia.com/departamentos-de-colombia/valle-del-cauca/relieve.html>
- [25] S. E. Fick y R. J. Hijmans, “WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas”, *Int. J. Climatol.*, vol. 37, núm. 12, pp. 4302–4315, oct. 2017, doi: 10.1002/joc.5086.
- [26] *scikit-learn: machine learning in Python*. Consultado: el 8 de octubre de 2023. [En línea]. Disponible en: <https://scikit-learn.org/stable/>
- [27] *sktime: A unified framework for machine learning with time series*. Consultado: el 8 de diciembre de 2023. [En línea]. Disponible en: <https://www.sktime.net/en/stable/>
- [28] J. Perktold *et al.*, *statsmodels/statsmodels: Release 0.14.1*. (el 14 de diciembre de 2023). doi: 10.5281/ZENODO.593847.
- [29] *Prophet: Forecasting at scale*. Consultado: el 17 de enero de 2024. [En línea]. Disponible en: <http://facebook.github.io/prophet/>
- [30] R. J. Hyndman y G. Athanasopoulos, *Forecasting: Principles and Practice (3rd ed)*. Consultado: el 14 de abril de 2024. [En línea]. Disponible en: <https://otexts.com/fpp3/>
- [31] *Keras: Deep Learning for humans*. Consultado: el 3 de diciembre de 2023. [En línea]. Disponible en: <https://keras.io/>
- [32] W. Wang, “The classification of daily rainfall intensity and the corresponding...”, ResearchGate. Consultado: el 20 de abril de 2024. [En línea]. Disponible en: [https://www.researchgate.net/figure/The-classification-of-daily-rainfall-intensity-and-the-corresponding-number-of-rain-day\\_tbl1\\_352152382](https://www.researchgate.net/figure/The-classification-of-daily-rainfall-intensity-and-the-corresponding-number-of-rain-day_tbl1_352152382)
- [33] H. Latupapua, A. I. Latupapua, A. Wahab, y M. Alaydrus, “Wireless Sensor Network Design for Earthquake’s and Landslide’s Early Warnings”, *Indones. J. Electr. Eng. Comput. Sci.*, vol. 11, núm. 2, Art. núm. 2, ago. 2018, doi: 10.11591/ijeecs.v11.i2.pp437-445.
- [34] Buthan National Center for Hydrology and Meteorology, “Rainfall Classification: Intensity of Rainfall in 24 Hours”. Consultado: el 21 de abril de 2024. [En línea]. Disponible en: <https://www.nchm.gov.bt/attachment/ckfinder/userfiles/files/Rainfall%20intensity%20classification.pdf>

## ANEXO A

### A1.1 ESTADÍSTICA DESCRIPTIVA DATOS ESTACIONES

Estación	count	mean	std	min	25%	50%	75%	max
26055070	11409.0	4.1	9.72	0.0	0.0	0.0	2.9	122.6
26060020	11631.0	3.46	9.39	0.0	0.0	0.0	1.0	123.0
26070110	11220.0	4.36	11.08	0.0	0.0	0.0	2.0	120.0
26070760	11685.0	4.4	11.12	0.0	0.0	0.0	3.0	146.0
26075010	10127.0	2.93	7.7	0.0	0.0	0.0	1.5	102.2
26075040	11486.0	2.41	6.6	0.0	0.0	0.0	1.2	89.6
26080070	11334.0	2.46	6.6	0.0	0.0	0.0	1.0	75.0
26080290	11532.0	2.97	8.45	0.0	0.0	0.0	0.0	100.0
26090460	11374.0	4.54	11.11	0.0	0.0	0.0	3.0	136.0
26090630	11493.0	2.66	7.24	0.0	0.0	0.0	0.0	95.0
26095080	10128.0	3.34	6.28	0.0	0.0	0.4	4.1	77.2
26100070	11559.0	3.25	8.8	0.0	0.0	0.0	0.0	98.0
26100300	11522.0	4.08	9.24	0.0	0.0	0.0	4.0	118.0
26100350	11343.0	5.27	10.77	0.0	0.0	0.0	6.0	113.0
26100400	10770.0	4.3	10.34	0.0	0.0	0.0	3.0	116.0
26100410	11285.0	4.35	10.0	0.0	0.0	0.0	4.0	110.0
26100690	11291.0	3.95	7.97	0.0	0.0	0.0	5.0	135.0
26100700	11027.0	4.96	12.67	0.0	0.0	0.0	4.0	586.0
26100770	11309.0	4.91	10.72	0.0	0.0	0.0	5.0	107.0
26100780	11410.0	3.38	8.3	0.0	0.0	0.0	2.0	92.0
26100790	8939.0	3.79	9.79	0.0	0.0	0.0	3.0	150.0
26100800	10686.0	3.21	6.88	0.0	0.0	0.0	4.0	90.0
26105150	10328.0	4.11	9.33	0.0	0.0	0.0	3.3	112.0
26105160	10386.0	3.63	8.63	0.0	0.0	0.0	2.7	108.0
26110040	11348.0	2.99	8.04	0.0	0.0	0.0	1.0	115.0
26110120	11184.0	7.14	11.74	0.0	0.0	0.0	10.0	175.0
26110150	11125.0	4.21	9.23	0.0	0.0	0.0	4.0	88.0
26110160	11440.0	3.86	9.02	0.0	0.0	0.0	3.0	125.0
26110210	10155.0	5.01	9.37	0.0	0.0	1.0	6.0	200.0
26110230	11251.0	2.86	7.27	0.0	0.0	0.0	2.0	107.0
26110290	11499.0	3.2	8.11	0.0	0.0	0.0	2.0	125.0
26110300	11506.0	2.89	7.47	0.0	0.0	0.0	2.0	92.0
26115040	11175.0	3.11	7.28	0.0	0.0	0.0	2.5	116.0
26120120	11303.0	5.13	11.65	0.0	0.0	0.0	3.8	125.0
26120130	11252.0	4.97	11.41	0.0	0.0	0.0	3.0	135.0
26120150	11348.0	5.51	11.71	0.0	0.0	0.0	5.3	150.0
26120180	11405.0	5.01	11.09	0.0	0.0	0.0	4.0	115.0
26125130	11119.0	6.82	11.77	0.0	0.0	1.0	8.7	95.5
53080010	10806.0	28.5	30.81	0.0	5.0	18.0	41.0	230.0
53090030	10892.0	20.08	30.23	0.0	0.0	8.0	30.0	226.0

Estación	count	mean	std	min	25%	50%	75%	max
53090040	11033.0	20.24	33.34	0.0	0.0	0.0	31.0	335.0
53100040	11284.0	3.47	7.68	0.0	0.0	0.0	3.0	122.0
53110030	11154.0	3.31	5.83	0.0	0.0	1.0	4.0	110.0
53110040	10848.0	2.71	7.03	0.0	0.0	0.0	1.0	125.0
53115010	11187.0	19.04	27.41	0.0	1.0	8.0	25.0	217.2
54030010	11482.0	3.81	8.02	0.0	0.0	0.0	4.0	120.0
54030030	10432.0	3.16	8.38	0.0	0.0	0.0	2.0	133.0
54070030	10859.0	19.91	29.45	0.0	0.0	8.0	26.0	272.0
54070150	10687.0	19.92	28.58	0.0	0.0	9.0	29.0	280.0
54075020	8503.0	20.38	27.98	0.0	1.7	9.0	27.25	209.0
54075040	9807.0	19.9	29.05	0.0	1.0	8.0	26.0	241.1
26070170	10922.0	3.38	10.0	0.0	0.0	0.0	0.0	124.0
26080280	11668.0	4.79	10.83	0.0	0.0	0.0	4.0	137.0
26090060	11654.0	4.23	9.16	0.0	0.0	0.0	5.0	135.0
26095230	10369.0	3.22	7.84	0.0	0.0	0.0	2.5	108.0
26100830	11379.0	4.1	9.65	0.0	0.0	0.0	3.0	241.0
53110100	10990.0	3.51	9.03	0.0	0.0	0.0	2.0	110.0
26105230	10546.0	3.65	8.53	0.0	0.0	0.0	2.8	102.5

## A1.2 ESTADÍSTICA DESCRIPTIVA DATOS SATELITALES

Estación	count	mean	std	min	25%	50%	75%	max
26055070	11688.0	4.51	8.5	0.0	0.0	0.0	6.41	81.25
26060020	11688.0	3.87	7.72	0.0	0.0	0.0	5.2	70.32
26070110	11688.0	3.39	7.18	0.0	0.0	0.0	4.03	81.95
26070760	11688.0	4.31	9.24	0.0	0.0	0.0	5.3	176.38
26075010	11688.0	2.98	6.19	0.0	0.0	0.0	3.72	66.18
26075040	11688.0	3.01	6.18	0.0	0.0	0.0	3.84	64.99
26080070	11688.0	3.02	5.95	0.0	0.0	0.0	4.18	58.99
26080290	11688.0	3.89	7.29	0.0	0.0	0.0	5.64	74.33
26090460	11688.0	3.76	8.02	0.0	0.0	0.0	4.36	106.82
26090630	11688.0	3.69	7.25	0.0	0.0	0.0	5.15	85.01
26095080	11688.0	4.23	8.98	0.0	0.0	0.0	5.41	126.88
26100070	11688.0	4.41	7.76	0.0	0.0	0.0	6.65	138.54
26100300	11688.0	4.43	7.37	0.0	0.0	0.0	6.92	129.23
26100350	11688.0	5.25	10.02	0.0	0.0	0.0	7.52	121.11
26100400	11688.0	4.54	7.85	0.0	0.0	0.0	6.94	98.11
26100410	11688.0	4.54	8.09	0.0	0.0	0.0	6.81	90.27
26100690	11688.0	4.67	9.27	0.0	0.0	0.0	6.34	99.65
26100700	11688.0	4.6	8.43	0.0	0.0	0.0	6.83	119.66
26100770	11688.0	4.83	8.9	0.0	0.0	0.0	7.09	88.55
26100780	11688.0	4.22	7.22	0.0	0.0	0.0	6.63	139.16
26100790	11688.0	4.77	9.55	0.0	0.0	0.0	6.83	191.91
26100800	11688.0	4.67	9.41	0.0	0.0	0.0	6.48	99.62
26105150	11688.0	4.75	8.37	0.0	0.0	0.0	7.12	109.38

Estación	count	mean	std	min	25%	50%	75%	max
26105160	11688.0	4.51	8.02	0.0	0.0	0.0	6.8	107.14
26110040	11688.0	4.32	7.43	0.0	0.0	0.0	6.44	90.94
26110120	11688.0	6.66	10.3	0.0	0.0	0.0	10.29	213.12
26110150	11688.0	5.35	8.53	0.0	0.0	0.0	8.26	80.03
26110160	11688.0	4.69	7.72	0.0	0.0	0.0	7.3	107.66
26110210	11688.0	4.52	7.08	0.0	0.0	0.0	7.12	94.74
26110230	11688.0	4.29	7.3	0.0	0.0	0.0	6.52	80.88
26110290	11688.0	4.38	7.38	0.0	0.0	0.0	6.83	116.41
26110300	11688.0	4.54	7.42	0.0	0.0	0.0	7.11	80.5
26115040	11688.0	4.38	7.38	0.0	0.0	0.0	6.83	116.41
26120120	11688.0	5.35	9.63	0.0	0.0	0.0	8.03	115.49
26120130	11688.0	4.76	8.11	0.0	0.0	0.0	7.45	97.03
26120150	11688.0	5.64	8.98	0.0	0.0	0.0	8.89	102.78
26120180	11688.0	4.98	8.65	0.0	0.0	0.0	7.68	122.99
26125130	11688.0	5.2	9.68	0.0	0.0	0.0	7.72	100.26
53080010	11688.0	22.31	31.46	0.0	0.0	10.86	34.73	435.31
53090030	11688.0	20.43	29.1	0.0	0.0	9.06	31.68	298.35
53090040	11688.0	16.72	23.98	0.0	0.0	7.68	25.2	245.26
53100040	11688.0	5.64	10.08	0.0	0.0	0.0	8.28	136.61
53110030	11688.0	5.27	9.37	0.0	0.0	0.0	7.71	114.74
53110040	11688.0	3.17	6.08	0.0	0.0	0.0	4.45	61.31
53115010	11688.0	16.67	23.72	0.0	0.0	8.65	25.13	311.1
54030010	11688.0	4.05	6.62	0.0	0.0	0.0	6.27	78.06
54030030	11688.0	5.23	8.72	0.0	0.0	0.0	7.61	117.75
54070030	11688.0	20.85	27.6	0.0	0.0	11.48	32.3	229.03
54070150	11688.0	20.35	25.81	0.0	0.0	12.94	30.66	304.74
54075020	11688.0	18.87	24.92	0.0	0.0	11.51	28.75	303.13
54075040	11688.0	21.35	28.67	0.0	0.0	11.0	33.27	284.04
26070170	11688.0	3.34	7.01	0.0	0.0	0.0	4.02	92.49
26080280	11688.0	6.58	11.76	0.0	0.0	0.0	9.65	134.97
26090060	11688.0	3.93	7.95	0.0	0.0	0.0	5.21	85.27
26095230	11688.0	3.73	7.29	0.0	0.0	0.0	5.15	88.56
26100830	11688.0	4.86	7.97	0.0	0.0	0.0	7.63	112.46
53110100	11688.0	3.14	5.92	0.0	0.0	0.0	4.51	61.2
26105230	11688.0	4.29	7.84	0.0	0.0	0.0	6.33	101.85

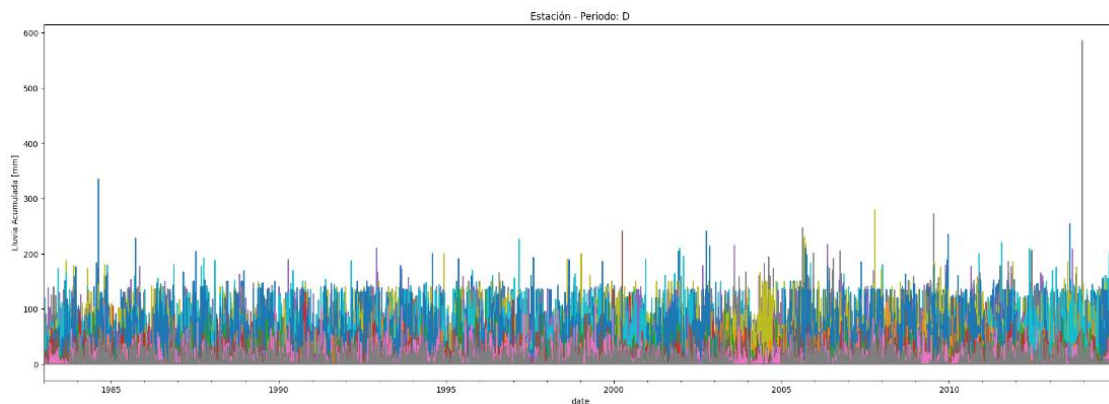
## A2. VISUALIZACIÓN DE DATOS

A continuación, se presentan gráficas para los datos de lluvia acorde a las series de tiempo, el comportamiento mensual y acumulado anual de los conjuntos de datos, así como diagramas de cajas para evidenciar la dispersión y posibles datos atípicos de los datos y unos histogramas y diagramas de densidad para comparar el comportamiento y rango de los datos de precipitación.

### A2.1 SERIES DE TIEMPO

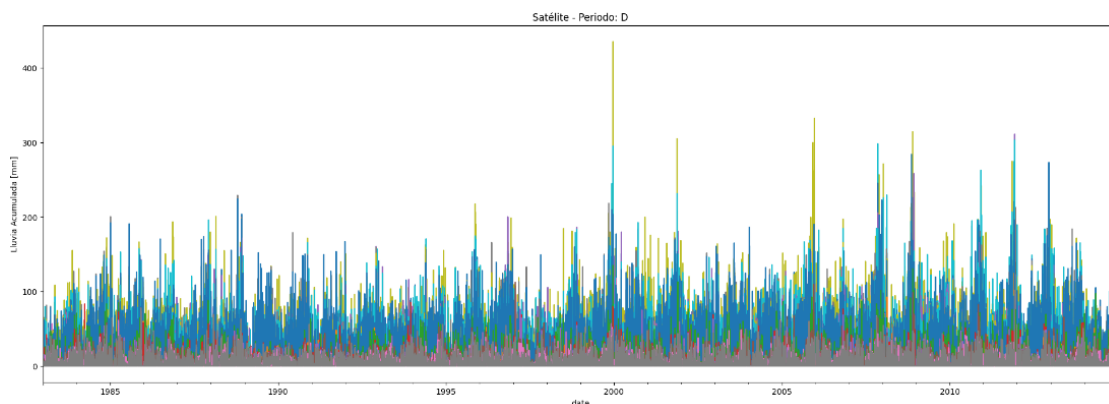
Las gráficas de línea siguientes presentan el comportamiento histórico diario y anual para las Estaciones y Satélites (tener presente que la lluvia / precipitación se suma y no se promedia al analizarla en periodos diarios, semanales, mensuales o anuales).

#### A2.1.1 SERIE DE TIEMPO DIARIA POR ESTACIONES



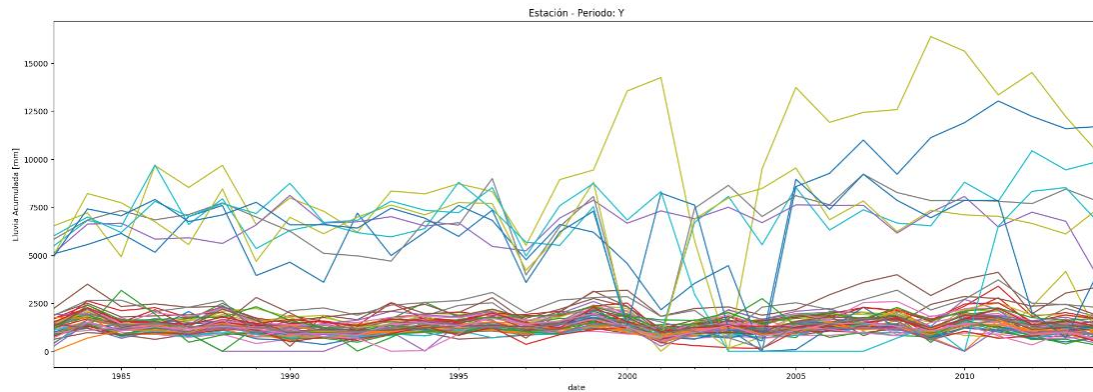
Nota: elaboración propia.

#### A2.1.2 SERIE DE TIEMPO DIARIA POR SATÉLITE

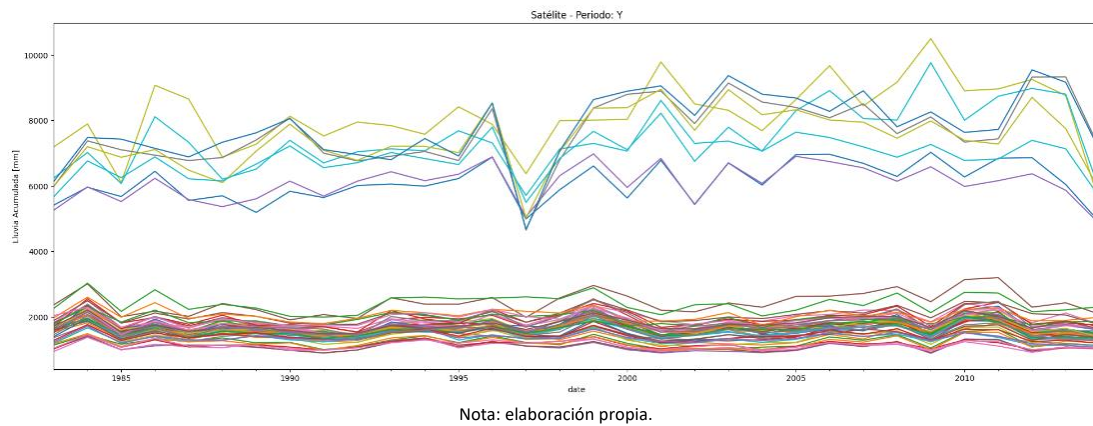


Nota: elaboración propia.

### A2.1.3 SERIE DE TIEMPO ANUAL POR ESTACIONES

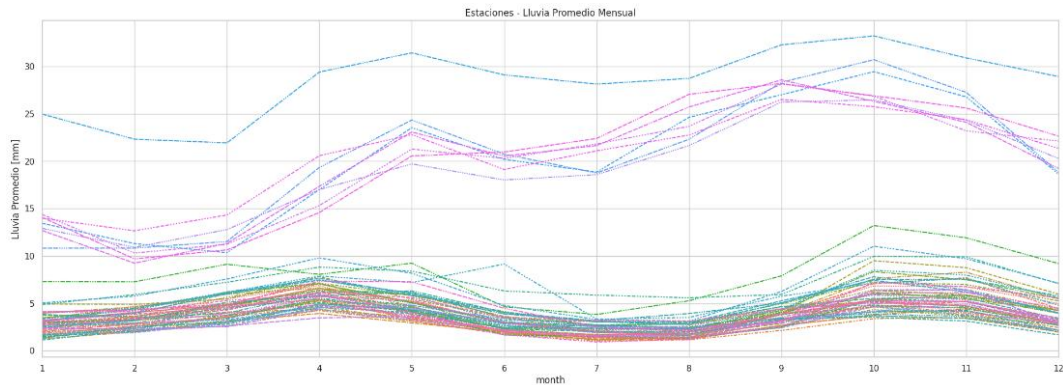


### A2.1.4 SERIE DE TIEMPO ANUAL POR SATÉLITES



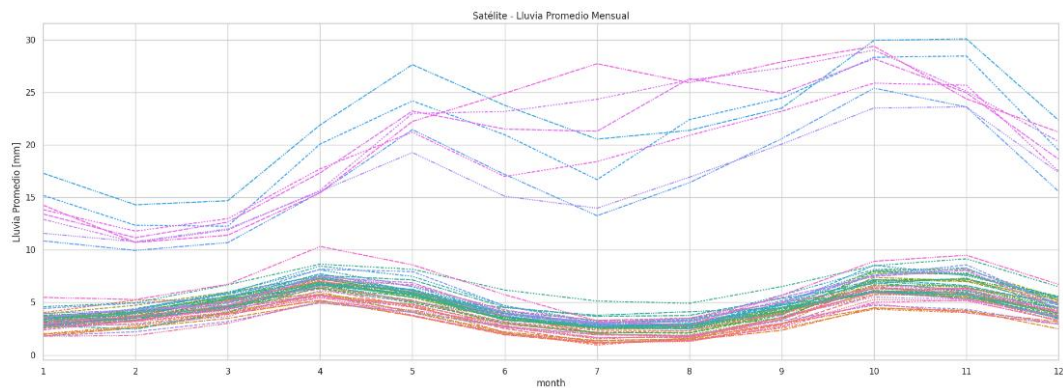
En las figuras siguientes se puede observar la lluvia promedio mensual para los datos de las Estaciones y Satelitales, como era de esperar y acorde al estado del arte y el área de estudio, el conjunto de datos analizado presenta las mayores precipitaciones en los meses de abril-mayo y octubre-noviembre, lo cual corresponde a un comportamiento bimodal en el departamento del Valle del Cauca, lo cual coincide con el patrón de lluvias a nivel nacional.

### A2.1.5 LLUVIA PROMEDIO MENSUAL POR ESTACIONES



Nota: elaboración propia.

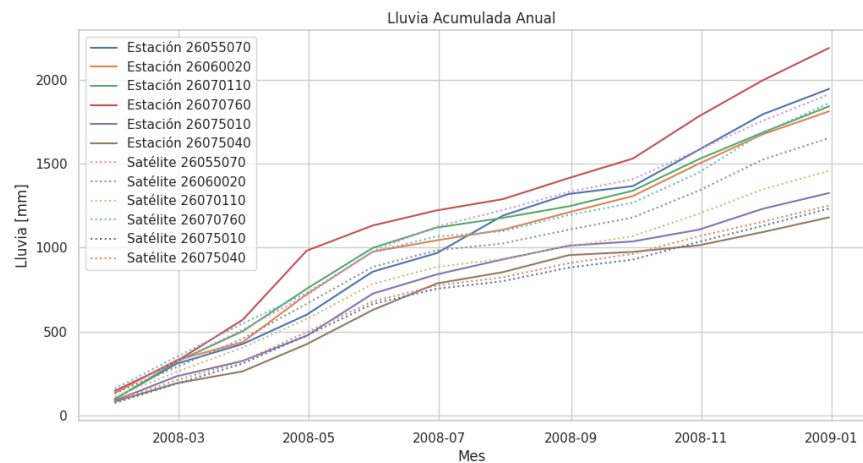
### A2.1.6 LLUVIA PROMEDIO MENSUAL POR SATÉLITE



Nota: elaboración propia.

La figura siguiente muestra la diferencia entre un subconjunto de estaciones para la lluvia acumulada en el año 2008.

### A2.1.7 LLUVIA ACUMULADA ANUAL 2008

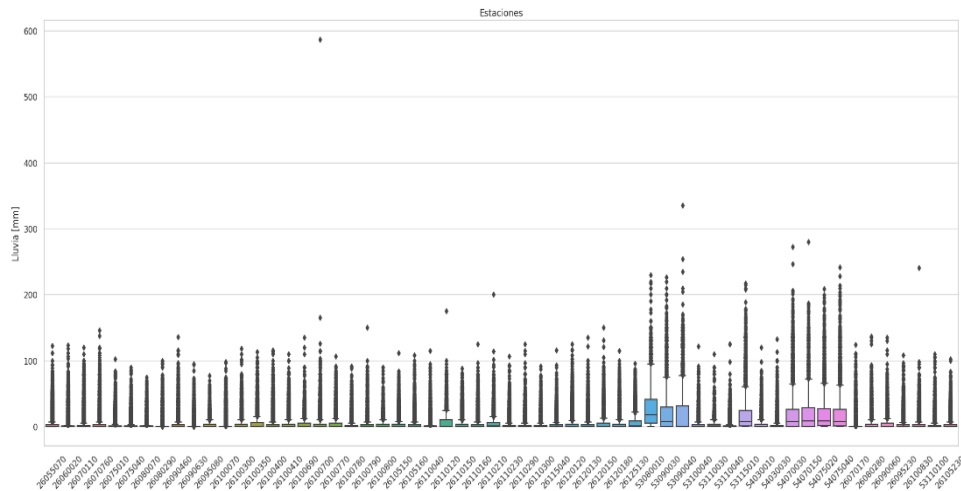


Nota: elaboración propia.

## A2.2 DIAGRAMAS DE CAJAS

En las figuras siguientes se presentan los diagramas de caja (*boxplots*) para los datos de Estaciones y Satélite.

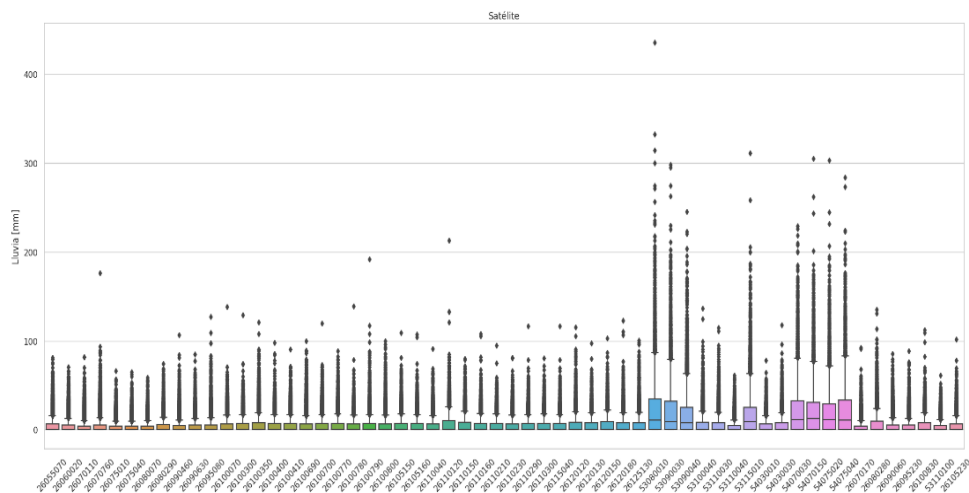
### A2.2.1 DIAGRAMAS DE CAJA POR ESTACIONES



Nota: elaboración propia.

Se observa que todas las estaciones presentan valores por fuera del rango intercuartílico. Al presentar este fenómeno en todas las estaciones y analizar la simetría de los valores atípicos, también se puede asumir que la presencia de datos inusuales en la variable precipitación es normal y que se debe considerar al analizar los resultados de los modelos. Con respecto a los cuartiles, se puede analizar que la dispersión de los datos alrededor de la mediana en la mayoría de las estaciones es mínima, lo que sugiere que el 50% de esos datos presentan valores muy cercanos a la mediana. Existen ocho estaciones con una dispersión muy grande en el 50% de sus datos.

### A2.2.2 DIAGRAMAS DE CAJA POR SATÉLITE



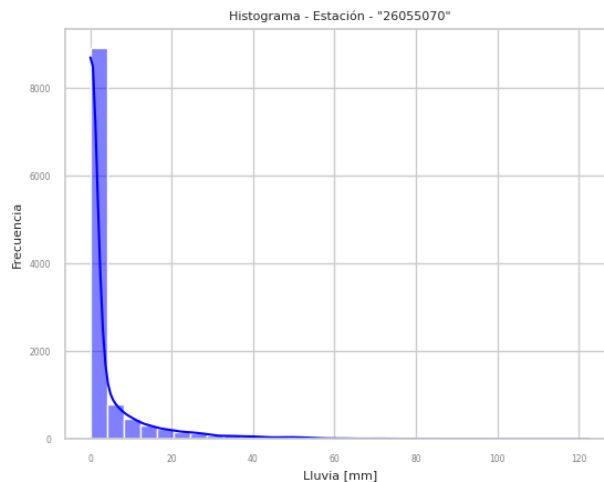
Nota: elaboración propia.

Con respecto a los diagramas de caja por satélite se puede confirmar la simetría de los valores atípicos, esto sugiere que las observaciones de las estaciones se encuentran acordes con la realidad y las condiciones meteorológicas de la zona de estudio. También se observa un crecimiento en la dispersión del 50% de los datos respecto de la mediana, esto puede ser debido a la forma de captura de los datos por los sensores remotos (Satélites). Adicionalmente, se puede confirmar que las ocho estaciones señalan de tener una dispersión mayor de los datos alrededor de su mediana comparadas con el resto de las estaciones, presentan una variabilidad similar con la fuente de datos suministrada por los satélites.

### A2.3 HISTOGRAMAS

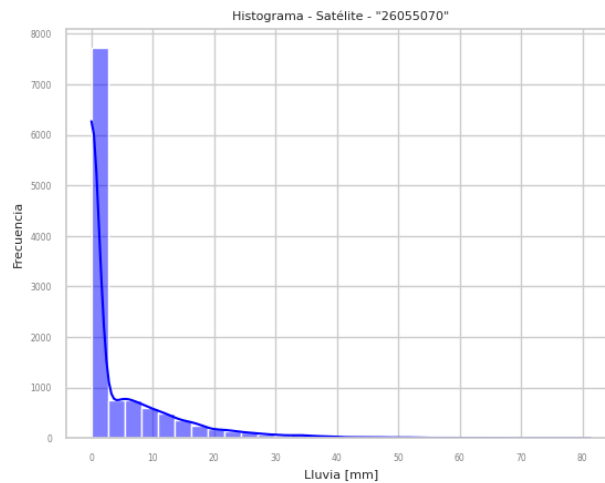
Las figuras siguientes muestran un comportamiento similar en cuanto a la distribución de los datos entre una estación seleccionada y su correspondiente satélite. No obstante, y como era de esperarse, al tomar la totalidad de los datos, se reflejan diferencias que serán abordadas más adelante.

#### A2.3.1 HISTOGRAMA ESTACIÓN 26055070



Nota: elaboración propia.

### A2.3.2 HISTOGRAMA SATÉLITE 26055070

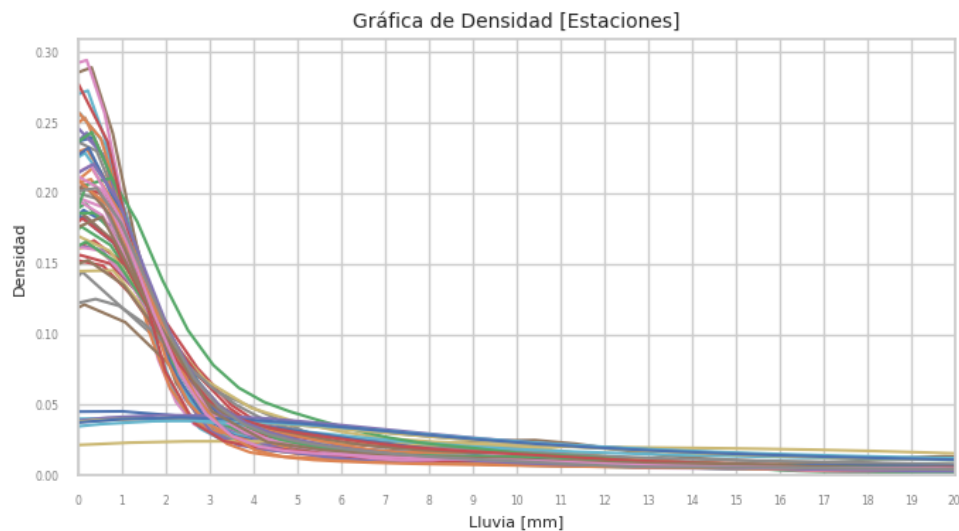


Nota: elaboración propia.

### A2.4 GRÁFICAS DE DENSIDAD

En las figuras siguientes se presentan las gráficas de densidad, es decir, la forma de distribución, para los datos de Estaciones y Satélite.

#### A2.4.1 GRÁFICA DE DENSIDAD POR ESTACIÓN

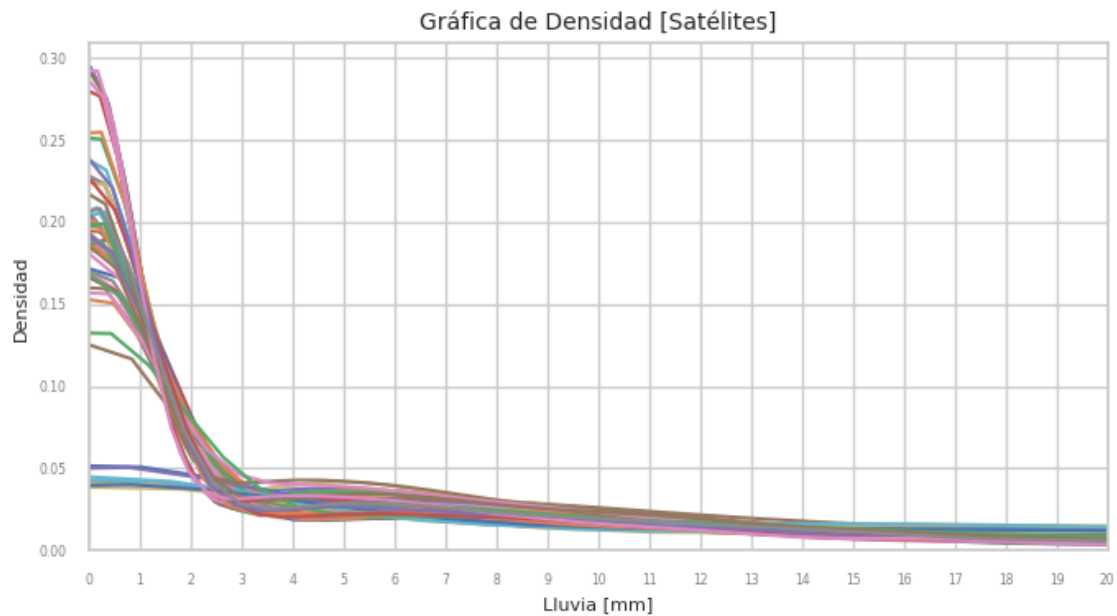


Nota: elaboración propia.

Como primera medida, los datos de la mayoría de las estaciones en tierra presentan una distribución similar, pero en valores de cero o nulos se observan variaciones, lo que indica que la pérdida de datos no es similar en todas las estaciones. Se observa un pequeño grupo de estaciones con una baja pérdida de datos, lo que nos sugiere que pueden ser las mismas estaciones que presentan una mayor distribución del 50% de sus datos respecto de la mediana comparadas con el resto de las estaciones. Analizando la forma de la curva de la densidad de la gran mayoría de estaciones se puede

decir que tienen una distribución sesgada a la derecha, por ser una curva asimétrica, con cola hacia la derecha y con una media desplazada hacia la izquierda y datos agrupados hacia el lado izquierdo. Este tipo de distribución implica que la mayoría de los valores se concentran en la parte inferior o más baja de la escala, para el caso de estudio, milímetros de lluvia, mientras que hay valores extremadamente altos que estiran la distribución hacia la derecha.

#### A2.4.2 GRÁFICA DE DENSIDAD POR SATÉLITE



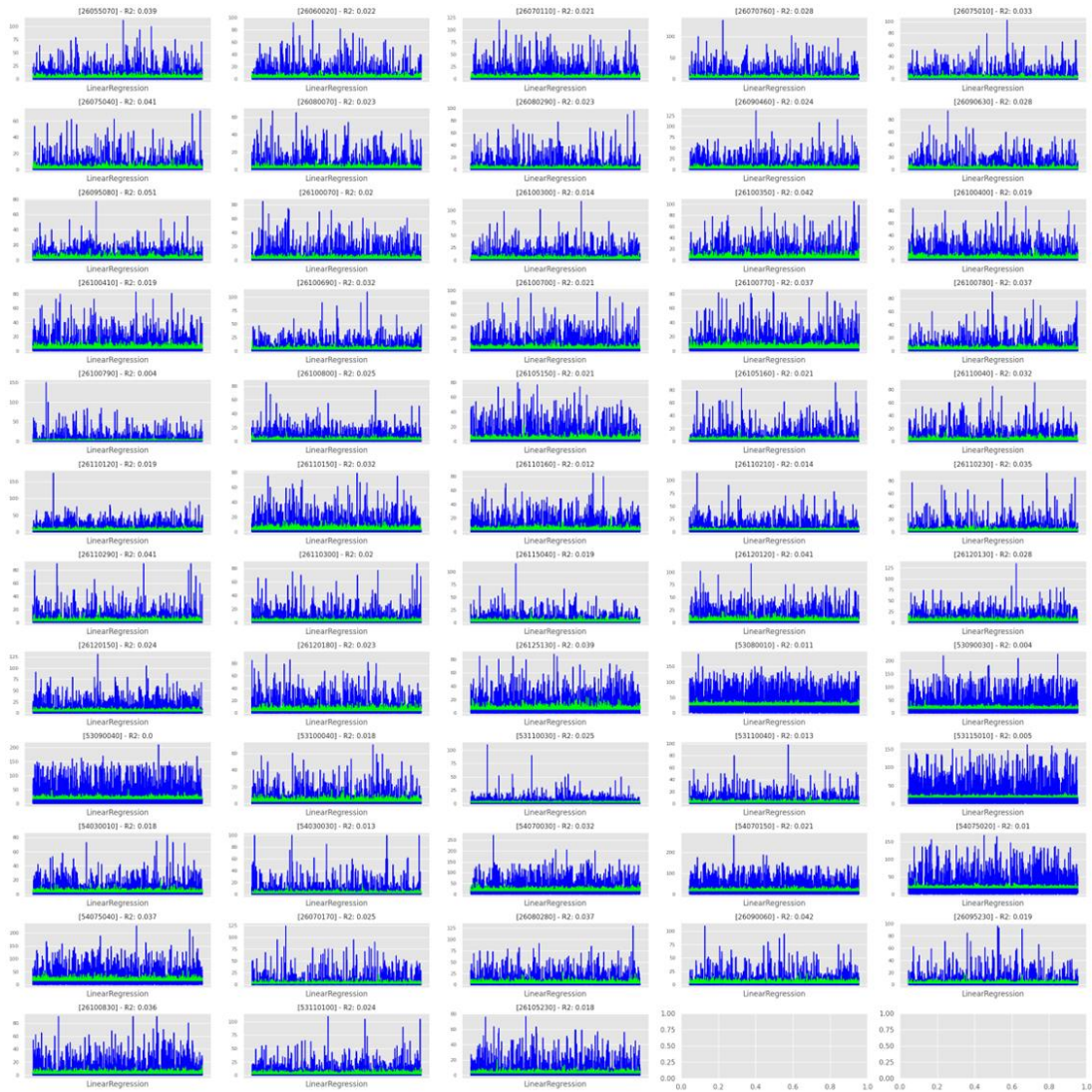
Nota: elaboración propia.

De la misma manera, en los datos proporcionados por los satélites, su distribución para la mayoría es similar y presentan las mismas características que los datos de las estaciones en tierra.

## ANEXO B

### PREDICCIONES SEGMENTADAS POR: MODELO Y FRACCIÓN DE TIEMPO PARA CADA ESTACIÓN

#### B1.1 RESULTADO R<sup>2</sup> PREDICCIONES DIARIAS - OLS POR ESTACIÓN



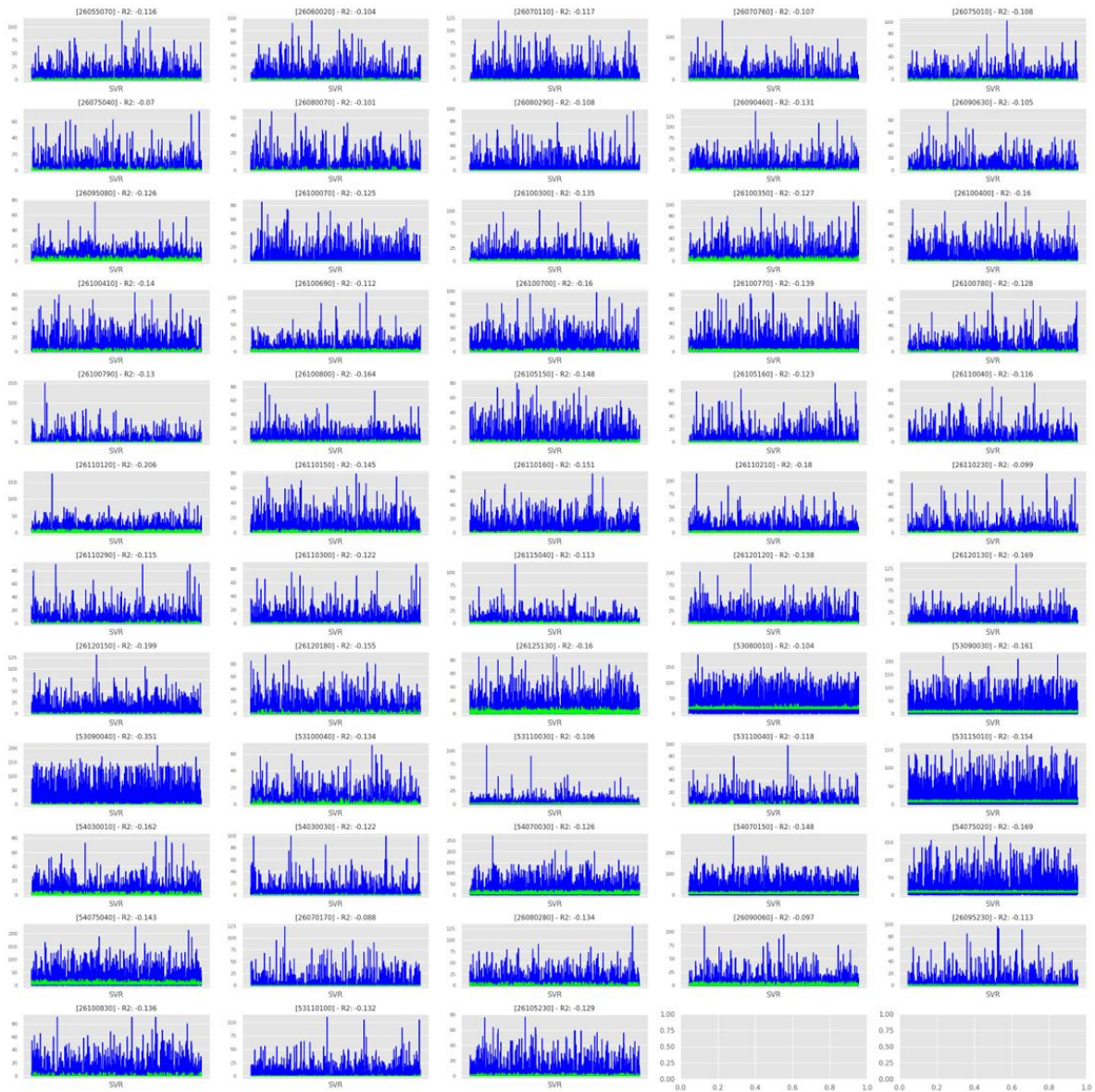
## B1.2 RESULTADO R<sup>2</sup> PREDICCIONES SEMANALES - OLS POR ESTACIÓN



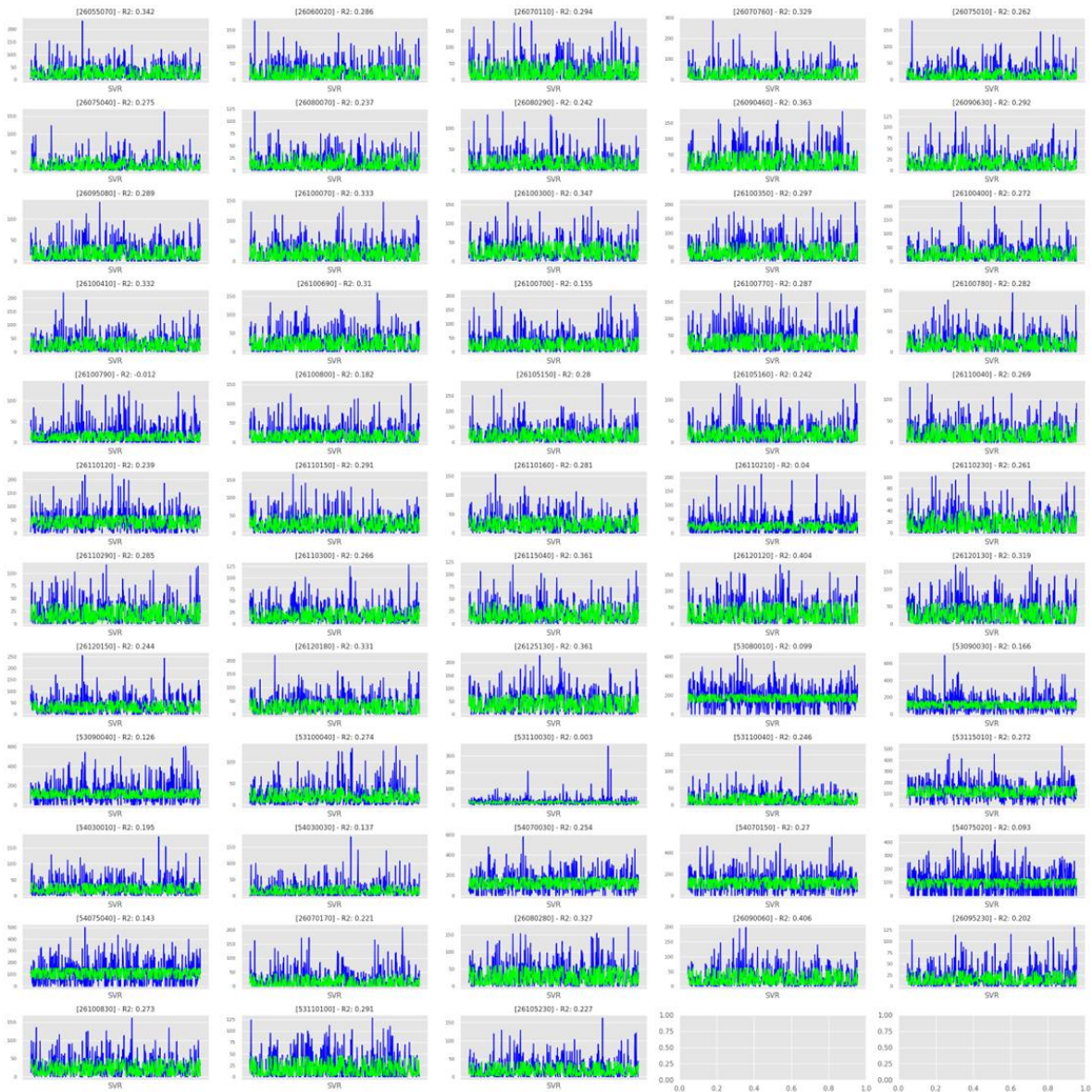
### B1.3 RESULTADO R<sup>2</sup> PREDICCIONES MENSUALES - OLS POR ESTACIÓN



## B2.1 RESULTADO $R^2$ PREDICCIONES DIARIAS - SVR POR ESTACIÓN



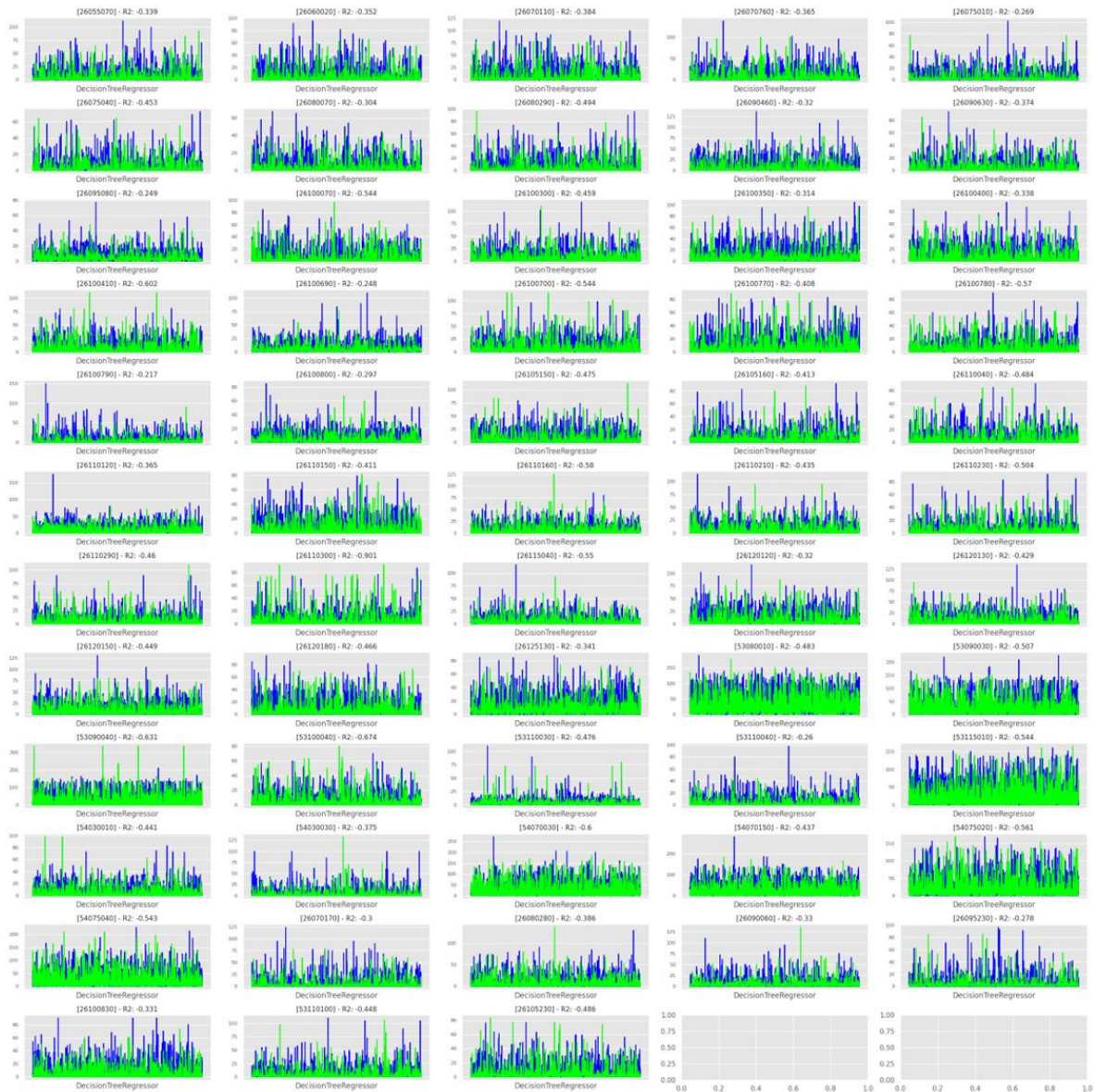
## B2.2 RESULTADO R<sup>2</sup> PREDICCIONES SEMANALES - SVR POR ESTACIÓN



### B2.3 RESULTADO R<sup>2</sup> PREDICCIONES MENSUALES - SVR POR ESTACIÓN



### B3.1 RESULTADO $R^2$ PREDICCIONES DIARIAS - DTR POR ESTACIÓN



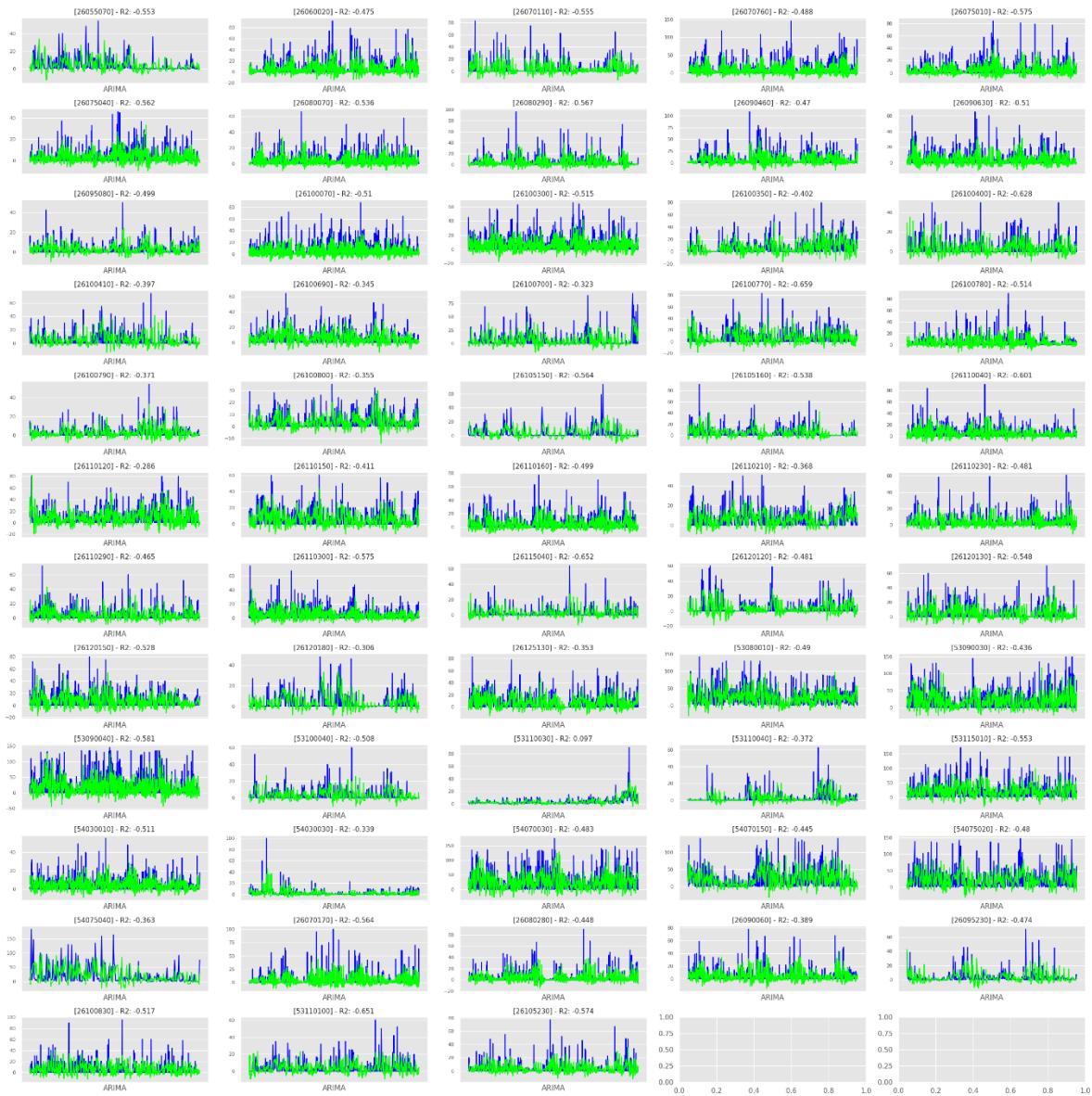
### B3.2 RESULTADO R<sup>2</sup> PREDICCIONES SEMANALES - DTR POR ESTACIÓN



### B3.3 RESULTADO R<sup>2</sup> PREDICIONES MENSUALES - DTR POR ESTACIÓN



## B4. RESULTADOS $R^2$ DE LAS PREDICCIONES DEL **MODELO ARIMA** PARA TODAS LAS ESTACIONES



## ANEXO C

### MÉTRICAS DE LOS MODELOS DE APRENDIZAJE AUTOMÁTICO Y ESTADÍSTICO

#### C1. MÉTRICAS DEL MODELO ARIMA

Estación	Parámetros del modelo	Dickey Fuller (p – value)	R <sup>2</sup>	MSE	MAE
26055070	ARIMA(4,1,0)(2,1,0)[12]	3.276891e-30	-0.553086266	10.11181656	6.807569565
26060020	ARIMA(4,1,0)(2,1,0)[12]	9.133591e-24	-0.475073408	156.6592238	7.617965993
26070110	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.555457361	125.0361073	6.627963254
26070760	ARIMA(4,1,0)(2,1,0)[12]	3.326200e-23	-0.487745512	239.1194634	9.322276303
26075010	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.574722117	116.7683112	6.496345675
26075040	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.561644161	49.93957674	4.580594631
26080070	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.536378147	60.91586175	4.838536033
26080290	ARIMA(4,1,0)(2,1,0)[12]	3.853067e-28	-0.567253898	108.4613502	5.992574187
26090460	ARIMA(4,1,0)(2,1,0)[12]	4.005024e-25	-0.469692563	173.668617	8.008182423
26090630	ARIMA(4,1,0)(2,1,0)[12]	1.385820e-27	-0.510153211	74.1936398	5.413417296
26095080	ARIMA(4,1,0)(2,1,0)[12]	6.252807e-20	-0.498873676	37.27064901	3.948196337
26100070	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.509635622	115.7026916	6.888568945
26100300	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.514578843	130.6081217	7.801236002
26100350	ARIMA(4,1,0)(2,1,0)[12]	8.910916e-13	-0.40241494	179.7001871	9.150009775
26100400	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.62790629	89.69704021	6.353743595
26100410	ARIMA(4,1,0)(2,1,0)[12]	4.999765e-28	-0.397290376	131.5646085	7.934670571
26100690	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.344971073	77.55543216	6.020479823
26100700	ARIMA(4,1,0)(2,1,0)[12]	1.031376e-28	-0.323084557	220.004349	9.615920054
26100770	ARIMA(4,1,0)(2,1,0)[12]	1.074485e-26	-0.659231764	230.1381092	10.34171259
26100780	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.514426813	97.578409	5.867918721
26100790	ARIMA(4,1,0)(2,1,0)[12]	3.926810e-30	-0.371272594	37.76911381	37.76911381
26100800	ARIMA(4,1,0)(2,1,0)[12]	2.212321e-17	-0.354655762	49.41931022	5.144588476
26105150	ARIMA(4,1,0)(2,1,0)[12]	1.059230e-25	-0.563887925	118.6569867	7.206070888
26105160	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.538016184	124.9269146	7.167555654
26110040	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.601398287	111.5941509	6.752283334
26110120	ARIMA(4,1,0)(2,1,0)[12]	1.736205e-14	-0.285678801	214.0645221	10.28440327
26110150	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.411320392	135.7089568	8.233773431
26110160	ARIMA(4,1,0)(2,1,0)[12]	6.404536e-29	-0.499277645	124.2224612	7.2693263
26110210	ARIMA(4,1,0)(2,1,0)[12]	7.931968e-28	-0.368043747	106.573545	7.644272881
26110230	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.480509723	65.57523471	5.129997937
26110290	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.464568722	91.86226337	6.066048826
26110300	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.574964373	87.3863197	6.063489792
26115040	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.652127401	63.76697955	5.149256137
26120120	ARIMA(4,1,0)(2,1,0)[12]	5.259709e-08	-0.481418956	110.0377584	6.548247107
26120130	ARIMA(4,1,0)(2,1,0)[12]	3.898010e-29	-0.548007965	127.6555405	7.506920577
26120150	ARIMA(4,1,0)(2,1,0)[12]	1.488759e-28	-0.528202675	208.2463261	10.3035529

Estación	Parámetros del modelo	Dickey Fuller (p – value)	R <sup>2</sup>	MSE	MAE
26120180	ARIMA(4,1,0)(2,1,0)[12]	6.267425e-28	-0.305557785	73.98817878	5.587848846
26125130	ARIMA(4,1,0)(2,1,0)[12]	1.041011e-17	-0.352722686	179.2250885	9.270588346
53080010	ARIMA(4,1,0)(2,1,0)[12]	3.901991e-15	-0.490338123	895.383071	22.05112817
53090030	ARIMA(4,1,0)(2,1,0)[12]	4.631006e-12	-0.436064946	1032.034646	23.59260669
53090040	ARIMA(4,1,0)(2,1,0)[12]	8.869010e-15	-0.581144011	1366.903752	25.30584433
53100040	ARIMA(4,1,0)(2,1,0)[12]	3.395118e-25	-0.507752979	74.65260541	5.843570098
53110030	ARIMA(4,1,0)(2,1,0)[12]	3.219996e-05	0.096734406	28.83985023	3.001939744
53110040	ARIMA(4,1,0)(2,1,0)[12]	1.458465e-06	-0.372001405	59.26726554	4.369767647
53115010	ARIMA(4,1,0)(2,1,0)[12]	1.760115e-21	-0.5532847	1188.45912	24.51836127
54030010	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.51115574	68.14234081	5.596917419
54030030	ARIMA(4,1,0)(2,1,0)[12]	1.804931e-29	-0.339305254	33.48894439	2.914744175
54070030	ARIMA(4,1,0)(2,1,0)[12]	2.095422e-09	-0.48287424	1423.005056	27.92502033
54070150	ARIMA(4,1,0)(2,1,0)[12]	9.722383e-10	-0.44499684	1185.964222	24.57196112
54075020	ARIMA(4,1,0)(2,1,0)[12]	2.016766e-27	-0.480375592	1425.259096	27.94378839
54075040	ARIMA(4,1,0)(2,1,0)[12]	2.016924e-02	-0.362992168	1446.082858	26.21217485
26070170	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.563525889	157.6509373	7.17337678
26080280	ARIMA(4,1,0)(2,1,0)[12]	6.846525e-18	-0.447680049	103.8108975	6.252468701
26090060	ARIMA(4,1,0)(2,1,0)[12]	1.841707e-12	-0.388892672	101.7556927	6.24047642
26095230	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.474047084	100.6088921	6.079507869
26100830	ARIMA(4,1,0)(2,1,0)[12]	0.000000e+00	-0.517173516	184.5442274	9.32873354
53110100	ARIMA(4,1,0)(2,1,0)[12]	2.385004e-18	-0.650615706	101.9073671	6.895528257
26105230	ARIMA(4,1,0)(2,1,0)[12]	3.690091e-19	-0.573823416	96.73266988	5.923613079

## C2.1 MÉTRICAS MODELO REGRESIÓN LINEAL OLS DIARIO

y	R2	MAE	MSE	MaxError
26055070_y	0.03869502	5.56602243	94.0547583	108.911915
26060020_y	0.02239651	5.05864674	77.9584917	90.935378
26070110_y	0.02098188	6.60585162	139.612151	116.581065
26070760_y	0.02818647	5.97911824	117.766959	128.002676
26075010_y	0.03254634	4.15756661	59.7199071	97.2505488
26075040_y	0.0414612	3.29214478	38.0218079	67.2199031
26080070_y	0.02335971	3.61081585	39.9968928	63.8762039
26080290_y	0.02240278	4.52338254	67.4339295	92.1970504
26090460_y	0.02381177	6.46393428	121.67469	132.402085
26090630_y	0.02845549	3.99353969	53.8021033	91.9821914
26095080_y	0.0510485	3.97294143	38.9688166	72.5166507
26100070_y	0.02001497	4.9888118	74.4737084	75.5058528
26100300_y	0.0142694	5.7277236	91.180491	113.224466
26100350_y	0.04230518	6.7377993	119.700233	101.10639
26100400_y	0.01935038	6.32402995	107.3167	87.6368035
26100410_y	0.01945567	5.94932118	88.7967726	79.6108344
26100690_y	0.03216178	4.88761633	64.5083528	106.812616
26100700_y	0.02109063	6.73576423	113.032368	93.8789811
26100770_y	0.03692278	6.39495502	101.698085	79.1204755
26100780_y	0.03693031	4.70170311	60.4789098	86.8921852
26100790_y	0.00409122	5.52926796	101.615995	146.265365
26100800_y	0.02568637	4.40717901	47.4285388	82.3705758
26105150_y	0.0208552	5.70529667	90.3543202	76.7195103
26105160_y	0.02103527	4.9062732	69.9811861	86.590905
26110040_y	0.03259446	4.55313057	64.9952853	88.8642887
26110120_y	0.01933875	8.21081212	144.520978	167.075668
26110150_y	0.03180564	5.73031204	82.1597223	72.0499874
26110160_y	0.01191025	5.32740647	72.0032107	80.7741239
26110210_y	0.01401612	5.78410082	82.9997194	107.803581
26110230_y	0.03460589	3.95898524	53.719303	85.8783951
26110290_y	0.04126025	4.47531971	60.9259801	87.6334148
26110300_y	0.0200712	4.25696896	57.3263819	84.7997007
26115040_y	0.01950562	4.17698928	52.1104299	113.801215
26120120_y	0.04059017	6.95241013	125.86954	104.060116
26120130_y	0.02841336	6.89211963	112.855324	128.182978
26120150_y	0.02355824	7.58452656	139.422009	124.687002
26120180_y	0.02261021	6.69984633	108.834102	91.1908491
26125130_y	0.03894915	7.96898149	131.447983	78.7850166
53080010_y	0.011082	22.9490001	882.591594	158.168917
53090030_y	0.00381083	21.5316708	912.037465	207.91591
53090040_y	0.00052672	25.0936379	1114.48702	188.65579
53100040_y	0.01839279	4.34950026	48.2520187	67.362111
53110030_y	0.02497826	3.26704063	29.2124057	104.773442
53110040_y	0.01297428	3.986813	49.0881771	94.7890382
53115010_y	0.00508344	19.2466732	718.74529	140.880395
54030010_y	0.01760836	4.7939557	57.247119	79.248992
54030030_y	0.01319423	4.59849503	73.2492298	97.5158373
54070030_y	0.03176787	20.4644535	850.130991	247.063882
54070150_y	0.02148103	20.7164965	855.370358	263.368319
54075020_y	0.00990082	19.7928251	744.510267	150.22622
54075040_y	0.03669631	19.4844443	761.20272	208.867514
26070170_y	0.02469224	5.25400958	98.3868789	121.248112
26080280_y	0.03710901	6.48714426	118.333941	121.655779
26090060_y	0.04215586	5.25199043	78.941315	104.713204
26095230_y	0.0186833	4.32853636	65.6992393	88.5944547
26100830_y	0.03645222	5.83850512	94.6204338	84.7512057
53110100_y	0.02443024	5.06564829	78.0401094	107.376805
26105230_y	0.01791903	4.95801455	72.0366307	73.9400687

## C2.2 MÉTRICAS MODELO REGRESIÓN LINEAL OLS SEMANAL

y	R2	MAE	MSE	MaxError
26055070_y	0.32864963	16.9986357	560.938055	90.1025389
26060020_y	0.34100352	17.1363537	585.820663	164.93098
26070110_y	0.44778332	18.82787	682.694629	119.228919
26070760_y	0.33018229	19.8532626	821.466434	179.460167
26075010_y	0.34081619	12.8539517	324.493418	88.7546994
26075040_y	0.32899614	10.6211572	227.898445	71.1578602
26080070_y	0.3115496	11.3683896	236.873121	65.1995029
26080290_y	0.34743879	14.1854362	366.038535	91.276504
26090460_y	0.39146814	20.5019799	758.525403	98.7460077
26090630_y	0.32683508	12.7050532	366.986817	164.823141
26095080_y	0.33509955	13.6443839	384.6094	133.313342
26100070_y	0.43567738	14.501509	359.616584	64.6201571
26100300_y	0.35319831	16.7066065	520.115267	115.710917
26100350_y	0.39965967	21.2689466	909.935216	155.799487
26100400_y	0.3255719	18.1979362	591.750203	98.4971992
26100410_y	0.41786489	18.1186604	609.197973	121.442773
26100690_y	0.32718268	16.9687397	530.736681	116.510939
26100700_y	0.21985163	23.65484	1099.31173	233.149549
26100770_y	0.34219523	21.0362629	806.285397	125.121007
26100780_y	0.3486167	14.8515446	382.150013	81.4165052
26100790_y	0.03852564	16.3763939	496.07762	164.696792
26100800_y	0.27798678	15.2783432	432.735933	88.8763744
26105150_y	0.3746509	17.0286689	558.393521	114.929573
26105160_y	0.29808679	15.9995599	480.268836	105.358762
26110040_y	0.34794193	13.3143605	347.109956	99.2983691
26110120_y	0.26861624	25.6058639	1207.41783	156.434438
26110150_y	0.35386386	16.9656564	536.7673	89.8510323
26110160_y	0.30039309	16.1589706	519.944983	120.490231
26110210_y	0.10546363	22.8847287	1137.54078	187.745311
26110230_y	0.31335018	12.9331017	329.694401	119.577807
26110290_y	0.31731374	14.3174823	610.550488	318.287225
26110300_y	0.3184522	12.6717003	287.058352	78.9546611
26115040_y	0.40268881	12.3681961	269.361654	75.0488498
26120120_y	0.43371654	21.1538168	829.810152	119.380692
26120130_y	0.33190589	21.3687977	849.884359	130.005277
26120150_y	0.27830567	23.3487568	963.0261	122.200642
26120180_y	0.35476334	20.6841171	736.744958	96.1519995
26125130_y	0.4388146	23.6042039	975.294716	145.740206
53080010_y	0.12536027	92.8626122	14104.1579	535.415918
53090030_y	0.11758464	74.6055663	11269.8903	754.300883
53090040_y	0.20692966	74.0507777	9879.74417	561.604855
53100040_y	0.31691437	14.6185264	421.683991	95.5888636
53110030_y	0.07954322	14.8114619	546.510746	264.231084
53110040_y	0.24342907	12.9908084	337.830137	135.731146
53115010_y	0.28002113	52.6140501	4695.99582	325.055468
54030010_y	0.22504278	16.471227	602.402056	204.782897
54030030_y	0.20606456	15.7961686	525.44186	136.177956
54070030_y	0.30918649	60.5600633	6361.4259	319.486543
54070150_y	0.27227478	64.275497	7280.09139	382.225008
54075020_y	0.08324349	73.225948	7895.2551	280.631696
54075040_y	0.19696039	68.1898167	7965.73559	333.312639
26070170_y	0.26266564	17.4113475	638.351012	158.176395
26080280_y	0.38865208	18.8696452	645.533023	99.0595969
26090060_y	0.38737407	16.0876638	585.897364	178.777487
26095230_y	0.22185441	14.1327839	397.976452	106.729387
26100830_y	0.29209649	17.3123451	613.040346	131.408014
53110100_y	0.31878599	15.5849944	402.619919	85.4242153
26105230_y	0.25997389	15.6720742	442.16032	88.2919524

### C2.3 MÉTRICAS MODELO REGRESIÓN LINEAL OLS MENSUAL

y	R2	MAE	MSE	MaxError
26055070_y	0.65447321	32.9372434	2325.54988	211.828315
26060020_y	0.7300919	32.9185495	1792.94339	174.288174
26070110_y	0.78139215	33.4997936	2247.90366	166.474143
26070760_y	0.75014324	37.6384799	2518.31503	181.056617
26075010_y	0.49497867	33.5093207	2251.73838	184.247352
26075040_y	0.5977065	22.493538	1195.60842	188.878343
26080070_y	0.53914573	23.475183	1117.68861	156.213672
26080290_y	0.55859312	29.7166161	1827.90586	193.50858
26090460_y	0.80060716	32.9070447	1987.94637	123.125873
26090630_y	0.65612409	26.0226274	1113.02979	98.6985674
26095080_y	0.50093956	33.3085667	2395.38894	172.204599
26100070_y	0.68717148	30.3126069	1728.38342	179.47715
26100300_y	0.67288729	28.4497586	1423.15781	114.695846
26100350_y	0.70757765	41.6514923	4418.45358	342.783289
26100400_y	0.49354691	44.5992131	4067.80083	247.589136
26100410_y	0.69413904	35.2767299	2674.45782	248.428349
26100690_y	0.61941779	35.3356434	2460.00409	169.400759
26100700_y	0.55870419	41.2615137	3795.89893	317.386094
26100770_y	0.71612242	39.2230854	3031.62656	227.248787
26100780_y	0.57262467	32.5671258	1792.08298	144.483258
26100790_y	0.38776409	39.3012846	3412.41636	375.740706
26100800_y	0.38399248	37.739858	2908.66868	245.946832
26105150_y	0.5809619	40.5354232	3168.97328	242.101146
26105160_y	0.60096778	34.0968642	1919.50116	154.002105
26110040_y	0.70937664	26.8013935	1321.10251	178.662466
26110120_y	0.6187752	54.0031943	5670.48578	264.085782
26110150_y	0.57759918	36.6728852	2474.23474	150.474519
26110160_y	0.59365988	27.217347	1300.05262	121.906097
26110210_y	0.19685073	61.3770077	6457.6469	274.690849
26110230_y	0.48241146	31.0565096	1812.65383	155.015344
26110290_y	0.51501518	28.0395029	3212.67158	512.072235
26110300_y	0.50857308	29.6575755	1494.01467	150.652045
26115040_y	0.68501118	22.6227577	845.057912	93.4677984
26120120_y	0.75929184	36.2616415	2220.22287	142.421164
26120130_y	0.57793476	40.6566841	3104.78103	198.972823
26120150_y	0.56475087	47.2919544	4251.88881	251.275793
26120180_y	0.5897584	40.1354744	2964.12131	190.811612
26125130_y	0.66471863	48.203163	4432.04019	235.513026
53080010_y	0.21320752	270.88184	140825.996	1289.46713
53090030_y	0.33289165	160.623398	58317.6171	799.355464
53090040_y	0.42532754	202.49645	72378.2781	766.382602
53100040_y	0.50406129	30.823213	2099.18375	219.4574
53110030_y	0.2685536	48.672193	4690.22667	304.813067
53110040_y	0.31897924	32.0106794	2576.38918	288.322636
53115010_y	0.57480861	102.536429	22702.3131	657.924397
54030010_y	0.38707215	33.0369986	3381.67637	423.131574
54030030_y	0.44807983	45.7421067	3961.31646	214.626822
54070030_y	0.41667968	147.008693	51972.2498	892.464833
54070150_y	0.39671903	153.753762	53613.3791	839.758781
54075020_y	0.06738154	249.774105	96248.9746	711.232372
54075040_y	0.20203972	189.736687	72944.9947	809.740551
26070170_y	0.5631475	40.2563582	3506.64032	221.594828
26080280_y	0.70388459	37.3993528	2169.9729	160.19717
26090060_y	0.754293	33.3903408	2247.48609	238.880165
26095230_y	0.33432832	34.3078706	2790.38596	238.919259
26100830_y	0.53217885	35.7937979	2339.51163	211.705103
53110100_y	0.69041035	28.2831385	1521.24281	126.46196
26105230_y	0.45856953	39.8461562	3290.92407	212.944261

### C3.1 MÉTRICAS MODELO REGRESIÓN SVR DIARIO

y	R2	MAE	MSE	MaxError
26055070_y	-0.11614083	4.15295155	109.204007	111.80001
26060020_y	-0.10422182	3.3855554	88.0556062	95.7975309
26070110_y	-0.11709773	4.63193156	159.30289	119.799849
26070760_y	-0.10686618	4.40078551	134.133	134.016555
26075010_y	-0.10801073	3.11999139	68.3963491	101.229254
26075040_y	-0.06961528	2.29421737	42.4278146	70.789356
26080070_y	-0.10141136	2.48028636	45.106712	66.8000014
26080290_y	-0.10782507	3.03287633	76.4169493	95.7999208
26090460_y	-0.13069356	4.695184	140.932645	135.800189
26090630_y	-0.10515933	2.80372745	61.2014124	93.800135
26095080_y	-0.12551697	3.31215392	46.2195005	74.6583138
26100070_y	-0.12489902	3.39197788	85.4864092	84.7882922
26100300_y	-0.13441131	4.22230513	104.933518	117.799377
26100350_y	-0.12653339	5.39344266	140.803006	104.799839
26100400_y	-0.16001396	4.52521288	126.945311	94.7996429
26100410_y	-0.1403695	4.23390118	103.270325	82.7999936
26100690_y	-0.11223732	3.84589202	74.132842	109.799829
26100700_y	-0.15963485	4.8828171	133.900315	97.7997503
26100770_y	-0.13863907	4.77640079	120.236893	82.7998537
26100780_y	-0.12814729	3.30558489	70.8454635	89.8005091
26100790_y	-0.12963391	3.9876062	115.26043	149.799717
26100800_y	-0.1638965	3.4068359	56.6572287	84.7999855
26105150_y	-0.14816109	4.27755859	105.950942	79.7002046
26105160_y	-0.12326911	3.5583205	80.2967689	90.4349273
26110040_y	-0.11569359	3.25018219	74.9580402	90.800134
26110120_y	-0.2058112	7.01421042	177.70154	173.229748
26110150_y	-0.1447843	4.19210156	97.1449169	77.394806
26110160_y	-0.15138152	3.7963093	83.9024653	84.8002311
26110210_y	-0.17956396	4.84990367	99.2952117	111.833912
26110230_y	-0.09936985	2.86579705	61.1743755	90.022866
26110290_y	-0.11479937	3.20936882	70.8432543	89.7999276
26110300_y	-0.12206466	3.0594041	65.6414091	86.8002618
26115040_y	-0.11343219	3.12305783	59.1756884	115.800007
26120120_y	-0.13821848	5.15319885	149.328299	114.602204
26120130_y	-0.16921811	4.81106232	135.811345	134.799777
26120150_y	-0.19881482	5.69668555	171.173722	130.799455
26120180_y	-0.15473049	4.83001062	128.581307	94.7997169
26125130_y	-0.16041663	6.42374632	158.716291	84.3003547
53080010_y	-0.1040799	21.3894689	985.371523	169.018867
53090030_y	-0.16180244	18.9389009	1063.66079	220.944806
53090040_y	-0.35132215	20.1790073	1506.82466	209.79948
53100040_y	-0.13424662	3.31078803	55.7551826	69.7996936
53110030_y	-0.10608922	2.85783801	33.1392887	107.502535
53110040_y	-0.11838293	2.83030503	55.621022	97.8002208
53115010_y	-0.15394624	16.697487	833.631136	151.87278
54030010_y	-0.16233098	3.69060067	67.7327627	82.796124
54030030_y	-0.12174933	3.34760649	83.265904	99.7996856
54070030_y	-0.1257692	17.7585243	988.452306	259.45742
54070150_y	-0.14832645	18.3562433	1003.80722	274.19983
54075020_y	-0.16850907	17.6725154	878.666519	161.799523
54075040_y	-0.14271522	16.9481745	902.973736	221.379796
26070170_y	-0.08836157	3.32708004	109.791496	123.799872
26080280_y	-0.13388185	4.9973234	139.347765	128.672583
26090060_y	-0.09664848	4.07925263	90.3809605	109.449384
26095230_y	-0.11326524	3.30304161	74.5332056	94.52481
26100830_y	-0.13606194	4.32448768	111.561332	89.800058
53110100_y	-0.13179046	3.58982478	90.5368888	109.800244
26105230_y	-0.12912218	3.69950479	82.8222518	76.6001548

### C3.2 MÉTRICAS MODELO REGRESIÓN SVR SEMANAL

y	R2	MAE	MSE	MaxError
26055070 y	0.33509362	15.7631098	555.553866	96.3480648
26060020 y	0.25798806	15.8989412	659.617985	170.548097
26070110 y	0.342982	18.3824968	812.258438	137.114937
26070760 y	0.34561296	17.8505153	802.54222	180.28587
26075010 y	0.2945395	12.1509182	347.273834	104.484834
26075040 y	0.29176613	9.90806696	240.543174	90.2173784
26080070 y	0.2523417	10.7334483	257.24461	69.3124725
26080290 y	0.26525837	13.5201055	412.135666	98.5984896
26090460 y	0.38575825	18.3735783	765.642683	103.569439
26090630 y	0.2832949	12.120676	390.723452	172.012697
26095080 y	0.3165153	12.9562144	395.359392	136.300271
26100070 y	0.36714667	14.167173	403.288027	78.0607739
26100300 y	0.28916366	16.6850763	571.607712	118.740381
26100350 y	0.34517429	20.2583598	992.518658	161.33766
26100400 y	0.29324552	17.4936489	620.113698	103.524973
26100410 y	0.35411401	17.6013951	675.912562	135.704237
26100690 y	0.30622745	16.290559	547.266738	118.998638
26100700 y	0.12704029	22.8587088	1230.09274	246.344632
26100770 y	0.2898367	20.3416539	870.462359	127.384909
26100780 y	0.28272221	14.1407614	420.808631	88.380206
26100790 y	0.01790518	14.6844608	506.716854	175.515661
26100800 y	0.22252974	14.4160682	465.973904	96.7933931
26105150 y	0.28015184	17.126026	642.774649	125.797748
26105160 y	0.23182132	15.4020878	525.609542	110.651701
26110040 y	0.32288579	12.6140649	360.448085	99.4312821
26110120 y	0.23437288	25.1751531	1263.94908	163.996785
26110150 y	0.33784933	16.3145318	550.071113	93.4935081
26110160 y	0.27734332	15.6225015	537.075484	130.748255
26110210 y	0.02957407	22.7234209	1234.04604	195.931524
26110230 y	0.25888773	12.0749371	355.844504	123.257338
26110290 y	0.25975272	13.5518346	662.029341	328.811887
26110300 y	0.28428154	11.8617545	301.450553	84.492677
26115040 y	0.38059116	11.6897479	279.326745	80.8992352
26120120 y	0.39479627	20.4414384	886.842437	122.609687
26120130 y	0.29493284	20.4511344	896.917866	132.343639
26120150 y	0.25422627	22.7866259	995.157561	146.960698
26120180 y	0.32915574	19.4582438	765.984262	106.100391
26125130 y	0.38198215	23.0618132	1074.06492	151.366634
53080010 y	0.09839435	94.1382054	14539.0015	546.393386
53090030 y	0.13480612	73.2464935	11049.9437	763.547759
53090040 y	0.12470639	75.6003721	10904.0479	573.922948
53100040 y	0.22800894	14.533836	476.567299	118.557632
53110030 y	0.01958267	13.6258926	582.111638	270.964593
53110040 y	0.19919987	12.0729157	357.579713	142.139341
53115010 y	0.27974441	52.8129331	4697.8007	304.837316
54030010 y	0.18741552	15.7782671	631.651072	209.029318
54030030 y	0.14782924	15.1630429	563.983119	146.573564
54070030 y	0.26632554	61.9613366	6756.11524	457.183167
54070150 y	0.23296058	64.8656016	7673.38679	467.764608
54075020 y	0.06847721	73.9150374	8022.42464	279.345762
54075040 y	0.16440325	68.5610062	8288.68554	474.602229
26070170 y	0.19932453	15.8735099	693.188904	165.436524
26080280 y	0.35669116	18.0597113	679.281122	104.857433
26090060 y	0.36309649	15.1262139	609.115731	184.929917
26095230 y	0.20474259	13.1492373	406.728158	111.842228
26100830 y	0.2061227	17.0686684	687.493156	130.199858
53110100 y	0.28747166	14.5444018	421.127716	91.3007062
26105230 y	0.23438345	14.7717107	457.450425	93.1212942

### C3.3 MÉTRICAS MODELO REGRESIÓN SVR MENSUAL

y	R2	MAE	MSE	MaxError
26055070 y	0.393304	46.8076578	4083.33549	338.703916
26060020 y	0.35545704	49.6525302	4281.5649	218.458949
26070110 y	0.32522908	56.5013966	6938.54311	362.411344
26070760 y	0.36638198	56.2823294	6386.25818	321.80174
26075010 y	0.30643658	40.6373412	3092.39089	308.487979
26075040 y	0.37159227	27.9457284	1867.61549	303.489212
26080070 y	0.4263919	27.4038226	1391.14526	140.006693
26080290 y	0.27713284	37.4292027	2993.4584	308.469717
26090460 y	0.40334087	52.7502186	5948.69081	432.93717
26090630 y	0.36130333	34.2605696	2067.28183	211.477905
26095080 y	0.34796964	40.4759891	3129.61352	168.900565
26100070 y	0.28766415	40.652258	3935.66889	345.391392
26100300 y	0.34155998	38.7971796	2864.65192	182.51425
26100350 y	0.23155157	65.6999858	11611.1292	624.487921
26100400 y	0.28171598	55.1364436	5769.21416	358.446747
26100410 y	0.22791941	52.0420979	6751.09693	353.378669
26100690 y	0.37028037	46.7549318	4070.37649	210.683946
26100700 y	0.30396801	54.2195336	5987.06598	378.750067
26100770 y	0.31530539	59.7838545	7312.08984	365.264818
26100780 y	0.2539656	40.9427886	3128.29368	313.137676
26100790 y	0.16620641	46.6626779	4647.31133	380.076433
26100800 y	0.29632478	43.6754167	3322.61869	219.646164
26105150 y	0.30414371	54.8351238	5262.40932	334.623407
26105160 y	0.34197618	41.897598	3165.3521	204.965945
26110040 y	0.29472716	36.6382449	3205.99732	293.717906
26110120 y	0.22394815	78.4195352	11543.2967	444.839101
26110150 y	0.26765249	46.7201265	4289.76356	247.720726
26110160 y	0.35124755	33.4407323	2075.63142	155.537606
26110210 y	0.11546929	67.5111023	7111.98672	275.345563
26110230 y	0.24961582	35.8949686	2627.93058	300.2714
26110290 y	0.23266324	37.9474627	5083.04778	613.519092
26110300 y	0.28056756	32.7814249	2187.18712	233.008324
26115040 y	0.40801005	28.6353935	1588.20172	157.373548
26120120 y	0.40613062	53.5125464	5477.68038	315.327002
26120130 y	0.2928031	53.692007	5202.25625	223.286892
26120150 y	0.22175846	60.3832587	7602.53444	483.947516
26120180 y	0.36355668	49.2829812	4598.49803	277.439334
26125130 y	0.35631738	66.7577052	8508.75552	362.638284
53080010 y	0.00417932	329.359792	178239.426	1454.59929
53090030 y	0.09051664	209.961487	79505.6789	902.106853
53090040 y	0.07381404	272.391292	116650.351	1126.91188
53100040 y	0.26438124	38.8975016	3113.68907	192.279853
53110030 y	0.03920678	50.6369566	6160.85879	427.47428
53110040 y	0.22250405	35.4706187	2941.36722	284.452211
53115010 y	0.18755043	153.587514	43379.2515	591.820985
54030010 y	0.19186501	41.1608241	4458.683	440.154286
54030030 y	0.14120925	53.2438128	6163.82969	314.591343
54070030 y	0.15012629	211.011901	75721.4305	672.342439
54070150 y	0.12386274	213.759625	77862.0262	906.253841
54075020 y	-0.08517919	263.215808	111993.691	572.466908
54075040 y	0.13208461	224.717192	79339.8937	728.881754
26070170 y	0.23786105	50.289706	6117.73347	424.837652
26080280 y	0.40741282	52.1298235	4342.55725	198.155214
26090060 y	0.37211547	49.4735809	5743.27039	424.884059
26095230 y	0.3007899	38.8585277	2930.97347	205.731717
26100830 y	0.25805442	44.367603	3710.37163	263.427818
53110100 y	0.35552212	37.706061	3166.79622	248.269974
26105230 y	0.24891001	47.0337142	4565.27708	268.643455

### C4.1 MÉTRICAS MODELO REGRESIÓN DTR DIARIO

y	R2	MAE	MSE	MaxError
26055070 y	-0.33862966	6.09047924	130.972472	108.81943
26060020 y	-0.3518347	5.36180701	107.801369	96
26070110 y	-0.38380209	7.31215822	197.336067	116.441751
26070760 y	-0.36468139	6.59916181	165.375736	136
26075010 y	-0.26953645	4.44333918	78.3671633	96.2
26075040 y	-0.45276328	3.69474282	57.6259262	67.1707286
26080070 y	-0.3039575	3.82768283	53.4016964	62.9681089
26080290 y	-0.4940889	4.95528905	103.061141	96
26090460 y	-0.32045675	7.07134134	164.585232	132.33723
26090630 y	-0.37366689	4.29886952	76.0707995	91.9512934
26095080 y	-0.24908715	4.34000426	51.2939261	77.2
26100070 y	-0.54435911	5.47223244	117.36317	88
26100300 y	-0.45932151	6.34538256	134.987847	118
26100350 y	-0.31359126	7.33760593	164.182969	100.865264
26100400 y	-0.33844547	6.78159507	146.471837	87
26100410 y	-0.60221177	6.61545411	145.094138	110
26100690 y	-0.24792383	5.31755769	83.1766189	106.795043
26100700 y	-0.54363225	7.38643686	178.239594	110
26100770 y	-0.40845222	6.94594693	148.728359	80
26100780 y	-0.56981067	5.25164146	98.5810677	85
26100790 y	-0.21724908	5.77514431	124.200107	147
26100800 y	-0.29696738	4.66527408	63.1349761	82.3632385
26105150 y	-0.47427688	6.41303565	136.04452	106
26105160 y	-0.41251048	5.3808964	100.973156	87.7
26110040 y	-0.48406502	5.07406896	99.7071296	88.689693
26110120 y	-0.36506602	9.09699479	201.171073	172
26110150 y	-0.41076447	6.12942143	119.71565	81.2
26110160 y	-0.57965074	5.94029577	115.110925	124
26110210 y	-0.43485521	6.45104879	120.785524	114
26110230 y	-0.50437629	4.46139169	83.7109368	85.8
26110290 y	-0.46019224	5.02375128	92.7922757	87.4545435
26110300 y	-0.90057826	5.01241491	111.184889	92
26115040 y	-0.54966955	4.76931855	82.3604374	113.74355
26120120 y	-0.32013249	7.41039189	173.194462	117
26120130 y	-0.42907423	7.47611287	165.995115	135
26120150 y	-0.44853866	8.39051927	206.830739	125
26120180 y	-0.46639995	7.27275655	163.28626	90.9822977
26125130 y	-0.34080553	8.83454194	183.38903	83.5
53080010 y	-0.48263559	26.6702061	1323.22569	150
53090030 y	-0.50669262	24.9503422	1379.41684	220
53090040 y	-0.63132527	28.2346795	1819.04896	335
53100040 y	-0.67376908	5.03413824	82.2760223	67.1671271
53110030 y	-0.47679126	3.7239534	44.2458085	101
53110040 y	-0.26016552	4.14728925	62.6723567	96
53115010 y	-0.54397048	22.5026064	1115.39153	159.8
54030010 y	-0.44063264	5.23755101	83.9502945	83
54030030 y	-0.37508449	4.82125966	102.070624	133
54070030 y	-0.60033583	24.3966503	1405.13317	233
54070150 y	-0.43678925	23.7667529	1255.96639	262.810705
54075020 y	-0.56089563	23.3410385	1173.72365	158.5
54075040 y	-0.54268788	22.8227823	1219.03219	208.2
26070170 y	-0.29981597	5.48114934	131.122546	121.204402
26080280 y	-0.38607565	7.14701623	170.340978	132
26090060 y	-0.32997866	5.76915931	109.611011	118
26095230 y	-0.27863458	4.61385999	85.6046977	91.6
26100830 y	-0.33113863	6.27213204	130.717871	90
53110100 y	-0.44836875	5.49822289	115.861376	107.28752
26105230 y	-0.48581977	5.54573906	108.986381	82.2

## C4.2 MÉTRICAS MODELO REGRESIÓN DTR SEMANAL

y	R2	MAE	MSE	MaxError
26055070_y	-0.25120675	22.4242279	1045.4295	148.3
26060020_y	-0.38075739	23.3881373	1227.43632	183
26070110_y	0.04824808	24.3599942	1176.6322	147
26070760_y	-0.23103373	26.7514811	1509.74345	201
26075010_y	-0.3651432	17.0734785	672.012835	184.5
26075040_y	-0.41801536	14.9886633	481.611979	117.1
26080070_y	-0.40557504	15.5903063	483.6121	93
26080290_y	-0.28717509	18.2340493	722.009946	118
26090460_y	-0.15131033	26.5306982	1435.09023	123
26090630_y	-0.13171076	16.6485454	616.970546	153
26095080_y	-0.11587835	17.7340353	645.476028	130.264762
26100070_y	-0.18244443	19.3030558	753.5169	124
26100300_y	-0.28316906	23.0969208	1031.83995	143
26100350_y	-0.37189093	30.9141854	2079.37366	182
26100400_y	-0.45263519	25.0160183	1274.55716	179
26100410_y	-0.26605107	25.553531	1324.90848	186
26100690_y	-0.1162935	21.6500812	880.562814	110.854749
26100700_y	-0.26838891	29.164423	1787.29439	250
26100770_y	-0.33083818	28.7644976	1631.23686	147
26100780_y	-0.46674372	20.5968	860.501225	125
26100790_y	-1.41929112	22.6210869	1248.24565	195
26100800_y	-0.3125683	19.771434	786.682925	106
26105150_y	-0.24865422	22.918801	1114.96192	150
26105160_y	-0.38699747	21.6533719	949.022832	114.1
26110040_y	-0.39652767	18.3965183	743.413327	114
26110120_y	-0.24856096	33.7368937	2061.20896	180
26110150_y	-0.28965141	23.6867453	1071.35735	113
26110160_y	-0.17835517	21.5482378	875.748732	134
26110210_y	-0.71371151	31.5471402	2179.24813	201
26110230_y	-0.55881496	18.7532138	748.463841	141.5
26110290_y	-0.05412476	19.1477619	942.741084	299
26110300_y	-0.52420594	17.4649651	641.974114	104
26115040_y	-0.19439736	16.8649379	538.621837	91.6
26120120_y	-0.05013045	27.6848011	1538.82105	162
26120130_y	-0.18383858	27.5823303	1505.96432	157
26120150_y	-0.25883357	30.671411	1679.78262	188.3
26120180_y	-0.30105825	26.6494833	1485.57602	149
26125130_y	0.08118243	29.7456093	1596.83044	176
53080010_y	-0.68467193	127.692497	27166.4755	721
53090030_y	-0.45880234	93.8572342	18631.2966	830
53090040_y	-0.53619351	103.543622	19137.267	654
53100040_y	-0.39759578	20.5910707	862.767045	135
53110030_y	-0.525439	19.1818441	905.712059	267
53110040_y	-0.24130865	16.5861123	554.279119	139
53115010_y	-0.37511516	73.2018718	8969.06185	388.7
54030010_y	-0.40019916	22.0825121	1088.42505	213
54030030_y	-0.8035446	21.7787459	1193.62077	179
54070030_y	-0.42011641	86.3284238	13077.2852	408
54070150_y	-0.27709298	84.2665452	12775.9123	418
54075020_y	-0.76677394	93.5957124	15215.7424	385
54075040_y	-0.57842407	94.2773089	15657.1464	440
26070170_y	-0.41565247	22.8982454	1225.60841	173
26080280_y	-0.19341891	25.2627362	1260.15202	145
26090060_y	-0.14859731	21.7104715	1098.48458	188
26095230_y	-0.58565528	19.6756252	810.970945	126.9
26100830_y	-0.19710256	22.7995856	1036.6839	129
53110100_y	-0.39101362	20.2357113	822.134863	110
26105230_y	-0.62648033	21.7512507	971.810397	125.3

### C4.3 MÉTRICAS MODELO REGRESIÓN DTR MENSUAL

y	R2	MAE	MSE	MaxError
26055070_y	0.45276331	45.7568966	3683.14776	195
26060020_y	0.52209024	42.9310345	3174.65517	199
26070110_y	0.51134019	53.9741379	5024.79741	211
26070760_y	0.49776685	50.0094828	5062.02543	233
26075010_y	-0.10409664	47.312069	4922.83517	384.5
26075040_y	-0.15610217	36.0801724	3435.91302	381.2
26080070_y	0.20672009	32.4396552	1923.90517	199
26080290_y	0.29947239	41.5517241	2900.94828	161
26090460_y	0.39607162	56.487069	6021.16526	300
26090630_y	0.28363598	36.0525862	2318.66922	155.4
26095080_y	-0.00835974	52.7215517	4839.92233	205.2
26100070_y	0.41760066	41.2844828	3217.76724	209
26100300_y	0.33788709	40.1853448	2880.63147	163
26100350_y	0.46412432	60.3336207	8096.99319	337.5
26100400_y	0.05523409	63.512069	7588.3031	294
26100410_y	0.50316634	49.1172414	4344.32914	212
26100690_y	0.11307262	52.362069	5732.91379	274
26100700_y	-0.40172927	72.2474138	12057.2699	427
26100770_y	0.3877259	58.3181034	6538.68629	260
26100780_y	0.15712219	46.587931	3534.38034	160
26100790_y	-0.51704094	54.2810345	8455.52379	599
26100800_y	-0.25849364	52.4008621	5942.36422	383
26105150_y	0.31542811	54.1422414	5177.07112	312.4
26105160_y	0.16473548	46.4387931	4017.94922	188.6
26110040_y	0.42020814	37.8344828	2635.59155	205
26110120_y	0.0130316	89.8913793	14680.5514	499
26110150_y	0.26658719	47.8086207	4296.00362	210.7
26110160_y	0.09606183	41.9232759	2892.07767	206
26110210_y	-1.78721308	105.415517	22410.3271	619
26110230_y	-0.05139406	40.9086207	3682.10138	384.5
26110290_y	0.1262645	47.7413793	5787.86207	504
26110300_y	0.03603744	39.2396552	2930.5969	231
26115040_y	0.45201276	29.825	1470.15043	135.2
26120120_y	0.42658302	57.1491379	5289.03336	207
26120130_y	0.07623388	59.0112069	6795.37491	337
26120150_y	0.16645531	67.7637931	8142.78328	283.8
26120180_y	0.19700746	56.6913793	5801.86724	244
26125130_y	0.55439609	58.7698276	5890.37922	237.4
53080010_y	-0.428606	363.387069	255702.575	2347
53090030_y	-0.50053528	263.343103	131174.556	1083
53090040_y	-0.52222792	307.962069	191720.053	1680
53100040_y	-0.0401074	47.3560345	4402.51284	229
53110030_y	-1.20833059	77.5	14160.3966	768
53110040_y	0.0287122	42	3674.50672	265
53115010_y	-0.09477735	167.983621	58453.6247	832.2
54030010_y	-0.21207773	44.3189655	6687.33621	541
54030030_y	-0.19316916	66.5172414	8563.77586	290
54070030_y	0.06948448	202.165517	82906.396	1008
54070150_y	-0.41782185	241.883621	126001.356	1190
54075020_y	-1.29982081	373.190517	237348.286	1101.5
54075040_y	-0.63395803	249.322414	149367.159	1286.1
26070170_y	0.41786976	47.6551724	4672.7931	235
26080280_y	0.41031793	51.8646552	4321.26819	162
26090060_y	0.61105426	42.7758621	3557.69328	250
26095230_y	-0.10638985	48.7077586	4637.80388	310.1
26100830_y	-0.10852669	56.9405172	5543.59526	223
53110100_y	0.41249672	40.9844828	2886.83793	161
26105230_y	0.06764817	56.962069	5667.02328	202.9

### C5.1 MÉTRICAS MODELO CLASIFICACIÓN LOGIT DIARIO

y	Accuracy	Recall	Precision	F1
26055070_y	0.58511548	0.44562022	0.59685039	0.5102659
26060020_y	0.74165954	0	0	0
26070110_y	0.71457086	0	0	0
26070760_y	0.63815227	0.44427711	0.52631579	0.48182932
26075010_y	0.58283433	0.38330394	0.61163227	0.47126852
26075040_y	0.61819219	0.41126761	0.53725851	0.46589549
26080070_y	0.70516111	0	0	0
26080290_y	0.74964357	0	0	0
26090460_y	0.68976333	0	0	0
26090630_y	0.74650699	0	0	0
26095080_y	0.61847733	1	0.61847733	0.76427061
26100070_y	0.76703735	0	0	0
26100300_y	0.65269461	0	0	0
26100350_y	0.61534075	0.46802721	0.54820717	0.50495413
26100400_y	0.66923296	0	0	0
26100410_y	0.66866267	0	0	0
26100690_y	0.61648132	0.43781095	0.61484716	0.51144206
26100700_y	0.61961791	0	0	0
26100770_y	0.60136869	0	0	0
26100780_y	0.70116909	0	0	0
26100790_y	0.53949244	0.37988827	0.49837134	0.43113772
26100800_y	0.59908754	0	0	0
26105150_y	0.57798688	0.46456693	0.56272927	0.5089582
26105160_y	0.55489022	0.44197138	0.62146051	0.5165686
26110040_y	0.71799259	0	0	0
26110120_y	0.55631594	0.55039638	0.56055363	0.55542857
26110150_y	0.63615626	0	0	0
26110160_y	0.6720844	0	0	0
26110210_y	0.605931	1	0.605931	0.75461648
26110230_y	0.6104933	0	0	0
26110290_y	0.69689193	0	0	0
26110300_y	0.67892786	0	0	0
26115040_y	0.57770174	0.49550706	0.52624404	0.51041322
26120120_y	0.59965783	0.49043232	0.50254176	0.4964132
26120130_y	0.67978329	0	0	0
26120150_y	0.63416025	0	0	0
26120180_y	0.63587111	0	0	0
26125130_y	0.61961791	1	0.61961791	0.76514085
53080010_y	0.85970915	1	0.85970915	0.92456302
53090030_y	0.61876248	1	0.61876248	0.76448829
53090040_y	0.54719133	0	0	0
53100040_y	0.56372968	0.45688073	0.53779698	0.49404762
53110030_y	0.72141431	1	0.72141431	0.83816465
53110040_y	0.68434559	0	0	0
53115010_y	0.81323068	1	0.81323068	0.89699638
54030010_y	0.58426005	0	0	0
54030030_y	0.62589107	0	0	0
54070030_y	0.75220987	1	0.75220987	0.85858421
54070150_y	0.75876818	1	0.75876818	0.86284047
54075020_y	0.88964927	1	0.88964927	0.94160254
54075040_y	0.83290562	1	0.83290562	0.90883634
26070170_y	0.75220987	0	0	0
26080280_y	0.62902766	0	0	0
26090060_y	0.61077844	0.43215031	0.53076923	0.47640967
26095230_y	0.55317936	0.38816504	0.61904762	0.47714381
26100830_y	0.64528087	0	0	0
53110100_y	0.679213	0	0	0
26105230_y	0.58026804	0.43790465	0.59000793	0.5027027

## C5.2 MÉTRICAS MODELO CLASIFICACIÓN LOGIT SEMANAL

y	Accuracy	Recall	Precision	F1
26055070 y	0.88844622	1	0.88844622	0.94092827
26060020 y	0.72908367	0.89182058	0.78060046	0.83251232
26070110 y	0.73306773	0.87165775	0.79126214	0.82951654
26070760 y	0.83266932	1	0.83266932	0.90869565
26075010 y	0.73904382	1	0.73904382	0.84994273
26075040 y	0.87051793	1	0.87051793	0.93077742
26080070 y	0.75697211	1	0.75697211	0.861678
26080290 y	0.79681275	1	0.79681275	0.88691796
26090460 y	0.70916335	1	0.70916335	0.82983683
26090630 y	0.74900398	1	0.74900398	0.85649203
26095080 y	0.79681275	1	0.79681275	0.88691796
26100070 y	0.74103586	0.96388889	0.74784483	0.84223301
26100300 y	0.83067729	1	0.83067729	0.90750816
26100350 y	0.83067729	1	0.83067729	0.90750816
26100400 y	0.7310757	1	0.7310757	0.84464902
26100410 y	0.79880478	1	0.79880478	0.88815061
26100690 y	0.82071713	1	0.82071713	0.90153173
26100700 y	0.78087649	1	0.78087649	0.87695749
26100770 y	0.82868526	1	0.82868526	0.90631808
26100780 y	0.78087649	0.94278607	0.81330472	0.87327189
26100790 y	0.64940239	1	0.64940239	0.78743961
26100800 y	0.71314741	1	0.71314741	0.83255814
26105150 y	0.80677291	1	0.80677291	0.89305402
26105160 y	0.84262948	1	0.84262948	0.91459459
26110040 y	0.74302789	1	0.74302789	0.85257143
26110120 y	0.90239044	1	0.90239044	0.9486911
26110150 y	0.85458167	1	0.85458167	0.92158969
26110160 y	0.83665339	1	0.83665339	0.91106291
26110210 y	0.79681275	1	0.79681275	0.88691796
26110230 y	0.80278884	1	0.80278884	0.89060773
26110290 y	0.79880478	1	0.79880478	0.88815061
26110300 y	0.82071713	0.95683453	0.84713376	0.89864865
26115040 y	0.87450199	1	0.87450199	0.93304995
26120120 y	0.83864542	1	0.83864542	0.91224269
26120130 y	0.79880478	0.955	0.82150538	0.88323699
26120150 y	0.81474104	1	0.81474104	0.89791438
26120180 y	0.84262948	1	0.84262948	0.91459459
26125130 y	0.88645418	1	0.88645418	0.93980993
53080010 y	0.92629482	1	0.92629482	0.96173733
53090030 y	0.9123506	1	0.9123506	0.95416667
53090040 y	0.91434263	1	0.91434263	0.95525494
53100040 y	0.87450199	1	0.87450199	0.93304995
53110030 y	0.93824701	1	0.93824701	0.96813977
53110040 y	0.70119522	1	0.70119522	0.82435597
53115010 y	0.96215139	1	0.96215139	0.98071066
54030010 y	0.88047809	1	0.88047809	0.93644068
54030030 y	0.70916335	1	0.70916335	0.82983683
54070030 y	0.93625498	1	0.93625498	0.96707819
54070150 y	0.92430279	1	0.92430279	0.96066253
54075020 y	0.7310757	1	0.7310757	0.84464902
54075040 y	0.85856574	1	0.85856574	0.92390139
26070170 y	0.64342629	0.88673139	0.65550239	0.75378267
26080280 y	0.85258964	1	0.85258964	0.92043011
26090060 y	0.86454183	1	0.86454183	0.92735043
26095230 y	0.81474104	1	0.81474104	0.89791438
26100830 y	0.84262948	1	0.84262948	0.91459459
53110100 y	0.77091633	1	0.77091633	0.87064117
26105230 y	0.82270916	1	0.82270916	0.90273224

### C5.3 MÉTRICAS MODELO CLASIFICACIÓN LOGIT MENSUAL

y	Accuracy	Recall	Precision	F1
26055070 y	0.99137931	1	0.99137931	0.995671
26060020 y	0.97413793	1	0.97413793	0.98689956
26070110 y	0.93103448	1	0.93103448	0.96428571
26070760 y	0.98275862	1	0.98275862	0.99130435
26075010 y	0.84482759	1	0.84482759	0.91588785
26075040 y	0.99137931	1	0.99137931	0.995671
26080070 y	0.97413793	1	0.97413793	0.98689956
26080290 y	0.96551724	1	0.96551724	0.98245614
26090460 y	0.93965517	1	0.93965517	0.96888889
26090630 y	0.94827586	1	0.94827586	0.97345133
26095080 y	0.89655172	1	0.89655172	0.94545455
26100070 y	1	1	1	1
26100300 y	0.99137931	1	0.99137931	0.995671
26100350 y	0.97413793	1	0.97413793	0.98689956
26100400 y	0.89655172	1	0.89655172	0.94545455
26100410 y	0.96551724	1	0.96551724	0.98245614
26100690 y	0.96551724	1	0.96551724	0.98245614
26100700 y	0.93965517	1	0.93965517	0.96888889
26100770 y	0.96551724	1	0.96551724	0.98245614
26100780 y	0.99137931	1	0.99137931	0.995671
26100790 y	0.90517241	1	0.90517241	0.95022624
26100800 y	0.92241379	1	0.92241379	0.95964126
26105150 y	0.87931034	1	0.87931034	0.93577982
26105160 y	0.87068966	1	0.87068966	0.93087558
26110040 y	0.96551724	1	0.96551724	0.98245614
26110120 y	0.97413793	1	0.97413793	0.98689956
26110150 y	0.93965517	1	0.93965517	0.96888889
26110160 y	0.99137931	1	0.99137931	0.995671
26110210 y	0.89655172	1	0.89655172	0.94545455
26110230 y	0.95689655	1	0.95689655	0.97797357
26110290 y	0.94827586	1	0.94827586	0.97345133
26110300 y	0.95689655	1	0.95689655	0.97797357
26115040 y	0.96551724	1	0.96551724	0.98245614
26120120 y	0.96551724	1	0.96551724	0.98245614
26120130 y	0.95689655	1	0.95689655	0.97797357
26120150 y	0.99137931	1	0.99137931	0.995671
26120180 y	1	1	1	1
26125130 y	0.95689655	1	0.95689655	0.97797357
53080010 y	0.92241379	1	0.92241379	0.95964126
53090030 y	0.93103448	1	0.93103448	0.96428571
53090040 y	0.94827586	1	0.94827586	0.97345133
53100040 y	0.96551724	1	0.96551724	0.98245614
53110030 y	0.95689655	1	0.95689655	0.97797357
53110040 y	0.87068966	1	0.87068966	0.93087558
53115010 y	0.97413793	1	0.97413793	0.98689956
54030010 y	0.99137931	1	0.99137931	0.995671
54030030 y	0.87068966	1	0.87068966	0.93087558
54070030 y	0.9137931	1	0.9137931	0.95495495
54070150 y	0.90517241	1	0.90517241	0.95022624
54075020 y	0.75	1	0.75	0.85714286
54075040 y	0.87068966	1	0.87068966	0.93087558
26070170 y	0.88793103	1	0.88793103	0.94063927
26080280 y	1	1	1	1
26090060 y	1	1	1	1
26095230 y	0.9137931	1	0.9137931	0.95495495
26100830 y	0.98275862	1	0.98275862	0.99130435
53110100 y	0.93103448	1	0.93103448	0.96428571
26105230 y	0.89655172	1	0.89655172	0.94545455

### C6.1 MÉTRICAS MODELO CLASIFICACIÓN SVC DIARIO

y	Accuracy	Recall	Precision	F1
26055070 y	0.58140861	0.42650602	0.57843137	0.49098474
26060020 y	0.74650699	0	0	0
26070110 y	0.70402053	0	0	0
26070760 y	0.63216424	0.41937926	0.51439183	0.46205171
26075010 y	0.5819789	0.38489209	0.59334566	0.46690909
26075040 y	0.61819219	0.42196133	0.54896676	0.47715736
26080070 y	0.71171942	0	0	0
26080290 y	0.74907328	0	0	0
26090460 y	0.69232963	0	0	0
26090630 y	0.74508127	0	0	0
26095080 y	0.61420017	1	0.61420017	0.76099629
26100070 y	0.77302538	0	0	0
26100300 y	0.64955803	0	0	0
26100350 y	0.59709153	0.45276653	0.52710134	0.48711434
26100400 y	0.6603935	0	0	0
26100410 y	0.66894782	0	0	0
26100690 y	0.59509552	0.42368588	0.56743003	0.48513416
26100700 y	0.60393499	0	0	0
26100770 y	0.61334474	0	0	0
26100780 y	0.69261477	0	0	0
26100790 y	0.5463359	0.382134	0.50867052	0.43641516
26100800 y	0.60764186	0	0	0
26105150 y	0.57057314	0.46543504	0.5618705	0.50912647
26105160 y	0.56686627	0.46008403	0.64081931	0.53561602
26110040 y	0.71628172	0	0	0
26110120 y	0.5489022	0.54381006	0.55382844	0.54877353
26110150 y	0.63330482	0	0	0
26110160 y	0.66295979	0	0	0
26110210 y	0.60678643	1	0.60678643	0.7552795
26110230 y	0.60479042	0	0	0
26110290 y	0.69289991	0	0	0
26110300 y	0.69175934	0	0	0
26115040 y	0.59024808	0.51462766	0.52261985	0.51859296
26120120 y	0.60536071	0.49723375	0.52252907	0.50956768
26120130 y	0.68149415	0	0	0
26120150 y	0.64556601	0	0	0
26120180 y	0.62788708	0	0	0
26125130 y	0.61248931	1	0.61248931	0.7596817
53080010 y	0.85486171	1	0.85486171	0.9217525
53090030 y	0.61534075	1	0.61534075	0.76187114
53090040 y	0.53264899	0	0	0
53100040 y	0.56287425	0.46666667	0.53581461	0.49885584
53110030 y	0.69803251	1	0.69803251	0.82216625
53110040 y	0.68006843	0	0	0
53115010 y	0.81437126	1	0.81437126	0.89768977
54030010 y	0.57513544	0	0	0
54030030 y	0.61362988	0	0	0
54070030 y	0.75591674	1	0.75591674	0.86099383
54070150 y	0.75278015	1	0.75278015	0.85895559
54075020 y	0.88651269	1	0.88651269	0.93984281
54075040 y	0.83632735	1	0.83632735	0.91086957
26070170 y	0.7459367	0	0	0
26080280 y	0.63045338	0	0	0
26090060 y	0.61961791	0.41724618	0.54744526	0.47355959
26095230 y	0.54833191	0.3869239	0.62134251	0.47688243
26100830 y	0.64271457	0	0	0
53110100 y	0.67522099	0	0	0
26105230 y	0.57199886	0.44051074	0.58564815	0.5028155

## C6.2 MÉTRICAS MODELO CLASIFICACIÓN SVC SEMANAL

y	Accuracy	Recall	Precision	F1
26055070 y	0.89840637	1	0.89840637	0.94648478
26060020 y	0.71713147	0.89944134	0.75233645	0.81933842
26070110 y	0.71713147	1	0.71713147	0.83526682
26070760 y	0.84262948	1	0.84262948	0.91459459
26075010 y	0.75298805	1	0.75298805	0.85909091
26075040 y	0.83864542	1	0.83864542	0.91224269
26080070 y	0.73705179	1	0.73705179	0.84862385
26080290 y	0.7749004	1	0.7749004	0.87317621
26090460 y	0.75298805	1	0.75298805	0.85909091
26090630 y	0.73705179	1	0.73705179	0.84862385
26095080 y	0.84262948	1	0.84262948	0.91459459
26100070 y	0.73705179	0.9403794	0.75929978	0.8401937
26100300 y	0.85458167	1	0.85458167	0.92158969
26100350 y	0.83665339	1	0.83665339	0.91106291
26100400 y	0.75498008	0.96226415	0.76609442	0.85304659
26100410 y	0.78685259	1	0.78685259	0.88071349
26100690 y	0.84462151	1	0.84462151	0.91576674
26100700 y	0.7749004	1	0.7749004	0.87317621
26100770 y	0.83466135	1	0.83466135	0.90988056
26100780 y	0.80677291	1	0.80677291	0.89305402
26100790 y	0.71513944	1	0.71513944	0.83391405
26100800 y	0.70916335	1	0.70916335	0.82983683
26105150 y	0.78087649	1	0.78087649	0.87695749
26105160 y	0.8187251	1	0.8187251	0.90032859
26110040 y	0.78884462	0.95431472	0.81034483	0.87645688
26110120 y	0.88844622	1	0.88844622	0.94092827
26110150 y	0.812749	1	0.812749	0.8967033
26110160 y	0.84262948	1	0.84262948	0.91459459
26110210 y	0.79681275	1	0.79681275	0.88691796
26110230 y	0.82669323	1	0.82669323	0.90512541
26110290 y	0.78685259	1	0.78685259	0.88071349
26110300 y	0.84262948	1	0.84262948	0.91459459
26115040 y	0.8625498	1	0.8625498	0.92620321
26120120 y	0.81673307	1	0.81673307	0.89912281
26120130 y	0.75896414	1	0.75896414	0.86296716
26120150 y	0.83466135	1	0.83466135	0.90988056
26120180 y	0.83067729	1	0.83067729	0.90750816
26125130 y	0.9123506	1	0.9123506	0.95416667
53080010 y	0.92629482	1	0.92629482	0.96173733
53090030 y	0.92231076	1	0.92231076	0.95958549
53090040 y	0.88844622	1	0.88844622	0.94092827
53100040 y	0.88844622	1	0.88844622	0.94092827
53110030 y	0.93027888	1	0.93027888	0.96388029
53110040 y	0.69521912	1	0.69521912	0.82021152
53115010 y	0.97609562	1	0.97609562	0.98790323
54030010 y	0.87250996	1	0.87250996	0.93191489
54030030 y	0.74103586	1	0.74103586	0.85125858
54070030 y	0.92231076	1	0.92231076	0.95958549
54070150 y	0.92231076	1	0.92231076	0.95958549
54075020 y	0.71912351	1	0.71912351	0.83661645
54075040 y	0.85657371	1	0.85657371	0.92274678
26070170 y	0.64940239	0.88141026	0.66425121	0.75757576
26080280 y	0.84860558	1	0.84860558	0.91810345
26090060 y	0.83466135	1	0.83466135	0.90988056
26095230 y	0.83067729	1	0.83067729	0.90750816
26100830 y	0.84661355	1	0.84661355	0.91693635
53110100 y	0.76294821	1	0.76294821	0.86553672
26105230 y	0.84661355	1	0.84661355	0.91693635

### C6.3 MÉTRICAS MODELO CLASIFICACIÓN SVC MENSUAL

y	Accuracy	Recall	Precision	F1
26055070 y	0.99137931	1	0.99137931	0.995671
26060020 y	0.98275862	1	0.98275862	0.99130435
26070110 y	0.95689655	1	0.95689655	0.97797357
26070760 y	0.98275862	1	0.98275862	0.99130435
26075010 y	0.90517241	1	0.90517241	0.95022624
26075040 y	0.97413793	1	0.97413793	0.98689956
26080070 y	0.97413793	1	0.97413793	0.98689956
26080290 y	0.99137931	1	0.99137931	0.995671
26090460 y	0.98275862	1	0.98275862	0.99130435
26090630 y	0.93965517	1	0.93965517	0.96888889
26095080 y	0.86206897	1	0.86206897	0.92592593
26100070 y	0.99137931	1	0.99137931	0.995671
26100300 y	0.97413793	1	0.97413793	0.98689956
26100350 y	0.96551724	1	0.96551724	0.98245614
26100400 y	0.9137931	1	0.9137931	0.95495495
26100410 y	0.96551724	1	0.96551724	0.98245614
26100690 y	0.98275862	1	0.98275862	0.99130435
26100700 y	0.93965517	1	0.93965517	0.96888889
26100770 y	0.93965517	1	0.93965517	0.96888889
26100780 y	0.97413793	1	0.97413793	0.98689956
26100790 y	0.93965517	1	0.93965517	0.96888889
26100800 y	0.93103448	1	0.93103448	0.96428571
26105150 y	0.90517241	1	0.90517241	0.95022624
26105160 y	0.85344828	1	0.85344828	0.92093023
26110040 y	0.97413793	1	0.97413793	0.98689956
26110120 y	0.96551724	1	0.96551724	0.98245614
26110150 y	0.98275862	1	0.98275862	0.99130435
26110160 y	0.96551724	1	0.96551724	0.98245614
26110210 y	0.87068966	1	0.87068966	0.93087558
26110230 y	0.94827586	1	0.94827586	0.97345133
26110290 y	0.99137931	1	0.99137931	0.995671
26110300 y	0.96551724	1	0.96551724	0.98245614
26115040 y	0.94827586	1	0.94827586	0.97345133
26120120 y	0.95689655	1	0.95689655	0.97797357
26120130 y	0.98275862	1	0.98275862	0.99130435
26120150 y	0.98275862	1	0.98275862	0.99130435
26120180 y	1	1	1	1
26125130 y	0.93965517	1	0.93965517	0.96888889
53080010 y	0.93965517	1	0.93965517	0.96888889
53090030 y	0.96551724	1	0.96551724	0.98245614
53090040 y	0.94827586	1	0.94827586	0.97345133
53100040 y	0.97413793	1	0.97413793	0.98689956
53110030 y	0.93965517	1	0.93965517	0.96888889
53110040 y	0.89655172	1	0.89655172	0.94545455
53115010 y	0.97413793	1	0.97413793	0.98689956
54030010 y	0.96551724	1	0.96551724	0.98245614
54030030 y	0.88793103	1	0.88793103	0.94063927
54070030 y	0.93103448	1	0.93103448	0.96428571
54070150 y	0.93965517	1	0.93965517	0.96888889
54075020 y	0.73275862	1	0.73275862	0.84577114
54075040 y	0.86206897	1	0.86206897	0.92592593
26070170 y	0.90517241	1	0.90517241	0.95022624
26080280 y	1	1	1	1
26090060 y	1	1	1	1
26095230 y	0.87931034	1	0.87931034	0.93577982
26100830 y	0.97413793	1	0.97413793	0.98689956
53110100 y	0.96551724	1	0.96551724	0.98245614
26105230 y	0.89655172	1	0.89655172	0.94545455

### C7.1 MÉTRICAS MODELO CLASIFICACIÓN DTC DIARIO

y	Accuracy	Recall	Precision	F1
26055070 y	0.58825207	0.45679758	0.58109147	0.51150203
26060020 y	0.74479612	0	0	0
26070110 y	0.70886798	0	0	0
26070760 y	0.63701169	0.41673122	0.50850662	0.45806726
26075010 y	0.57399487	0.3625	0.59011628	0.44911504
26075040 y	0.62132877	0.41184573	0.55783582	0.47385103
26080070 y	0.72398061	0	0	0
26080290 y	0.74536641	0	0	0
26090460 y	0.68263473	0	0	0
26090630 y	0.75392073	0	0	0
26095080 y	0.60906758	1	0.60906758	0.75704413
26100070 y	0.76703735	0	0	0
26100300 y	0.64984317	0	0	0
26100350 y	0.60250927	0.46143617	0.54303599	0.49892164
26100400 y	0.6720844	0	0	0
26100410 y	0.6675221	0	0	0
26100690 y	0.5988024	0.42014112	0.5656304	0.48214943
26100700 y	0.61819219	0	0	0
26100770 y	0.59281437	0	0	0
26100780 y	0.68919304	0	0	0
26100790 y	0.55374964	0.37587302	0.50425894	0.43070207
26100800 y	0.61362988	0	0	0
26105150 y	0.58112347	0.47257653	0.53579176	0.50220264
26105160 y	0.56544055	0.45914602	0.63623083	0.53337416
26110040 y	0.73510123	0	0	0
26110120 y	0.5560308	0.53942144	0.56073113	0.5498699
26110150 y	0.62646136	0	0	0
26110160 y	0.66096379	0	0	0
26110210 y	0.60536071	1	0.60536071	0.75417407
26110230 y	0.61248931	0	0	0
26110290 y	0.68691189	0	0	0
26110300 y	0.70230967	0	0	0
26115040 y	0.5794126	0.50064599	0.52471225	0.51239669
26120120 y	0.60136869	0.48830617	0.5047619	0.49639769
26120130 y	0.68520103	0	0	0
26120150 y	0.64528087	0	0	0
26120180 y	0.63672655	0	0	0
26125130 y	0.60650128	1	0.60650128	0.75505857
53080010 y	0.86170516	1	0.86170516	0.92571604
53090030 y	0.62959795	1	0.62959795	0.77270341
53090040 y	0.55260907	0	0	0
53100040 y	0.56914742	0.46491763	0.54584527	0.50214168
53110030 y	0.69746222	1	0.69746222	0.82177054
53110040 y	0.69118905	0	0	0
53115010 y	0.81437126	1	0.81437126	0.89768977
54030010 y	0.5891075	0	0	0
54030030 y	0.62161392	0	0	0
54070030 y	0.74736242	1	0.74736242	0.85541775
54070150 y	0.76589678	1	0.76589678	0.86743097
54075020 y	0.88765326	1	0.88765326	0.94048338
54075040 y	0.84031936	1	0.84031936	0.9132321
26070170 y	0.74793271	0	0	0
26080280 y	0.63301967	0	0	0
26090060 y	0.61277445	0.41487144	0.53638814	0.46786834
26095230 y	0.55831195	0.40310078	0.60717264	0.48452579
26100830 y	0.64870259	0	0	0
53110100 y	0.67579127	0	0	0
26105230 y	0.58825207	0.44809689	0.61471519	0.51834556

## C7.2 MÉTRICAS MODELO CLASIFICACIÓN DTC SEMANAL

y	Accuracy	Recall	Precision	F1
26055070 y	0.89243028	1	0.89243028	0.94315789
26060020 y	0.71912351	0.89915966	0.75352113	0.81992337
26070110 y	0.69920319	1	0.69920319	0.82297773
26070760 y	0.8187251	1	0.8187251	0.90032859
26075010 y	0.74501992	1	0.74501992	0.85388128
26075040 y	0.86454183	1	0.86454183	0.92735043
26080070 y	0.78087649	1	0.78087649	0.87695749
26080290 y	0.76095618	1	0.76095618	0.86425339
26090460 y	0.71713147	1	0.71713147	0.83526682
26090630 y	0.75498008	1	0.75498008	0.86038593
26095080 y	0.85059761	1	0.85059761	0.91926803
26100070 y	0.74701195	0.95616438	0.75869565	0.84606061
26100300 y	0.84262948	1	0.84262948	0.91459459
26100350 y	0.84262948	1	0.84262948	0.91459459
26100400 y	0.7689243	0.95287958	0.78787879	0.86255924
26100410 y	0.80677291	1	0.80677291	0.89305402
26100690 y	0.80278884	1	0.80278884	0.89060773
26100700 y	0.75896414	1	0.75896414	0.86296716
26100770 y	0.812749	1	0.812749	0.8967033
26100780 y	0.78486056	1	0.78486056	0.87946429
26100790 y	0.69521912	1	0.69521912	0.82021152
26100800 y	0.72709163	1	0.72709163	0.84198385
26105150 y	0.78087649	1	0.78087649	0.87695749
26105160 y	0.82868526	1	0.82868526	0.90631808
26110040 y	0.77689243	1	0.77689243	0.87443946
26110120 y	0.89243028	1	0.89243028	0.94315789
26110150 y	0.82071713	1	0.82071713	0.90153173
26110160 y	0.83665339	1	0.83665339	0.91106291
26110210 y	0.81075697	1	0.81075697	0.89548955
26110230 y	0.83466135	1	0.83466135	0.90988056
26110290 y	0.78685259	0.94334975	0.82012848	0.87743414
26110300 y	0.80079681	1	0.80079681	0.88938053
26115040 y	0.87450199	1	0.87450199	0.93304995
26120120 y	0.81474104	1	0.81474104	0.89791438
26120130 y	0.76494024	0.95778364	0.78064516	0.86018957
26120150 y	0.82071713	1	0.82071713	0.90153173
26120180 y	0.84262948	1	0.84262948	0.91459459
26125130 y	0.90239044	1	0.90239044	0.9486911
53080010 y	0.90836653	1	0.90836653	0.9519833
53090030 y	0.92828685	1	0.92828685	0.96280992
53090040 y	0.90438247	1	0.90438247	0.94979079
53100040 y	0.86653386	1	0.86653386	0.9284952
53110030 y	0.95418327	1	0.95418327	0.97655454
53110040 y	0.70119522	1	0.70119522	0.82435597
53115010 y	0.97211155	1	0.97211155	0.98585859
54030010 y	0.8625498	1	0.8625498	0.92620321
54030030 y	0.71513944	1	0.71513944	0.83391405
54070030 y	0.92430279	1	0.92430279	0.96066253
54070150 y	0.92430279	1	0.92430279	0.96066253
54075020 y	0.71314741	1	0.71314741	0.83255814
54075040 y	0.84661355	1	0.84661355	0.91693635
26070170 y	0.66733068	0.90789474	0.66506024	0.76773296
26080280 y	0.85258964	1	0.85258964	0.92043011
26090060 y	0.85458167	1	0.85458167	0.92158969
26095230 y	0.83466135	1	0.83466135	0.90988056
26100830 y	0.82669323	1	0.82669323	0.90512541
53110100 y	0.76693227	1	0.76693227	0.8680947
26105230 y	0.82071713	1	0.82071713	0.90153173

### C7.3 MÉTRICAS MODELO CLASIFICACIÓN DTC MENSUAL

y	Accuracy	Recall	Precision	F1
26055070 y	1	1	1	1
26060020 y	0.99137931	1	0.99137931	0.995671
26070110 y	0.94827586	1	0.94827586	0.97345133
26070760 y	0.98275862	1	0.98275862	0.99130435
26075010 y	0.87068966	1	0.87068966	0.93087558
26075040 y	0.97413793	1	0.97413793	0.98689956
26080070 y	0.96551724	1	0.96551724	0.98245614
26080290 y	0.98275862	1	0.98275862	0.99130435
26090460 y	0.96551724	1	0.96551724	0.98245614
26090630 y	0.96551724	1	0.96551724	0.98245614
26095080 y	0.8362069	1	0.8362069	0.91079812
26100070 y	0.97413793	1	0.97413793	0.98689956
26100300 y	0.98275862	1	0.98275862	0.99130435
26100350 y	0.95689655	1	0.95689655	0.97797357
26100400 y	0.9137931	1	0.9137931	0.95495495
26100410 y	0.97413793	1	0.97413793	0.98689956
26100690 y	0.98275862	1	0.98275862	0.99130435
26100700 y	0.94827586	1	0.94827586	0.97345133
26100770 y	0.97413793	1	0.97413793	0.98689956
26100780 y	0.98275862	1	0.98275862	0.99130435
26100790 y	0.95689655	1	0.95689655	0.97797357
26100800 y	0.90517241	1	0.90517241	0.95022624
26105150 y	0.89655172	1	0.89655172	0.94545455
26105160 y	0.89655172	1	0.89655172	0.94545455
26110040 y	0.93965517	1	0.93965517	0.96888889
26110120 y	0.94827586	1	0.94827586	0.97345133
26110150 y	0.95689655	1	0.95689655	0.97797357
26110160 y	0.98275862	1	0.98275862	0.99130435
26110210 y	0.92241379	1	0.92241379	0.95964126
26110230 y	0.97413793	1	0.97413793	0.98689956
26110290 y	0.95689655	1	0.95689655	0.97797357
26110300 y	0.97413793	1	0.97413793	0.98689956
26115040 y	0.93965517	1	0.93965517	0.96888889
26120120 y	0.98275862	1	0.98275862	0.99130435
26120130 y	0.94827586	1	0.94827586	0.97345133
26120150 y	0.94827586	1	0.94827586	0.97345133
26120180 y	0.99137931	1	0.99137931	0.995671
26125130 y	0.9137931	1	0.9137931	0.95495495
53080010 y	0.95689655	1	0.95689655	0.97797357
53090030 y	0.92241379	1	0.92241379	0.95964126
53090040 y	0.96551724	1	0.96551724	0.98245614
53100040 y	0.96551724	1	0.96551724	0.98245614
53110030 y	0.94827586	1	0.94827586	0.97345133
53110040 y	0.89655172	1	0.89655172	0.94545455
53115010 y	0.96551724	1	0.96551724	0.98245614
54030010 y	0.98275862	1	0.98275862	0.99130435
54030030 y	0.88793103	1	0.88793103	0.94063927
54070030 y	0.9137931	1	0.9137931	0.95495495
54070150 y	0.93103448	1	0.93103448	0.96428571
54075020 y	0.73275862	1	0.73275862	0.84577114
54075040 y	0.87068966	1	0.87068966	0.93087558
26070170 y	0.89655172	1	0.89655172	0.94545455
26080280 y	1	1	1	1
26090060 y	0.99137931	1	0.99137931	0.995671
26095230 y	0.89655172	1	0.89655172	0.94545455
26100830 y	0.97413793	1	0.97413793	0.98689956
53110100 y	0.93965517	1	0.93965517	0.96888889
26105230 y	0.96551724	1	0.96551724	0.98245614

### C8.1 MÉTRICAS MODELO CLASIFICACIÓN LOGIT INTENSIDAD DIARIA 4 CATEGORÍAS

y	Accuracy	Recall	Precision	F1
26055070 y	0.77958369	0.25132824	0.27023878	0.22228428
26060020 y	0.81722227	0.25	0.20430567	0.22485486
26070110 y	0.76846307	0.25	0.19211577	0.21726862
26070760 y	0.78043912	0.25047747	0.24266417	0.2220791
26075010 y	0.72939835	0.25	0.18234959	0.21088211
26075040 y	0.84488167	0.25058229	0.33626961	0.23029967
26080070 y	0.82862846	0.25	0.20715711	0.22657103
26080290 y	0.82748788	0.25	0.20687197	0.22640037
26090460 y	0.74679213	0.25001078	0.2148847	0.2148118
26090630 y	0.82919875	0.25	0.20729969	0.22665627
26095080 y	0.68662675	0.25295247	0.27728151	0.21035412
26100070 y	0.82292558	0.25	0.20573139	0.22571563
26100300 y	0.75220987	0.25	0.18805247	0.21464605
26100350 y	0.68006843	0.25323327	0.26880928	0.21018802
26100400 y	0.71884802	0.25	0.179712	0.2091075
26100410 y	0.73538637	0.24997227	0.21985714	0.21280292
26100690 y	0.71542629	0.25389923	0.32193058	0.21703518
26100700 y	0.71656687	0.25	0.17914172	0.20872093
26100770 y	0.7245509	0.25057069	0.23928933	0.21274951
26100780 y	0.78243513	0.25	0.19560878	0.21948488
26100790 y	0.58682635	0.25	0.14670659	0.18490566
26100800 y	0.70915312	0.25	0.17728828	0.20745746
26105150 y	0.69888794	0.25	0.17472198	0.20568983
26105160 y	0.71628172	0.25	0.17907043	0.20867254
26110040 y	0.81693755	0.25	0.20423439	0.22481168
26110120 y	0.58882236	0.25077636	0.2256472	0.18829992
26110150 y	0.73367551	0.25	0.18341888	0.21159539
26110160 y	0.76646707	0.25	0.19161677	0.21694915
26110210 y	0.6130596	0.25035754	0.20907185	0.19151685
26110230 y	0.80153978	0.25	0.20038494	0.22245964
26110290 y	0.78956373	0.25	0.19739093	0.22060229
26110300 y	0.82292558	0.25	0.20573139	0.22571563
26115040 y	0.77359567	0.25131864	0.28737675	0.22111991
26120120 y	0.75648703	0.25	0.18912176	0.21534091
26120130 y	0.73025378	0.25	0.18256344	0.21102505
26120150 y	0.71371543	0.25	0.17842886	0.20823627
26120180 y	0.73481608	0.25	0.18370402	0.21178501
26125130 y	0.62617622	0.25103193	0.22643785	0.19601292
53080010 y	0.27573425	0.25	0.06893356	0.10806884
53090030 y	0.40205304	0.25	0.10051326	0.14338011
53090040 y	0.55032792	0.25	0.13758198	0.17748759
53100040 y	0.74964357	0.25	0.18741089	0.21422751
53110030 y	0.73025378	0.25	0.18256344	0.21102505
53110040 y	0.76846307	0.25	0.19211577	0.21726862
53115010 y	0.39264328	0.25	0.09816082	0.14097052
54030010 y	0.74365555	0.25075988	0.31098344	0.21480244
54030030 y	0.73795267	0.25	0.18448817	0.21230517
54070030 y	0.37867123	0.25	0.09466781	0.13733195
54070150 y	0.37496436	0.26781823	0.17145629	0.18792144
54075020 y	0.37696036	0.25	0.09424009	0.13688134
54075040 y	0.39435415	0.30164395	0.19142856	0.23348657
26070170 y	0.77730254	0.25	0.19432563	0.2186748
26080280 y	0.75135443	0.25	0.18783861	0.21450668
26090060 y	0.75363559	0.25142275	0.35511282	0.2178748
26095230 y	0.72226975	0.25	0.18056744	0.20968543
26100830 y	0.7530653	0.25	0.18826632	0.2147853
53110100 y	0.75819789	0.25	0.18954947	0.2156179
26105230 y	0.73310522	0.25	0.1832763	0.21150049

## C8.2 MÉTRICAS MODELO CLASIFICACIÓN LOGIT INTENSIDAD DIARIA 5 CATEGORÍAS

y	Accuracy	Recall	Precision	F1
26055070 y	0.7815797	0.20169703	0.21561436	0.179622
26060020 y	0.81123467	0.2	0.16224693	0.17915617
26070110 y	0.76504135	0.2	0.15300827	0.17337641
26070760 y	0.79555175	0.2	0.15911035	0.17722725
26075010 y	0.73595666	0.2	0.14719133	0.1695795
26075040 y	0.84716282	0.20052632	0.36942384	0.18449644
26080070 y	0.8172227	0.20049261	0.26348074	0.1808862
26080290 y	0.82349587	0.2	0.16469917	0.18064113
26090460 y	0.73738238	0.2	0.14747648	0.16976859
26090630 y	0.82834331	0.2	0.16566866	0.18122271
26095080 y	0.67094383	0.20173584	0.22009734	0.16469745
26100070 y	0.81779299	0.2	0.1635586	0.17995294
26100300 y	0.76133447	0.2002861	0.25229672	0.17361011
26100350 y	0.69090391	0.2010929	0.1790279	0.1672899
26100400 y	0.71799259	0.2	0.14359852	0.16717012
26100410 y	0.73310522	0.2	0.14662104	0.16920039
26100690 y	0.71571143	0.20164507	0.23668958	0.17086667
26100700 y	0.71057884	0.19991978	0.1421563	0.16616103
26100770 y	0.72341032	0.2	0.14468206	0.16790205
26100780 y	0.79184488	0.20151172	0.25850243	0.18016197
26100790 y	0.60051326	0.2	0.12010265	0.15008017
26100800 y	0.71685201	0.2	0.1433704	0.16701545
26105150 y	0.69232963	0.2	0.13846593	0.16363943
26105160 y	0.71428571	0.20039635	0.2428816	0.16760344
26110040 y	0.79355575	0.2	0.15871115	0.17697933
26110120 y	0.57827203	0.20077355	0.1636833	0.14906084
26110150 y	0.7362418	0.2	0.14724836	0.16961734
26110160 y	0.75563159	0.19992456	0.15116942	0.17216177
26110210 y	0.60165384	0.2	0.12033077	0.15025814
26110230 y	0.80439122	0.2	0.16087824	0.17831858
26110290 y	0.79441118	0.20097276	0.25898315	0.17941212
26110300 y	0.81522669	0.2	0.16304534	0.17964185
26115040 y	0.77758768	0.20019594	0.18081595	0.17589828
26120120 y	0.74793271	0.2	0.14958654	0.17115824
26120130 y	0.74878814	0.2	0.14975763	0.17127018
26120150 y	0.7219846	0.20037106	0.34438106	0.16843993
26120180 y	0.72740234	0.2	0.14552196	0.16846624
26125130 y	0.6033647	0.20156236	0.2097258	0.15449777
53080010 y	0.27915597	0.20306435	0.11118654	0.1259686
53090030 y	0.39520958	0.2	0.07904192	0.11330472
53090040 y	0.5437696	0.2	0.10875392	0.14089398
53100040 y	0.75990875	0.20034722	0.35196805	0.17339989
53110030 y	0.73937839	0.2	0.14787568	0.17003279
53110040 y	0.78015398	0.2	0.1560308	0.17530034
53115010 y	0.38522954	0.19970042	0.12929176	0.11669989
54030010 y	0.75078415	0.2	0.15015683	0.17153094
54030030 y	0.73025378	0.2	0.14605076	0.16882004
54070030 y	0.37581979	0.20287601	0.124784	0.14186353
54070150 y	0.36298831	0.19864035	0.08857436	0.10896867
54075020 y	0.2683205	0.18937113	0.10809667	0.13702633
54075040 y	0.36184773	0.22284906	0.19603032	0.15769424
26070170 y	0.78842315	0.2	0.15768463	0.17633929
26080280 y	0.75449102	0.2	0.15094124	0.17204161
26090060 y	0.7362418	0.20054377	0.22241355	0.17130302
26095230 y	0.7362418	0.2	0.14724836	0.16961734
26100830 y	0.75163958	0.2	0.15032792	0.17164252
53110100 y	0.76161962	0.2	0.15232392	0.17293623
26105230 y	0.73396065	0.2	0.14679213	0.16931426

### C8.3 MÉTRICAS MODELO CLASIFICACIÓN LOGIT INTENSIDAD DIARIA 7 CATEGORÍAS

y	Accuracy	Recall	Precision	F1
26055070 y	0.52723125	0.14805203	0.12514354	0.11027127
26060020 y	0.73652695	0.16666667	0.12275449	0.14137931
26070110 y	0.70259481	0.14285714	0.10037069	0.1179032
26070760 y	0.6130596	0.17426605	0.18142504	0.14330563
26075010 y	0.53093812	0.14476552	0.11515974	0.10338582
26075040 y	0.58426005	0.14401196	0.13365767	0.1083459
26080070 y	0.71856287	0.2	0.14371257	0.16724739
26080290 y	0.75591674	0.14285714	0.10798811	0.12299912
26090460 y	0.67978329	0.14285714	0.0971119	0.11562432
26090630 y	0.75192472	0.14285714	0.10741782	0.12262835
26095080 y	0.41631024	0.20100229	0.13367401	0.15884836
26100070 y	0.77929855	0.16666667	0.12988309	0.14599359
26100300 y	0.64841745	0.14285714	0.09263106	0.11238787
26100350 y	0.58254919	0.16658513	0.0972024	0.12276906
26100400 y	0.67579127	0.14285714	0.09654161	0.11521914
26100410 y	0.66609638	0.14285714	0.09515663	0.11422704
26100690 y	0.56572569	0.19338563	0.16859333	0.17243102
26100700 y	0.61534075	0.14285714	0.08790582	0.10883873
26100770 y	0.59823211	0.14285714	0.08546173	0.10694533
26100780 y	0.69831765	0.14285714	0.09975966	0.11748057
26100790 y	0.54975763	0.16666667	0.09162627	0.11824594
26100800 y	0.61191902	0.2	0.1223838	0.15184858
26105150 y	0.53264899	0.14285714	0.07611442	0.09931416
26105160 y	0.47105788	0.14555208	0.11169075	0.0989522
26110040 y	0.7245509	0.16666667	0.12075848	0.1400463
26110120 y	0.48132307	0.16666667	0.08022051	0.10830927
26110150 y	0.62246935	0.14285714	0.08892419	0.10961587
26110160 y	0.66181922	0.16666667	0.1103032	0.13274994
26110210 y	0.39806102	0.15539382	0.10916102	0.11616388
26110230 y	0.607927	0.14285714	0.08684671	0.10802321
26110290 y	0.69432563	0.16666667	0.11572094	0.13659823
26110300 y	0.70088395	0.14285714	0.10012628	0.1177344
26115040 y	0.55460508	0.14407948	0.11804066	0.10597629
26120120 y	0.59737668	0.14285714	0.08533953	0.10684959
26120130 y	0.67436555	0.14285714	0.09633794	0.11507396
26120150 y	0.64414029	0.14285714	0.09202004	0.11193697
26120180 y	0.64185914	0.14285714	0.09169416	0.11169553
26125130 y	0.41146279	0.17180828	0.11156954	0.13519756
53080010 y	0.29141717	0.16666667	0.04856953	0.07521896
53090030 y	0.39235814	0.16666667	0.06539302	0.09393133
53090040 y	0.53692615	0.16666667	0.08948769	0.11645022
53100040 y	0.53378956	0.17020331	0.15059861	0.12916203
53110030 y	0.57599088	0.16666667	0.09599848	0.12182619
53110040 y	0.68691189	0.14285714	0.09813027	0.11634309
53115010 y	0.28229256	0.14285714	0.04032751	0.06289908
54030010 y	0.59252923	0.16666667	0.09875487	0.12402268
54030030 y	0.60821215	0.16666667	0.10136869	0.12606383
54070030 y	0.27060165	0.16347446	0.07769297	0.10513759
54070150 y	0.26603935	0.15825879	0.07685263	0.10308424
54075020 y	0.27003137	0.14285714	0.03857591	0.06074796
54075040 y	0.26404334	0.14285714	0.03772048	0.05968225
26070170 y	0.7576276	0.16666667	0.12627127	0.14368376
26080280 y	0.63416025	0.14285714	0.09059432	0.11087569
26090060 y	0.58825207	0.14588961	0.13953043	0.11315098
26095230 y	0.47704591	0.1450456	0.13538534	0.09723657
26100830 y	0.65155403	0.14285714	0.09307915	0.11271705
53110100 y	0.67892786	0.14285714	0.09698969	0.11553766
26105230 y	0.50527516	0.1466981	0.13426582	0.10453676

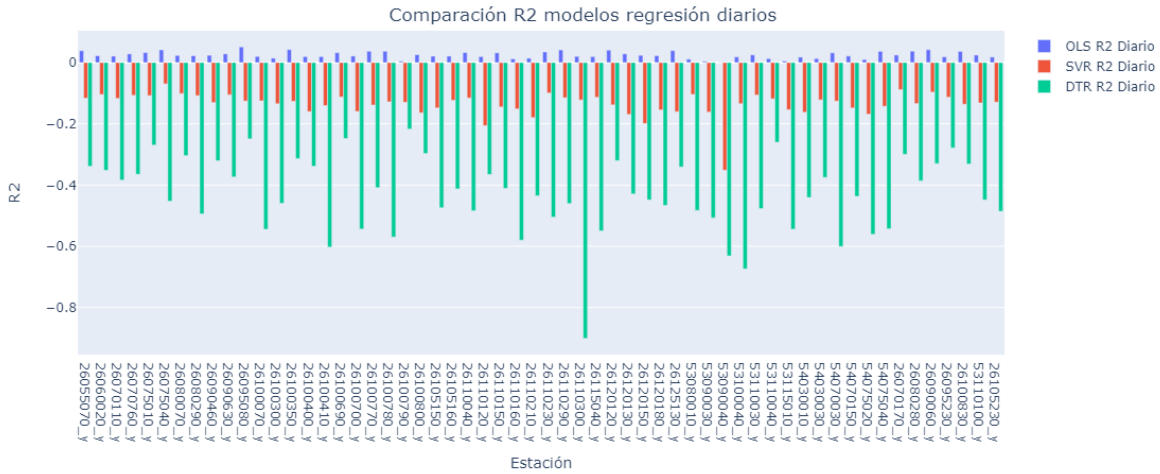
#### C8.4 MÉTRICAS MODELO CLASIFICACIÓN LOGIT INTENSIDAD MENSUAL 5 CATEGORÍAS

y	Accuracy	Recall	Precision	F1
26055070_y	0.84482759	0.55330113	0.58220762	0.55881691
26060020_y	0.80172414	0.54021572	0.55937922	0.53543678
26070110_y	0.77586207	0.5585397	0.51712089	0.53677419
26070760_y	0.82758621	0.58969887	0.55920709	0.57268336
26075010_y	0.75862069	0.7625	0.74181818	0.7458529
26075040_y	0.75862069	0.80208333	0.75862069	0.74961455
26080070_y	0.72413793	0.77813051	0.73502235	0.71568627
26080290_y	0.74137931	0.50505051	0.54009979	0.49026025
26090460_y	0.78448276	0.55581921	0.52585859	0.54
26090630_y	0.68103448	0.73188406	0.7797619	0.67561031
26095080_y	0.76724138	0.77440476	0.82522046	0.75935459
26100070_y	0.78448276	0.50690579	0.55555556	0.50963633
26100300_y	0.78448276	0.71723577	0.80253623	0.73284201
26100350_y	0.75	0.40244391	0.40563878	0.39200832
26100400_y	0.70689655	0.47053763	0.51339031	0.46052632
26100410_y	0.74137931	0.50990437	0.53639847	0.499532
26100690_y	0.76724138	0.493083	0.53931624	0.49577718
26100700_y	0.79310345	0.51186791	0.53864734	0.51559934
26100770_y	0.78448276	0.61734103	0.89637681	0.67300415
26100780_y	0.68965517	0.41303884	0.48049119	0.4048216
26100790_y	0.63793103	0.72	0.74698795	0.63620072
26100800_y	0.68965517	0.47654321	0.50499684	0.45452868
26105150_y	0.68103448	0.4296153	0.50960586	0.41483586
26105160_y	0.69827586	0.68809524	0.79546517	0.66410193
26110040_y	0.67241379	0.45392216	0.51955743	0.42995885
26110120_y	0.74137931	0.40263749	0.52604167	0.43479471
26110150_y	0.69827586	0.42962963	0.55339806	0.4145325
26110160_y	0.72413793	0.65590062	0.81375	0.64781784
26110210_y	0.63793103	0.31346154	0.35346535	0.2958379
26110230_y	0.64655172	0.45278795	0.49590789	0.4215918
26110290_y	0.64655172	0.4516129	0.5212766	0.41496599
26110300_y	0.65517241	0.44949495	0.51238897	0.41995007
26115040_y	0.69827586	0.69827586	0.81182796	0.66805658
26120120_y	0.80172414	0.44968487	0.50954433	0.46645663
26120130_y	0.80172414	0.52790462	0.56280193	0.53131313
26120150_y	0.74137931	0.46141975	0.46821153	0.45663957
26120180_y	0.8362069	0.52255109	0.56611661	0.53369947
26125130_y	0.70689655	0.46243789	0.53787879	0.44875776
53080010_y	0.44827586	0.19939504	0.14830745	0.16956522
53090030_y	0.60344828	0.47479532	0.57474747	0.49226423
53090040_y	0.40517241	0.34994479	0.31130435	0.30039487
53100040_y	0.6637931	0.43636364	0.53535354	0.40897973
53110030_y	0.52586207	0.34567901	0.39626352	0.25207398
53110040_y	0.80172414	0.81818182	0.79973238	0.79835235
53115010_y	0.5862069	0.46810777	0.56560593	0.46679077
54030010_y	0.75862069	0.7225099	0.80399501	0.72604588
54030030_y	0.64655172	0.33382066	0.28888889	0.28007478
54070030_y	0.56034483	0.47786822	0.46876194	0.46126064
54070150_y	0.62068966	0.51077139	0.55439462	0.51327839
54075020_y	0.54310345	0.45068418	0.38225957	0.39326087
54075040_y	0.54310345	0.47389597	0.46934685	0.44178527
26070170_y	0.69827586	0.4896862	0.46973317	0.47535467
26080280_y	0.76724138	0.46608947	0.54166667	0.47296207
26090060_y	0.80172414	0.56013986	0.54210526	0.54880108
26095230_y	0.69827586	0.49275362	0.4941064	0.46743697
26100830_y	0.80172414	0.70411606	0.85947712	0.72593734
53110100_y	0.82758621	0.56038798	0.54878534	0.55267834
26105230_y	0.77586207	0.50459559	0.56006494	0.50416389

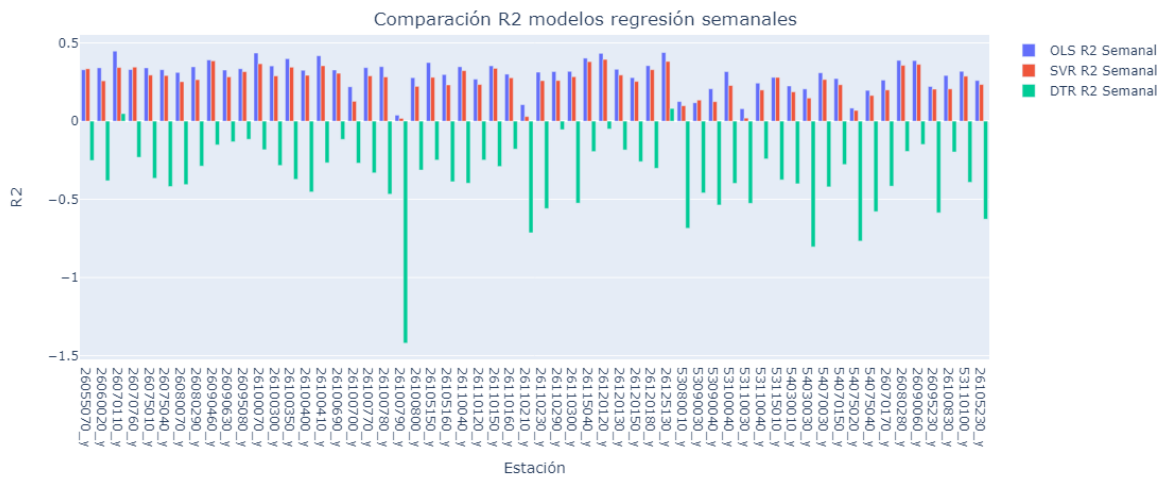
## ANEXO D

### COMPARACIÓN MÉTRICAS MODELOS DE APRENDIZAJE AUTOMÁTICO

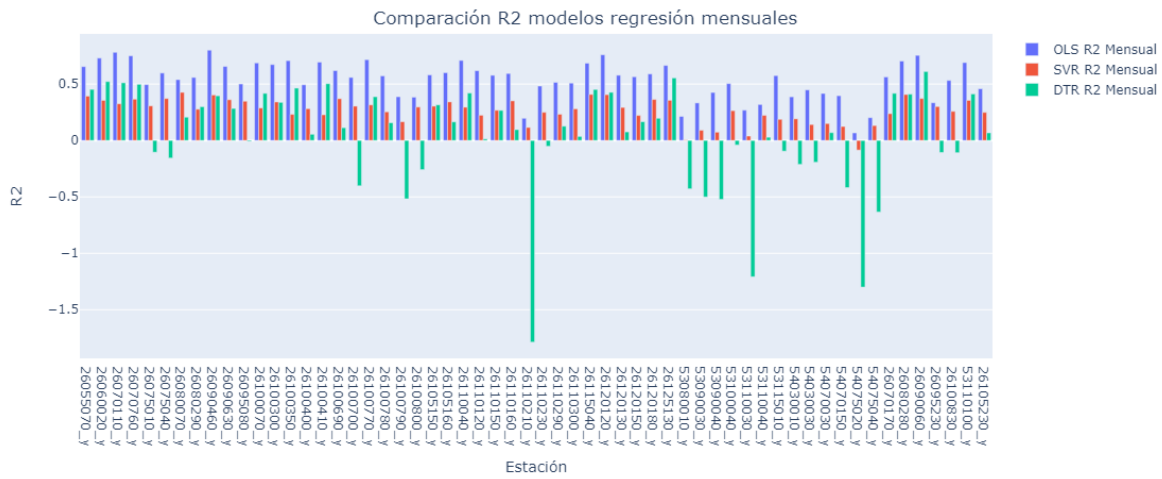
#### D1.1 COMPARACIÓN R2 MODELOS REGRESIÓN DIARIOS



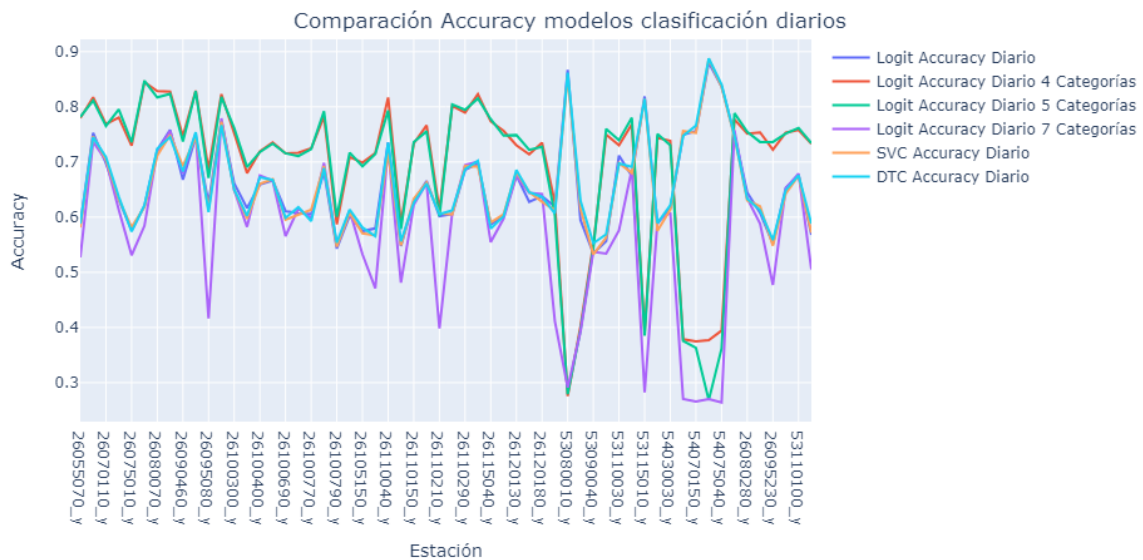
#### D1.2 COMPARACIÓN R2 MODELOS REGRESIÓN SEMANALES



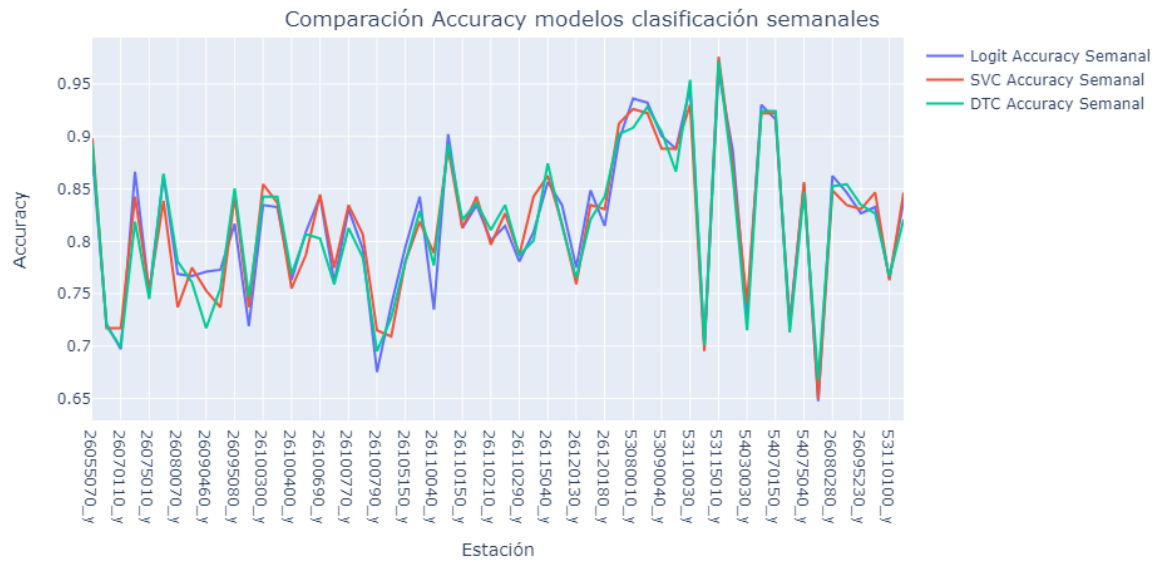
### D1.3 COMPARACIÓN R2 MODELOS REGRESIÓN MENSUALES



### D2.1 COMPARACIÓN ACCURACY MODELOS CLASIFICACIÓN DIARIOS



## D2.2 COMPARACIÓN ACCURACY MODELOS CLASIFICACIÓN SEMANALES



## D2.3 COMPARACIÓN ACCURACY MODELOS CLASIFICACIÓN MENSUALES

