Santiago de Cali, 15 de Enero del 2026

Doctor
Diego Luis Linares O.
Director Maestría en Ciencia de Datos
Facultad de Ingeniería y Ciencias
Pontificia Universidad Javeriana de Cali

Asunto: Presentación para evaluación del proyecto aplicado

Cordial Saludo,

Con el fin de cumplir con los requisitos exigidos por la Universidad para optar por el título de Magíster en Ciencia de Datos, nos permitimos presentar a su consideración el proyecto denominado "Predicción del Tráfico de Datos de las Zonas Wifi Públicas de Santiago de Cali", el cual fue realizado por los estudiantes Paulo Andrés Martínez Méndez y Edier Guzmán Morales con códigos 0183321 y 9016769 respectivamente, pertenecientes a la Maestría en Ciencia de Datos, bajo la dirección de Diego Luis Linares Ospina y la codirección de Gloria Inés Álvarez Vargas.

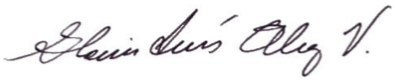
El suscrito director del Proyecto Aplicado autoriza para que se proceda a hacer la evaluación de este proyecto, toda vez que ha revisado cuidadosamente el documento y avala que ya se encuentra listo para ser presentado y sustentado oficialmente.

Atentamente,

Paulo Andrés Martínez Méndez
C.C. 94.424.446

Diego Luis Linares Ospina
C.C. 16.739.144

Edier Guzmán Morales
C.C. 1.130.618.441



Gloria Inés Álvarez Vargas
C.C. 30.306.105

Documentación anexa:
Resumen del Proyecto Aplicado en formato digital (máximo 1 página).
Una copia digital (PDF) del documento del proyecto aplicado

FICHA RESUMEN
PROYECTO APLICADO – MAESTRÍA EN CIENCIA DE DATOS

TÍTULO:

1. **ÁREA DE TRABAJO:** Ciencia de Datos
2. **TIPO DE PROYECTO:** Proyecto Aplicado
3. **ESTUDIANTE(S):** Paulo Andrés Martínez Méndez, Edier Guzmán Morales
4. **CORREO ELECTRÓNICO:** pmartinez46@javerianacali.edu.co, edguzman@javerianacali.edu.co
5. **DIRECCIÓN Y TELEFONO:**
   Paulo Andrés Martínez
   Carrera 47 #5E-57
   317 254 8474

   Edier Guzmán
   Carrera 2 # 10 – 03
   300 360 5681
6. **DIRECTOR:** Diego Luis Linares Ospina
7. **VINCULACIÓN DEL DIRECTOR:** Por medio de la universidad
8. **CORREO ELECTRÓNICO DEL DIRECTOR:** dlinares@javerianacali.edu.co
9. **CO-DIRECTORA**: Gloria Inés Álvarez Vargas
10. **CORREO ELECTRÓNICO DE LA CO-DIRECTORA:** galvarez@javerianacali.edu.co
11. **PALABRAS CLAVE (al menos 5)**: Robust Scaler, Random Forest, Modelo Base, Multicolinealidad, Ventaneo, Serie Temporal
12. **FECHA DE INICIO:** Noviembre 2024
13. **FECHA DE FINALIZACIÓN:** Enero 2026
14. **RESUMEN:**

Las zonas WiFi de Cali se vienen operando desde hace años en la ciudad simplemente con el servicio estándar de internet. No existe una preparación del ancho de banda y la infraestructura necesaria para dicho servicio dependiendo de la demanda del área, la población y visitantes en ciertos periodos de tiempo. Este trabajo busca, desde la ciencia de datos, predecir el tráfico que se necesita en determinadas épocas para así adecuar el ancho de banda de la zona y la infraestructura necesaria. Lo anterior se logra alimentando modelos de aprendizaje con los datos del tráfico de 63 zonas WiFi de la ciudad que la alcaldía de Cali ha recolectado a lo largo de dos años.

Pontificia Universidad
JAVERIANA
Cali

PREDICCIÓN DEL TRÁFICO DE DATOS DE LAS ZONAS WIFI PÚBLICAS DE SANTIAGO DE CALI

Paulo Andrés Martínez Méndez
Código estudiante: 0183321

Edier Guzmán Morales
Código estudiante: 9015769

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

Director
Diego Luis Linares Ospina

Codirectora
Gloria Inés Álvarez Vargas

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI, ENERO 1 DE 2025

TABLA DE CONTENIDO

Contenido

INTRODUCCIÓN

En la ciudad de Cali, donde actualmente se tienen aproximadamente de 2 millones de habitantes, se hace imprescindible para el gobierno local mantener a su población conectada en distintos sitios públicos, entre ellos parques o zonas recreativas. Es por eso que se han instalado en diferentes zonas de la ciudad zonas WiFi públicas que ayudan a tener conectada a la población.

Sin embargo, la alta demanda de este servicio trae consigo problemas de rendimiento en la red, incluyendo congestión en horas pico, desconexiones frecuentes o zonas que, aunque tienen infraestructura WiFi, no se utilizan. Estos problemas no solo afectan la satisfacción del ciudadano, sino que también limitan el potencial de aprovechar el WiFi como una herramienta que pueda conectar a los ciudadanos, brindando soluciones como acercarse al conocimiento, tener oportunidades de empleo, acceder a servicios públicos en la red y así cerrar la brecha digital.

Para abordar estos desafíos, el presente proyecto busca implementar un enfoque de ciencia de datos que permita analizar en profundidad el uso de las redes WiFi en diferentes zonas públicas de la ciudad de Cali. A través de la recopilación de datos de conexión, tráfico de datos y tiempos de uso, se pretende realizar un análisis que identifique horarios y zonas de alta demanda, áreas de congestión o baja conectividad.

Estos hallazgos permitirán proponer ajustes estratégicos en la red, como la reubicación de puntos de acceso y la adaptación dinámica del ancho de banda según la demanda.
La solución se basa en el uso de modelos de aprendizaje supervisado para anticipar la carga de la red en diferentes momentos, lo que permitirá tomar decisiones informadas y ajustar los recursos de manera eficiente para identificar picos de demanda inesperados con el fin de responder rápidamente y minimizar los impactos en el servicio.

Con este proyecto, se espera mejorar significativamente la calidad del servicio WiFi, optimizando su distribución y funcionamiento en base a datos reales y patrones de uso. Los resultados esperados incluyen un aumento en la satisfacción del ciudadano, una reducción de los problemas de conectividad, y una herramienta de análisis que permitirá al distrito gestionar de forma proactiva la infraestructura WiFi. Además, el análisis de datos ofrecerá conocimientos valiosos que podrán ser aprovechados para iniciativas comerciales, eventos públicos y el crecimiento urbano entorno a dichas zonas.

1 DEFINICIÓN DEL PROBLEMA

1.1 PLANTEAMIENTO DEL PROBLEMA

Actualmente las zonas WiFi públicas en Cali se ven sujetas a una serie de sobrecostos, ya que hay zonas wifi a la que se les ha asignado más recursos y que no son tan demandadas como otras.

Lo anterior es producido por la falta de un sistema que permita predecir el tráfico de datos en diferentes ubicaciones y horarios, esta ausencia limita la capacidad de anticipar la demanda, lo que resulta en una asignación ineficiente del ancho de banda, sobrecarga en puntos de acceso, y una experiencia deficiente para los ciudadanos.

1.2 FORMULACIÓN DEL PROBLEMA

Pregunta de formulación:

¿Cómo predecir el tráfico de datos para las zonas WiFi públicas de Cali?

Preguntas de sistematización:

¿Cómo preparar los datos?
¿Cómo entrenar los modelos predictivos?
¿Cómo evaluar el desempeño de los modelos entrenados?
¿Cómo poner el modelo en funcionamiento?

2 OBJETIVOS DEL PROYECTO

2.1 OBJETIVO GENERAL

Desarrollar un sistema para predecir el tráfico de datos de las zonas WiFi públicas de Cali aplicando técnicas de aprendizaje automático supervisado.

2.2 OBJETIVOS ESPECÍFICOS

- Preparar los datos con herramientas de limpieza y transformación.
- Entrenar los modelos predictivos de aprendizaje supervisado.
- Evaluar el desempeño de los modelos entrenados.
- Implementar un prototipo para el modelo en funcionamiento.

3 MARCO TEÓRICO Y ANTECEDENTES

3.1 PALABRAS CLAVE

DATIC:
Departamento Administrativo de Tecnologías de la Información y las Comunicaciones de la Alcaldía de Santiago de Cali.

Conexiones totales:
Número total de sesiones o dispositivos que se conectaron a una zona WiFi durante un periodo de tiempo determinado (en este caso, por día).

Acceso Point (AP):
Dispositivo de red inalámbrica que permite a los usuarios conectarse a una red WiFi. En cada zona WiFi pueden existir uno o más AP que distribuyen el acceso a Internet.

Dashboard:
Interfaz gráfica interactiva que permite visualizar indicadores clave, métricas y resultados del modelo de predicción de forma clara y comprensible para el usuario final.

Tráfico de datos (kB):
Cantidad de información transmitida a través de la red WiFi, medida en kilobytes, durante un periodo.

Zona WiFi:
Área geográfica delimitada que ofrece conectividad inalámbrica al público, generalmente equipada con infraestructura de red y puntos de acceso.

Variable binaria:
Tipo de variable que solo puede tomar dos valores (0 o 1), utilizada para representar condiciones lógicas como festivo, laboral o fin de semana.

Variable objetivo (target):
Variable que el modelo busca predecir. En este trabajo corresponde al uso total de datos (USAGE_KB) por zona y día.

Variables explicativas:
Conjunto de variables de entrada que el modelo utiliza para realizar la predicción, tales como número de conexiones, día de la semana y tipo de día.

3.2 **MARCO TEÓRICO**

El desarrollo de este proyecto requiere un fundamento teórico sólido que abarque los aspectos esenciales relacionados con el análisis y predicción del tráfico en las zonas WiFi públicas de Cali. En esta sección se abordarán tres temas principales: primero, se analizará el contexto y los retos actuales de las zonas WiFi públicas en Cali, destacando su importancia en la reducción de la brecha digital.
En segundo lugar, se explorarán los conceptos y técnicas del aprendizaje supervisado, centrándose en los algoritmos de predicción más adecuados para este proyecto, como la regresión lineal, los bosques aleatorios (Random Forest) y las redes neuronales. Finalmente, se discutirán las métricas de evaluación empleadas para medir el desempeño de los modelos desarrollados y se presentarán antecedentes relevantes que aportan conocimiento y precedentes valiosos para la implementación de esta propuesta.
Esta introducción permite estructurar la discusión y sentar las bases teóricas necesarias para garantizar la viabilidad y la efectividad del proyecto.

- **Zonas WiFi Públicas en Cali**

  La ciudad de Santiago de Cali cuenta con una red de zonas WiFi públicas distribuidas estratégicamente en diferentes sectores urbanos, cuyo propósito principal es facilitar el acceso gratuito a internet y contribuir a la reducción de la brecha digital. Estas zonas están ubicadas en parques, espacios públicos, escenarios deportivos y áreas de alta afluencia ciudadana, y se constituyen como un componente relevante dentro de las políticas de inclusión digital y acceso a la información.

  No obstante, la operación de estas zonas WiFi enfrenta diversos retos técnicos, operativos y de gestión, derivados tanto del crecimiento de la demanda como de la variabilidad en los patrones de uso. El tráfico de datos en estas zonas no se comporta de manera homogénea, sino que presenta fluctuaciones significativas asociadas a factores como el día de la semana, la hora, la ubicación geográfica y la naturaleza de las actividades que se desarrollan en cada entorno. Estas variaciones generan escenarios de congestión en determinados periodos, mientras que en otros momentos se observa una subutilización de los recursos disponibles.

  Desde el punto de vista técnico, uno de los principales desafíos consiste en la gestión eficiente del ancho de banda, dado que una asignación estática de capacidad no responde adecuadamente a la dinámica real del consumo. La falta de mecanismos predictivos limita la capacidad de anticipar picos de tráfico, lo que puede impactar negativamente la calidad del servicio ofrecido a los usuarios finales, manifestándose en bajas velocidades de conexión, interrupciones o degradación de la experiencia de uso.

  Adicionalmente, desde una perspectiva operativa y de planificación, la ausencia de herramientas analíticas que permitan comprender y anticipar el comportamiento del tráfico dificulta la toma de decisiones relacionadas con la expansión, el refuerzo o la priorización de determinadas zonas. En este contexto, la gestión de las zonas WiFi públicas suele basarse en

análisis descriptivos o reactivos, los cuales resultan insuficientes para enfrentar escenarios de alta demanda y crecimiento sostenido del uso del servicio.

En este sentido, la incorporación de técnicas de ciencia de datos y modelos predictivos se presenta como una alternativa viable para abordar estos retos, al permitir analizar el comportamiento histórico del tráfico y generar estimaciones anticipadas sobre su evolución. Este enfoque no solo aporta valor desde el punto de vista técnico, sino que también ofrece insumos relevantes para la toma de decisiones estratégicas orientadas a mejorar la calidad, eficiencia y sostenibilidad del servicio de WiFi público en la ciudad de Cali.

A continuación se muestran algunos conceptos importantes, desde lo contextual, sobre las infraestructuras WiFi: [27]

802.11g: Estandar para la radiación de señales inalámbricas. Muy usado en zonas WiFi públicas. El 802.11g es el tercer estándar de modulación del conjunto IEEE 802.11. Éste utiliza la banda de 2.4 Ghz.

LAN: Siglas en inglés de Red de Área Local; son redes que ocupan una distancia corta, un edificio, casa o grupo de oficinas

SSID: (Service Set IDentifier). Es un código incluido en todos los paquetes de una red inalámbrica (Wi-Fi), para identificarlos como parte de esa red. El código consiste en un máximo de 32 caracteres alfanuméricos.

Backhaul: Enlace que conecta los Access Points con la red troncal o el proveedor de Internet.

Access Point: Punto de acceso inalámbrico de una red WiFi. Es un dispositivo que da señal de internet a un área determinada en el espacio.

Ancho de Banda: Capacidad máxima de transmisión de datos de una red, usualmente medida en Mbps.

Latencia: Tiempo que tarda un paquete de datos en viajar desde el origen al destino.

Tráfico de datos: Volumen de información transmitida y recibida en la red durante un intervalo de tiempo.

Consumo de datos: Cantidad total de información utilizada por un usuario o por la red.

Picos de tráfico: Intervalos en el que la demanda de datos alcanza su valor máximo.

Portal Cautivo: Página web donde el usuario debe aceptar algunos términos y condiciones o se debe autenticar antes de acceder a Internet.

Autenticación: Proceso mediante el cual se valida la identidad del usuario.

Logs de conexión: Registros históricos de eventos de acceso y uso de la red.

Dirección IP: Identificador numérico asignado a cada dispositivo conectado a la red.

Disponibilidad: Porcentaje de tiempo en que la red se encuentra operativa.

Aprendizaje Supervisado en Predicción

El aprendizaje supervisado constituye uno de los enfoques más utilizados dentro del campo del aprendizaje automático, especialmente en problemas de predicción y estimación de variables continuas. Este paradigma se basa en el uso de datos etiquetados, es decir, conjuntos de datos en los que cada instancia de entrada se encuentra asociada a un valor objetivo conocido. A partir de esta relación, los modelos aprenden patrones que permiten generalizar y realizar predicciones sobre datos no observados.

En el contexto de la predicción del tráfico de datos en zonas WiFi públicas, el aprendizaje supervisado resulta particularmente adecuado, dado que se dispone de información histórica estructurada, en la cual el volumen de tráfico registrado para cada periodo temporal es conocido. Esta disponibilidad de datos etiquetados permite formular el problema como una tarea de regresión, donde el objetivo es estimar el valor futuro del tráfico a partir de observaciones pasadas y variables explicativas asociadas.

Una de las principales ventajas del aprendizaje supervisado en este tipo de problemas es su capacidad para capturar relaciones complejas entre variables, incluyendo dependencias temporales, patrones recurrentes y comportamientos no lineales. En el caso de las zonas WiFi públicas, el consumo de datos no sigue una tendencia uniforme, sino que está influenciado por múltiples factores, como el día de la semana, la ubicación de la zona, la cantidad de usuarios conectados y eventos específicos que pueden alterar significativamente la demanda. Los modelos supervisados permiten incorporar estas variables y aprender su impacto sobre el tráfico de manera sistemática.

No obstante, la efectividad del aprendizaje supervisado depende en gran medida de la calidad del proceso de preparación de los datos, así como de la correcta selección y configuración de los modelos. Aspectos como la limpieza de datos, el escalado de variables, la generación de ventanas temporales y la partición adecuada entre conjuntos de entrenamiento y prueba son determinantes para evitar problemas como el sobreajuste o la filtración de información del futuro hacia el pasado.

En este proyecto, el aprendizaje supervisado se emplea como la base metodológica para la construcción de modelos predictivos de tráfico de datos, utilizando distintos algoritmos de regresión con capacidades diversas para modelar relaciones lineales y no lineales. La comparación entre modelos simples y modelos más complejos permite evaluar cuál enfoque se adapta mejor a

las características específicas de cada zona WiFi, reconociendo que no existe un único modelo óptimo para todos los escenarios. [11]

En síntesis, el uso del aprendizaje supervisado en la predicción del tráfico de zonas WiFi públicas no solo se justifica por la disponibilidad de datos históricos etiquetados, sino también por su capacidad para ofrecer estimaciones anticipadas confiables, que pueden ser utilizadas como insumo para la toma de decisiones técnicas y operativas. Este enfoque constituye el fundamento sobre el cual se desarrollan las etapas posteriores de modelado, evaluación y análisis presentadas en este trabajo.

Regresión lineal:

La regresión lineal es un método estadístico y de aprendizaje supervisado utilizado para modelar y analizar la relación entre una variable dependiente y una o más variables independientes. Su objetivo principal es describir cómo cambia la variable de interés en función de las variables explicativas, bajo el supuesto de que existe una relación aproximadamente lineal entre ellas. Este modelo se representa mediante una ecuación matemática en la que los coeficientes indican la contribución individual de cada variable independiente sobre la variable objetivo.

El proceso de estimación de la regresión lineal se basa comúnmente en el método de mínimos cuadrados, el cual busca minimizar la suma de los errores cuadráticos entre los valores observados y los valores predichos por el modelo. Esta característica permite obtener un ajuste óptimo desde el punto de vista estadístico y facilita la interpretación de los resultados.

Una de las principales ventajas de la regresión lineal es su simplicidad y transparencia, ya que permite identificar de manera clara la influencia de cada variable sobre la predicción. Sin embargo, este modelo asume condiciones como linealidad, independencia de los errores y homocedasticidad, las cuales no siempre se cumplen en fenómenos reales complejos. En contextos donde existen comportamientos no lineales o alta variabilidad, la capacidad predictiva de la regresión lineal puede verse limitada.

En el marco de este proyecto, la regresión lineal se emplea como un modelo base, que permite establecer un punto de referencia inicial para evaluar el desempeño de técnicas más avanzadas en la predicción del tráfico de datos de las zonas WiFi públicas. [12]

Ecuación General:
La relación se expresa como una línea recta:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon$$

- Y: Variable dependiente (resultado o predicción).
- $X_1, X_2, \ldots, X_n$: Variables independientes (predictores o inputs).
- β_0: Intercepto (valor de Y cuando $X = 0$).
- $\beta_1, \beta_2, \ldots, \beta_n$: Coeficientes que indican la influencia de cada X en Y.
- ϵ: Error residual (diferencia entre el valor observado y el predicho).

Ecuación 1. Ecuación General

Random Forest: Random Forest es un algoritmo de aprendizaje supervisado basado en el enfoque de ensambles, cuyo principio fundamental consiste en la construcción de múltiples árboles de decisión independientes durante la fase de entrenamiento. Cada árbol es generado a partir de una muestra aleatoria del conjunto de datos original y utiliza un subconjunto aleatorio de variables en cada división, lo que introduce diversidad entre los modelos individuales.

La predicción final del modelo se obtiene mediante la agregación de las predicciones de todos los árboles, generalmente a través del promedio en problemas de regresión o votación mayoritaria en problemas de clasificación. Este mecanismo permite reducir la varianza del modelo y mejorar su capacidad de generalización frente a datos no observados.

Una de las principales ventajas de Random Forest es su capacidad para modelar relaciones no lineales complejas y manejar conjuntos de datos con alta dimensionalidad, así como su robustez frente al ruido y los valores atípicos. No obstante, su interpretabilidad es menor en comparación con modelos lineales, dado que la predicción resulta de la combinación de múltiples árboles. En este proyecto, Random Forest se utiliza como un modelo robusto para capturar la variabilidad del tráfico de datos en las zonas WiFi públicas. [13]

Ecuación conceptual:

Para Regresión: La predicción final ($\hat{y}$) es el promedio de las predicciones individuales ($\hat{y}_i$) de los árboles (T):

$$\hat{y} = \frac{1}{T} \sum_{i=1}^{T} \hat{y}_i$$

- T: Número total de árboles en el bosque.
- $\hat{y}_i$: Predicción del árbol i.

Ecuación 2. Ecuación conceptual de la Regresión

Redes neuronales: Las redes neuronales artificiales son modelos de aprendizaje supervisado inspirados en el funcionamiento del sistema nervioso biológico, diseñados para identificar patrones complejos a partir de datos. Estas redes están compuestas por un conjunto de unidades básicas denominadas neuronas artificiales, organizadas en capas: una capa de entrada, una o varias capas ocultas y una capa de salida. Cada neurona recibe señales de entrada ponderadas mediante pesos, las combina y aplica una función de activación que determina su salida.

El proceso de aprendizaje de una red neuronal consiste en ajustar los pesos de las conexiones con el objetivo de minimizar el error entre las salidas predichas y los valores reales, comúnmente mediante algoritmos de optimización como el descenso por gradiente y el método de retropropagación del error (*backpropagation*). Esta capacidad de ajuste permite a las redes neuronales modelar relaciones no lineales complejas entre las variables de entrada y la variable objetivo.

Las redes neuronales han demostrado ser especialmente efectivas en problemas donde los patrones subyacentes no pueden ser representados adecuadamente mediante modelos lineales. Sin embargo, su desempeño depende de factores como la cantidad de datos disponibles, la arquitectura de la red y la correcta selección de hiperparámetros. En este proyecto, las redes neuronales se emplean como una alternativa no lineal para la predicción del tráfico de datos en las zonas WiFi públicas, permitiendo capturar comportamientos complejos y variaciones temporales en el consumo. [14]

Support Vector Regression (SVR): es una técnica de aprendizaje supervisado derivada de las máquinas de soporte vectorial, orientada a la resolución de problemas de regresión. Su objetivo principal es encontrar una función que aproxime la relación entre las variables de entrada y la variable objetivo, permitiendo un margen de error controlado. A diferencia de otros modelos de regresión, SVR busca maximizar la capacidad de generalización mediante la definición de un margen de tolerancia ε, dentro del cual los errores no son penalizados.

El modelo se construye a partir de un subconjunto de los datos de entrenamiento, conocidos como vectores de soporte, que determinan la forma de la función de regresión. Mediante el uso de funciones kernel, SVR puede modelar relaciones no lineales complejas transformando los datos a espacios de mayor dimensión. No obstante, su desempeño depende significativamente de la correcta selección de hiperparámetros y del tipo de kernel utilizado. En este proyecto, SVR se emplea como un modelo no lineal para capturar patrones complejos en la predicción del tráfico de datos de las zonas WiFi públicas. [15]

Estructura básica:

$$z = \sum_{i=1}^{n} w_i x_i + b$$

$$a = \sigma(z)$$

- x_i: Entrada a la neurona i.
- w_i: Peso asociado a la entrada x_i.
- b: Sesgo (bias), un término adicional para ajustar la salida.
- σ: Función de activación (ReLU, Sigmoide, etc.).
- a: Salida de la neurona.

Ecuación 3. Ecuación básica de neurona

Estas técnicas son ampliamente utilizadas en problemas similares debido a su capacidad para identificar patrones y realizar predicciones precisas basadas en datos históricos.

Métricas de Evaluación de Modelos

Para evaluar el desempeño de los modelos predictivos, se emplearán las siguientes métricas:

Error Cuadrático Medio (MSE) El MSE mide la media de los errores al cuadrado entre las predicciones del modelo y los valores reales. Penaliza más los errores grandes debido al término cuadrático, lo que lo hace útil para identificar grandes desviaciones. Se muestra cómo es su calculo en la ecuación 4. [10]

Ecuación:

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

- y_i: Valor real de la variable dependiente.
- $\hat{y}_i$: Predicción del modelo.
- n: Número total de observaciones.

Ecuación 4. Ecuación del MSE (Mean Squared Error)

Error Absoluto Medio (MAE): El MAE mide la media de los valores absolutos de las diferencias entre las predicciones del modelo y los valores reales. Es una métrica menos sensible a los errores extremos que el MSE, lo que la hace adecuada para medir la precisión general.

Ecuación:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Ecuación 5. Ecuación del MAE (Mean Absolute Error)

Coeficiente de Determinación (R^2): Evalúa la proporción de la varianza total de la variable dependiente que es explicada por el modelo. Un valor de **R^2** cercano a 1 indica un buen ajuste del modelo, mientras que valores cercanos a 0 indican que el modelo no explica bien la varianza de los datos.

Ecuación:

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

- $\bar{y}$: Media de los valores reales.

Ecuación 6. Ecuación del R^2 (Coeficiente de Determinación)

Error porcentual absoluto medio (MAPE): El error de porcentaje medio absoluto es una medida de error relativa que utiliza valores absolutos para evitar que los errores positivos y negativos se cancelen entre sí y utiliza errores relativos para permitirle comparar la precisión de previsión entre métodos de serie de tiempo. [9]

$$\text{MAPE} = 100\frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

Ecuación 7. MAPE (Mean Absolute Percentage Error)

3.3 **ANTECEDENTES**

Los antecedentes proporcionan una base de trabajos previos relacionados con la temática del proyecto.

Wi-Fi Data Analysis Based on Machine Learning

El trabajo *Wi-Fi Data Analysis Based on Machine Learning* propone el uso de técnicas de aprendizaje automático para fortalecer la seguridad en redes Wi-Fi mediante la detección de anomalías y posibles intrusiones.

La metodología se fundamenta en la recolección de datos de red a través del protocolo SNMP, seguida de un proceso de preprocesamiento que permite estructurar la información para su análisis. En el estudio se implementaron y evaluaron principalmente modelos basados en redes neuronales profundas (DNN) y máquinas de soporte vectorial (SVM), considerando métricas como precisión, recall y estabilidad del modelo.

Los resultados evidenciaron que las redes neuronales profundas presentan una alta capacidad para identificar patrones complejos de comportamiento anómalo, mientras que los modelos SVM demostraron una mayor consistencia y robustez en diferentes escenarios de red. Adicionalmente, el artículo destaca que en la literatura previa se han empleado algoritmos como Random Forest y K-Nearest Neighbors para problemas similares, lo que resalta la relevancia de estos enfoques en el análisis de datos Wi-Fi. Este antecedente resulta pertinente para el presente proyecto al evidenciar la aplicabilidad del aprendizaje automático en el análisis y la seguridad de redes inalámbricas.

Fuente: https://aircconline.com/ijcsit/V15N4/15423ijcsit04.pdf

Internet Usage Modeling of Large Wireless Networks Using Self-Organizing Maps

(Modelado del uso de internet en redes inalámbricas grandes mediante mapas auto-organizados) El trabajo "Internet Usage Modeling of Large Wireless Networks Using Self-Organizing Maps" propone un enfoque sistemático para el análisis y modelado del comportamiento de uso de Internet en redes inalámbricas de gran escala, a partir del estudio de tendencias colectivas de usuarios móviles.

Los autores emplean mapas auto-organizados (Self-Organizing Maps, SOM) como técnica de aprendizaje no supervisado para descubrir, organizar y visualizar patrones de acceso a dominios web, utilizando un conjunto masivo de datos compuesto por miles de usuarios y billones de registros provenientes de trazas WLAN, netflow y DHCP. La metodología incluye la recolección y correlación de datos de tráfico, la agregación temporal del tiempo de uso por usuario y dominio, y la normalización de las variables para permitir un análisis multidimensional eficiente.

A través del entrenamiento del SOM, se identifican tendencias menores representadas por neuronas individuales, las cuales posteriormente se agrupan en tendencias mayores mediante técnicas de clustering aplicadas sobre el mapa. Los resultados muestran que los patrones de uso de Internet pueden ser modelados de forma organizada, preservando la similitud entre comportamientos de usuarios y revelando correlaciones entre dominios web relacionados. Este enfoque permite visualizar tendencias de acceso, identificar grupos de comportamiento y establecer relaciones entre categorías de sitios web, aportando una base sólida para aplicaciones como la planificación de redes, el modelado de tráfico, la simulación de redes inalámbricas y el análisis del comportamiento de usuarios en entornos móviles de gran escala.

Fuente: https://www.cise.ufl.edu/~helmy/papers/scenes-cameraready.pdf

Kaseya Center Installing Enterprise-Class Guest WiFi to Capture Data and Improve the Fan Experience

(Kaseya Center: implementación de WiFi público de clase empresarial para la captura de datos y mejora de la experiencia de los aficionados)

El caso de estudio Kaseya Center Installing Enterprise-Class Guest WiFi to Capture Data and Improve the Fan Experience describe la implementación de una infraestructura de WiFi público de clase empresarial en el estadio del equipo Miami Heat, con el objetivo de capturar datos de uso y mejorar la experiencia de los asistentes durante eventos deportivos y de entretenimiento. La solución permitió analizar patrones de tráfico y comportamiento de los usuarios en tiempo real, facilitando la optimización de servicios y la toma de decisiones orientadas a la experiencia del aficionado.

La metodología se basó en la captura de información demográfica y de comportamiento de los asistentes, quienes en su mayoría eran desconocidos debido a la compra de entradas a través de terceros. Para ello, se emplearon portales cautivos (splash pages) personalizados que permitían a los usuarios registrarse de forma sencilla mediante redes sociales como Facebook o a través de formularios breves.

Como resultado, en los primeros doce meses se registraron más de 290.000 usuarios únicos conectados a la red WiFi, identificándose aproximadamente 700.000 dispositivos únicos. El análisis reveló que cerca del 25 % de los asistentes a los eventos utilizaban el WiFi durante cada partido o concierto y que aproximadamente el 90 % correspondían a visitantes únicos. Estos resultados permitieron personalizar campañas de comunicación y ofertas, generando un incremento en las interacciones con los asistentes y un aumento en las ventas de alimentos, bebidas y productos oficiales del equipo. Este antecedente evidencia el valor estratégico del WiFi público no solo como servicio de conectividad, sino como una fuente clave de datos para el análisis de comportamiento y la optimización de experiencias en entornos de alta concurrencia.

Fuente: https://purple.ai/case-studies/miami-heat/

4 PREPARACIÓN DE DATOS

La preparación de los datos constituye una etapa fundamental dentro del desarrollo del presente proyecto aplicado, dado que la calidad de los datos de entrada condiciona directamente el desempeño, la confiabilidad y la validez de los modelos predictivos construidos. En el contexto de este trabajo, los datos corresponden al tráfico histórico de las zonas WiFi públicas de Santiago de Cali, recolectados de forma continua durante un periodo aproximado de un año.

Este capítulo describe la metodología seguida para la depuración, transformación y estructuración de los datos, garantizando que estos cumplieran con los requisitos necesarios para su uso en modelos de predicción basados en series de tiempo. Las actividades desarrolladas en esta etapa fueron aplicadas de forma homogénea a todas las zonas WiFi analizadas, con el fin de asegurar consistencia metodológica y comparabilidad de resultados.

A continuación, se muestra un diagrama de bloques de la metodología aplicada. Posteriormente se explicará de manera detallada cada bloque (paso) en los puntos del 4.1 al 4.11.

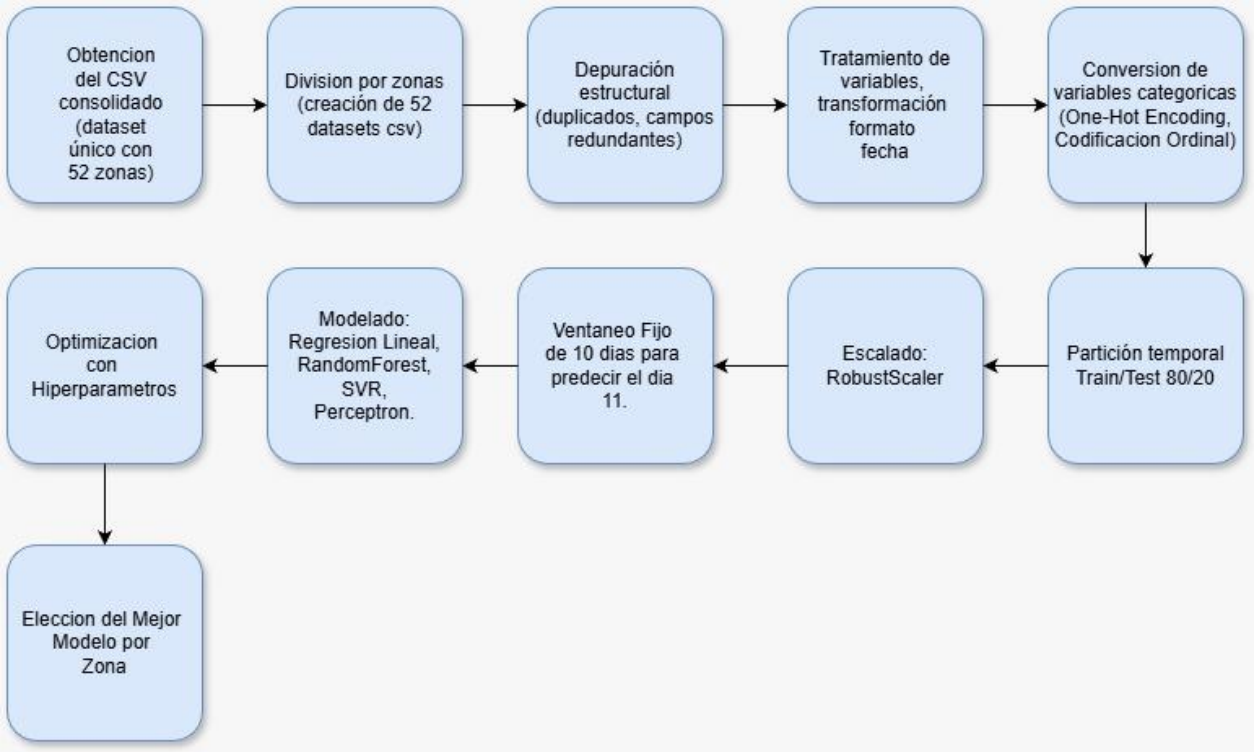


Figura 1. Diagrama de Bloques Metodología

4.1 Obtención CSV Consolidado

Como punto de partida, la data proporcionada por el DATIC consistía en un solo dataset en formato CSV con todas las estadísticas de todas las zonas WiFi (52 en total), tal como se observa en la tabla 1.

FECHA CONEXIÓN	AREA	NOMBRE ZONA	COMUNA	MODEL	NÚMERO CONEXIONES	USAGE (kB)	PORCENTAJE USO
1-ene-24	URBAN	001_ZW Parque Ingenio	17	MR76	42	3709424	4,244
1-ene-24	URBAN	001_ZW Parque Ingenio	17	MR76	49	2550212	2,917
1-ene-24	URBAN	002_ZW Canchas Panamericanas	19	MR76	8	1419251	1,624
1-ene-24	URBAN	003_ZW Parque del Perro	19	MR76	31	558191	0,639
1-ene-24	URBAN	003_ZW Parque del Perro	19	MR76	26	279273	0,319
1-ene-24	URBAN	005_ZW Parque Barrio Obrero	9	MR76	131	1791964	2,05
1-ene-24	URBAN	005_ZW Parque Barrio Obrero	9	MR76	99	631069	0,722
1-ene-24	URBAN	006_ZW Parque Vallado	15	MR76	110	38318	0,044
1-ene-24	URBAN	010_ZW Parque Antonio Nariño	16	MR76	31	5614	0,006

Tabla 1. Estructura inicial de datos

4.2 **División por zonas (Creación de 52 datasets)**

Dado que cada zona WiFi tiene un comportamiento diferente se optó por dividir dicho dataset en 52 datasets los cuales iban a ser uno a uno depurados.

4.3 **Depuración Estructural**

Una vez obtenidos los 52 dataset, se realizó una exploración de su contenido para identificar outliers, datos redundantes y variables que no aportaban información relevante al proceso de modelado.

Durante esta etapa se identificaron columnas cuyo valor permanecía constante para todos los registros de una misma zona, tales como el nombre de la zona, el tipo de área (urbana), modelo del Access Point, la latitud y la longitud. Dado que el modelado se realiza de manera independiente por zona y que dichas variables no presentan variabilidad temporal, se concluyó que no aportaban capacidad discriminativa al modelo y, por tanto, fueron eliminadas.

4.4 **Tratamiento de variables, transformación formato fecha**

Se convirtieron atributos relacionados con fechas en formato calendario que no podían ser utilizados directamente como variables de entrada por lo que se optó cambiar dichas fechas al formato *yyyy-mm-dd*.

4.5 **Conversión de variables categóricas**

Una vez depurados los datos, se procedió a la transformación de las variables categóricas en representaciones numéricas, requisito indispensable para su uso en algoritmos de aprendizaje automático.

En particular, se incorporaron variables relacionadas con la temporalidad del consumo, tales como el tipo de día (laboral, fin de semana o festivo). Estas variables fueron codificadas mediante técnicas de codificación binaria (one-hot encoding), permitiendo que los modelos capturaran patrones de

comportamiento asociados a distintos tipos de días sin introducir relaciones ordinales artificiales. También se codificaron los días de la semana de forma ordinal, de 0 a 6. Asignando el 0 al Lunes y el 6 al Domingo.

Al final de la preparación de los datos quedaron las siguientes variables en cada Dataset:

FECHA_CONEXION	DIA_SEMANA	LABORAL	FIN_DE_SEMANA	FESTIVO	PORCENTAJE_USO	NUMERO_CONEXIONES	USAGE_KB
1/01/2024	0	0	0	1	1233	350	10774630
2/01/2024	1	1	0	0	2865	280	24119120
3/01/2024	2	1	0	0	4708	360	57052340
4/01/2024	3	1	0	0	1386	290	14644230
5/01/2024	4	1	0	0	277	410	3003800
6/01/2024	5	0	1	0	180	370	1377260
7/01/2024	6	0	1	0	71	390	715450
8/01/2024	0	0	0	1	1507	310	23754920
9/01/2024	1	1	0	0	5	310	23754920

Tabla 2. Variables definitivas por dataset y su configuración

- **FECHA_CONEXION:**
  Fecha calendario en la que se registraron las mediciones de tráfico y conexiones de la zona WiFi. Se utiliza para garantizar el orden cronológico de la serie de tiempo, aunque no se emplea directamente como variable predictora.
- **DIA_SEMANA:**
  Variable numérica que representa el día de la semana en que se realizó la medición, codificada de 0 a 6 (lunes a domingo).
  Permite capturar patrones semanales recurrentes en el uso de la red WiFi.
- **LABORAL**
  Variable binaria que indica si la fecha corresponde a un día hábil (1) o no (0). Se utiliza para diferenciar el comportamiento del tráfico entre días laborales y no laborales.
- **FIN_DE_SEMANA**
  Variable binaria que indica si la medición corresponde a un sábado o domingo. Ayuda a identificar cambios en el patrón de uso asociados al comportamiento recreativo de los usuarios.
- **FESTIVO**
  Variable binaria que señala si la fecha corresponde a un día festivo oficial. Es relevante para capturar picos o reducciones atípicas en el tráfico WiFi debido a eventos no laborales.
- **PORCENTAJE_USO**
  Indicador que representa el nivel de utilización de la capacidad disponible de la zona WiFi durante el día.
  Se emplea como variable explicativa del comportamiento del tráfico y debe ser escalado antes del modelado.

- **NUMERO_CONEXIONES**
  Cantidad total de conexiones registradas en la zona WiFi durante el día. Refleja la demanda de usuarios y se utiliza como variable de entrada del modelo predictivo.
- **USAGE_KB**
  Cantidad total de datos transmitidos en la zona WiFi durante el día, medida en kilobytes (kB). Constituye la variable objetivo del modelo de predicción de tráfico.

La composición completa del conjunto de datos de la primera zona WiFi (Parque del Ingenio) se muestra como ejemplo en el Anexo 1.

4.6 **Partición temporal Train/Test 80/20**

Dado que el problema abordado corresponde a una tarea de predicción sobre series de tiempo, la partición de los datos se realizó respetando estrictamente el orden temporal de los registros.
Para cada zona WiFi, el conjunto de datos fue dividido en:

- **80 % de los registros iniciales** para entrenamiento.
- **20 % de los registros finales** para probar cada modelo.

Esta estrategia garantiza que los modelos sean evaluados sobre datos futuros no observados durante el entrenamiento, reproduciendo un escenario realista de predicción y evitando la filtración de información del futuro hacia el pasado.

Como resultado del proceso de preparación de datos, cada zona WiFi quedó representada por un conjunto de datos estructurado, compuesto exclusivamente por variables numéricas y listo para ser utilizado en los algoritmos de modelado.
Los atributos finales incluyen:

- Variables temporales codificadas.
- Valores históricos del tráfico de datos.
- Número de conexiones activas.
- Variables derivadas necesarias para el proceso de ventaneo.

El tamaño final de los datasets varía ligeramente entre zonas, dependiendo de la disponibilidad histórica de datos, pero mantiene una estructura homogénea que facilita la aplicación de los mismos modelos y métricas de evaluación en todas ellas.

Con la finalización de esta etapa, los datos quedaron adecuadamente preparados para el desarrollo del capítulo siguiente, en el cual se abordan los procesos de modelado predictivo, comenzando por la construcción de modelos base y continuando con su posterior optimización y evaluación.

La Tabla 3 presenta, para cada zona WiFi analizada, el número total de registros disponibles y su respectiva partición en conjuntos de entrenamiento y prueba, utilizando una división temporal del 80 % para entrenamiento y 20 % para prueba. Esta estrategia garantiza una cantidad suficiente de datos históricos para el entrenamiento de los modelos, manteniendo al mismo tiempo un conjunto de prueba representativo para su evaluación.

Se tiene un total de 17900 registros entre todos los dataset y un promedio de 344.23 registros por cada dataset.
Se tiene una tabla de todas las zonas con el total de registros de cada zona y cómo quedó su división 80/20 en el Anexo 2.

4.7 **Escalado de datos con Robust Scaler**

Posteriormente, se llevó a cabo el escalado de las variables numéricas con el propósito de reducir el impacto de valores atípicos y mejorar la estabilidad de los modelos predictivos.

Dado que el tráfico de datos y el número de conexiones presentan distribuciones asimétricas y valores extremos frecuentes, se optó por el uso del ***Robust Scaler***, el cual utiliza estadísticas robustas como la mediana y el rango intercuartílico. Este método resulta más adecuado que técnicas como la normalización o la estandarización clásica en escenarios donde los outliers son comunes.

Es importante destacar que el escalador fue ajustado exclusivamente utilizando los datos del conjunto de entrenamiento, y posteriormente aplicado al conjunto de prueba, siguiendo las buenas prácticas en aprendizaje automático y evitando la introducción de sesgos en la evaluación del modelo.

4.8 **Ventaneo fijo de 10 días para predecir el 11**

Con los datos ya depurados, transformados y escalados, se procedió a la construcción de las ventanas temporales necesarias para el modelado predictivo.

El ventaneo consiste en transformar la serie temporal original en un conjunto de instancias supervisadas, donde cada observación se construye a partir de valores históricos de la serie. En este proyecto se adoptó una estrategia de ventanas deslizantes de tamaño fijo de 10 días, en la cual se utilizan los valores correspondientes a diez días consecutivos para predecir el valor del tráfico del día siguiente, como se muestra en la figura 2.

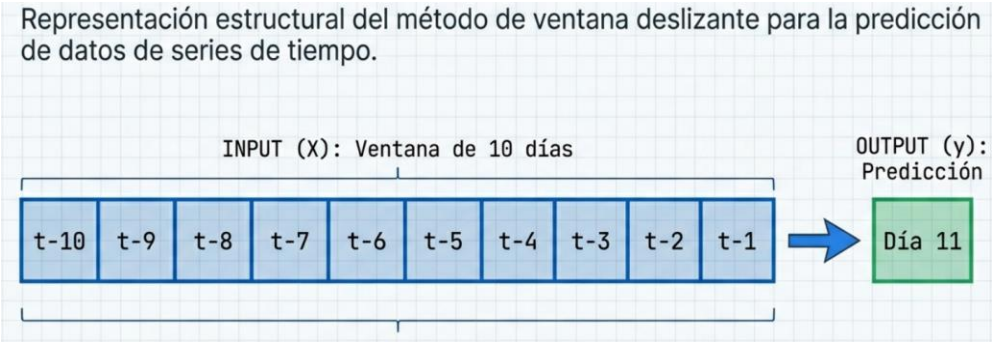


Figura 2. Ventaneo deslizante de 10 días

Esta configuración permite capturar patrones semanales recurrentes en el uso del servicio WiFi, al tiempo que mantiene un equilibrio entre la cantidad de información histórica utilizada y la complejidad del modelo.

El proceso de ventaneo se realizó de manera independiente sobre los conjuntos de entrenamiento y prueba, preservando el orden temporal y asegurando que no existiera solapamiento de información entre ambos.
A continuación se muestra un ejemplo del ventaneo resultante de una de las Zonas, así queda el dataset ventaneado tanto de Train como de Test:

DIA_SEMANA_t-10	LABORAL_t-10	FIN_DE_SEMANA_t-10	FESTIVO_t-10	PORCENTAJE_USO_t-10	NUMERO_CONEXIONES_t-10	USAGE_KB_t-10
0	1	0	0	-2.825.732.899.022.800	-4.827.586.206.896.550	-129.543.487.497.112
1	1	0	0	-2.833.876.221.498.370	-5.517.241.379.310.340	-1.545.713.372.131.270
2	1	0	0	-2.825.732.899.022.800	-5.517.241.379.310.340	-1.474.043.055.587.380
3	1	0	0	-2.817.589.576.547.230	-5.517.241.379.310.340	-1.227.590.675.290.700
4	1	0	0	-2.817.589.576.547.230	-5.517.241.379.310.340	-1.068.340.123.289.250
5	0	1	0	-2.809.446.254.071.660	2.068.965.517.241.370	432.351.205.573.685
6	0	1	0	-2.825.732.899.022.800	7.586.206.896.551.720	845.609.667.546.183
0	0	0	1	-2.858.306.188.925.080	6.896.551.724.137.930	-911.536.604.270.829
1	1	0	0	-2.825.732.899.022.800	-4.827.586.206.896.550	-1.190.872.566.973.330
2	1	0	0	-2.809.446.254.071.660	-5.517.241.379.310.340	-1.007.950.544.706.590
3	1	0	0	-2.809.446.254.071.660	-5.517.241.379.310.340	-976.580.590.950.056
4	1	0	0	-2.809.446.254.071.660	-4.827.586.206.896.550	-804.623.967.342.621
5	0	1	0	-2.833.876.221.498.370	0.0	-792.926.981.502.827
6	0	1	0	-2.833.876.221.498.370	9.655.172.413.793.100	799.376.721.358.414
0	1	0	0	-2.817.589.576.547.230	-4.137.931.034.482.750	-918.453.046.293.281
1	1	0	0	-2.809.446.254.071.660	-4.827.586.206.896.550	-929.880.942.595.423
2	1	0	0	-2.809.446.254.071.660	-5.517.241.379.310.340	-882.891.182.083.009
3	1	0	0	-2.793.159.609.120.520	-5.517.241.379.310.340	-551.183.668.115.055
4	1	0	0	-2.793.159.609.120.520	-4.827.586.206.896.550	-392.076.069.930.484
5	0	1	0	-2.817.589.576.547.230	1.379.310.344.827.580	-68.298.288.275.201
6	0	1	0	-2.833.876.221.498.370	896.551.724.137.931	737.877.148.432.942
0	1	0	0	-2.809.446.254.071.660	-4.137.931.034.482.750	-678.799.381.346.334
1	1	0	0	-2.801.302.931.596.090	-4.827.586.206.896.550	-67.585.621.452.827
2	1	0	0	-2.793.159.609.120.520	-5.517.241.379.310.340	-577.184.444.690.642

Tabla 3. Ventaneo del día 10

Se muestra que se ha tomado el valor de cada variable para 10 días atrás, esto se reproduce de igual manera para los siguientes días de esta manera:

valor de las variables de 9 días atrás:
DIA_SEMANA_t-9
LABORAL_t-9
FIN_DE_SEMANA_t-9
FESTIVO_t-9
PORCENTAJE_USO_t-9
NUMERO_CONEXIONES_t-8
USAGE_KB_t-9

valor de las variables de 8 días atrás:
DIA_SEMANA_t-8
LABORAL_t-8
FIN_DE_SEMANA_t-8
FESTIVO_t-8
PORCENTAJE_USO_t-8
NUMERO_CONEXIONES_t-8
USAGE_KB_t-8

Así sucesivamente hasta el t-0 donde se tiene el target, el cual es el valor de la variable USAGE_KB a predecir para el día siguiente:

DIA_SEMANA_t-0
LABORAL_t-0
FIN_DE_SEMANA_t-0
FESTIVO_t-0
PORCENTAJE_USO_t-0
NUMERO_CONEXIONES_t-0
TARGET

4.9 **Modelado: Regresión Lineal, Random Forest, SVR, Perceptrón**

- **Regresión Lineal**: utilizada como modelo de referencia simple para evaluar la capacidad de los enfoques más complejos.
- **Random Forest Regressor**: basado en ensambles de árboles de decisión, capaz de modelar relaciones no lineales y manejar interacciones complejas entre variables.
- **Support Vector Regression (SVR)**: que permite capturar relaciones no lineales mediante el uso de funciones kernel.
- **Perceptrón Multicapa (MLP Regressor)**: representativo de los modelos de redes neuronales aplicados a problemas de regresión.

4.10 **Optimización con Hiperparámetros**

Por búsqueda de grilla, se buscan los mejores hiperparámetros de cada modelo con tal de encontrar mejores predicciones de las zonas.

4.11 **Elección del mejor modelo por zona**

Usando métricas de evaluación como el MAPE, RMSE, MAE, R^2, se elige el mejor modelo para cada una de las 52 zonas. El proceso de modelado se explica con mayor detalle en el capitulo 5.

5 MODELADO

En este capítulo se describe a profundidad el proceso de construcción, entrenamiento y evaluación de los modelos predictivos empleados para estimar el tráfico de datos en las zonas WiFi públicas de Santiago de Cali. El objetivo principal de esta etapa es identificar modelos capaces de capturar los patrones temporales presentes en los datos históricos, garantizando un adecuado desempeño predictivo y permitiendo comparaciones consistentes entre distintas técnicas de aprendizaje automático supervisados.

El proceso de modelado se desarrolló de manera sistemática, iniciando con la implementación de modelos base para cada zona WiFi y continuando con la optimización de dichos modelos mediante la exploración de diferentes configuraciones de hiperparámetros. Esta estrategia permitió establecer un punto de referencia inicial y evaluar de forma objetiva las mejoras obtenidas a través del ajuste de los modelos.

5.1 Modelos base

Como primer paso, se implementaron modelos base con parámetros por defecto, con el propósito de obtener una línea base de desempeño que sirviera como referencia para posteriores comparaciones. Estos modelos fueron entrenados utilizando exclusivamente los conjuntos de datos preparados en el capítulo anterior, respetando la partición temporal entre entrenamiento y prueba.

Los modelos seleccionados pertenecen a distintas familias de algoritmos de aprendizaje automático, con el fin de capturar diferentes tipos de relaciones entre las variables de entrada y la variable objetivo. Entre los modelos base evaluados se incluyen:

- **Regresión Lineal**: utilizada como modelo de referencia simple para evaluar la capacidad de los enfoques más complejos.
- **Random Forest Regressor**: basado en ensambles de árboles de decisión, capaz de modelar relaciones no lineales y manejar interacciones complejas entre variables.
- **Support Vector Regression (SVR)**: que permite capturar relaciones no lineales mediante el uso de funciones kernel.
- **Perceptrón Multicapa (MLP Regressor)**: representativo de los modelos de redes neuronales aplicados a problemas de regresión.

Cada uno de estos modelos fue entrenado de forma independiente para cada zona WiFi, utilizando los datos ventaneados correspondientes, con el fin de capturar las particularidades temporales propias de cada ubicación.

Con el propósito de evaluar de manera estructurada el desempeño inicial de los modelos predictivos propuestos, se presentan a continuación los resultados correspondientes a los modelos base de Random Forest, Regresión Lineal y Support Vector Regression (SVR), aplicados de forma consistente sobre las distintas zonas WiFi analizadas.

Para cada modelo se reportan las métricas de desempeño MAE, RMSE, MAPE y R^2, calculadas sobre el conjunto de prueba, lo que permite obtener una visión integral del nivel de error y de la capacidad explicativa alcanzada en cada zona.

Con el fin de mejorar la claridad y facilitar la interpretación de los resultados, las métricas se presentan en tablas separadas por modelo, permitiendo comparar el comportamiento de cada enfoque predictivo a lo largo de las diferentes zonas WiFi. (Se muestran las métricas de las primeras 10 zonas en cada modelo. Para una visualización completa, por favor referirse a los anexos 3, 4, 5 y 6)
Los resultados constituyen la línea base a partir de la cual se justifica y orienta el posterior proceso de optimización de hiperparámetros y selección de modelos

Zona WiFi	MAE	RMSE	MAPE (%)	R^2
001_ZW Parque Ingenio	1.041.531,47	1.434.059,00	25,80	0,667
002_ZW Canchas Panamericanas	600.874,54	839.877,00	21,97	0,675
003_ZW Parque del Perro	692.025,95	893.662,00	24,48	0,498
004_ZW Parque San Nicolas	2.166.368,31	2.900.307,00	20,05	0,546
005_ZW Parque Barrio Obrero	3.359.334,15	4.438.344,00	17,91	0,663
006_ZW Parque Pizamos	765.846,47	1.325.091,00	20,78	0,723
007_ZW Parque Alfonso Lopez	2.935.948,85	3.519.269,00	31,52	0,279
008_ZW Parque Antonio Nariño	753.761,88	914.938,00	16,85	0,763
009_ZW Parque Santa Rosa Poblado	1.902.469,62	2.554.255,00	54,07	0,105
010_ZW Polideportivo Los Naranjos	1.628.934,74	2.483.142,00	31,12	0,400

Tabla 4. Métricas de desempeño – Random Forest (modelo base)

Zona WiFi	MAE	RMSE	MAPE (%)	R^2
001_ZW Parque Ingenio	1.333.157,15	1.808.586,11	38,95	0,495
002_ZW Canchas Panamericanas	723.666,69	971.803,44	27,98	0,604
003_ZW Parque del Perro	846.340,99	1.117.579,78	26,15	0,317
004_ZW Parque San Nicolas	2.588.836,28	3.298.815,08	20,01	0,618
005_ZW Parque Barrio Obrero	4.679.565,87	5.661.379,22	34,06	0,524
006_ZW Parque Pizamos	818.176,86	1.066.200,58	20,56	0,856
007_ZW Parque Alfonso Lopez	3.209.979,79	3.778.141,65	30,35	0,346
008_ZW Parque Antonio Nariño	722.121,47	924.698,10	14,92	0,777
009_ZW Parque Santa Rosa Poblado	17.602.258,36	23.131.209,33	639,18	-1,440
010_ZW Polideportivo Los Naranjos	2.163.544,77	3.088.639,42	36,97	0,361

Tabla 5. Métricas de desempeño – Regresión Lineal (modelo base)

Zona WiFi	MAE	RMSE	MAPE (%)	R^2
001_ZW Parque Ingenio	1.078.031,66	1.670.629,85	33,45	0,426
002_ZW Canchas Panamericanas	792.958,84	1.150.524,25	40,13	0,318
003_ZW Parque del Perro	1.053.117,13	1.555.317,83	30,65	0,360
004_ZW Parque San Nicolas	4.164.597,84	5.807.695,62	28,01	-0,185
005_ZW Parque Barrio Obrero	7.178.427,65	8.947.742,39	28,73	-0,041
006_ZW Parque Pizamos	1.572.501,81	2.157.869,43	45,84	0,246
007_ZW Parque Alfonso Lopez	3.047.779,64	3.955.758,16	29,93	0,410
008_ZW Parque Antonio Nariño	1.312.741,94	1.756.855,36	21,46	0,461
009_ZW Parque Santa Rosa Poblado	14.743.922,36	21.582.367,08	258,94	-0,863
010_ZW Polideportivo Los Naranjos	2.744.256,90	4.046.963,61	104,77	-0,157

Tabla 6. Métricas de desempeño – SVR (modelo base)

Zona WiFi	MAE	RMSE	MAPE (%)	R^2
001_ZW Parque Ingenio	6.231.274,72	6.706.314,45	269,73	-8,252
002_ZW Canchas Panamericanas	551.104,81	750.546,97	31,57	0,710
003_ZW Parque del Perro	1.057.232,48	1.428.209,74	33,81	0,460
004_ZW Parque San Nicolas	2.863.132,89	3.918.591,31	20,83	0,460
005_ZW Parque Barrio Obrero	4.300.606,96	5.047.431,39	21,79	0,669
006_ZW Parque Pizamos	942.269,20	1.405.260,34	20,39	0,680
007_ZW Parque Alfonso Lopez	2.833.231,21	3.559.343,42	32,05	0,523
008_ZW Parque Antonio Nariño	1.017.913,75	1.246.428,05	21,42	0,729
009_ZW Parque Santa Rosa Poblado	11.209.602,60	17.074.890,28	1609,65	-0,166
010_ZW Polideportivo Los Naranjos	1.723.843,02	2.613.352,62	82,48	0,518

Tabla 7. Métricas de desempeño – Perceptron (modelo base)

5.2 **Optimización de modelos**

Una vez obtenidos los resultados de los modelos base, se procedió a la etapa de optimización, cuyo objetivo fue mejorar el desempeño predictivo mediante el ajuste de los hiperparámetros más relevantes de cada algoritmo.

La optimización se realizó utilizando técnicas de búsqueda de grilla (*grid search*), explorando combinaciones de parámetros previamente definidas en función de las características de cada modelo. Este proceso se llevó a cabo exclusivamente sobre el conjunto de entrenamiento, empleando validación cruzada interna cuando fue pertinente, con el fin de evitar sobreajuste.

Entre los hiperparámetros ajustados se incluyen, entre otros:

- **Random Forest Regressor**: Número de árboles y profundidad máxima en Random Forest.
- **Support Vector Regression (SVR)**: Parámetros del kernel y regularización en SVR.
- **Perceptrón Multicapa (MLP Regressor)**: Número de capas ocultas, número de neuronas y tasa de aprendizaje en el Perceptrón Multicapa.

El modelo optimizado para cada zona fue seleccionado con base en el mejor desempeño obtenido según las métricas definidas, priorizando un equilibrio entre precisión y capacidad de generalización.
Para la optimización de hiperparámetros se excluyó el modelo de Regresión Lineal, ya que este no tiene esta característica.

Se muestra en la tabla 8 los hiperparámetros que fueron usados para encontrar los modelos optimizados:

MODELO	HIPERPARAMETROS	VALORES
Random Forest	n_estimators	[50, 100, 150, 250, 350]
	max_depth	[5, 10, 20, 30, 40]
Perceptron	hidden_layer_sizes	[(50,), (100,), (50, 50), (100, 50), (100, 100), (50, 25, 10)]
	activation	['relu', 'tanh']
	solver	adam
	alpha	[0.0001, 0.001, 0.01]
	learning_rate	['constant', 'adaptive']
	learning_rate_init	[0.001, 0.01]
SVR	kernel	['rbf']
	C	[0.1, 1.0, 10.0, 50, 100]
	epsilon	[0.1, 0.01, 0.004]
	gamma	['scale', 'auto', 0.5, 0.05, 0.005]

Tabla 8. Valores usados en la búsqueda de hiperparámetros

Para cada modelo, aplicado a cada una de las 52 zonas, se hizo una búsqueda de grilla con la función **grid_search_forecaster** de la biblioteca ScikitLearn de Python.
La grilla evalúa combinaciones de cada modelo, según se ve en la tabla 8. Por ejemplo, para Random Forest, variando **n_estimators** entre 50, 100, 150, 250, 350 y max_depth entre 5, 10, 20, 30, 40.
Esto nos lleva, en el caso de Random Forest, a tener 25 combinaciones (5×5) por zona, con los lags o tamaño de ventana que fija los rezagos del target en 10.

La validación interna se define con la función TimeSeriesFold, respetando el orden temporal de la serie.

En cada fold, el modelo se entrena con un bloque inicial y se valida prediciendo un horizonte de steps (igual al tamaño del test, 80% del total de registros del dataset).

La función **grid_search_forecaster** prueba cada combinación (25 combinaciones) de hiperparámetros mas los lags y calcula la métrica objetivo.

La métrica usada para seleccionar el mejor set es MAPE.

Finalmente, los resultados completos de la grilla se guardan en un CSV por zona y por modelo. Es decir, se tendrá 4 CSV por zona, uno para Random Forest, otro para SVR, otro para Regresión Lineal y otro para Perceptrón multicapa.

Se muestra un ejemplo de lo dicho en el código de Python compartido en el anexo 11.

5.3 **Resultados de la optimización**

A continuación, se presentan los resultados completos de la evaluación del desempeño de los modelos predictivos considerados en este estudio, aplicados a la totalidad de las zonas WiFi analizadas. En particular, se reportan los resultados correspondientes a los modelos de Support Vector Regression (SVR) y Perceptrón, tanto en su versión base como en su versión optimizada, así como los resultados del modelo Random Forest en su configuración base.

Para cada técnica y para cada zona WiFi se presentan de manera conjunta las métricas MAE, RMSE, MAPE y R^2, calculadas sobre el conjunto de prueba. Esta presentación integral permite evaluar de forma objetiva el nivel de error, la estabilidad de las predicciones y la capacidad explicativa de los modelos, así como analizar el impacto del proceso de optimización de hiperparámetros en el desempeño predictivo.

Con el fin de mejorar la claridad y facilitar la interpretación de los resultados, las métricas se organizan en tablas separadas por técnica, manteniendo en todos los casos una estructura homogénea y consistente entre zonas. Estos resultados constituyen la base para el análisis comparativo posterior y sustentan las conclusiones sobre el comportamiento relativo de los modelos evaluados.
Tal como se hizo con las tablas del capitulo 5.1 (modelos base), se muestran las primeras 10 zonas, las tablas completas se encuentran en los anexos 7, 8 y 9.

Zona WiFi	MAE Base	MAE Opt	RMSE Base	RMSE Opt	MAPE (%) Base	MAPE (%) Opt	R^2 Base	R^2 Opt
Parque Ingenio	1.078.031,66	1.212.908,21	1.670.629,85	1.810.270,40	33,45	35,37	0,426	0,326
Canchas Panamericanas	792.958,84	594.268,45	1.150.524,25	819.178,20	40,13	26,68	0,318	0,654
Parque del Perro	1.053.117,13	880.912,35	1.555.317,83	1.183.474,98	30,65	27,51	0,360	0,629
Parque San Nicolas	4.164.597,84	3.409.988,88	5.807.695,62	4.383.302,51	28,01	26,16	-0,185	0,325
Parque Barrio Obrero	7.178.427,65	4.777.467,53	8.947.742,39	5.848.322,43	28,73	20,60	-0,041	0,555
Parque Pizamos	1.572.501,81	912.416,31	2.157.869,43	1.254.878,11	45,84	22,29	0,246	0,745
Parque Alfonso Lopez	3.047.779,64	2.471.210,70	3.955.758,16	3.027.176,76	29,93	25,97	0,410	0,655
Parque Antonio Nariño	1.312.741,94	1.019.245,81	1.756.855,36	1.273.795,59	21,46	18,25	0,461	0,716
Parque Santa Rosa Poblado	14.743.922,3	11.450.847,6	21.582.367,08	17.760.614,05	258,94	633,99	-0,863	-0,262
Polideportivo Los Naranjos	2.744.256,90	2.635.657,76	4.046.963,61	3.943.596,06	104,77	94,99	-0,157	-0,099

Tabla 9. SVR base vs SVR optimizado

Zona WiFi	MAE Base	MAE Opt	RMSE Base	RMSE Opt	MAPE (%) Base	MAPE (%) Opt	R^2 Base	R^2 Opt
Parque Ingenio	6.231.274,72	1.321.993,78	6.706.314,45	1.759.345,77	269,73	42,94	-8,252	0,363
Canchas Panamericanas	551.104,81	609.670,19	750.546,97	834.375,68	31,57	32,11	0,710	0,641
Parque del Perro	1.057.232,48	1.101.335,72	1.428.209,74	1.438.148,35	33,81	34,47	0,460	0,453
Parque San Nicolas	2.863.132,89	3.717.160,52	3.918.591,31	4.697.175,19	20,83	29,39	0,460	0,225
Parque Barrio Obrero	4.300.606,96	4.979.891,73	5.047.431,39	6.235.258,68	21,79	20,56	0,669	0,494
Parque Pizamos	942.269,20	1.164.503,36	1.405.260,34	1.530.186,00	20,39	27,38	0,680	0,621
Parque Alfonso Lopez	2.833.231,21	2.852.702,35	3.559.343,42	3.520.371,83	32,05	31,26	0,523	0,533
Parque Antonio Nariño	1.017.913,75	1.047.455,41	1.246.428,05	1.344.555,65	21,42	18,49	0,729	0,684
Parque Santa Rosa Poblado	11.209.602,60	7.987.164,00	17.074.890,28	12.286.097,42	1609,65	2186,57	-0,166	0,396
Polideportivo Los Naranjos	1.723.843,02	3.175.521,62	2.613.352,62	4.519.302,55	82,48	164,90	0,518	-0,443

Tabla 10. perceptrón base vs perceptrón optimizado

Zona WiFi	MAE Base	MAE Opt	RMSE Base	RMSE Opt	MAPE (%) Base	MAPE (%) Opt	R^2 Base	R^2 Opt
Parque Ingenio	1.041.531,47	1.076.748,99	1.434.059,00	1.489.884,00	25,80	26,40	0,667	0,640
Canchas Panamericanas	600.874,54	582.352,50	839.877,00	817.877,00	21,97	21,54	0,675	0,692
Parque del Perro	692.025,95	697.841,76	893.662,00	900.166,00	24,48	24,75	0,498	0,491
Parque San Nicolas	2.166.368,31	2.162.611,55	2.900.307,00	2.854.942,00	20,05	20,38	0,546	0,560
Parque Barrio Obrero	3.359.334,15	3.330.300,58	4.438.344,00	4.370.968,00	17,91	17,70	0,663	0,674
Parque Pizamos	765.846,47	755.150,13	1.325.091,00	1.335.905,00	20,78	20,63	0,723	0,718
Parque Alfonso Lopez	2.935.948,85	2.893.594,39	3.519.269,00	3.455.481,00	31,52	30,86	0,279	0,305
Parque Antonio Nariño	753.761,88	759.194,82	914.938,00	919.569,00	16,85	16,94	0,763	0,761
Parque Santa Rosa Poblado	1.902.469,62	1.905.681,01	2.554.255,00	2.559.401,00	54,07	54,12	0,105	0,101
Polideportivo Los Naranjos	1.628.934,74	1.575.362,00	2.483.142,00	2.409.040,00	31,12	30,36	0,400	0,436

Tabla 11. Random Forest base vs Random Forest optimizado

Los resultados obtenidos tras el proceso de optimización evidencian mejoras consistentes frente a los modelos base en la mayoría de las zonas analizadas. En particular, los modelos basados en vectores de soporte y redes neuronales mostraron una mayor capacidad para capturar patrones no lineales presentes en el tráfico de datos, especialmente en zonas con alta variabilidad en el consumo.

Las grillas de búsqueda permitieron identificar configuraciones óptimas de hiperparámetros, las cuales fueron posteriormente utilizadas para entrenar los modelos finales evaluados en el conjunto de prueba. Estos resultados constituyen la base para el análisis comparativo que se presenta en el capítulo siguiente.

En conclusión, el proceso de modelado permitió establecer un marco sólido para la predicción del tráfico de datos en las zonas WiFi públicas, partiendo de modelos base y avanzando hacia configuraciones optimizadas que maximizan el desempeño predictivo. En el siguiente capítulo se presenta un análisis detallado de los resultados obtenidos, comparando el comportamiento de los modelos entre distintas zonas y técnicas, así como su impacto desde una perspectiva de negocio.

Las zonas que mostraron un mayor beneficio tras la búsqueda de hiperparámetros fueron aquellas cuyos modelos de predicción se basaron en SVR y Perceptrón, evidenciando una mejora significativa en su desempeño predictivo.

En contraste, para los conjuntos de datos modelados con Random Forest, el rendimiento se mantuvo prácticamente constante antes y después del ajuste de hiperparámetros, lo cual puede atribuirse al comportamiento intrínseco de las series temporales asociadas al tráfico WiFi analizado.

Finalmente, mostramos los modelos con mejor desempeño para cada zona wifi con sus hiperparametros respectivos. Se muestran los siete primeros, para una visión completa porfavor referirse al anexo 10.

Zona	Modelo	Hiperparámetros
001_ZW Parque Ingenio	Random Forest	n_estimators: 100 max_depth: None random_state: 123 kernel: rbf
002_ZW Canchas Panamericanas	Random Forest	n_estimators: 350 max_depth: 10 random_state: 123 kernel: rbf
003_ZW Parque del Perro	Random Forest	n_estimators: 100 max_depth: None random_state: 123 kernel: rbf
005_ZW Parque Barrio Obrero	Random Forest	n_estimators: 50 max_depth: 10 random_state: 123 kernel: rbf
006_ZW Parque Pizamos	Perceptrón	hidden_layer_sizes: (100) activation: relu solver: adam max_iter: 200 alpha: 0.0001 learning_rate: constant learning_rate_init: 0.001
007_ZW Parque Alfonso Lopez	SVR	kernel: rbf C: 10.0 epsilon: 0.1 gamma: 0.005

Tabla 12. Hiperparametros por Zona.

6 ANÁLISIS DE LOS RESULTADOS

En este capítulo se analizan los resultados obtenidos a partir de los modelos predictivos construidos y optimizados en el capítulo anterior. El análisis se aborda desde tres perspectivas complementarias: la comparación del desempeño entre zonas WiFi, la comparación entre las diferentes técnicas de modelado empleadas y, finalmente, la interpretación de los resultados desde un enfoque de negocio y operación del servicio.

Este análisis permite evaluar no solo la capacidad predictiva de los modelos, sino también su utilidad práctica como herramienta de apoyo para la toma de decisiones relacionadas con la gestión del ancho de banda y la planificación de la infraestructura de las zonas WiFi públicas de Santiago de Cali.

6.1 Comparación de resultados entre zonas WiFi

El desempeño de los modelos mostró variaciones significativas entre las diferentes zonas WiFi analizadas, lo cual refleja la heterogeneidad en los patrones de uso del servicio a lo largo de la ciudad. Zonas con un comportamiento de consumo más estable y patrones temporales regulares presentaron menores errores de predicción, mientras que aquellas con alta variabilidad diaria o picos de tráfico esporádicos evidenciaron mayores dificultades para ser modeladas con precisión.

Estas diferencias ponen de manifiesto la influencia de factores externos, como la ubicación de la zona, la afluencia de usuarios y la naturaleza de las actividades que se desarrollan en cada sector. En zonas con mayor previsibilidad del consumo, los modelos lograron capturar de manera efectiva las tendencias y estacionalidades presentes en la serie temporal, obteniendo valores más bajos de error absoluto medio y raíz del error cuadrático medio.

Por el contrario, en zonas con patrones de uso más irregulares, los errores de predicción tendieron a incrementarse, lo que sugiere la necesidad de considerar información adicional o modelos más complejos para capturar adecuadamente dichos comportamientos.

6.2 **Comparación entre técnicas de modelado**

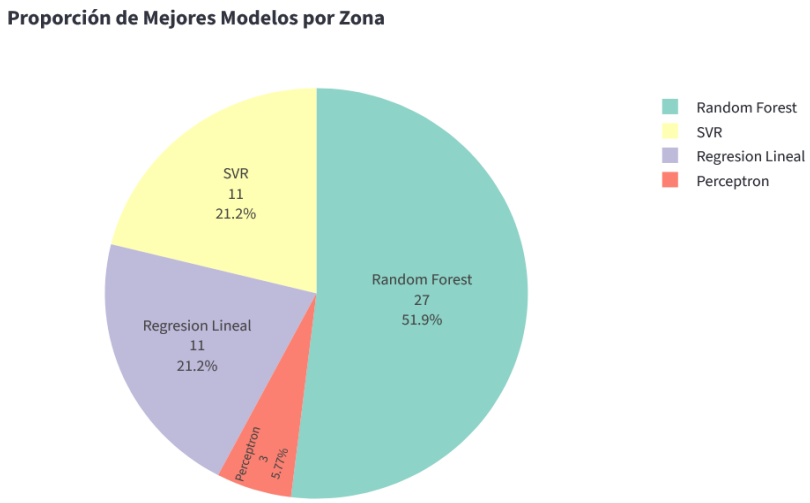


Figura 3. Grafico de torta mostrando la proporción de los mejores modelos por zona

Al comparar las diferentes técnicas de modelado utilizadas, se evidencia que los modelos más simples, como la regresión lineal, presentan un desempeño limitado frente a aquellos capaces de capturar relaciones no lineales. Si bien la regresión lineal ofrece ventajas en términos de simplicidad, interpretabilidad y bajo costo computacional, su capacidad para modelar la complejidad inherente al tráfico de datos resulta insuficiente en escenarios con alta variabilidad temporal.

Los resultados obtenidos muestran que los modelos basados en ensambles, particularmente Random Forest, presentan el mejor desempeño global, al ser el modelo con menor error en 27 de las 52 zonas WiFi analizadas. Este comportamiento se explica por su capacidad para capturar interacciones complejas entre variables, modelar relaciones no lineales y reducir la varianza mediante la combinación de múltiples árboles de decisión, lo que le otorga mayor estabilidad frente a cambios en los patrones de consumo.

Por su parte, los modelos de soporte vectorial (SVR) y la regresión lineal se destacaron como el mejor modelo en 11 zonas cada uno. En el caso de SVR, su desempeño refleja la ventaja de incorporar funciones kernel para modelar relaciones no lineales; sin embargo, al tratarse de un modelo individual, su rendimiento mostró mayor sensibilidad a la selección de hiperparámetros y a la naturaleza específica de los datos de cada zona. En contraste, la regresión lineal presentó un mejor ajuste en aquellas zonas con patrones de consumo más estables y comportamiento cercano a la linealidad.

Por último el Perceptrón Multicapa (MLP) se desempeñó como el mejor modelo en solo 3 zonas.

6.3 **Análisis de las métricas de evaluación**

El uso conjunto de múltiples métricas permitió realizar una evaluación robusta del desempeño de los modelos predictivos. Se emplearon el Error Absoluto Medio (MAE), la Raíz del Error Cuadrático Medio (RMSE), el Coeficiente de Determinación (R^2) y el Error Porcentual Absoluto Medio (MAPE).

El MAE proporcionó una medida directa del error absoluto cometido por los modelos en las mismas unidades del tráfico de datos, mientras que el RMSE permitió identificar la sensibilidad de los modelos frente a errores de gran magnitud. Sin embargo, dado que la variable objetivo del proyecto corresponde al volumen de tráfico medido en gigabytes, estas métricas tendieron a arrojar valores numéricamente elevados, incluso en escenarios donde el ajuste relativo del modelo era adecuado. Esta característica dificulta su interpretación comparativa entre zonas con diferentes niveles de consumo.

El coeficiente de determinación R^2 ofreció una visión global de la capacidad explicativa de los modelos, indicando qué proporción de la variabilidad del tráfico observado es capturada por las variables de entrada. Se identificaron casos en los que un modelo presentaba un buen R^2, pero errores absolutos elevados, lo que evidencia que esta métrica, por sí sola, no es suficiente para evaluar la precisión práctica de las predicciones.

En este contexto, el MAPE se consolidó como la métrica más adecuada para la comparación y selección de modelos, al expresar el error de predicción en términos porcentuales respecto al valor real del tráfico.

Aunque el MAE, el RMSE y el R^2 aportan información relevante sobre distintos aspectos del ajuste de los modelos, el MAPE fue la métrica principal utilizada para evaluar y comparar el desempeño predictivo, debido a su mayor pertinencia en un contexto donde la variable objetivo presenta valores elevados en gigabytes.

A continuación, se muestran un análisis del top 10 de las zonas con menor y mayor MAPE dependiendo del mejor modelo encontrado para cada zona:

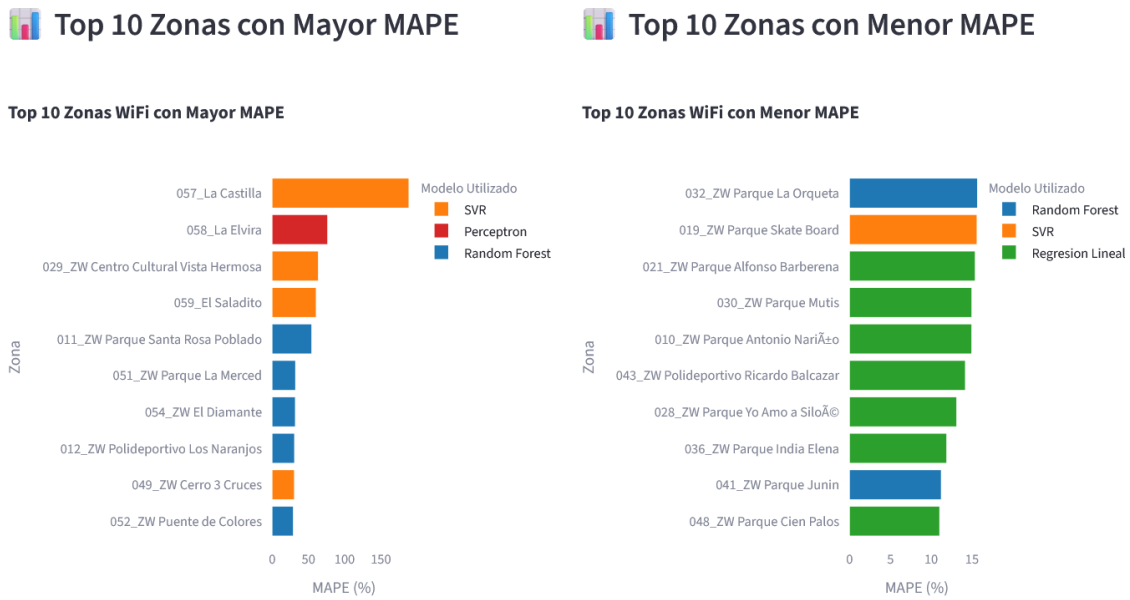


Figura 4. Zonas con mejores predicciones según el MAPE evaluado de su correspondiente modelo (entre menor MAPE, mejor)

6.4 **Análisis desde la perspectiva del negocio**

Desde el punto de vista operativo y de negocio, los resultados obtenidos tienen implicaciones directas en la gestión de las zonas WiFi públicas. La capacidad de anticipar el tráfico de datos permite a los operadores ajustar de manera proactiva el ancho de banda disponible, reduciendo la probabilidad de congestión en periodos de alta demanda y evitando la sobreprovisión de recursos en momentos de bajo uso.

Asimismo, la identificación de zonas con patrones de consumo altamente variables permite priorizar intervenciones específicas, como la ampliación de capacidad o la instalación de nuevos de access points o la implementación de políticas de gestión del tráfico. En este sentido, los modelos predictivos se convierten en una herramienta estratégica para optimizar el uso de los recursos disponibles, enfocarlos en zonas con mayor demanda a futuro y disminuirlos en zonas donde la demanda de proyecta a ser baja.

6.5 **Reproductibilidad de los Resultados**

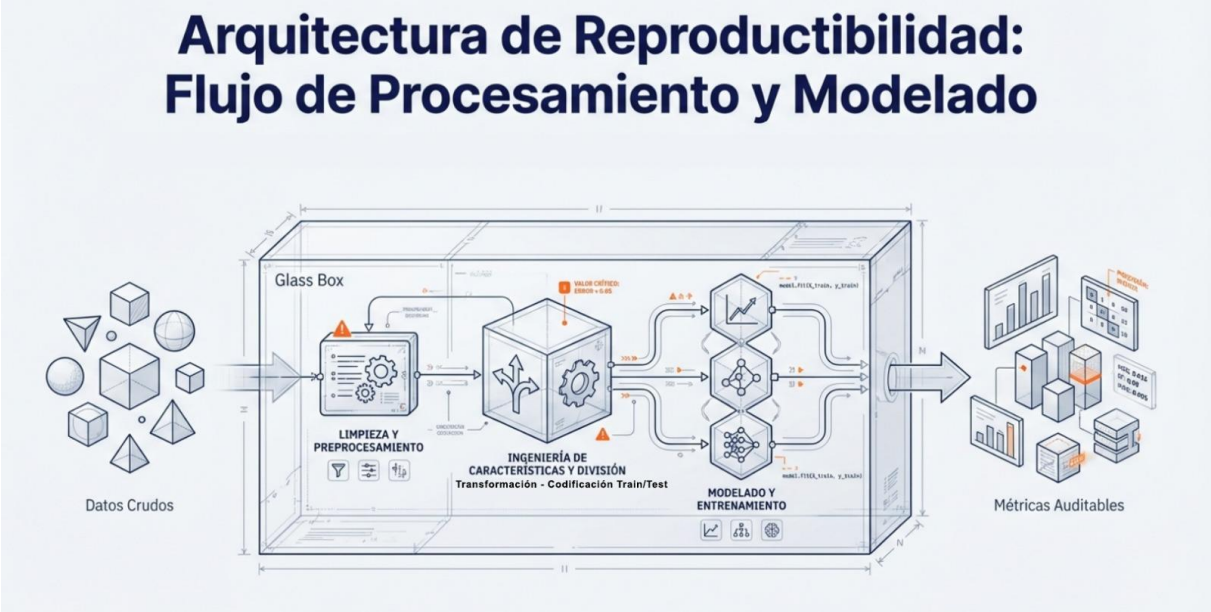


Figura 5. Flujo de Procesamiento y Modelado para la Reproductibilidad

- Entorno de Ejecución:
  Sistema Operativo Windows o Linux
  Python versión 3.13.2 o superior
  PIP versión 25 o superior
  Librerías de Python para Ciencia de Datos (Se muestran en el Anexo 12)
  Hardware: CPU Intel i5 o superior, RAM 16GB o superior, Disco duro con al menos 7GB de espacio libre.
- Datos, Fuente y Fecha de Extracción:
  Datos Abiertos Alcaldía de Cali
  Fecha de Extracción Septiembre de 2025 o mas reciente.
  Frecuencia Temporal de los datos: Diario
  Número de zonas: 52
  Estructura de archivos:
  csv-zonas-wifi-separados
  csv-train
  csv-test
  csv-windowed
  Perceptron_Graficas_TrainTest
  Perceptron_Metricas
  Perceptron_Modelos_Guardados
  RandomForest_Graficas_TrainTest

RandomForest _Metricas
RandomForest _Modelos_Guardados
RegresionLineal _Graficas_TrainTest
RegresionLineal _Metricas
RegresionLineal _Modelos_Guardados
SVR_Graficas_TrainTest
SVR _Metricas
SVR _Modelos_Guardados

- Dependencias de los paquetes de Python para la ejecución, se adjuntan en el anexo 12

7 CONSTRUCCION DEL PROTOTIPO

7.1 Objetivo del prototipo

El objetivo del prototipo es mostrar los resultados de las predicciones de cada zona, del proceso de modelado y análisis predictivo en una herramienta interactiva, accesible vía web, que permita visualizar, analizar y apoyar la toma de decisiones sobre el tráfico de datos de las zonas WiFi públicas de la ciudad de Cali.

Este prototipo busca:

- Facilitar la interpretación de los resultados de los modelos predictivos.
- Permitir la exploración comparativa entre zonas y modelos.
- Ofrecer una interfaz práctica para usuarios técnicos y no técnicos.
- Servir como prueba de concepto funcional de un sistema de predicción aplicado a infraestructura WiFi urbana.

7.2 Arquitectura general del prototipo

El prototipo fue desarrollado como una aplicación web interactiva utilizando la librería Streamlit de Python, desplegada en un entorno virtual.

La arquitectura lógica del prototipo se compone de los siguientes elementos:

1. **Capa de datos**
    - Archivos CSV que contienen:
        - Métricas de evaluación por modelo y zona.
        - Resultados de los mejores modelos por zona.
        - Resultados globales de desempeño de los modelos.
    - Modelos de cada zona.
        - Archivos del mejor modelo de cada zona en formato .joblib, se cargan en streamlit y dependiendo de la zona y variables de entrada arroja una predicción.

2. **Capa de procesamiento**
    - Lectura, limpieza y normalización de datos.
    - Cálculo de métricas agregadas (promedios, rankings, mejoras).
    - Filtrado dinámico por modelo y zona.
    - Lógica de simulación de predicción interactiva.
3. **Capa de visualización**
    - Gráficos interactivos (barras, líneas, dispersión, torta).
    - Tablas dinámicas.
    - Visualización de imágenes de predicción temporal.
4. **Capa de interacción**
    - Panel lateral de navegación.
    - Controles de selección (zonas, modelos, parámetros).
    - Navegación entre secciones del dashboard.

7.3 **Librerías y herramientas utilizadas**

El prototipo fue implementado principalmente en **Python**, apoyándose en las siguientes librerías:

- **Streamlit**: desarrollo de la interfaz web interactiva.
- **Pandas** y **NumPy**: manipulación y análisis de datos.
- **Plotly**: generación de visualizaciones interactivas.
- **Pillow (PIL)**: carga y despliegue de imágenes.
- **Pathlib y Glob**: gestión de archivos y directorios.

Estas herramientas permitieron construir un prototipo **ligero, reproducible y fácilmente desplegable**, sin necesidad de infraestructura compleja.

7.4 **Estructura funcional del dashboard**

El dashboard se organiza en **tres secciones principales**, accesibles desde el panel lateral izquierdo:

7.4.1 **Predicción interactiva**

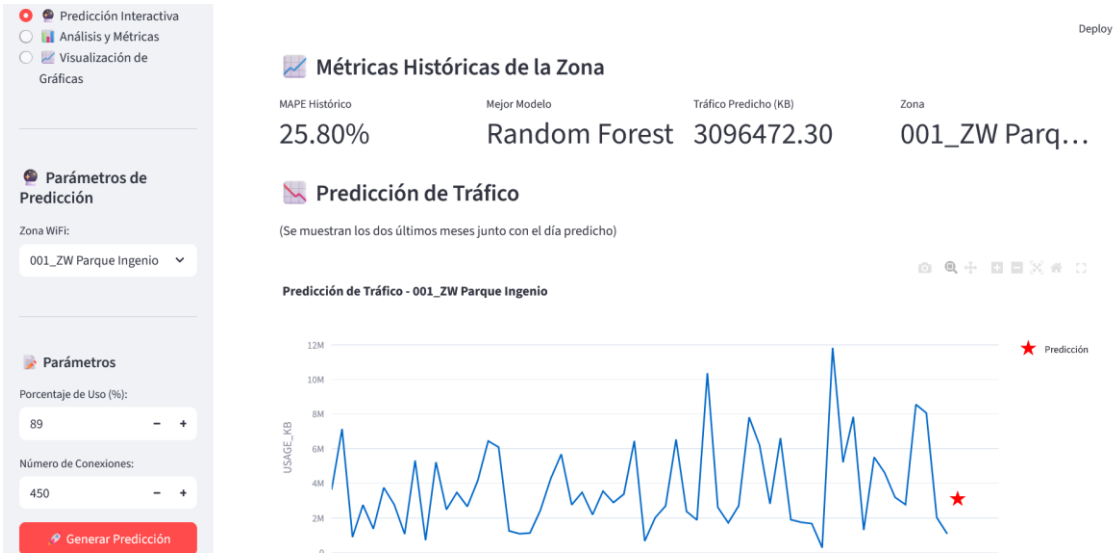


Figura 6. Interfaz de predicción interactiva, para generar la predicción de cada zona

En la figura 5 se muestra la interfaz de la vista de Predicción Interactiva.
Se tiene una barra lateral izquierda donde se selecciona la zona a la cual se le quiere hacer la predicción y los parámetros de entrada que son “Porcentaje de Uso” y “Número de Conexiones”. Teniendo esto el usuario da click en “Generar Predicción” y automáticamente se muestra una gráfica donde se marca el tráfico en el día predicho.

El modelo guardado en .joblib tiene almacenada la información de los 10 días anteriores (ventaneo de 10 días como se explicó en el capítulo 4.8), el programa consigue, mediante librerías de calendario de python, saber si el siguiente día es laboral, día de semana o festivo y le ingresa dicha data al modelo. Así este nos da la salido en términos de Bytes para el día siguiente.

En resumen, al generar la predicción la interfaz presenta:

- Tabla de resultados de la predicción.
- Métricas históricas de la zona (MAPE resultante del modelo entrenado).
- Gráfico de tendencia temporal con el punto de predicción resaltado.

Esta funcionalidad está orientada a **simular escenarios operativos**, más que a realizar inferencia en tiempo real.

7.4.2 **Análisis y métricas**

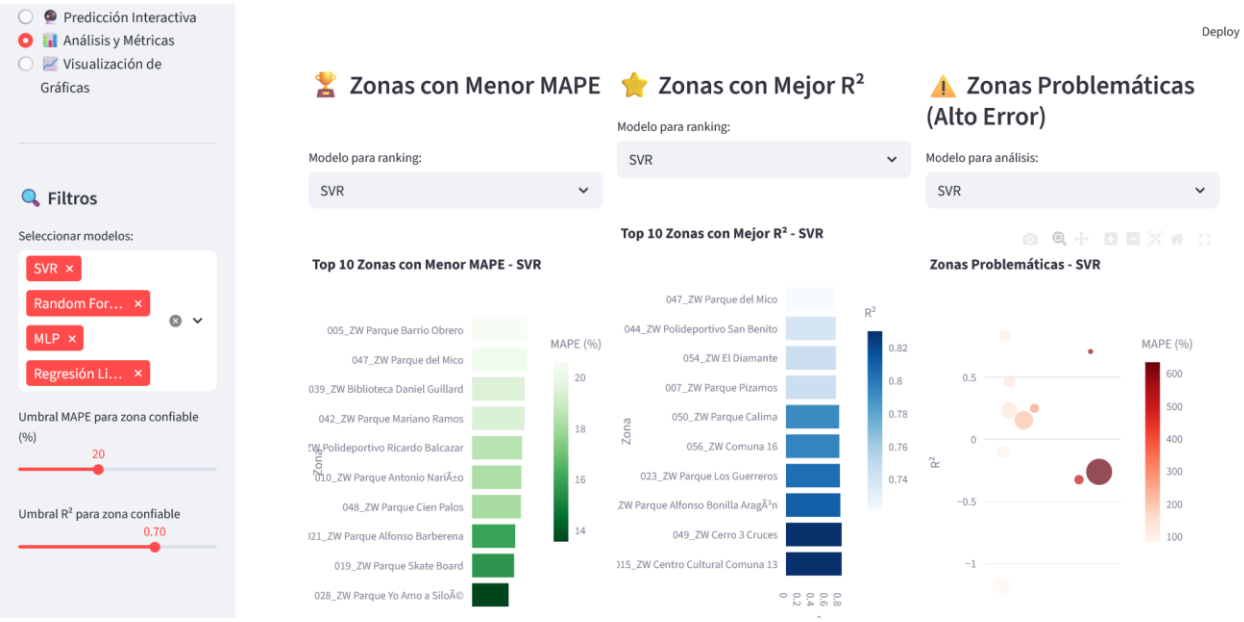


Figura 7. Análisis de métricas por zona

A continuación se detalla lo mostrado en la figura 6. Vista de Análisis y Métricas del prototipo.
En esta sección se consolidan los resultados del proceso de modelado, se tiene:

- Comparación de los modelos Base y Optimizado por modelo.
- Se visualiza mejora promedio del MAPE.
- Se tienen gráficos de barras con las zonas con menor error.

Además, se incorporan umbrales configurables para clasificar zonas confiables, apoyando decisiones operativas basadas en evidencia.

7.5 **Despliegue del prototipo**

El prototipo fue desplegado en la url https://wifianalytics.noesislabs.cloud/

El despliegue permite:

- Acceso con instalación local.
- Validación funcional de los resultados del proyecto.

El entorno de despliegue garantiza la estabilidad necesaria para la demostración académica del sistema.

7.6 **Aporte del prototipo al proyecto**

El prototipo constituye un puente entre el análisis académico y la aplicación práctica, permitiendo:

- Validar visualmente la calidad de los modelos predictivos.
- Traducir métricas técnicas en información comprensible.
- Apoyar decisiones de gestión de infraestructura WiFi.
- Demostrar la viabilidad de aplicar técnicas de ciencia de datos en entornos urbanos reales.

Más allá de su valor académico, el prototipo sienta las bases para una posible evolución hacia sistemas de monitoreo y planeación operativa a mayor escala.

7.7 **Puntos a tener en cuenta del prototipo**

Aunque funcional, el prototipo presenta algunas limitaciones:

- Las predicciones interactivas se basan en simulación, no en inferencia en tiempo real.
- No se integra directamente con sistemas productivos de red.
- El análisis se apoya en datos históricos preprocesados.

Estas limitaciones no afectan el objetivo del proyecto, dado que el enfoque principal es analítico y demostrativo.

8 CONCLUSIONES

El desarrollo del presente proyecto permitió abordar de manera integral el problema de la predicción del tráfico de datos en las zonas WiFi públicas de la ciudad de Santiago de Cali, combinando técnicas de ciencia de datos, modelado predictivo y visualización interactiva. A partir de los resultados obtenidos, se derivan las siguientes conclusiones:

1. Se alcanzaron satisfactoriamente los objetivos planteados en el proyecto.
   Fue posible preparar y estructurar los datos históricos de tráfico, construir y evaluar modelos predictivos basados en series de tiempo, y desarrollar un prototipo funcional que permite visualizar y analizar las predicciones de manera interactiva, cumpliendo así el objetivo general y los objetivos específicos definidos.
2. El proceso de preparación de datos resultó determinante para el desempeño de los modelos predictivos. Las etapas de depuración, transformación, escalado y ventaneo permitieron convertir datos crudos en conjuntos de datos estructurados y adecuados para el aprendizaje automático, evidenciando que una correcta ingeniería de datos es un factor crítico en proyectos de predicción sobre series temporales.
3. Los resultados del modelado predictivo muestran que los algoritmos capaces de capturar relaciones no lineales, como los modelos basados en vectores y redes neuronales, presentan un mejor desempeño frente a modelos más simples, especialmente en zonas con alta variabilidad en el consumo de datos.
4. El análisis comparativo entre zonas WiFi evidenció que no todas presentan el mismo nivel de predictibilidad. Las predicciones del Parque Santa Rosa, El Saladito, Centro Cultural Vista Hermosa y La Elvira obtuvieron errores MAPE por encima del 50% y La Castilla obtuvo una predicción con error MAPE del 188%.
   Aunque a cada zona se les evaluó varios modelos y se les clasificó con uno en particular con el que obtuvieron mejores métricas, las predicciones de estas zonas no tuvieron gran mejoría debido a su tráfico errático, con picos bruscos que van desde unos pocos bytes hasta los Gigabytes.
   También afecta el tráfico que se comporta casi como una línea recta evidenciando muy poco tráfico sin variaciones en periodos grandes de tiempo dentro del dataset.
5. Desde la perspectiva del negocio y la operación, la capacidad de anticipar el tráfico de datos constituye un insumo valioso para la gestión de la infraestructura WiFi pública. Las predicciones permiten apoyar decisiones relacionadas con la asignación de recursos, la prevención de congestión y la priorización de zonas que requieren intervenciones específicas de infraestructura y ancho de banda.
6. La construcción del prototipo demostró la viabilidad de integrar los resultados analíticos en una herramienta interactiva accesible vía web. El dashboard desarrollado facilita la interpretación de métricas técnicas y traduce los resultados del modelado en información comprensible y accionable para diferentes tipos de usuarios.
7. En conjunto, el proyecto evidencia que la aplicación de técnicas de ciencia de datos a problemas reales de infraestructura urbana es factible y aporta valor tanto desde un enfoque académico como práctico, sentando bases sólidas para futuras implementaciones más avanzadas en el ámbito de la gestión de redes WiFi públicas.

9 TRABAJOS FUTUROS

Si bien el proyecto permitió cumplir los objetivos planteados y obtener resultados relevantes en la predicción del tráfico de datos de las zonas WiFi públicas de la ciudad de Santiago de Cali, se identifican diversas líneas de trabajo que podrían explorarse para ampliar o profundizar los alcances del estudio.

Una posible línea de extensión consiste en la incorporación de nuevas fuentes de información, tales como variables climáticas, eventos especiales, calendarios académicos o actividades culturales, que podrían contribuir a explicar variaciones abruptas en el consumo de datos y mejorar la capacidad predictiva de los modelos, especialmente en zonas con alta variabilidad.

Otra línea de trabajo corresponde a la integración del sistema predictivo con datos en tiempo casi real, lo cual permitiría evaluar el comportamiento de los modelos bajo escenarios dinámicos y analizar su aplicabilidad en contextos operativos más exigentes. Este enfoque requeriría considerar aspectos adicionales relacionados con la infraestructura tecnológica y la disponibilidad de datos.

Desde el punto de vista del prototipo desarrollado, podrían explorarse mejoras orientadas a la experiencia de usuario, tales como la incorporación de visualizaciones adicionales con mapas interactivos, filtros más avanzados o mecanismos de comparación automática entre zonas y periodos de tiempo, con el objetivo de facilitar aún más el análisis por parte de los usuarios finales.

Finalmente, el enfoque metodológico empleado en este proyecto podría ser replicado o adaptado a otros contextos urbanos o tipos de infraestructura, sistemas de monitoreo energético o servicios digitales municipales, permitiendo evaluar la generalización de la metodología y su utilidad en otros escenarios de gestión pública.

10. REFERENCIAS BIBLIOGRÁFICAS

[1] International Organization for Standardization, “ISO 9241-210:2019 — Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems,” ISO, Geneva, Switzerland, 2019. [Online]. Available: https://www.iso.org/standard/77520.html

[2] V. K. Gurbani, E. Burger, T. Anjali, H. Abdelnur y O. Festor, *The Common Log Format (CLF) for the Session Initiation Protocol (SIP): Framework and Information Model*, RFC 6872, Standards Track, Internet Engineering Task Force, Feb. 2013, doi: 10.17487/RFC6872. [Online]. Available: https://www.rfc-editor.org/info/rfc6872

[3] D. Stanley, M. Montemurro y P. R. Calhoun, *Control and Provisioning of Wireless Access Points (CAPWAP) Protocol Binding for IEEE 802.11*, RFC 5416, Proposed Standard, Internet Engineering Task Force, Mar. 2009, doi: 10.17487/RFC5416. [Online]. Available: https://www.rfc-editor.org/info/rfc5416

[4] V. K. Gurbani, E. Burger, T. Anjali, H. Abdelnur y O. Festor, *The Common Log Format (CLF) for the Session Initiation Protocol (SIP): Framework and Information Model*, RFC 6872, Standards Track, Internet Engineering Task Force, Feb. 2013, doi: 10.17487/RFC6872. [Online]. Available: https://www.rfc-editor.org/info/rfc6872

[5] Cisco Wireless. (2022, Noviembre 11). Cisco Embedded Wireless Controller on Catalyst Access Points (EWC) White Paper. [Online]. Available: https://www.cisco.com/c/en/us/products/collateral/wireless/embedded-wireless-controller-catalyst-access-points/white-paper-c11-743398.html

[6] B. Y. Fraser, *Site Security Handbook*, RFC 2196, Informational, Internet Engineering Task Force, Sep. 1997. doi: 10.17487/RFC2196. [Online]. Available: https://www.rfc-editor.org/info/rfc2196

[7] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in Proc. 23rd Int. Conf. Mach. Learn., Pittsburgh, PA, USA, 2006, pp. 161-168, doi: 10.1145/1143844.1143865. [Online]. Available: https://www.cs.cornell.edu/~caruana/ctp/ct.papers/caruana.icml06.pdf

[8] N. R. Draper and H. Smith, "Applied Regression Analysis, 3rd ed.," New York, NY, USA: Wiley, 1998. doi: 10.1002/0471722554.
[Online]. Available:
https://blog.minitab.com/es/blog/analisis-de-regresion-como-puedo-interpretar-el-r-cuadrado-y-evaluar-la-bondad-de-ajuste

[9] A. De Myttenaere, B. Golden, B. Le Grand y F. Rossi, "Mean Absolute Percentage Error for regression models," Neurocomputing, vol. 192, pp. 38-48, jun. 2016, doi: 10.1016/j.neucom.2015.12.114.
[Online]. Available:
https://www.sciencedirect.com/science/article/abs/pii/S0925231216003325

[10] S. Hodson, "Mean Squared Error, Deconstructed," Journal of Advances in Modeling Earth Systems, vol. 14, no. 1, ene. 2022, Art. no. e2021MS002681, doi: 10.1029/2021MS002681.
[Online]. Available:
https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021MS002681

[11] A. Aplicación del aprendizaje supervisado para reconocer patrones de datos de ciberamenazas y aprendizaje no supervisado para analizar y agrupar conjuntos de datos, ResearchGate, 2024.
[Online]. Available:
https://www.researchgate.net/publication/380293092_APLICACION_DEL_APRENDIZAJE_SUPERVISADO_PARA_RECONOCER_PATRONES_DE_DATOS_DE_CIBERAMENAZAS_Y_APRENDIZAJE_NO_SUPERVISADO_PARA_ANALIZAR_Y_AGRUPAR_CONJUNTOS_DE_DATOS

[12] Universidad de los Andes, “Regresión lineal,” Blog Uniandes.
[Online]. Available:
https://programas.uniandes.edu.co/blog/regresion-lineal

[13] J. E. Andrade Andrade, “Aplicación de los algoritmos K-Means y Random Forest para la segmentación de potenciales estudiantes del programa de maestría en estadística con mención en ciencia de datos e inteligencia artificial,” Univ. Nac. de Chimborazo, Riobamba, Ecuador, 2025.
[Online]. Available:
http://dspace.unach.edu.ec/bitstream/51000/14997/1/Andrade%20Andrade%2C%20Jes%C3%BAs%20E.%20%282025%29%20%E2%80%9CAPLICACI%C3%93N%20DE%20LOS%20ALGORITMOS%20K-MEANS%20Y%20RANDOM%20FOREST%20PARA%20LA%20SEGMENTACI%C3%93N%20DE%20POTENCIALES%20ESTUDIANTES%20DEL%20PROGRAMA%20DE%20MAESTR%C3%8DA%20EN%20ESTAD%C3%8DSTICA%20CON%20MENCI%C3%93N%20EN%20CIENCIA%20DE%20DATOS%20E%20INTELI.pdf

[14] Universidad de Valladolid, “Redes neuronales artificiales,” Trabajo Fin de Grado, UVaDOC, 2023.
[Online]. Available:
https://uvadoc.uva.es/bitstream/handle/10324/79665/TFG-G7735.pdf

[15] Universidad Nacional de La Plata, “Máquinas de soporte vectorial para regresión,” SEDICI, UNLP.
[Online]. Available:
https://sedici.unlp.edu.ar/bitstream/handle/10915/90892/Documento_completo.pdf-PDFA.pdf

[16] Modeling WiFi Traffic for White Space Prediction in Wireless Sensor Networks
[Online]. Available:
https://arxiv.org/pdf/1709.08950

[17] Aneja, N., Aneja, S., & Bhargava, B. (2023). AI-Enabled Learning Architecture Using Network Traffic Traces over IoT Network: A Comprehensive Review. Wireless Communications and Mobile Computing. Hindawi Limited.
[Online]. Available:
https://doi.org/10.1155/2023/8658278

[18] Information theory based clustering of cellular network usage data for the identification of representative urban areas.
[Online]. Available:
https://www.sciencedirect.com/science/article/pii/S2352864823001207?via%3Dihub

[19] Adoption of public WiFi using UTAUT2: An exploration in an emerging economy
[Online]. Available:
https://www.sciencedirect.com/science/article/pii/S1877050918309141?via%3Dihub

[20] Economics of public WiFi:
[Online]. Available:
https://telsoc.org/journal/ajtde-v2-n1/a20

[21] M. Krishna, M. S. Shashi, and D. J. D. Sarma, “Modeling WiFi Traffic for White Space Prediction in Wireless Sensor Networks,” arXiv preprint arXiv:1709.08950, Sep. 2017.
[Online]. Available:
https://arxiv.org/pdf/1709.08950

[22] N. Aneja, S. Aneja, and B. Bhargava, “AI-Enabled Learning Architecture Using Network Traffic Traces over IoT Network: A Comprehensive Review,” Wireless Communications and Mobile Computing, vol. 2023, Article ID 8658278, 2023.
[Online]. Available:
https://doi.org/10.1155/2023/8658278

[23] R. Morstyn, S. Pfenninger, and M. D. McCulloch, “Information theory based clustering of cellular network usage data for the identification of representative urban areas,” Data in Brief, vol. 51, 2023.
[Online]. Available:
https://www.sciencedirect.com/science/article/pii/S2352864823001207

[24] R. K. Singh and V. Sinha, “Adoption of public WiFi using UTAUT2: An exploration in an emerging economy,” Procedia Computer Science, vol. 132, pp. 250–257, 2018.
[Online]. Available:
https://www.sciencedirect.com/science/article/pii/S1877050918309141

[25] M. Biggs and J. Kelly, “Economics of public WiFi,” Australian Journal of Telecommunications and the Digital Economy, vol. 2, no. 1, pp. 20–30, Mar. 2014.
[Online]. Available:
https://telsoc.org/journal/ajtde-v2-n1/a20

[26] E. O. (Autor no identificado), “Diseño e implementación de una red inalámbrica de área metropolitana para distribución de Internet en medios suburbanos, utilizando el protocolo IEEE 802.11b”, Universidad de San Carlos de Guatemala, Guatemala, Tesis de grado. [Online]. Available:
http://biblioteca.usac.edu.gt/tesis/08/08_0178_EO.pdf
. Accessed: Feb. 4, 2026.

[27] Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC), “Zonas Digitales en Espacios Públicos – Anexo técnico del proyecto tipo”, Gobierno de Colombia, 2024. [Online]. Available:
https://www.mintic.gov.co/portal/715/articles-326715_Anexo_tecnico_proyecto_tipo_de_Zonas_Digitales_U20240829.pdf
. Accessed: Feb. 4, 2026.

ANEXOS

FECHA_CONEXION	DIA_SEMANA	LABORAL	FIN_DE_SEMANA	FESTIVO	PORCENTAJE_USO	NUMERO_CONEXIONES	USAGE_KB
2024-10-10	3	1	0	0	3.41	120	21396
2024-10-11	4	1	0	0	0.2	27	1378
2024-10-12	5	0	1	0	82.23	29	854109
2024-10-13	6	0	1	0	257.68	41	2166288
2024-10-14	0	0	0	1	262.27	31	2090416
2024-10-15	1	1	0	0	93.96	37	509146
2024-10-16	2	1	0	0	176.59	50	932943
2024-10-17	3	1	0	0	894.47	110	5154126
2024-10-18	4	1	0	0	3.1	35	14446
2024-10-19	5	0	1	0	251.56	55	2222443
2024-10-20	6	0	1	0	763.8	43	5754904
2024-10-21	0	1	0	0	859.74	44	5726575
2024-10-22	1	1	0	0	215.32	42	1580316
2024-10-23	2	1	0	0	2.35	54	16193
2024-10-24	3	1	0	0	04.06	103	26808
2024-10-25	4	1	0	0	683.31	17	3058883
2024-10-26	5	0	1	0	206.2	24	1592314
2024-10-27	6	0	1	0	418.08	67	4161923
2024-10-28	0	1	0	0	0.28	14	1715
2024-10-29	1	1	0	0	1.12	30	7104
2024-10-30	2	1	0	0	1	8	4897
2024-10-31	3	1	0	0	28.57	20	134019
2024-11-01	4	1	0	0	111.95	12	619084
2024-11-02	5	0	1	0	261.32	42	1623766
2024-11-03	6	0	1	0	736.42	61	4332021
2024-11-04	0	0	0	1	614.92	32	4079280
2024-11-05	1	1	0	0	110.89	23	809726
2024-11-06	2	1	0	0	7.42	21	46441
2024-11-07	3	1	0	0	178.28	30	1585495
2024-11-08	4	1	0	0	103.09	18	1274896
2024-11-09	5	0	1	0	168.55	34	2806486
2024-11-10	6	0	1	0	150	95	2905734
2024-11-11	0	0	0	1	137.43	67	2413513
2024-11-12	1	1	0	0	123.01	33	2351318
2024-11-13	2	1	0	0	360.82	33	5845641

2024-11-14	**3**	**1**	**0**	**0**	**107.35**	**28**	**2060840**
2024-11-15	**4**	**1**	**0**	**0**	**151.55**	**28**	**2760723**
2024-11-16	**5**	**0**	**1**	**0**	**97.35**	**39**	**1792471**
2024-11-17	**6**	**0**	**1**	**0**	**98.61**	**84**	**1694110**
2024-11-18	**0**	**1**	**0**	**0**	**20.3**	**35**	**326245**
2024-11-19	**1**	**1**	**0**	**0**	**18.43**	**19**	**294381**
2024-11-20	**2**	**1**	**0**	**0**	**79.26**	**35**	**1512590**
2024-11-21	**3**	**1**	**0**	**0**	**19.2**	**32**	**346674**
2024-11-22	**4**	**1**	**0**	**0**	**98.13**	**35**	**1981044**
2024-11-23	**5**	**0**	**1**	**0**	**110.58**	**35**	**2119095**
2024-11-24	**6**	**0**	**1**	**0**	**108.64**	**96**	**804336**
2024-11-25	**0**	**1**	**0**	**0**	**23.27**	**27**	**95693**
2024-11-26	**1**	**1**	**0**	**0**	**5.46**	**41**	**42990**
2024-11-27	**2**	**1**	**0**	**0**	**51.36**	**29**	**997413**

Anexo 1. Composición del conjunto de datos de la Zona WiFi Parque del Ingenio (primeros 50 registros)

Zona WiFi	Total de registros	Train (80 %)	Test (20 %)
ZW Parque Ingenio	356	284	72
ZW Canchas Panamericanas	327	261	66
ZW Parque del Perro	341	272	69
ZW Parque San Nicolás	334	267	67
ZW Parque Barrio Obrero	382	305	77
ZW Parque Pízamos	124	99	25
ZW Parque Alfonso López	341	272	69
ZW Parque Antonio Nariño	331	264	67
ZW Parque Santa Rosa Poblado	380	304	76
ZW Polideportivo Los Naranjos	402	321	81
ZW Parque Llano Verde	366	292	74
ZW Centro Cultural Comuna 13	340	272	68
ZW Conjunto Habitacional Ramali	341	272	69
ZW Parque Skate Board	335	268	67
ZW Parque Alfonso Barberena	341	272	69
ZW Polideportivo Los Farallones	341	272	69
ZW Parque Los Guerreros	352	281	71

ZW Museo La Tertulia	417	333	84
ZW Parque San Marino	376	300	76
ZW Parque La Flora	341	272	69
ZW Parque Alfonso Bonilla Aragón	342	273	69
ZW Parque Yo Amo a Siloé	341	272	69
ZW Centro Cultural Vista Hermosa	511	408	103
ZW Parque Mutis	341	272	69
ZW Cancha Los Azules	305	244	61
ZW Parque La Orqueta	385	308	77
ZW Parque Villa Colombia	458	366	92
ZW Parque Colseguros	341	272	69
ZW Parque India Elena	326	260	66
ZW Parque Tequendama	378	302	76
ZW Parque Sector Amarillo Skate Park	329	263	66
ZW Biblioteca Daniel Guillard	354	283	71
ZW Polideportivo Torres Comfandi	298	238	60
ZW Parque Junín	205	164	41
ZW Parque Mariano Ramos	365	292	73
ZW Polideportivo Ricardo Balcázar	264	211	53
ZW Polideportivo San Benito	341	272	69
ZW Parque Santa Anita	488	390	98
ZW Sebastián de Belalcázar	539	431	108
ZW Parque del Mico	324	259	65
ZW Parque Cien Palos	197	157	40
ZW Cerro 3 Cruces	315	252	63
ZW Parque Calima	335	268	67
ZW Parque La Merced	462	369	93
ZW Puente de Colores	394	315	79
ZW Polideportivo Laureano Gómez	364	291	73
ZW El Diamante	385	308	77
ZW Polideportivo Petecuy	273	218	55
ZW Comuna 16	282	225	57
La Castilla	285	228	57
La Elvira	241	192	49
El Saladito	264	211	53

Anexo 2. Total, de registros por dataset y partición Train/Test (80/20)

Zona WiFi	MAE	RMSE	MAPE (%)	R^2
001_ZW Parque Ingenio	1.041.531,47	1.434.059,00	25,80	0,667
002_ZW Canchas Panamericanas	600.874,54	839.877,00	21,97	0,675
003_ZW Parque del Perro	692.025,95	893.662,00	24,48	0,498
004_ZW Parque San Nicolas	2.166.368,31	2.900.307,00	20,05	0,546
005_ZW Parque Barrio Obrero	3.359.334,15	4.438.344,00	17,91	0,663
006_ZW Parque Pizamos	765.846,47	1.325.091,00	20,78	0,723
007_ZW Parque Alfonso Lopez	2.935.948,85	3.519.269,00	31,52	0,279
008_ZW Parque Antonio Nariño	753.761,88	914.938,00	16,85	0,763
009_ZW Parque Santa Rosa Poblado	1.902.469,62	2.554.255,00	54,07	0,105
010_ZW Polideportivo Los Naranjos	1.628.934,74	2.483.142,00	31,12	0,400
011_ZW Parque Llano Verde	811.251,94	1.020.262,00	23,34	0,354
012_ZW Centro Cultural Comuna 13	1.356.253,35	1.880.866,00	22,44	0,888
013_ZW Conjunto Habitacional Ramali	4.840.442,01	7.683.316,00	25,21	0,400
014_ZW Parque Skate Board	2.158.449,99	2.572.112,00	25,00	0,181
015_ZW Parque Alfonso Barberena	1.438.238,46	1.786.836,00	19,82	0,649
016_ZW Polideportivo Los Farallones	1.049.463,93	1.663.346,00	26,60	0,653
017_ZW Parque Los Guerreros	944.571,73	1.230.534,00	23,27	0,575
018_ZW Museo La Tertulia	256.054,24	473.914,00	27,16	0,764
019_ZW Parque San Marino	1.328.944,05	1.546.633,00	28,30	0,686
020_ZW Parque La Flora	1.464.655,09	2.065.924,00	24,30	0,529
021_ZW Parque Alfonso Bonilla Aragón	753.981,87	976.329,00	22,84	0,821
022_ZW Parque Yo Amo a Siloé	3.816.150,47	4.523.177,00	18,22	-0,080
023_ZW Centro Cultural Vista Hermosa	1.192.950,95	1.848.013,00	14977,03	0,540
024_ZW Parque Mutis	1.348.414,45	1.770.931,00	21,45	0,663
025_ZW Cancha Los Azules	3.278.059,95	3.961.082,00	23,19	0,478
026_ZW Parque La Orqueta	2.849.583,84	3.698.491,00	15,64	0,543
027_ZW Parque Villa Colombia	1.113.487,50	1.506.139,00	27,27	0,443
028_ZW Parque Colseguros	1.238.262,80	1.564.623,00	17,47	0,610
029_ZW Parque India Elena	1.616.556,43	1.862.000,00	20,55	0,702
030_ZW Parque Tequendama	1.609.187,87	2.026.114,00	21,78	0,601
031_ZW Parque Sector Amarillo Skate Park	1.079.713,16	1.450.176,00	19,71	0,672
032_ZW Biblioteca Daniel Guillard	1.678.754,54	2.062.479,00	18,73	0,670
033_ZW Polideportivo Torres Comfandi	1.907.529,47	3.242.404,00	26,59	0,259
034_ZW Parque Junin	832.090,80	1.201.016,00	11,27	0,835
035_ZW Parque Mariano Ramos	1.337.949,09	1.545.469,00	22,79	0,426
036_ZW Polideportivo Ricardo Balcazar	1.868.567,03	2.203.301,00	19,66	0,463
037_ZW Polideportivo San Benito	1.239.763,42	1.697.408,00	23,25	0,767
038_ZW Parque Santa Anita	919.488,48	1.161.592,00	25,37	0,399

039_ZW Sebastian Belalcazar	676.233,79	987.831,00	24,52	0,480
040_ZW Parque del Mico	1.098.313,19	1.362.113,00	22,94	0,762
041_ZW Parque Cien Palos	1.422.678,61	1.729.461,00	17,70	0,428
042_ZW Cerro 3 Cruces	830.494,20	1.138.297,00	31,54	0,875
043_ZW Parque Calima	1.049.547,98	1.511.996,00	25,39	0,795
044_ZW Parque La Merced	642.362,99	1.061.839,00	31,77	0,449
045_ZW Puente de Colores	1.082.413,15	1.664.723,00	28,68	0,398
046_ZW Polideportivo Laureano Gomez	1.047.908,20	1.708.221,00	16,17	0,778
047_ZW El Diamante	639.656,75	1.095.225,00	31,78	0,541
048_ZW Polideportivo Petecuy	995.049,89	1.402.487,00	20,12	0,644
049_ZW Comuna 16	133.503,20	277.108,00	24,65	0,799
050_La Castilla	597.506,43	875.485,00	3979,81	0,555
051_La Elvira	1.659.301,03	2.977.944,00	232,75	0,534
052_El Saladito	1.556.688,83	1.948.527,00	85,48	0,224

Anexo 3. Métricas de desempeño – Random Forest (modelo base)

Zona WiFi	MAE	RMSE	MAPE (%)	R^2
001_ZW Parque Ingenio	1.333.157,15	1.808.586,11	38,95	0,495
002_ZW Canchas Panamericanas	723.666,69	971.803,44	27,98	0,604
003_ZW Parque del Perro	846.340,99	1.117.579,78	26,15	0,317
004_ZW Parque San Nicolas	2.588.836,28	3.298.815,08	20,01	0,618
005_ZW Parque Barrio Obrero	4.679.565,87	5.661.379,22	34,06	0,524
006_ZW Parque Pizamos	818.176,86	1.066.200,58	20,56	0,856
007_ZW Parque Alfonso Lopez	3.209.979,79	3.778.141,65	30,35	0,346
008_ZW Parque Antonio Nariño	722.121,47	924.698,10	14,92	0,777
009_ZW Parque Santa Rosa Poblado	17.602.258,36	23.131.209,33	639,18	-1,440
010_ZW Polideportivo Los Naranjos	2.163.544,77	3.088.639,42	36,97	0,361
011_ZW Parque Llano Verde	996.845,75	1.305.779,24	25,46	0,197
012_ZW Centro Cultural Comuna 13	1.860.201,91	2.533.190,51	53,94	0,848
013_ZW Conjunto Habitacional Ramali	4.420.515,06	6.460.768,92	26,70	0,707
014_ZW Parque Skate Board	1.660.293,55	2.054.584,08	22,74	0,286
015_ZW Parque Alfonso Barberena	1.289.969,30	1.703.975,28	15,35	0,704
016_ZW Polideportivo Los Farallones	1.566.306,25	2.113.731,82	450,67	0,540
017_ZW Parque Los Guerreros	1.644.453,67	2.233.427,77	40,08	-0,284

018_ZW Museo La Tertulia	489.776,94	662.275,54	619,60	0,563
019_ZW Parque San Marino	4.218.982,71	5.412.492,85	337,07	-2,746
020_ZW Parque La Flora	1.289.999,10	1.704.445,17	22,53	0,696
021_ZW Parque Alfonso Bonilla Aragón	681.275,52	967.332,28	23,17	0,841
022_ZW Parque Yo Amo a Siloé	3.219.242,73	4.185.895,53	13,08	0,298
023_ZW Centro Cultural Vista Hermosa	2.430.659,20	3.473.875,24	130,01	-0,208
024_ZW Parque Mutis	859.524,19	1.177.045,61	14,93	0,828
025_ZW Cancha Los Azules	3.340.686,18	4.038.285,41	21,57	0,506
026_ZW Parque La Orqueta	5.213.714,42	6.670.994,18	32,10	-0,036
027_ZW Parque Villa Colombia	2.075.997,14	2.718.344,80	37,41	0,187
028_ZW Parque Colseguros	1.445.039,01	1.848.374,50	19,03	0,409
029_ZW Parque India Elena	937.645,58	1.176.643,57	11,85	0,877
030_ZW Parque Tequendama	2.183.877,96	3.018.036,43	27,03	0,085
031_ZW Parque Sector Amarillo Skate Park	1.114.815,16	1.335.836,85	30,49	0,624
032_ZW Biblioteca Daniel Guillard	5.555.077,55	8.370.587,59	88,48	-3,825
033_ZW Polideportivo Torres Comfandi	2.368.615,16	3.260.962,59	64,54	0,351
034_ZW Parque Junin	1.175.186,39	1.534.382,58	14,91	0,716
035_ZW Parque Mariano Ramos	2.340.648,96	3.004.073,58	34,74	-1,033
036_ZW Polideportivo Ricardo Balcazar	1.573.729,00	1.984.783,99	14,14	0,715
037_ZW Polideportivo San Benito	923.991,95	1.264.378,50	20,64	0,845
038_ZW Parque Santa Anita	1.550.430,65	2.042.777,16	37,09	-0,709
039_ZW Sebastian Belalcazar	2.007.376,70	2.561.566,58	54,00	-0,657
040_ZW Parque del Mico	1.115.050,75	1.436.339,33	25,14	0,732
041_ZW Parque Cien Palos	993.836,48	1.254.673,24	11,00	0,763
042_ZW Cerro 3 Cruces	1.016.775,81	1.440.574,53	68,24	0,849
043_ZW Parque Calima	1.250.205,25	1.785.739,24	30,30	0,720
044_ZW Parque La Merced	1.017.747,97	1.622.540,20	62,43	-0,117
045_ZW Puente de Colores	1.978.328,20	2.591.034,95	67,84	-0,311
046_ZW Polideportivo Laureano Gomez	2.001.516,53	2.803.942,43	28,79	0,762
047_ZW El Diamante	883.015,28	1.173.986,19	176,32	0,625
048_ZW Polideportivo Petecuy	1.042.203,73	1.462.709,23	20,67	0,423
049_ZW Comuna 16	213.597,39	315.890,24	266,69	0,783
050_La Castilla	6.143.862,48	7.149.344,85	3060,20	-25,578
051_La Elvira	2.343.721,16	3.336.201,41	275,96	0,317
052_El Saladito	2.558.926,65	3.324.284,29	69,26	-0,939

Anexo 4. Métricas de desempeño – Regresión Lineal (modelo base)

Zona WiFi	MAE	RMSE	MAPE (%)	R^2
001_ZW Parque Ingenio	1.078.031,66	1.670.629,85	33,45	0,426
002_ZW Canchas Panamericanas	792.958,84	1.150.524,25	40,13	0,318
003_ZW Parque del Perro	1.053.117,13	1.555.317,83	30,65	0,360
004_ZW Parque San Nicolas	4.164.597,84	5.807.695,62	28,01	-0,185
005_ZW Parque Barrio Obrero	7.178.427,65	8.947.742,39	28,73	-0,041
006_ZW Parque Pizamos	1.572.501,81	2.157.869,43	45,84	0,246
007_ZW Parque Alfonso Lopez	3.047.779,64	3.955.758,16	29,93	0,410
008_ZW Parque Antonio Nariño	1.312.741,94	1.756.855,36	21,46	0,461
009_ZW Parque Santa Rosa Poblado	14.743.922,36	21.582.367,08	258,94	-0,863
010_ZW Polideportivo Los Naranjos	2.744.256,90	4.046.963,61	104,77	-0,157
011_ZW Parque Llano Verde	1.470.739,96	1.744.797,22	37,10	-0,546
012_ZW Centro Cultural Comuna 13	2.818.422,68	3.955.006,30	143,41	0,618
013_ZW Conjunto Habitacional Ramali	6.285.107,37	9.886.917,50	47,68	0,211
014_ZW Parque Skate Board	1.747.547,01	2.400.351,64	20,70	0,382
015_ZW Parque Alfonso Barberena	2.169.617,05	2.806.239,32	25,86	0,105
016_ZW Polideportivo Los Farallones	1.404.156,47	2.233.080,00	286,80	0,284
017_ZW Parque Los Guerreros	2.342.812,55	3.133.103,90	48,61	-0,785
018_ZW Museo La Tertulia	562.367,28	724.367,22	722,85	0,524
019_ZW Parque San Marino	5.582.349,16	8.345.127,04	579,89	-0,225
020_ZW Parque La Flora	2.102.348,81	2.751.191,73	32,29	0,216
021_ZW Parque Alfonso Bonilla Aragón	1.252.067,69	1.789.628,35	37,48	0,588
022_ZW Parque Yo Amo a Siloé	6.622.281,67	7.965.328,34	24,40	-0,731
023_ZW Centro Cultural Vista Hermosa	2.991.384,92	4.095.655,24	70,70	-0,763
024_ZW Parque Mutis	1.350.573,15	2.104.637,93	21,63	0,645
025_ZW Cancha Los Azules	4.170.357,30	5.688.012,17	26,64	0,184
026_ZW Parque La Orqueta	11.740.065,29	13.797.615,70	631,75	-2,240
027_ZW Parque Villa Colombia	2.764.053,03	3.759.242,23	58,80	-0,388
028_ZW Parque Colseguros	1.842.968,88	2.377.621,84	23,85	0,022
029_ZW Parque India Elena	2.626.546,31	3.370.473,97	26,95	0,250
030_ZW Parque Tequendama	2.559.346,70	3.215.315,96	28,32	0,056
031_ZW Parque Sector Amarillo Skate Park	1.444.884,31	1.743.530,50	71,11	0,284
032_ZW Biblioteca Daniel Guillard	3.544.196,27	4.668.956,05	34,34	-0,170
033_ZW Polideportivo Torres Comfandi	3.811.598,57	5.350.874,57	52,87	-0,551
034_ZW Parque Junin	3.056.053,83	3.775.732,31	28,34	-0,161
035_ZW Parque Mariano Ramos	1.709.257,94	2.241.406,92	25,80	-0,453
036_ZW Polideportivo Ricardo Balcazar	2.513.865,16	3.406.837,80	22,23	0,120

037_ZW Polideportivo San Benito	3.811.805,79	6.318.237,27	48,11	0,277
038_ZW Parque Santa Anita	2.041.752,81	2.561.684,81	45,39	-1,409
039_ZW Sebastian Belalcazar	3.392.592,19	4.304.977,74	125,38	-1,535
040_ZW Parque del Mico	1.392.045,42	1.824.569,63	30,92	0,484
041_ZW Parque Cien Palos	2.894.202,28	3.838.165,23	24,38	-0,014
042_ZW Cerro 3 Cruces	1.505.894,50	2.159.658,99	61,68	0,738
043_ZW Parque Calima	1.339.537,07	2.128.008,29	28,27	0,556
044_ZW Parque La Merced	1.030.450,77	1.498.509,95	78,00	-0,047
045_ZW Puente de Colores	1.987.869,07	2.790.387,94	52,73	-0,175
046_ZW Polideportivo Laureano Gomez	4.994.240,37	7.282.773,75	47,49	-0,102
047_ZW El Diamante	951.110,88	1.385.624,30	55,54	0,342
048_ZW Polideportivo Petecuy	954.568,09	1.387.683,92	23,48	0,233
049_ZW Comuna 16	298.684,85	492.738,26	73,17	0,448
050_La Castilla	825.543,38	1.425.363,56	188,15	0,172
051_La Elvira	2.501.040,64	3.663.751,81	132,12	0,369
052_El Saladito	2.538.781,69	3.459.820,85	60,24	-0,403

Anexo 5. Métricas de desempeño – SVR (modelo base)

Zona WiFi	**MAE**	**RMSE**	**MAPE (%)**	**R^2**
001_ZW Parque Ingenio	6.231.274,72	6.706.314,45	269,73	-8,252
002_ZW Canchas Panamericanas	551.104,81	750.546,97	31,57	0,710
003_ZW Parque del Perro	1.057.232,48	1.428.209,74	33,81	0,460
004_ZW Parque San Nicolas	2.863.132,89	3.918.591,31	20,83	0,460
005_ZW Parque Barrio Obrero	4.300.606,96	5.047.431,39	21,79	0,669
006_ZW Parque Pizamos	942.269,20	1.405.260,34	20,39	0,680
007_ZW Parque Alfonso Lopez	2.833.231,21	3.559.343,42	32,05	0,523
008_ZW Parque Antonio Nariño	1.017.913,75	1.246.428,05	21,42	0,729
009_ZW Parque Santa Rosa Poblado	11.209.602,60	17.074.890,28	1609,65	-0,166
010_ZW Polideportivo Los Naranjos	1.723.843,02	2.613.352,62	82,48	0,518
011_ZW Parque Llano Verde	1.147.587,77	1.394.759,70	30,70	0,012
012_ZW Centro Cultural Comuna 13	2.149.644,77	3.033.566,43	86,88	0,775
013_ZW Conjunto Habitacional Ramali	6.681.711,37	9.306.385,19	48,98	0,301
014_ZW Parque Skate Board	1.719.320,18	2.256.043,12	19,03	0,454

015_ZW Parque Alfonso Barberena	1.421.567,16	1.940.057,24	18,28	0,572
016_ZW Polideportivo Los Farallones	1.205.997,89	1.749.945,49	248,33	0,560
017_ZW Parque Los Guerreros	2.170.307,29	2.786.892,82	64,60	-0,412
018_ZW Museo La Tertulia	770.074,74	900.806,24	1078,52	0,264
019_ZW Parque San Marino	7.050.187,21	8.757.283,92	300,30	-0,349
020_ZW Parque La Flora	1.699.178,97	2.293.723,09	28,98	0,455
021_ZW Parque Alfonso Bonilla Aragón	1.147.962,25	1.569.654,72	44,97	0,683
022_ZW Parque Yo Amo a Siloé	4.331.432,04	5.459.143,87	16,70	0,187
023_ZW Centro Cultural Vista Hermosa	3.052.551,43	3.961.689,29	100,65	-0,649
024_ZW Parque Mutis	1.813.744,20	2.214.331,23	34,00	0,607
025_ZW Cancha Los Azules	3.591.561,56	4.597.839,19	27,87	0,467
026_ZW Parque La Orqueta	14.257.939,43	16.725.285,27	968,14	-3,760
027_ZW Parque Villa Colombia	1.982.615,80	2.739.156,52	51,28	0,263
028_ZW Parque Colseguros	2.366.553,41	2.833.112,79	32,56	-0,388
029_ZW Parque India Elena	2.339.662,17	2.733.962,10	26,98	0,507
030_ZW Parque Tequendama	1.690.522,73	2.255.329,33	21,64	0,535
031_ZW Parque Sector Amarillo Skate Park	1.172.453,77	1.471.592,95	46,28	0,490
032_ZW Biblioteca Daniel Guillard	2.531.882,92	3.349.208,22	26,25	0,398
033_ZW Polideportivo Torres Comfandi	4.506.263,48	5.802.683,90	71,57	-0,824
034_ZW Parque Junin	2.065.600,89	2.628.757,95	21,81	0,437
035_ZW Parque Mariano Ramos	1.667.000,49	2.049.644,94	27,14	-0,215
036_ZW Polideportivo Ricardo Balcazar	2.048.237,82	2.861.041,63	20,42	0,379
037_ZW Polideportivo San Benito	2.624.801,57	3.784.281,63	37,43	0,740
038_ZW Parque Santa Anita	2.084.061,91	2.654.540,27	55,14	-1,587
039_ZW Sebastian Belalcazar	2.393.571,87	3.211.876,30	162,84	-0,411
040_ZW Parque del Mico	1.077.900,59	1.420.057,79	23,79	0,687
041_ZW Parque Cien Palos	2.071.665,65	2.640.959,80	19,00	0,520
042_ZW Cerro 3 Cruces	1.331.725,14	1.832.711,09	44,48	0,811
043_ZW Parque Calima	1.264.508,43	1.787.766,82	30,22	0,687
044_ZW Parque La Merced	291.017.040.639.514,00	571.945.333.431.136,00	25987054518,19	-152524615518870000,000
045_ZW Puente de Colores	1.633.675,85	1.950.272,80	93,07	0,426

046_ZW Polideportivo Laureano Gomez	3.127.446,29	4.583.145,48	33,83	0,564
047_ZW El Diamante	692.484,10	931.375,85	74,48	0,703
048_ZW Polideportivo Petecuy	1.475.694,38	1.906.922,82	36,48	-0,448
049_ZW Comuna 16	240.588,66	329.494,50	131,77	0,753
050_La Castilla	6.735.367,77	9.048.561,40	2278,14	-32,374
051_La Elvira	2.628.538,07	3.557.652,67	169,51	0,405
052_El Saladito	2.423.900,42	3.240.094,95	79,39	-0,230

Anexo 6. Métricas de Desempeño – Perceptrón (Modelo base)

Zona WiFi	MAE Base	MAE Opt	RMSE Base	RMSE Opt	MAPE (%) Base	MAPE (%) Opt	R^2 Base	R^2 Opt
Parque Ingenio	1.078.031,66	1.212.908,21	1.670.629,85	1.810.270,40	33,45	35,37	0,426	0,326
Canchas Panamericanas	792.958,84	594.268,45	1.150.524,25	819.178,20	40,13	26,68	0,318	0,654
Parque del Perro	1.053.117,13	880.912,35	1.555.317,83	1.183.474,98	30,65	27,51	0,360	0,629
Parque San Nicolas	4.164.597,84	3.409.988,88	5.807.695,62	4.383.302,51	28,01	26,16	-0,185	0,325
Parque Barrio Obrero	7.178.427,65	4.777.467,53	8.947.742,39	5.848.322,43	28,73	20,60	-0,041	0,555
Parque Pizamos	1.572.501,81	912.416,31	2.157.869,43	1.254.878,11	45,84	22,29	0,246	0,745
Parque Alfonso Lopez	3.047.779,64	2.471.210,70	3.955.758,16	3.027.176,76	29,93	25,97	0,410	0,655
Parque Antonio Nariño	1.312.741,94	1.019.245,81	1.756.855,36	1.273.795,59	21,46	18,25	0,461	0,716
Parque Santa Rosa Poblado	14.743.922,3	11.450.847,6	21.582.367,08	17.760.614,05	258,94	633,99	-0,863	-0,262
Polideportivo Los Naranjos	2.744.256,90	2.635.657,76	4.046.963,61	3.943.596,06	104,77	94,99	-0,157	-0,099
Parque Llano Verde	1.470.739,96	1.359.994,77	1.744.797,22	1.566.591,06	37,10	37,05	-0,546	-0,246
Centro Cultural Comuna 13	2.818.422,68	1.919.397,18	3.955.006,30	2.638.010,89	143,41	107,50	0,618	0,830
Conjunto Habitacional Ramali	6.285.107,37	5.265.971,54	9.886.917,50	8.315.372,34	47,68	35,85	0,211	0,442
Parque Skate Board	1.747.547,01	1.262.357,88	2.400.351,64	1.684.059,05	20,70	15,57	0,382	0,696
Parque Alfonso Barberena	2.169.617,05	1.334.484,76	2.806.239,32	1.841.898,26	25,86	15,97	0,105	0,615
Polideportivo Los Farallones	1.404.156,47	1.436.111,00	2.233.080,00	2.285.027,40	286,80	272,20	0,284	0,250
Parque Los Guerreros	2.342.812,55	775.102,21	3.133.103,90	1.038.514,07	48,61	22,57	-0,785	0,804
Museo La Tertulia	562.367,28	451.768,24	724.367,22	570.461,96	722,85	585,89	0,524	0,705
Parque San Marino	5.582.349,16	4.373.125,16	8.345.127,04	6.608.249,33	579,89	129,95	-0,225	0,232
Parque La Flora	2.102.348,81	1.452.733,27	2.751.191,73	1.940.843,12	32,29	23,75	0,216	0,610

Parque Alfonso Bonilla Aragón	1.252.067,69	856.367,89	1.789.628,35	1.213.863,83	37,48	37,65	0,588	0,810
Parque Yo Amo a Siloé	6.622.281,67	3.542.465,53	7.965.328,34	4.817.025,93	24,40	13,55	-0,731	0,367
Centro Cultural Vista Hermosa	2.991.384,92	2.252.262,93	4.095.655,24	2.905.562,16	70,70	63,22	-0,763	0,113
Parque Mutis	1.350.573,15	1.974.579,13	2.104.637,93	2.315.092,50	21,63	37,34	0,645	0,570
Cancha Los Azules	4.170.357,30	3.605.915,93	5.688.012,17	4.504.933,96	26,64	23,60	0,184	0,488
Parque La Orqueta	11.740.065,29	5.760.371,79	13.797.615,70	7.054.918,29	631,75	212,30	-2,240	0,153
Parque Villa Colombia	2.764.053,03	1.960.139,26	3.759.242,23	2.735.555,19	58,80	46,89	-0,388	0,265
Parque Colseguros	1.842.968,88	2.423.297,40	2.377.621,84	2.963.453,33	23,85	32,80	0,022	-0,519
Parque India Elena	2.626.546,31	2.048.019,15	3.370.473,97	2.399.113,41	26,95	23,55	0,250	0,620
Parque Tequendama	2.559.346,70	2.695.703,74	3.215.315,96	3.237.412,01	28,32	33,23	0,056	0,043
Parque Sector Amarillo Skate Park	1.444.884,31	927.099,53	1.743.530,50	1.245.609,77	71,11	38,05	0,284	0,635
Biblioteca Daniel Guillard	3.544.196,27	2.035.956,48	4.668.956,05	2.869.510,11	34,34	19,46	-0,170	0,558
Polideportivo Torres Comfandi	3.811.598,57	5.057.202,09	5.350.874,57	6.335.108,78	52,87	84,46	-0,551	-1,174
Parque Junin	3.056.053,83	2.647.574,97	3.775.732,31	3.448.965,76	28,34	26,54	-0,161	0,031
Parque Mariano Ramos	1.709.257,94	1.223.236,90	2.241.406,92	1.602.669,19	25,80	19,44	-0,453	0,257
Polideportivo Ricardo Balcazar	2.513.865,16	1.939.287,04	3.406.837,80	2.553.421,13	22,23	18,53	0,120	0,506
Polideportivo San Benito	3.811.805,79	2.695.400,09	6.318.237,27	3.787.322,27	48,11	36,19	0,277	0,740
Parque Santa Anita	2.041.752,81	850.898,05	2.561.684,81	1.163.980,69	45,39	22,80	-1,409	0,503
Sebastian Belalcazar	3.392.592,19	2.909.568,13	4.304.977,74	3.497.150,59	125,38	82,14	-1,535	-0,673
Parque del Mico	1.392.045,42	1.024.579,45	1.824.569,63	1.341.288,79	30,92	20,30	0,484	0,721
Parque Cien Palos	2.894.202,28	2.140.227,61	3.838.165,23	2.682.468,01	24,38	18,09	-0,014	0,505
Cerro 3 Cruces	1.505.894,50	1.243.160,79	2.159.658,99	1.737.407,88	61,68	30,17	0,738	0,830
Parque Calima	1.339.537,07	1.034.276,78	2.128.008,29	1.455.345,45	28,27	24,48	0,556	0,792
Parque La Merced	1.030.450,77	1.036.600,20	1.498.509,95	1.511.332,70	78,00	76,75	-0,047	-0,065
Puente de Colores	1.987.869,07	1.346.356,12	2.790.387,94	1.891.771,31	52,73	50,45	-0,175	0,460
Polideportivo Laureano Gomez	4.994.240,37	3.165.231,66	7.282.773,75	4.911.579,77	47,49	34,54	-0,102	0,499
El Diamante	951.110,88	645.521,96	1.385.624,30	862.768,25	55,54	57,72	0,342	0,745
Polideportivo Petecuy	954.568,09	1.216.185,27	1.387.683,92	1.498.669,52	23,48	29,78	0,233	0,106
Comuna 16	298.684,85	184.822,04	492.738,26	300.142,85	73,17	62,45	0,448	0,795
050_La Castilla	825.543,38	1.482.521,10	1.425.363,56	1.803.214,89	188,15	521,20	0,172	-0,325
051_La Elvira	2.501.040,64	2.311.158,93	3.663.751,81	3.376.654,76	132,12	131,98	0,369	0,464
052_El Saladito	2.538.781,69	2.304.829,60	3.459.820,85	3.137.605,91	60,24	66,31	-0,403	-0,154

Anexo 7. SVR base vs SVR optimizado

Zona WiFi	MAE Base	MAE Opt	RMSE Base	RMSE Opt	MAPE (%) Base	MAPE (%) Opt	R^2 Base	R^2 Opt
Parque Ingenio	6.231.274,72	1.321.993,78	6.706.314,45	1.759.345,77	269,73	42,94	-8,252	0,363
Canchas Panamericanas	551.104,81	609.670,19	750.546,97	834.375,68	31,57	32,11	0,710	0,641
Parque del Perro	1.057.232,48	1.101.335,72	1.428.209,74	1.438.148,35	33,81	34,47	0,460	0,453
Parque San Nicolas	2.863.132,89	3.717.160,52	3.918.591,31	4.697.175,19	20,83	29,39	0,460	0,225
Parque Barrio Obrero	4.300.606,96	4.979.891,73	5.047.431,39	6.235.258,68	21,79	20,56	0,669	0,494
Parque Pizamos	942.269,20	1.164.503,36	1.405.260,34	1.530.186,00	20,39	27,38	0,680	0,621
Parque Alfonso Lopez	2.833.231,21	2.852.702,35	3.559.343,42	3.520.371,83	32,05	31,26	0,523	0,533
Parque Antonio Nariño	1.017.913,75	1.047.455,41	1.246.428,05	1.344.555,65	21,42	18,49	0,729	0,684
Parque Santa Rosa Poblado	11.209.602,60	7.987.164,00	17.074.890,28	12.286.097,42	1609,65	2186,57	-0,166	0,396
Polideportivo Los Naranjos	1.723.843,02	3.175.521,62	2.613.352,62	4.519.302,55	82,48	164,90	0,518	-0,443
Parque Llano Verde	1.147.587,77	1.522.222,34	1.394.759,70	1.777.881,98	30,70	40,68	0,012	-0,605
Centro Cultural Comuna 13	2.149.644,77	2.081.356,36	3.033.566,43	2.968.629,15	86,88	81,13	0,775	0,785
Conjunto Habitacional Ramali	6.681.711,37	5.489.522,38	9.306.385,19	8.509.188,99	48,98	47,30	0,301	0,416
Parque Skate Board	1.719.320,18	12.766.933,93	2.256.043,12	14.700.211,71	19,03	164,08	0,454	-22,162
Parque Alfonso Barberena	1.421.567,16	1.290.007,14	1.940.057,24	1.778.540,97	18,28	17,13	0,572	0,641
Polideportivo Los Farallones	1.205.997,89	1.459.004,43	1.749.945,49	2.052.788,01	248,33	223,27	0,560	0,395
Parque Los Guerreros	2.170.307,29	1.845.649,28	2.786.892,82	2.615.097,27	64,60	43,69	-0,412	-0,244
Museo La Tertulia	770.074,74	530.832.484,42	900.806,24	1.776.380.775,95	1078,52	2609241,84	0,264	-2863977,310
Parque San Marino	7.050.187,21	3.946.997,67	8.757.283,92	6.097.251,28	300,30	335,70	-0,349	0,346
Parque La Flora	1.699.178,97	2.339.405,23	2.293.723,09	2.897.139,10	28,98	39,53	0,455	0,130
Parque Alfonso Bonilla Aragón	1.147.962,25	845.002,56	1.569.654,72	1.151.233,36	44,97	40,32	0,683	0,829
Parque Yo Amo a Siloé	4.331.432,04	6.343.883,71	5.459.143,87	7.865.359,14	16,70	24,16	0,187	-0,688
Centro Cultural Vista Hermosa	3.052.551,43	2.643.361,25	3.961.689,29	3.747.743,81	100,65	73,95	-0,649	-0,476
Parque Mutis	1.813.744,20	1.884.707,82	2.214.331,23	2.347.738,48	34,00	31,36	0,607	0,558
Cancha Los Azules	3.591.561,56	3.671.075,58	4.597.839,19	4.772.537,89	27,87	26,80	0,467	0,426
Parque La Orqueta	14.257.939,43	4.327.134,96	16.725.285,27	5.791.549,54	968,14	200,69	-3,760	0,429
Parque Villa Colombia	1.982.615,80	2.218.624,86	2.739.156,52	2.848.201,96	51,28	54,92	0,263	0,203
Parque Colseguros	2.366.553,41	1.832.479,28	2.833.112,79	2.391.392,03	32,56	24,95	-0,388	0,011
Parque India Elena	2.339.662,17	2.946.342,80	2.733.962,10	3.318.270,68	26,98	33,53	0,507	0,273
Parque Tequendama	1.690.522,73	2.428.845,87	2.255.329,33	3.064.654,71	21,64	28,93	0,535	0,142
Parque Sector Amarillo Skate Park	1.172.453,77	1.135.342,84	1.471.592,95	1.498.665,46	46,28	45,42	0,490	0,471

Biblioteca Daniel Guillard	2.531.882,92	2.974.507,45	3.349.208,22	3.972.204,77	26,25	30,68	0,398	0,153
Polideportivo Torres Comfandi	4.506.263,48	4.708.902,25	5.802.683,90	6.045.202,79	71,57	70,91	-0,824	-0,980
Parque Junin	2.065.600,89	4.306.981,14	2.628.757,95	4.836.992,56	21,81	44,22	0,437	-0,906
Parque Mariano Ramos	1.667.000,49	3.533.518,03	2.049.644,94	3.978.429,14	27,14	58,41	-0,215	-3,576
Polideportivo Ricardo Balcazar	2.048.237,82	1.783.187,17	2.861.041,63	2.314.469,17	20,42	18,37	0,379	0,594
Polideportivo San Benito	2.624.801,57	2.531.394,09	3.784.281,63	3.750.727,96	37,43	48,80	0,740	0,745
Parque Santa Anita	2.084.061,91	1.853.337,08	2.654.540,27	2.353.060,52	55,14	41,92	-1,587	-1,033
Sebastian Belalcazar	2.393.571,87	3.215.282,12	3.211.876,30	4.144.118,01	162,84	133,80	-0,411	-1,349
Parque del Mico	1.077.900,59	1.212.936,20	1.420.057,79	1.543.069,74	23,79	25,74	0,687	0,631
Parque Cien Palos	2.071.665,65	2.148.219,29	2.640.959,80	2.777.431,25	19,00	20,26	0,520	0,469
Cerro 3 Cruces	1.331.725,14	2.108.039,24	1.832.711,09	2.667.909,09	44,48	51,62	0,811	0,600
Parque Calima	1.264.508,43	1.221.873,93	1.787.766,82	1.736.049,39	30,22	28,89	0,687	0,704
Parque La Merced	291.017.040.639.514,00	382.234.641.460,32	571.945.333.431.136,00	498.733.771.395,64	25987054518,19	34663032,40	-15252461551887000000,000	-115976086931,714
Puente de Colores	1.633.675,85	1.839.985,86	1.950.272,80	2.316.286,80	93,07	60,78	0,426	0,190
Polideportivo Laureano Gomez	3.127.446,29	2.864.240,45	4.583.145,48	3.697.715,76	33,83	45,44	0,564	0,716
El Diamante	692.484,10	850.171,16	931.375,85	1.124.378,42	74,48	65,84	0,703	0,566
Polideportivo Petecuy	1.475.694,38	1.075.785,34	1.906.922,82	1.498.690,13	36,48	25,75	-0,448	0,106
Comuna 16	240.588,66	216.384,20	329.494,50	319.183,21	131,77	64,00	0,753	0,768
050_La Castilla	6.735.367,77	4.692.569,41	9.048.561,40	7.411.641,57	2278,14	1238,69	-32,374	-21,391
051_La Elvira	2.628.538,07	1.967.482,21	3.557.652,67	3.039.738,42	169,51	76,04	0,405	0,566
052_El Saladito	2.423.900,42	2.559.797,69	3.240.094,95	3.323.514,01	79,39	64,88	-0,230	-0,295

Anexo 8. perceptrón base vs perceptrón optimizado

Zona WiFi	MAE Base	MAE Opt	RMSE Base	RMSE Opt	MAPE (%) Base	MAPE (%) Opt	R^2 Base	R^2 Opt
Parque Ingenio	1.041.531,47	1.076.748,99	1.434.059,00	1.489.884,00	25,80	26,40	0,667	0,640
Canchas Panamericanas	600.874,54	582.352,50	839.877,00	817.877,00	21,97	21,54	0,675	0,692
Parque del Perro	692.025,95	697.841,76	893.662,00	900.166,00	24,48	24,75	0,498	0,491
Parque San Nicolas	2.166.368,31	2.162.611,55	2.900.307,00	2.854.942,00	20,05	20,38	0,546	0,560
Parque Barrio Obrero	3.359.334,15	3.330.300,58	4.438.344,00	4.370.968,00	17,91	17,70	0,663	0,674
Parque Pizamos	765.846,47	755.150,13	1.325.091,00	1.335.905,00	20,78	20,63	0,723	0,718
Parque Alfonso Lopez	2.935.948,85	2.893.594,39	3.519.269,00	3.455.481,00	31,52	30,86	0,279	0,305
Parque Antonio Nariño	753.761,88	759.194,82	914.938,00	919.569,00	16,85	16,94	0,763	0,761
Parque Santa Rosa Poblado	1.902.469,62	1.905.681,01	2.554.255,00	2.559.401,00	54,07	54,12	0,105	0,101
Polideportivo Los Naranjos	1.628.934,74	1.575.362,00	2.483.142,00	2.409.040,00	31,12	30,36	0,400	0,436
Parque Llano Verde	811.251,94	816.827,70	1.020.262,00	1.026.478,00	23,34	23,55	0,354	0,346
Centro Cultural Comuna 13	1.356.253,35	1.381.605,43	1.880.866,00	1.924.450,00	22,44	21,80	0,888	0,882
Conjunto Habitacional Ramali	4.840.442,01	4.932.653,45	7.683.316,00	7.789.346,00	25,21	25,68	0,400	0,383
Parque Skate Board	2.158.449,99	2.140.122,24	2.572.112,00	2.536.561,00	25,00	24,81	0,181	0,204
Parque Alfonso Barberena	1.438.238,46	1.437.480,96	1.786.836,00	1.801.224,00	19,82	19,75	0,649	0,643
Polideportivo Los Farallones	1.049.463,93	1.034.906,34	1.663.346,00	1.629.004,00	26,60	26,56	0,653	0,667
Parque Los Guerreros	944.571,73	959.920,50	1.230.534,00	1.244.676,00	23,27	23,48	0,575	0,565
Museo La Tertulia	256.054,24	251.605,47	473.914,00	487.681,00	27,16	26,75	0,764	0,750
Parque San Marino	1.328.944,05	1.308.280,05	1.546.633,00	1.522.624,00	28,30	27,85	0,686	0,695
Parque La Flora	1.464.655,09	1.468.210,84	2.065.924,00	2.059.087,00	24,30	24,45	0,529	0,533
Parque Alfonso Bonilla Aragón	753.981,87	753.981,87	976.329,00	976.329,00	22,84	22,84	0,821	0,821
Parque Yo Amo a Siloé	3.816.150,47	3.816.150,47	4.523.177,00	4.523.177,00	18,22	18,22	-0,080	-0,080
Centro Cultural Vista Hermosa	1.192.950,95	1.380.520,64	1.848.013,00	2.130.314,00	14977,03	32627,34	0,540	0,389
Parque Mutis	1.348.414,45	1.354.698,11	1.770.931,00	1.787.837,00	21,45	21,46	0,663	0,656
Cancha Los Azules	3.278.059,95	3.264.120,02	3.961.082,00	3.934.525,00	23,19	23,13	0,478	0,485
Parque La Orqueta	2.849.583,84	2.944.762,46	3.698.491,00	3.742.831,00	15,64	16,24	0,543	0,532
Parque Villa Colombia	1.113.487,50	1.079.780,76	1.506.139,00	1.459.426,00	27,27	26,75	0,443	0,477
Parque Colseguros	1.238.262,80	1.257.193,25	1.564.623,00	1.583.242,00	17,47	17,70	0,610	0,601
Parque India Elena	1.616.556,43	1.674.686,71	1.862.000,00	1.931.496,00	20,55	21,26	0,702	0,679
Parque Tequendama	1.609.187,87	1.620.461,30	2.026.114,00	2.026.647,00	21,78	21,92	0,601	0,601
Parque Sector Amarillo Skate Park	1.079.713,16	1.074.271,42	1.450.176,00	1.455.090,00	19,71	19,55	0,672	0,669
Biblioteca Daniel Guillard	1.678.754,54	1.660.763,44	2.062.479,00	2.021.791,00	18,73	18,72	0,670	0,683
Polideportivo Torres Comfandi	1.907.529,47	1.889.987,57	3.242.404,00	3.224.736,00	26,59	26,47	0,259	0,267
Parque Junin	832.090,80	840.154,67	1.201.016,00	1.224.621,00	11,27	11,19	0,835	0,829
Parque Mariano Ramos	1.337.949,09	1.340.769,50	1.545.469,00	1.555.552,00	22,79	22,73	0,426	0,419

Polideportivo Ricardo Balcazar	1.868.567,03	1.857.946,52	2.203.301,00	2.186.641,00	19,66	19,55	0,463	0,471
Polideportivo San Benito	1.239.763,42	1.245.552,09	1.697.408,00	1.702.866,00	23,25	23,37	0,767	0,766
Parque Santa Anita	919.488,48	906.824,51	1.161.592,00	1.140.504,00	25,37	25,18	0,399	0,421
Sebastian Belalcazar	676.233,79	653.659,75	987.831,00	954.484,00	24,52	24,28	0,480	0,515
Parque del Mico	1.098.313,19	1.096.752,05	1.362.113,00	1.370.129,00	22,94	22,80	0,762	0,760
Parque Cien Palos	1.422.678,61	1.370.421,36	1.729.461,00	1.636.156,00	17,70	17,07	0,428	0,488
Cerro 3 Cruces	830.494,20	845.414,57	1.138.297,00	1.144.260,00	31,54	31,24	0,875	0,874
Parque Calima	1.049.547,98	1.051.567,12	1.511.996,00	1.535.447,00	25,39	25,38	0,795	0,789
Parque La Merced	642.362,99	667.528,26	1.061.839,00	1.114.203,00	31,77	32,31	0,449	0,394
Puente de Colores	1.082.413,15	1.090.372,77	1.664.723,00	1.666.002,00	28,68	29,21	0,398	0,397
Polideportivo Laureano Gomez	1.047.908,20	1.082.996,98	1.708.221,00	1.771.588,00	16,17	16,72	0,778	0,761
El Diamante	639.656,75	635.259,00	1.095.225,00	1.094.182,00	31,78	31,46	0,541	0,542
Polideportivo Petecuy	995.049,89	1.030.595,11	1.402.487,00	1.428.335,00	20,12	21,04	0,644	0,631
Comuna 16	133.503,20	131.016,33	277.108,00	265.762,00	24,65	24,85	0,799	0,815
050_La Castilla	597.506,43	616.318,18	875.485,00	878.483,00	3979,81	4622,41	0,555	0,552
051_La Elvira	1.659.301,03	1.659.301,03	2.977.944,00	2.977.944,00	232,75	232,75	0,534	0,534
052_El Saladito	1.556.688,83	1.568.250,63	1.948.527,00	1.986.971,00	85,48	85,03	0,224	0,193

Anexo 9. Random Forest base vs Random Forest optimizado

Zona	Modelo	Hiperparámetros
001_ZW Parque Ingenio	Random Forest	n_estimators: 100 max_depth: None random_state: 123 kernel: rbf
002_ZW Canchas Panamericanas	Random Forest	n_estimators: 350 max_depth: 10 random_state: 123 kernel: rbf
003_ZW Parque del Perro	Random Forest	n_estimators: 100 max_depth: None random_state: 123 kernel: rbf
005_ZW Parque Barrio Obrero	Random Forest	n_estimators: 50 max_depth: 10 random_state: 123 kernel: rbf
006_ZW Parque Pizamos	Perceptrón	hidden_layer_sizes: (100) activation: relu solver: adam max_iter: 200 alpha: 0.0001 learning_rate: constant learning_rate_init: 0.001

007_ZW Parque Alfonso Lopez	SVR	kernel: rbf C: 10.0 epsilon: 0.1 gamma: 0.005
009_ZW Parque Santa Rosa Poblado	Random Forest	n_estimators: 100 max_depth: None random_state: 123 kernel: rbf
010_ZW Polideportivo Los Naranjos	Random Forest	n_estimators: 250 max_depth: 10 random_state: 123 kernel: rbf
011_ZW Parque Llano Verde	Random Forest	n_estimators: 100 max_depth: None random_state: 123 kernel: rbf
012_ZW Centro Cultural Comuna 13	Random Forest	n_estimators: 100 max_depth: 5 random_state: 123 kernel: rbf
013_ZW Conjunto Habitacional Ramali	Random Forest	n_estimators: 100 max_depth: None random_state: 123 kernel: rbf
014_ZW Parque Skate Board	SVR	kernel: rbf C: 10.0 epsilon: 0.01 gamma: 0.005
016_ZW Polideportivo Los Farallones	Random Forest	n_estimators: 350 max_depth: 10 random_state: 123 kernel: rbf
017_ZW Parque Los Guerreros	SVR	kernel: rbf C: 50.0 epsilon: 0.1 gamma: 0.005
018_ZW Museo La Tertulia	Random Forest	n_estimators: 50 max_depth: 10 random_state: 123 kernel: rbf
019_ZW Parque San Marino	Random Forest	n_estimators: 250 max_depth: 20 random_state: 123 kernel: rbf
021_ZW Parque Alfonso Bonilla Aragón	Random Forest	n_estimators: 100 max_depth: 20 random_state: 123 kernel: rbf
023_ZW Centro Cultural Vista Hermosa	SVR	kernel: rbf C: 50.0 epsilon: 0.01 gamma: 0.005
026_ZW Parque La Orqueta	Random Forest	n_estimators: 100 max_depth: None random_state: 123

		kernel: rbf
027_ZW Parque Villa Colombia	Random Forest	n_estimators: 100 max_depth: 10 random_state: 123 kernel: rbf
028_ZW Parque Colseguros	Random Forest	n_estimators: 100 max_depth: None random_state: 123 kernel: rbf
030_ZW Parque Tequendama	Perceptrón	hidden_layer_sizes: (100) activation: relu solver: adam max_iter: 200 alpha: 0.0001 learning_rate: constant learning_rate_init: 0.001
031_ZW Parque Sector Amarillo Skate Park	Random Forest	n_estimators: 100 max_depth: 10 random_state: 123 kernel: rbf
032_ZW Biblioteca Daniel Guillard	Random Forest	n_estimators: 350 max_depth: 10 random_state: 123 kernel: rbf
033_ZW Polideportivo Torres Comfandi	Random Forest	n_estimators: 150 max_depth: 20 random_state: 123 kernel: rbf
034_ZW Parque Junin	Random Forest	n_estimators: 350 max_depth: 5 random_state: 123 kernel: rbf
035_ZW Parque Mariano Ramos	SVR	kernel: rbf C: 50.0 epsilon: 0.1 gamma: scale
038_ZW Parque Santa Anita	SVR	kernel: rbf C: 100.0 epsilon: 0.1 gamma: 0.005
039_ZW Sebastian Belalcazar	Random Forest	n_estimators: 50 max_depth: 10 random_state: 123 kernel: rbf
040_ZW Parque del Mico	SVR	kernel: rbf C: 10.0 epsilon: 0.004 gamma: 0.005
042_ZW Cerro 3 Cruces	SVR	kernel: rbf C: 10.0 epsilon: 0.01 gamma: 0.005
043_ZW Parque Calima	SVR	kernel: rbf C: 10.0 epsilon: 0.004

		gamma: 0.005
044_ZW Parque La Merced	Random Forest	n_estimators: 100 max_depth: None random_state: 123 kernel: rbf
045_ZW Puente de Colores	Random Forest	n_estimators: 100 max_depth: None random_state: 123 kernel: rbf
046_ZW Polideportivo Laureano Gomez	Random Forest	n_estimators: 100 max_depth: None random_state: 123 kernel: rbf
047_ZW El Diamante	Random Forest	n_estimators: 150 max_depth: 10 random_state: 123 kernel: rbf
048_ZW Polideportivo Petecuy	Random Forest	n_estimators: 100 max_depth: None random_state: 123 kernel: rbf
049_ZW Comuna 16	Random Forest	n_estimators: 100 max_depth: None random_state: 123 kernel: rbf
050_La Castilla	SVR	kernel: rbf C: 1.0 epsilon: 0.1 gamma: scale
051_La Elvira	Perceptrón	hidden_layer_sizes: (50,25,10) activation: relu solver: adam max_iter: 10000 alpha: 0.0001 learning_rate: constant learning_rate_init: 0.01
052_El Saladito	SVR	kernel: rbf C: 1.0 epsilon: 0.1 gamma: scale

Anexo 10. Hiperparametros por Zona.

```
# ----------------------------------------------------------
    # Busqueda de Hiper-parámetros por zona:
    # ----------------------------------------------------------


    forecaster = ForecasterRecursive(
        regressor = RandomForestRegressor(random_state=123),
        lags      = 10 # Este valor será remplazado en el grid search
    )

    # particiones train y validacion
    cv = TimeSeriesFold(
        steps              = steps,
        initial_train_size = max(30, int(0.5 * len(df_train))),
        refit              = False,
        fixed_train_size   = False,
    )

    # Valores de lags para evaluar
    lags_grid = [10]

    # Valores a evaluar como hiperparámetros
    param_grid = {
        'n_estimators': [50, 100, 150, 250, 350],
        'max_depth': [5, 10, 20, 30, 40]
    }

    resultados_grid = grid_search_forecaster(
        forecaster  = forecaster,
        y           = df_train['USAGE_KB'],
        exog        = df_train[todas_variables_entrada],  # Variables exogenas
        cv          = cv,
        param_grid  = param_grid,
        lags_grid   = lags_grid,
        metric      = 'mean_absolute_percentage_error',
        return_best = True,
        n_jobs      = 1,  # ← Sin procesamiento paralelo para que no genere error
        verbose     = False
    )


    resultados_grid.to_csv(DESTINO_METRICAS / f"grilla_{nombre_zona}", index=False,
encoding='utf-8')
```

Anexo 11. Código en Python para la búsqueda de hiperparámetros en Random Forest

Libreria - Version	Libreria - Version	Libreria - Version
alembic 1.17.2	llvmlite 0.46.0	requests 2.32.5
altair 6.0.0	Mako 1.3.10	rich 14.2.0
attrs 25.4.0	markdown-it-py 4.0.0	rpds-py 0.30.0
blinker 1.9.0	MarkupSafe 3.0.3	scikit-learn 1.8.0
cachetools 6.2.3	matplotlib 3.10.8	scipy 1.16.3
certifi 2025.11.12	mdurl 0.1.2	seaborn 0.13.2
charset-normalizer 3.4.4	narwhals 2.13.0	shap 0.50.0
click 8.3.1	numba 0.63.1	six 1.17.0
cloudpickle 3.1.2	numpy 2.3.5	skforecast 0.19.1
colorama 0.4.6	optuna 4.6.0	slicer 0.0.8
colorlog 6.10.1	packaging 25.0	smmap 5.0.2
contourpy 1.3.3	pandas 2.3.3	SQLAlchemy 2.0.45
cycler 0.12.1	pillow 12.0.0	streamlit 1.52.1
fonttools 4.61.1	plotly 6.5.0	tenacity 9.1.2
gitdb 4.0.12	protobuf 6.33.2	threadpoolctl 3.6.0
GitPython 3.1.45	pyarrow 22.0.0	toml 0.10.2
greenlet 3.3.0	pydeck 0.9.1	tornado 6.5.3
holidays 0.86	Pygments 2.19.2	tqdm 4.67.1
idna 3.11	pyparsing 3.2.5	typing_extensions 4.15.0
Jinja2 3.1.6	python-dateutil 2.9.0.post0	tzdata 2025.3
joblib 1.5.2	pytz 2025.2	urllib3 2.6.2
jsonschema 4.25.1	PyYAML 6.0.3	watchdog 6.0.0
jsonschema-specifications 2025.9.1	referencing 0.37.0	
kiwisolver 1.4.9		

Anexo 12. Librerías de Python usadas para Ciencia de Datos