



Pontificia Universidad  
**JAVERIANA**  
Cali

**MODELO PREDICTIVO DEL TRÁFICO DE REDES DE COMUNICACIÓN 4G EN EVENTOS CULTURALES O SOCIALES EN MUNICIPIOS INTERMEDIOS DE CUNDINAMARCA.**

*Wilson Edilberto Balaguera Cubillos, Paula Ginette González Torres*

*Código 8975495, 8975386*

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

Director(a)  
*Luis Eduardo Tobón Llano*

FACULTAD DE INGENIERÍA Y CIENCIAS  
MAESTRÍA EN CIENCIA DE DATOS  
SANTIAGO DE CALI, DICIEMBRE 20 DE 2024

## TABLA DE CONTENIDO

	Pág.
INTRODUCCIÓN .....	6
1. DEFINICIÓN DEL PROBLEMA.....	7
1.1. PLANTEAMIENTO DEL PROBLEMA.....	7
1.2. FORMULACIÓN DEL PROBLEMA.....	7
2. OBJETIVOS DEL PROYECTO .....	9
2.1. OBJETIVO GENERAL .....	9
2.2. OBJETIVOS ESPECÍFICOS .....	9
3. MARCO TEÓRICO .....	10
3.1. MARCO TEÓRICO .....	10
3.1.1. SECTOR DE LAS TELECOMUNICACIONES .....	10
3.1.2. TECNOLOGÍA LTE (Long Term Evolution) .....	11
3.1.3. TRÁFICO EN REDES DE TELECOMUNICACIONES MÓVILES. ....	12
3.1.4. REGULACIÓN NACIONAL .....	12
3.1.5. PAYLOAD O TRÁFICO DE RED .....	13
3.1.6. TASA DE TRANSFERENCIA EFECTIVA.....	13
3.1.7. CATEGORIZACIÓN DANE .....	13
3.1.8. WEB SCRAPING.....	14
3.1.9. PROCESAMIENTO DE LENGUAJE NATURAL .....	14
3.1.10. MODELO DE DATOS .....	15
3.2. ANTECEDENTES .....	19
3.2.1. PREDICCIÓN DE COMPORTAMIENTO EN TRÁFICO DE RED LTE Y AJUSTE DE PARAMETRIZACIÓN PARA MAXIMIZAR PERFORMANCE DE RED .....	19
3.2.2. STOCK PRICE PREDICTION USING NEWS SENTIMENT ANALYSIS .....	20
3.2.3. DETECCIÓN DE PATRONES DE COMPORTAMIENTO DE PARÁMETROS DE RF EN REDES DE COMUNICACIÓN MÓVIL MEDIANTE MEDICIONES DE CAMPO Y TÉCNICAS DE MACHINE LEARNING. 20	20
3.2.4. ANÁLISIS DE SENTIMIENTOS DE DATOS DE REDES SOCIALES USANDO TÉCNICAS DE MACHINE LEARNING.....	21
4. METODOLOGÍA .....	22
5. DESARROLLO DEL PROYECTO .....	24
5.1. IDENTIFICACIÓN DE DATOS .....	24
5.1.1. CLASIFICACIÓN DE MUNICIPIOS .....	24
5.1.2. DETALLE DE TRÁFICO .....	24

5.2.	EXTRACCIÓN DE EVENTOS POR MEDIO DE WEB SCRAPING. ....	26
5.3.	COMPORTAMIENTO DE TRÁFICO INICIAL.....	26
6.	PREPARACIÓN DE LOS DATOS .....	30
6.1.	IMPUTACIÓN DE NULOS .....	30
6.2.	SUAVIZADO DE DATOS O DETECCIÓN DE VALORES ATÍPICOS.....	30
6.3.	ANÁLISIS DEL TRÁFICO.....	31
7.	IMPLEMENTACIÓN Y EVALUACIÓN DE LOS MODELOS .....	33
7.1.	FACEBOOK PROPHET .....	33
7.2.	RED NEURONAL ELMAN .....	35
7.3.	RED NEURONAL JORDAN.....	36
7.4.	NEURALPROPHET .....	37
7.4.1.	EVALUACIÓN DEL MODELO .....	42
8.	CONCLUSIONES Y TRABAJOS FUTUROS.....	45
8.1.	CONCLUSIONES.....	45
8.2.	TRABAJOS FUTUROS.....	45
	BIBLIOGRAFÍA .....	46
	ANEXOS.....	49

## LISTA DE FIGURAS

Ilustración 1. Arquitectura de red tecnología LTE [40] .	11
Ilustración 2 Arquitectura modelo Facebook Prophet [43]	15
Ilustración 3 Arquitectura red de Elman [41]	16
Ilustración 4 Arquitectura Red de Jordan [41]	17
Ilustración 5 Arquitectura modelo NeuralProphet [42]	19
Ilustración 6 Metodología Implementada en el Proyecto (Autores)	23
Ilustración 7 Comportamiento del tráfico año 2021. En línea azul el promedio diario y en gris la desviación estándar (Autores).	27
Ilustración 8 Comportamiento del tráfico año 2022. En línea azul el promedio diario y en gris la desviación estándar (Autores).	28
Ilustración 9 Comportamiento del tráfico año 2023. En línea azul el promedio diario y en gris la desviación estándar (Autores).	28
Ilustración 10 Comportamiento del tráfico año 2024. En línea azul el promedio diario y en gris la desviación estándar (Autores).	29
Ilustración 11 Serie de tiempo posterior a la imputación de nulos (Año 2024) (Autores).	30
Ilustración 12 Comportamiento del Payload (Tráfico en línea azul) y posterior suavizado de datos (línea naranja) (Autores).	31
Ilustración 13 Distribución Modal del Tráfico (Autores).	31
Ilustración 14 Descomposición Estacional del Tráfico (Autores).	32
Ilustración 15 Predicción del Tráfico Modelo Prophet (Autores).	34
Ilustración 16 Descomposición del Tráfico - Modelo Prophet (Autores).	34
Ilustración 17 Comparación Tráfico Actual vs Predicción - Red Neuronal Elman (Autores).	36
Ilustración 18 Comparación Tráfico Actual vs Predicción - Red Neuronal Jordan (Autores).	37
Ilustración 19 Comparación Tráfico Real vs Predicción del Modelo - NeuralProphet (Autores).	40
Ilustración 20 Análisis de Tráfico haciendo uso Modelo NeuralProphet (Autores).	40
Ilustración 21 Distribución de Residuos - NeuralProphet (Autores).	41
Ilustración 22 Tráfico real versus predicción con conjunto de datos de prueba. (Autores).	41
Ilustración 23 Predicción de Tráfico septiembre 2024 - noviembre 2024 (Autores).	42
Ilustración 24 Predicción de Tráfico para Festividades (Autores).	43
Ilustración 25 Comparación Predicción de Tráfico vs Tráfico Real del 10 al 14 de octubre 2024 (Autores).	44

## LISTA DE TABLAS

Tabla 1 Estructura de Datos - Clasificación Municipio (Autores).....	24
Tabla 2 Estructura de Datos - Tráfico de Red (Autores).....	25
Tabla 3 Estructura de Datos – Festividades (Autores).....	26
Tabla 4 Variables de entrada a los modelos (Autores).....	27
Tabla 5 Entradas Modelo Facebook Prophet (Autores).....	33
Tabla 6 Parámetros Modelo Facebook Prophet (Autores).....	33
Tabla 7 Entradas Modelo Elman (Autores).....	35
Tabla 8 Parámetros Modelo Elman (Autores).....	35
Tabla 9 Entradas Modelo Jordan (Autores).....	36
Tabla 10 Parámetros Modelo Jordan (Autores).....	36
Tabla 11 Entradas Modelo Neuralprophet (Autores).....	39

## INTRODUCCIÓN

Según un artículo de la revista Portafolio en 2010, la saturación de red es uno de los problemas más recurrentes para los proveedores de comunicación móvil en Colombia, ya que la demanda de la red aumenta al usar un servicio de voz o datos en un periodo específico [2]. El crecimiento paulatino presentado por la población, eventos exógenos, concentración de usuarios y la infraestructura de red implementada a nivel nacional han impactado de manera negativa el comportamiento de la red e indirectamente la experiencia del cliente, considerando que el diseño de las estaciones base se da acorde a los requerimientos iniciales en cada municipio o zona analizada. Esto, a su vez, tiene repercusiones económicas para las compañías de telecomunicaciones, ya que una mala experiencia del usuario puede resultar en la pérdida de clientes y en una disminución de ingresos.

Para evaluar dicha situación se puede acudir a la ciencia de datos, entendida como una combinación de métodos y tecnología para generar significado a partir de la información. Se puede resaltar a esta, como una herramienta potencial para la mejora continua de los procesos de diferentes sectores, pues como lo menciona el Departamento de Planeación de Colombia, los grupos de científico de datos son quienes, a partir de la formulación, ejecución y analítica de datos, permiten dar direcciones técnicas y crear valor a las diferentes entidades [3].

Este proyecto permite dar a conocer el comportamiento de la red en futuros eventos integrando información de esta, fechas de festividades y herramientas como la Ciencia de datos, como redes neuronales y modelos de aprendizaje mixto, a partir de la experticia que se tiene en el sector, el acceso a los datos de red existentes en la empresa de telecomunicaciones y los cronogramas de eventos obtenidos de entidades Gubernamentales. Con el fin de generar planes preventivos que aportan en la mejora de la experiencia del cliente en diferentes escenarios

Para un mayor entendimiento del lector a continuación se detallan los siguientes ítems: Definición del problema, objetivos del proyecto, marco teórico, alcance, metodología, desarrollo del proyecto, preparación de datos, implementación y evaluación de los modelos, conclusiones y trabajos futuros.

## 1. DEFINICIÓN DEL PROBLEMA

### 1.1. PLANTEAMIENTO DEL PROBLEMA

Como lo menciona en su página web el Programa de las Naciones Unidas para el Desarrollo (PNUD). Actualmente, existen más de 4000 millones de persona que aún no cuentan con acceso a internet [13], lo cual ha limitado el desarrollo de algunos países, el acceso a la información, el conocimiento y creación de nuevos emprendimientos.

Hoy en día, el sector de las telecomunicaciones se enfoca en desarrollar infraestructuras resilientes y sostenibles que reduzcan la brecha digital, especialmente en zonas rurales y remotas [18]. Priorizar redes de alta capacidad y baja latencia, permiten garantizar servicios accesibles confiables y sostenibles en el tiempo, con el fin, de satisfacer la demanda de conectividad que se requiere en las diferentes zonas del país. Esto implica invertir en nuevas tecnologías, optimizar el uso de recursos y garantizar la redundancia para prevenir fallos. Enfocas en mantener estándares de calidad que aseguren una experiencia uniforme, incluso en áreas remotas, fomentando así la inclusión digital y promoviendo el desarrollo económico y social mediante un acceso equitativo a servicios [1].

Unos de los objetivos principales de las últimas décadas en Colombia ha sido cerrar la brecha digital entre los colombianos e interconectar las diferentes zonas del país, ya que, el sector de las telecomunicaciones ha venido presentando cambios radicales, los cuales han contribuido al desarrollo de actividades diarias de las personas y al crecimiento de diferentes sectores económicos.

Brindar una cobertura amplia y de calidad en zonas de mediana población depende de las dimensiones de diseño de las estaciones base que soportan el tráfico de la tecnología 4G, las cuales son determinadas según la cantidad de habitantes y la disponibilidad de espectro radioeléctrico que permite el tránsito de voz, datos y vídeo [38]. La saturación de uno de estos elementos ya sea por causas naturales, actos conmemorativos o eventos sociales donde se aumenta significativamente la afluencia de público, impactan de forma directa el diseño inicial de la red, lo que lleva a una sobreocupación de la estación base afectando la experiencia de servicio de los residentes de la zona y en algunos casos impactando de manera económica la compañía.

### 1.2. FORMULACIÓN DEL PROBLEMA

Usando la información sobre la saturación de la red de telecomunicaciones en los diferentes municipios de Cundinamarca y haciendo uso de la ciencia de datos, se dio respuesta a los siguientes interrogantes:

¿Se puede predecir la afectación en los servicios móviles de un operador de telecomunicaciones en eventos sociales y culturales realizados en municipios de Cundinamarca usando técnicas de Ciencia de Datos?

#### **Preguntas de Sistematización**

- ¿Qué eventos sociales y culturales se han desarrollado en los municipios de Cundinamarca en el último año?

- ¿Son los eventos sociales o culturales la principal causa de saturación de la red de telecomunicaciones en los municipios intermedios de Cundinamarca?
- ¿Es posible desarrollar un modelo predictivo que permita conocer el tráfico en futuros eventos y así crear planes preventivos que permitan para garantizar la disponibilidad de la red?
- ¿Los resultados obtenidos se alinean con la experiencia que perciben los usuarios en los municipios de Cundinamarca?

## 2. OBJETIVOS DEL PROYECTO

### 2.1. OBJETIVO GENERAL

Desarrollar un modelo predictivo de tráfico en estaciones base con tecnología 4G usando técnicas de ciencias de datos e integrando el impacto de eventos sociales y/o culturales. Con el fin de tener una mejor cuantificación del tráfico por eventos externos en municipios intermedios del departamento de Cundinamarca.

### 2.2. OBJETIVOS ESPECÍFICOS

1. Identificar por medio de web scraping, el desarrollo de eventos sociales y culturales con el fin de integrar los datos.
2. Determinar con procesamiento de lenguaje natural los tipos de evento que impactan la red y, mediante gráficas, analizar los datos para identificar patrones únicos de comportamiento del tráfico de la red para las estaciones.
3. Diseñar un modelo predictivo que permita conocer el tráfico de futuros eventos en los municipios de Cundinamarca.
4. Verificar los resultados obtenidos en el modelo frente a la experiencia del cliente y la información recolectada inicialmente.

### 3. MARCO TEÓRICO

#### 3.1. MARCO TEÓRICO

##### 3.1.1. SECTOR DE LAS TELECOMUNICACIONES

Las telecomunicaciones son un conjunto de sistemas de recepción a distancia que integran componentes físicos y lógicos, capaces de proveer conexión a dos o más puntos de fijos o móviles [7]. Los productos móviles son servicios de voz y datos públicos de cubrimiento nacional que se componen de diferentes generaciones tecnológicas (2G, 3G, 4G y 5G) y permite la comunicación de usuarios en la red [8].

Para poder generar un ecosistema digital en el sector de las comunicaciones móviles, se debe garantizar una infraestructura de red en la cual se relacionen estaciones radioeléctricas o también conocidas como estaciones base, las cuales están compuestas por equipos transmisores, radiantes y/o receptores que se encuentran ubicados en azoteas, torres, zonas urbanas y zonas rurales.

Adicionalmente se debe tener una disponibilidad de espectro radioeléctrico, el cual es el insumo esencial en las telecomunicaciones inalámbricas, pero que a su vez está compartido tanto por la telefonía móvil, radio y televisión abierta [9].

Los proveedores de servicios de telecomunicaciones se encargan de cumplir con los mantenimientos, límites de exposición y garantizar la calidad y cobertura de los servicios adquiridos por los usuarios (voz o datos), para evitar sanciones impuestas por los entes regulatorios.

Una de las principales afectaciones evidenciadas en las estaciones base es la saturación de red, producida por el aumento de actividad o usuarios en la red en un rango de tiempo específico, debido a comportamientos inusuales en la red, las estaciones presentan un aumento en el tráfico donde algunas veces no pueden asimilar ese comportamiento, afectando las peticiones del usuario al realizar una conexión de voz o datos [4].

Según el último informe estadístico del Ministerio de las TIC para el primer trimestre del año 2023 en Colombia, colombianos adquirieron aproximadamente 76,9 millones de líneas móviles y 40,1 millones de colombianos tienen acceso a internet móvil, siendo la cuarta generación (4G) con 34,7 millones de usuarios, seguida por la tercera generación (3G) con 4,9 millones de usuarios y el restante corresponde a los usuarios que usan segunda generación (2G) [5].

Actualmente, en Colombia existen doce empresas proveedoras de servicio móvil pero solo cinco de ellas cuentan con una infraestructura física de red [10]. La Comisión Reguladora de Comunicaciones y el Ministerio de Tecnología de la Información y comunicaciones son algunos de los entes reguladores en Colombia, encargados de garantizar que cada prestador de servicios mencionado cumpla con los requisitos normativos, regulación existente y garantice la calidad del servicio en cada producto adquirido por los clientes.

En resumen, es claro la importancia de entender las causas que impactan la red y afectan a los habitantes de forma directa. Poder dar una guía para la toma de decisiones basado en los datos recopilados tiene grandes beneficios para empresa [11].

### 3.1.2. TECNOLOGÍA LTE (Long Term Evolution)

La tecnología LTE o conocida como la cuarta generación (4G) tiene como objetivo principal proporcionar acceso a internet de alta velocidad tanto de transmisión como de recepción, mejorando notablemente la capacidad y eficiencia de las redes de telefonía móvil.

Esta tecnología se distingue por su alta eficiencia espectral, baja latencia y soporte para alta movilidad, gracias a técnicas avanzadas como OFDMA (Orthogonal Frequency Division Multiple Access), MIMO (Multiple Input Multiple Output) en otras palabras es considerado un avance significativo en la evolución de las redes móviles y la cual permitió una migración eficiente al coexistir con tecnología previas (3G y HSPA) [20].

En la ilustración 1 se muestra la arquitectura de la tecnología LTE, donde sus componentes fundamentales son la red de acceso E-ULTRA y los paquetes EPC de cada una de las troncales y la evolución de los IMS de la tecnología previa (UMTS) [21].

- **E-UTRAN (Evolved Universal Terrestrial Radio Access Network):** Encargada de gestionar la comunicación inalámbrica entre el usuario y la red móvil, incluye las estaciones base (eNodeBs) las cuales son responsables de la transmisión y recepción de datos.
- **EPC (Evolved Packet Core):** Es el núcleo de la red que gestiona el tráfico de datos y facilita la conectividad a otras redes, tramitando los paquetes de datos, conexión a internet y la movilidad de los usuarios entre la red.

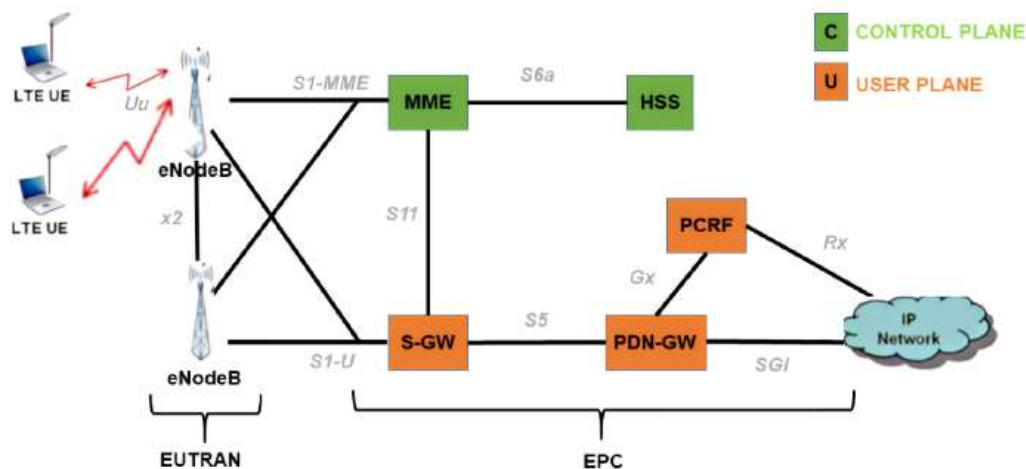


Ilustración 1. Arquitectura de red tecnología LTE [40].

Dentro de las características principales de esta tecnología son:

- **Velocidades de Transmisión:** De acuerdo con la configuración del espectro y la tecnología usada en el dispositivo permite alcanzar velocidades de descarga de hasta 300 Mbps y velocidades de subida de hasta 75 Mbps [22].

- **Baja Latencia:** La latencia comparada con otras generaciones es significativamente menor, lo que permite una experiencia más fluida en aplicaciones en tiempo real, como videojuegos y videoconferencias.
- **Uso Eficiente del Espectro:** Emplea técnicas avanzadas de modulación y multiplexión, como OFDM (Orthogonal Frequency Division Multiplexing), lo que permite un uso más eficiente del espectro radioeléctrico [23].
- **Conectividad Mejorada:** Mejora la cobertura en zonas rurales y facilita una mayor densidad de usuarios en áreas urbanas, ampliando el alcance de los servicios de comunicaciones y generando una mejor experiencia y confiabilidad en los usuarios.

### 3.1.3. TRÁFICO EN REDES DE TELECOMUNICACIONES MÓVILES.

Se hace referencia al flujo de datos y llamadas que se transmiten a través de redes de telecomunicaciones móviles, incluyendo las comunicaciones de voz y el tráfico de datos generado por el uso de aplicaciones, la navegación web, la mensajería y otros servicios en dispositivos móviles.

En Colombia, el tráfico se considera una dimensión de estudio que permite relacionar la cantidad de usuarios con las comunicaciones realizadas los mismos [19]. Por ello, la Comisión de Regulación de Comunicaciones (CRC) y el Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC) han desarrollado políticas y regulaciones que buscan mejorar la calidad del servicio y aumentar la cobertura de red, con el fin de gestionar el tráfico y garantizar una experiencia satisfactoria para los usuarios.

### 3.1.4. REGULACIÓN NACIONAL

Las telecomunicaciones son un aspecto fundamental en el desarrollo social y económico de los países, incluyendo Colombia. A lo largo de los años, el marco legal relacionado a este sector ha ido evolucionado y adaptándose a los avances tecnológicos y necesidades de la población. Un ejemplo de esto ha sido la **Ley 1341 de 2009** que establece el marco regulatorio para el uso y desarrollo de las telecomunicaciones, como el acceso, cobertura y calidad en los servicios en Colombia [32]. Además de la **Ley 1450 de 2011** y la **Ley 1978 de 2019** las cuales han reforzado la importancia de garantizar una infraestructura de red robusta que permita a todos los ciudadanos acceder a servicios de telecomunicaciones de calidad [33] [34].

Por otro lado, la Comisión de Regulación de Comunicaciones en Colombia emite las siguientes resoluciones que permite hacer seguimiento y evaluar el comportamiento a la calidad de la red móvil en Colombia:

- **Resolución 3066 de 2011 (CRC):** Define los parámetros de calidad del servicio que los operadores de telecomunicaciones deben cumplir [35].
- **Resolución 421 de 2015:** Busca garantizar que las redes móviles funcionen de manera óptima, modificando y complementa la regulación sobre calidad de servicio, estableciendo indicadores específicos para conocer el comportamiento de la red [36].
- **Resolución 124 de 2020 (CRC)** [37]: Establece y facilita la medición de la calidad del servicio integrando aspectos técnicos y procedimientos de control, asegurando que los operadores se mantengan en el margen de cumplimiento de acuerdo con los estándares

establecidos.

Finalmente, se encuentra el Plan Nacional de Desarrollo que incluye políticas orientadas a mejorar la infraestructura y garantizar la cobertura en todo el país [39] y junto al Reglamento Técnico de Calidad de Servicio [38] que define los diferentes estándares de calidad que deben cumplir los operadores en aspectos de disponibilidad y velocidad tanto de carga como de descarga de información [40], permiten generar un desarrollo equitativo en la prestación del servicio garantizando una calidad adecuada y la inclusión de zonas remotas del país.

### 3.1.5. PAYLOAD O TRÁFICO DE RED

El payload se refiere a los datos que se transmiten en uno paquetes útiles para el usuario final, excluyendo los datos de control o la sobrecarga de protocolo. Es esencialmente la información efectiva que se desea comunicar [24].

### 3.1.6. TASA DE TRANSFERENCIA EFECTIVA

La tasa de transferencia efectiva o Throughput es una medida del rendimiento de la red que indica la cantidad de datos transmitidos en un periodo de tiempo específico, la cual se expresa comúnmente en bits por segundo (bps) y puede su rendimiento puede verse afectado por diferentes factores como:

- Las condiciones de la red es un factor fundamental, ya que, la interferencia y congestión en la red pueden reducir la cantidad de datos que se transmiten.
- La retransmisión de datos por diferentes errores que pueden presentarse en la red.
- El tiempo que tarda un paquete de datos en viajar de un punto a otro, afectando la experiencia del usuario y/o la cantidad de datos que pueden enviar en un tiempo determinado.

### 3.1.7. CATEGORIZACIÓN DANE

De acuerdo con el Artículo 6 de la Ley 136 de 1994 que menciona que “*Los distritos y municipios se clasificarán atendiendo su población, ingresos corrientes de libre destinación y situación geográfica. Para efectos de lo previsto en la ley y las demás normas que expresamente lo dispongan, las categorías serán las siguientes:*” [7]

#### 3.1.7.1. Grandes Municipios (Primer Grupo)

##### Categoría Especial

Municipios con una población igual o superior a quinientos mil uno habitantes (500.001) y cuyo ingreso corriente de libre destinación anual sea superior a cuatrocientos mil SMLV (400.000).

##### Primera Categoría

Municipios con un rango de población superior a ciento un mil uno (101.001) hasta quinientos mil habitantes (500.000), cuyo ingreso corriente de libre destinación anual es superior a cien mil (100.000) hasta cuatrocientos mil (400.000) SMLV.

#### 3.1.7.2. Municipios Intermedio (Segundo Grupo)

##### Segunda Categoría

Municipios con un rango de población entre cincuenta mil uno (50.001) y cien mil habitantes (100.000) y cuyo ingreso corriente de libre destinación anual es superior a cincuenta mil (50.000) hasta cien mil (100.000) SMLV.

#### Tercera Categoría

Municipios con un rango de población entre treinta mil uno (30.001) y cincuenta mil habitantes (50.000) y cuyo ingreso corriente de libre destinación anual es superior a treinta mil (30.000) hasta cincuenta mil (50.000) SMLV.

#### Cuarta Categoría

Municipios con un rango de población entre veinte mil uno (20.001) y treinta mil habitantes (30.000) y cuyo ingreso corriente de libre destinación anual es superior a veinticinco mil (25.000) hasta treinta mil (30.000) SMLV.

#### 3.1.7.3. Municipios Básicos (Tercer Grupo)

#### Quinta Categoría

Municipios con un rango de población entre diez mil uno (10.001) y veinte mil habitantes (20.000) y cuyo ingreso corriente de libre destinación anual es superior a quince mil (15.000) hasta veinticinco mil (25.000) SMLV.

#### Sexta Categoría

Municipios cuya población es inferior a diez mil (10.000) habitantes y cuyo ingreso corriente de libre destinación anual no es superior a quince mil (15.000) SMLV.

#### 3.1.8. WEB SCRAPING

Técnica que consiste en extraer o raspar información de sitios web de manera automática haciendo uso de scripts o programas que permitan navegar por las páginas y extraer datos específicos, como textos, imágenes o enlaces. Esta técnica puede aplicarse en múltiples sectores y generar y generar diferentes beneficios, ya que facilita el acceso a grandes volúmenes de información en poco tiempo, automatizar tareas repetitivas, disminuir tiempos de ejecución y mejorar la calidad de los productos realizados y/o analizados [22].

#### 3.1.9. PROCESAMIENTO DE LENGUAJE NATURAL

El procesamiento de Lenguaje Natural o PLN es la combinación del lenguaje humano basado en reglas con aprendizaje automático y modelos estadísticos, con el objetivo de que los sistemas computacionales y dispositivos digitales generen un resultado ya sea de texto y/o voz [27].

Dentro de las actividades más relevantes que se pueden desarrollar con el PLN son:

- Respuesta a comandos de texto y voz.
- Reducción a grandes volúmenes de datos.
- Evaluar la intención o el sentimiento expresado en un texto o voz.
- Generar texto, gráficas, entre otro contenido a demanda.

Dentro de las tareas más relevantes de este procesamiento es el Etiquetado de los fragmentos de una oración, el cual permite determinar el papel que desempeña una palabra o fragmento particular de un texto en función del uso y contexto, ya sea como verbo o sustantivo dentro de la oración.

### 3.1.10. MODELO DE DATOS

Entiéndase como la como la relación entre variables y conjunto de datos basado en conceptos, reglas y convenciones que permite analizar e identificar comportamientos, patrones y restricciones de una población a partir de muestras de información.

#### 3.1.10.1. FACEBOOK PROPHET

Es una herramienta para el modelado de series de tiempo desarrollado por Facebook. Este modelo es utilizado para predecir series temporales univariadas y es eficaz en datos con tendencias estacionales y sucesos recurrentes, muy útil para la aplicación en entornos empresariales [25]

#### Características Principales

- Modelo aditivo Prophet define la serie en la descomposición en tendencia, estacionalidad y eventos. La tendencia modela los cambios a largo plazo, la estacionalidad captura los patrones semanales diario, y eventos como festivos que pueden afectar la serie de tiempo.
- Robustez a Falta de Datos y Outliers: Está diseñado para el manejo de datos faltantes y valores atípicos.
- Ajustable: Los usuarios pueden ajustar manualmente los días festivos o la flexibilidad de la tendencia, lo que permite una personalización que se alinea con el conocimiento específico del dominio del usuario.
- Facilidad de uso: Este modelo está diseñado para un fácil uso y aplicación, incluye una sintaxis sencilla y la capacidad de generar pronósticos sin tanta configuración inicial
- Diagnóstico y Validación: El modelo incluye funcionalidades para evaluar el modelo y la precisión del proceso permitiendo a los usuarios realizar validación cruzada y ajustar el modelo.

Su arquitectura esta basada en extrapolar las predicciones en intervalos de confianza como lo muestra la ilustración 2

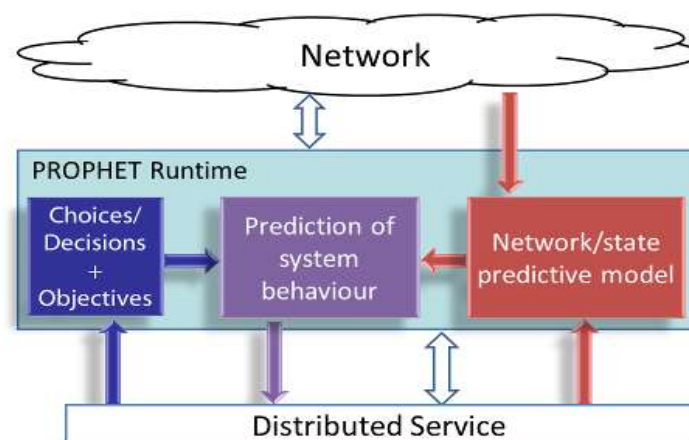


Ilustración 2 Arquitectura modelo Facebook Prophet [43]

### 3.1.10.2. REDES NEURONALES

Son modelos computacionales inspirados en el funcionamiento del cerebro humano, diseñados para reconocer patrones, aprender, clasificar y optimizar datos. Las Redes Neuronales consisten en capas de nodos interconectados entre sí, donde cada conexión tiene un peso que se ajusta durante el proceso de aprendizaje. Las redes neuronales son ampliamente utilizadas en tareas como la clasificación, regresión y reconocimiento de imágenes, entre otros [26].

#### 3.1.9.3.1 RED NEURONAL ELMAN

Las Redes Neuronales de Elman, también llamadas redes recurrentes de Elman, con un tipo de red neuronal recurrente (RNN) que incluye una capa de contexto para almacenar la información de los estados anterior (Ilustración 3). Este diseño permite modelar datos secuenciales y temporales, capturando dependencias a lo largo del tiempo [28].

#### Estructura de la red de Elman

- Capa de Entrada: Recibe el vector de entrada
- Capa Oculta: Procesa información, aplicando una función de activación
- Capa de contexto: Almacena los estados anteriores de la capa oculta y los retroalimenta a la capa oculta en el siguiente paso del tiempo.
- Capa de salida: produce el resultado de la red

#### Función de Activación

Las funciones de activación comunes utilizadas en las capas ocultas incluyen:

- Función sigmoide.
- Función tangente hiperbólica (tanh).
- Función ReLU (Rectified Linear Unit), aunque esta es más común en arquitecturas más modernas.

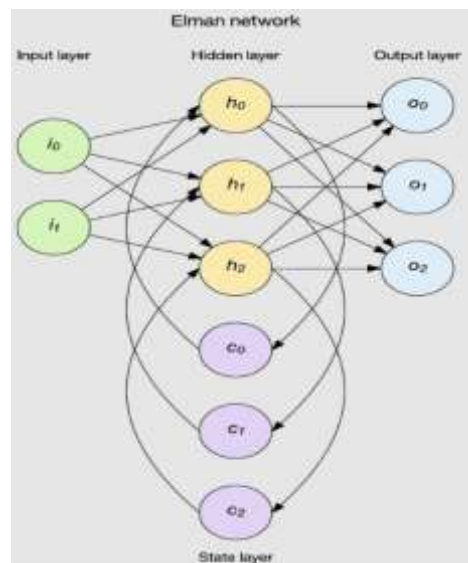


Ilustración 3 Arquitectura red de Elman [41]

### Entrenamiento de la Red de Elman

El entrenamiento de una red de Elman se realiza mediante el algoritmo de retro propagación a través del tiempo (Backpropagation Through Time, BPTT), el cual es una extensión del algoritmo de retro propagación estándar para redes feedforward. BPTT ajusta los pesos de la red considerando la secuencia completa de entradas y salidas.

#### 3.1.9.3.2 RED NEURONAL JORDAN

Las redes neuronales de Jordan también conocidas como Redes Recurrentes de Jordan, son un tipo de red neuronal recurrente (RNN) que similar a las redes de Elman incorporan retroalimentación, la diferencia es que la retroalimentación proviene de la capa de salida (Ilustración 4). Esta arquitectura permite que las redes neuronales manejen una información secuencial y temporal lo que se hace útil para el modelado de series de tiempo y lenguaje natural [29].

#### Estructura de la red de Jordan

- Capa de Entrada: Recibe el vector de entrada
- Capa Oculta: Procesa la información aplicando una función de activación
- Capa de Estado (Contexto): Almacena los estados de la capa de salida anterior y los retroalimenta a la capa oculta en el siguiente paso del tiempo
- Capa de Salida: Produce el resultado de la red.

#### Función de Activación

Al igual que en las redes de Elman, las funciones de activación comunes para la capa oculta incluyen:

- Función sigmoide.
- Función tangente hiperbólica (tanh).
- Función ReLU (Rectified Linear Unit).

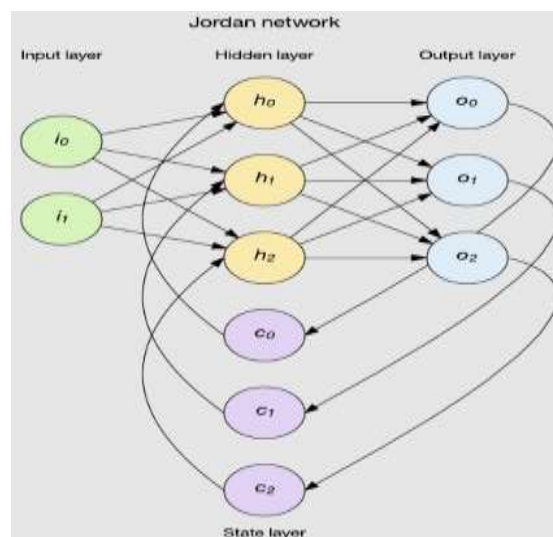


Ilustración 4 Arquitectura Red de Jordan [41]

## Entrenamiento de la red de Jordan

El entrenamiento de una red de Jordan también se realiza mediante el algoritmo de retro propagación a través del tiempo (Backpropagation Through Time, BPTT), ajustando los pesos de la red en función de la secuencia completa de entradas y salidas.

### 3.1.10.3. *NEURALPROPHET*

Modelo de predicción de series de tiempo, es una extensión del modelo Prophet desarrollado por Facebook. Mientras que Prophet es ampliamente utilizado por su facilidad y simplicidad, además su capacidad para manejar los componentes de tendencia, estacionalidad y eventos especiales, el NeuralProphet introduce una mejora significativa al incorporar redes neuronales, lo que permite realizar modelamiento más complejos y no lineales sobre las series temporales.

Este modelo se basa en la estructura descomponible de Prophet pero incorpora redes neuronales para manejar tendencias más complejas, estacionalidades y efectos de días festivos o eventos espaciales de forma no lineal (Ilustración 5). El modelo está diseñado para ser interpretable, escalable y fácil de usar, similar a Prophet, pero con mejoras en la precisión y capacidad para capturar patrones complejos.[30]

#### Características:

- **Modelo Híbrido:** Neural-Prophet combina la estructura modular y posee la facilidad de implementación del modelo Prophet con la capacidad de aprender patrones no lineales
- **Tendencia Flexible y Estacionales.** al igual Prophet incluye componentes como tendencia (Lineal y no lineal) y estacionalidad. La diferencia es que Neural-Prophet es que los componentes sean modelados a través de redes neuronales lo que representa una mayor flexibilidad en capturar patrones
- **Capacidad de Múltiples Temporalidades y Regresores:** Neural-Prophet maneja estacionalidades múltiples como la diaria y la anual. Además, puede incluir regresores (variables externas) que puedan afectar la serie temporal
- **Modelado de eventos espaciales:** Igual que el modelo Prophet, permite modelar el impacto de eventos espaciales sobre las predicciones. Estos eventos se pueden incluir como regresores adicionales, permitiendo que el modelo ajuste las predicciones en torno a algunos periodos específicos
- **Mejora Interpretación:** Los componentes del modelo como son tendencia estacionalidad y efectos regresores son interpretables y se pueden descomponer para que los usuarios puedan entender como cada componente contribuye a la predicción.
- **Ajuste Automático:** El modelo NeuralProphet incluye ajuste automático de los hiper parámetros, facilitando la optimización si necesidad de tener conocimientos profundos sobre ajustes de redes neuronales.

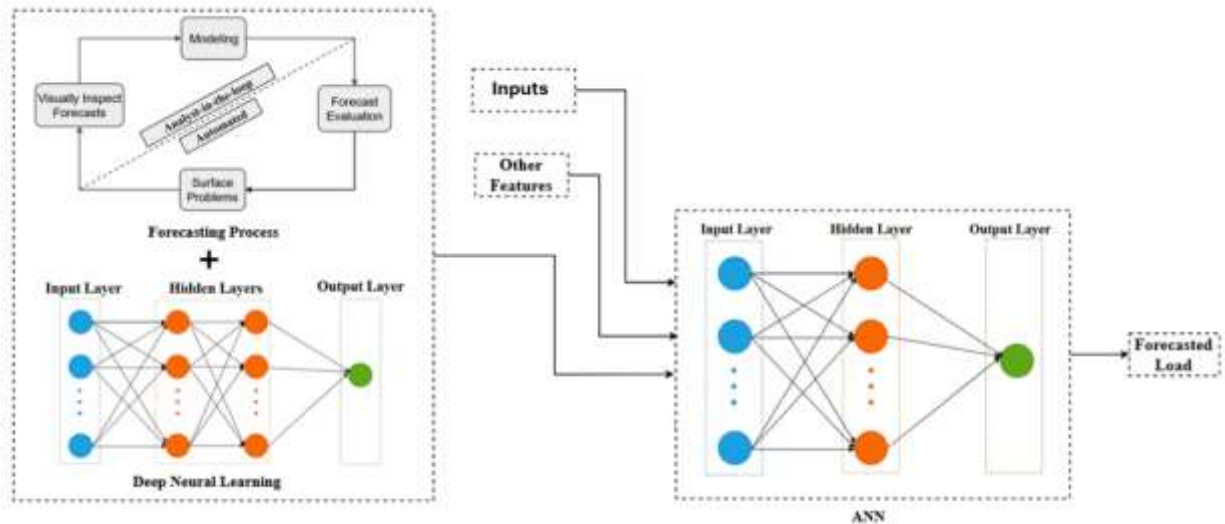


Ilustración 5 Arquitectura modelo NeuralProphet [42]

El Modelo NeuralProphet es usado en casos de uso donde se requiere alta precisión y donde los datos tienen patrones no lineales y complejos, por ejemplo, en pronósticos financieros, pronósticos de la demanda y análisis de series de tiempo con eventos irregulares

### 3.2. ANTECEDENTES

#### 3.2.1. PREDICCIÓN DE COMPORTAMIENTO EN TRÁFICO DE RED LTE Y AJUSTE DE PARAMETRIZACIÓN PARA MAXIMIZAR PERFORMANCE DE RED

Como primer antecedente se presenta el proyecto titulado “Predicción De comportamiento en tráfico de red LTE y ajuste de parametrización para maximizar performance de red”. En la propuesta el autor realizó un estudio a diferentes KPIs que permiten determinar que la naturaleza del indicador no se encuentra relacionada con el throughput (Tasa Efectiva de Transferencia de Datos), esto con el fin de crear correlaciones y pruebas de rendimiento en la red.

Para llevar a cabo la predicción de las variables en el modelo, se implementó series de tiempo asociado con las variables de la red, el objetivo era a partir de un análisis descriptivo de los datos ver su comportamiento de forma diaria; el modelo incluyó clasificadores los cuales generaban como resultado el valor futuro del throughput de cada una de las celdas, adicionalmente, se hizo uso de clústeres por colindancia y por características similares como por ejemplo un bajo throughput con la finalidad de realizar una regresión sobre las métricas de evaluación y así conocer la tasa de aciertos y error que genera el modelo.

Como resultado del modelo se concluyó que con una tasa del 91% de aciertos se logró predecir el valor de throughput con base al historial de los indicadores seleccionados para los sitios con tecnología LTE. Adicionalmente se clasificaron (alto o bajo) las celdas de la red de acuerdo con su rendimiento en el throughput [14].

### 3.2.2. STOCK PRICE PREDICTION USING NEWS SENTIMENT ANALYSIS

En el segundo antecedente se presenta como el mercado bursátil siempre ha tenido una naturaleza muy volátil y como estos pueden afectar en las personas. Mencionan en el documento que los modelos toman como insumo los valores históricos de los precios de las acciones, que al hacer una predicción no son muy acertados, en esta propuesta se incluyeron como entrada adicional noticias financieras para analizar sentimientos y demostrar una entrada más fuerte al predecir mediante la polaridad de los sentimientos. Del mismo modo se resalta la variabilidad del tráfico y el manejo en el análisis de noticias sobre eventos relevantes a través de tweets para contribuir a predecir patrones de tráfico con mayor precisión.

En la investigación se enfocaron en dos fuentes principales, los precios de las acciones de cierre de forma diaria (2013-2017), y en las noticias financiera nacionales e internaciones en esta misma temporalidad. Como es norma de todos los modelos de investigación limpiaron datos y en metodología usaron MAPE (Error porcentual absoluto medio), una medida de precisión de predicción del modelo.

En la fase experimental usaron varias metodologías donde tomaban la base del precio de las acciones y la polaridad textual extraída de las noticias, una de las utilizadas fue ARIMA para el análisis de series de tiempo, Facebook Prophet el cual es un modelo de regresión en una serie temporal de variable única, RNN LSTM Red Neuronal recurrente el cual no pone restricciones.

Como conclusiones lograron predecir el precio de las acciones usando en el modelo las entradas de las noticias financiera, demostrando así una fuerte relación, de los resultados más relevantes se dieron con el modelo RNN (Modelo de Red Neuronal Recurrente) logrando mayor confiabilidad al momento de relacionar el sentir de los inversionistas versus el comportamiento que se tienen las inversiones [15].

### 3.2.3. DETECCIÓN DE PATRONES DE COMPORTAMIENTO DE PARÁMETROS DE RF EN REDES DE COMUNICACIÓN MÓVIL MEDIANTE MEDICIONES DE CAMPO Y TÉCNICAS DE MACHINE LEARNING.

El cual se enfocó en un estudio de la relación que existe entre la velocidad de conexión y la velocidad que tiene un teléfono móvil al hacer uno de la red, en donde se usaron técnicas de Machine Learning delimitado en el centro histórico de la ciudad de Quito.

El objetivo de la investigación fue soportar si la hipótesis de que existe una relación directa entre la calidad de conexión y la velocidad del teléfono móvil era cierta. Para realizar el análisis correspondiente y confirmar lo mencionado anteriormente, se desarrolló un modelo el cual se dividía en diferentes fases, la fase recolección de datos utilizaron variables técnicas de la red que suministraba el teléfono como son: RSRP (Reference Signal Received power) RSRQ (Reference Signal Received Quality) SNR (Signal – to-noise ratio) y RSSI( Received Signal Strength Indicator), estas fueron definidas como parámetros de entrada y los parámetros de salida fue la velocidad de conexión que en este caso sería categórica , “Mala” si era menor a 1000kps, “Regular” si estaba en este rango (1000kbps-5000kbps) y “Excelente” si era mayor a 5000 kbps ,estas pruebas se realizaron en tres rutas del centro de Quito. Cabe mencionar que dentro de las técnicas de Machine Learning utilizaron aprendizaje supervisado usando arboles de decisión, donde con los parámetros de entrada antes mencionados predijeron por medio de un campo categórico la calidad de conexión.

Como conclusiones propusieron la metodología de recolección de datos para analizar la degradación en las velocidades de descarga para futuros trabajos de causa-raíz. Otra conclusión es la falta de relación

entre la calidad de la conexión y la velocidad del teléfono. Sin embargo, sí se encontraron unas constantes donde los parámetros de entrada garantizaban la calidad de conexión [16].

#### 3.2.4. ANÁLISIS DE SENTIMIENTOS DE DATOS DE REDES SOCIALES USANDO TÉCNICAS DE MACHINE LEARNING

Este proyecto muestra la relevancia del aprendizaje supervisado donde se entrena un algoritmo con un conjunto de etiquetas, por medio de los clasificadores que predice la categoría de los datos de entrada.

Dentro de la fase de recolección de datos, las descargas se realizaron por medio del código desarrollado en Python, así como la API de Twitter en múltiples idiomas, a través de la cual también realizaba la traducción en varios idiomas. Luego se utilizaron unos tweets subyacentes para clasificarlos manualmente como positivos o negativos. Para el diagnóstico de polaridad de sentimientos se utiliza un “Clasificador Naive Bayes” de probabilidad, este modelo utiliza un diccionario de frecuencia basado en los tweets clasificados y en base a encontrar una connotación de la palabra y se debe enfatizar en la importancia de limpiar los datos.

En las conclusiones del artículo esta la ventaja de poder tener un clasificador en texto no estructurado que se soporta con la connotación de tweet aportando información relevante al investigador. Otra barrera derrumbada fue el idioma de los tweets, ya que con traducción de api en Python enriquecían la base de datos [17].

## 4. METODOLOGÍA

Para dar cumplimiento a los objetivos planteados, se toma como base la metodología CRISP-DM (Cross Industry Standard Process for Data Mining) reconocido por la efectividad que tiene en los proyectos en ciencia de datos y el cual está compuesto por 6 etapas, las cuales se detallan a continuación y en la Ilustración 6:

- **Etapa 1. Comprensión del Negocio**

En la primera, se plantean los objetivos y se identifica el detalle al que se quiere llegar, los sistemas a usar en el desarrollo del proyecto y los posibles modelos que podían llevarse a cabo para optimizar los resultados.

- **Etapa 2. Comprensión de los Datos**

Posteriormente, se procede a identificar la relación que existe entre los datos que se van a implementar (Fechas de los eventos, información de la red móvil y clasificación del DANE), las variables acordes a la selección, el acceso a las fuentes de datos y a nivel técnico validar el comportamiento inicial del tráfico en las Estaciones Base con el fin de identificar patrones o irregularidades dentro de los datos.

- **Etapa 3. Preparación de los Datos**

Se desarrolla el proceso de Web Scraping y procesamiento de Lenguaje Natural, posteriormente se procede a evaluar las fuentes de información disponibles, realizar la limpieza de los datos y garantizar la calidad de los mismos. Siguiendo a esto se integran las fuentes de información y se realiza la preparación de la información para el entrenamiento de los modelos de datos.

- **Etapa 4. Modelado**

Una vez se tiene definido los parámetros que se van a analizar y garantizando la fuente de datos, se inicia seleccionando y evaluando los diferentes modelos para realizar la proyección del tráfico en futuros eventos, esto con el fin de seleccionar el modelo más apropiados acorde a las selecciones realizadas inicialmente.

- **Etapa 5. Evaluación**

Se procede a identificar posibles limitaciones o mejoras que se puedan implementar dentro del modelo y así evaluar la precisión y eficacia que se tiene.

- **Etapa 6. Verificación**

Para finalmente se realiza el despliegue del modelo desarrollado obteniendo un ecosistema analítico, que a través de tres pasos sencillos ingreso de la información, procesamiento y respuesta, permita conocer el tráfico en futuros eventos y tomar acciones preventivas en esas ocasiones, mejorando la experiencia del usuario.

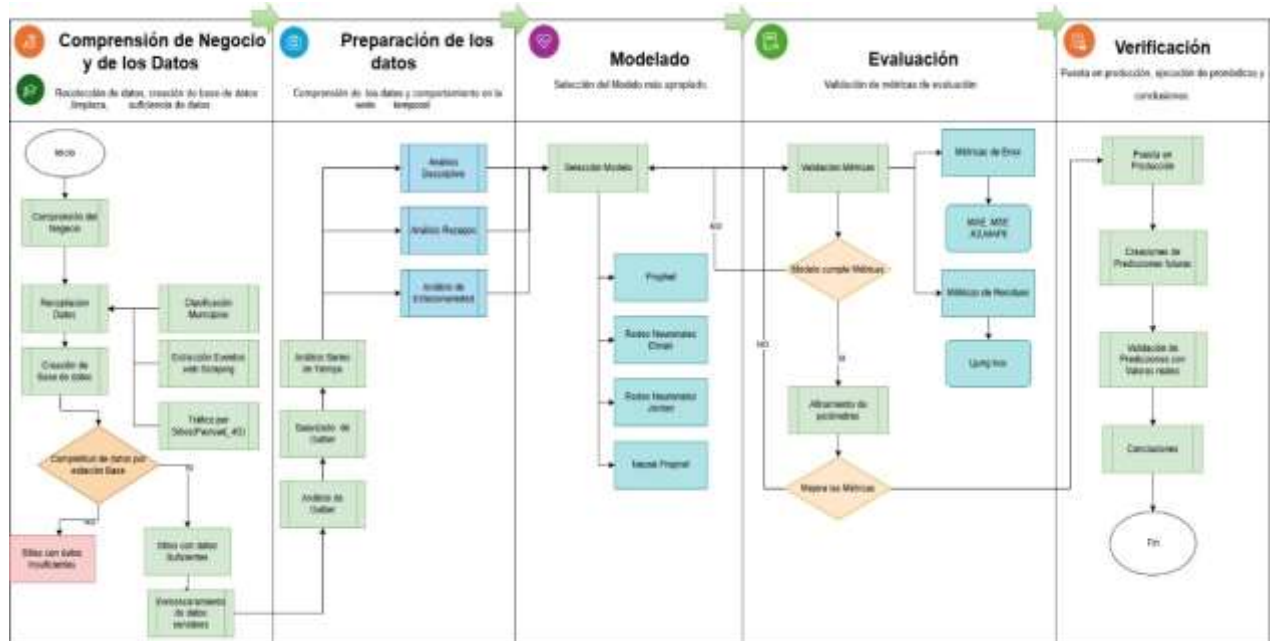


Ilustración 6 Metodología Implementada en el Proyecto (Autores)

## 5. DESARROLLO DEL PROYECTO

Con el propósito de comprender el comportamiento del servicio en las zonas intermedias del departamento de Cundinamarca, se desarrolló un modelo de datos que permite identificar de manera clara el tráfico, lo que facilita la toma de decisiones en eventos futuros. El objetivo es garantizar la calidad del servicio al cliente, mejorar su experiencia y prevenir problemas como mala publicidad, deserción y pérdida de ingresos económicos para la compañía de telecomunicaciones.

El proyecto se basa en diversas fuentes de información que se han procesado para asegurar la confiabilidad de los datos presentados en la herramienta. Esto se hace con el objetivo de garantizar la seguridad de la información y cumplir con las políticas de seguridad de la compañía.

### 5.1. IDENTIFICACIÓN DE DATOS

Dando cumplimiento a la fase 1 de “Comprensión del Negocio” y fase 2 “Comprensión de los datos” se despliegan las siguientes actividades.

#### 5.1.1. CLASIFICACIÓN DE MUNICIPIOS

Para obtener la información correspondiente a clasificación de los municipios a nivel nacional, se hace uso de la base consolidada y publicada por la Contaduría General de la Nación, la cual cuenta con la siguiente estructura (Tabla 1) y permite identificar los municipios Intermedios a partir de la cantidad de habitantes en cada uno de los municipios.

*Tabla 1 Estructura de Datos - Clasificación Municipio (Autores).*

ID	ETIQUETA	DESCRIPCIÓN	TIPO	FORMATO
D001	Código CGN	Código Contaduría General Nacional	DISCRETE	NUMBER
D002	Nombre	Nombre del Municipio	DISCRETE	CHARACTER
D003	Departamento	Nombre del Departamento	DISCRETE	CHARACTER
D004	Poblacion_DANE	Cantidad de Población por Municipio	DISCRETE	CHARACTER
D005	ICLD	Ingresos Corrientes de Libre Destino	CONTIN	NUMERIC
D006	Gastos Funcionamiento	Gastos de Funcionamiento por Municipio	CONTIN	NUMERIC
D007	% Gastos	Porcentaje de Gastos por Municipio	CONTIN	NUMERIC
D008	Categoría	Categoría de Clasificación	DISCRETE	NUMERIC

#### 5.1.2. DETALLE DE TRÁFICO

La Tabla 2, corresponde a las variables de negocio asociadas al tráfico de la red por cada estación base a nivel nacional, estos datos se encuentran desagregados por las Estaciones Base instaladas en cada uno de los municipios y maneja un histórico de información del año 2021 y 2024.

Tabla 2 Estructura de Datos - Tráfico de Red (Autores).

ID	ETIQUETA	DESCRIPCIÓN	TIPO	FORMATO
T001	SITIO	Nombre de la estación Base	DISCRETE	CHARACTER
T002	DEPTO	Nombre Departamento	DISCRETE	CHARACTER
T003	MUNICIPIO	Nombre Municipio	DISCRETE	CHARACTER
T004	DANE	Código del Municipio	DISCRETE	CHARACTER
T005	TIPO	Tipo de estación Base	DISCRETE	CHARACTER
T006	OBJETIVOCOBERTURA	Clasificación de la Estación Base	DISCRETE	CHARACTER
T007	MARKET	Tipo de Mercado Departamental	DISCRETE	CHARACTER
T008	FECHA	Fecha de la medición	DISCRETE	DATE
T009	LONGITUD	Coordenada Longitud	DISCRETE	NUMERIC
T010	LATITUD	Coordenada Latitud	DISCRETE	NUMERIC
T011	M_FECHA	Primer día del mes que se realiza la medición	DISCRETE	DATE
T012	VAL_SATURACION	Valor de Saturación	CONTIN	NUMERIC
T013	SATURACION	Categorizado como "Saturado" o "Normal"	CONTIN	CHARACTER
T014	SCTP_S1_SUCC	Latencia del Core	CONTIN	NUMERIC
T015	CAT_PORC_MUESTRAS	Categoría del valor de saturación	DISCRETE	CHARACTER
T016	THROUGHPUT_4G	Velocidad de transmisión	CONTIN	NUMERIC
T017	PAYLOAD_4G	Carga útil dado en bits	CONTIN	NUMERIC
T018	INCUMPLE_SATURACION	Incumplimiento en saturación	DISCRETE	BOOLEAN
T019	INCUMPLE_PAYLOAD_4G	Incumplimiento en el Tráfico.	DISCRETE	BOOLEAN

## 5.2. EXTRACCIÓN DE EVENTOS POR MEDIO DE WEB SCRAPING.

Para realizar la recolección de los eventos sociales y culturales que se han realizado en los municipios intermedios de Cundinamarca. Inicialmente, se intentó implementar técnicas de web scraping en las redes sociales del grupo Meta (Facebook e Instagram) y en la plataforma X (anteriormente Twitter). Sin embargo, los cambios que se han venido presentando en las políticas de estas plataformas han restringido el acceso a datos mediante scraping. Por lo cual, para llevar a cabo esta actividad, es necesario solicitar un usuario desarrollador a cada plataforma, el cual se solicitó y hasta la fecha no hemos recibido respuesta.

De acuerdo con lo mencionado anteriormente, se realizó un análisis de que tan viables era realizar la extracción de dicha información directamente en las páginas web de los municipios. Donde se identificó que no existe una estandarización en la estructura y navegación de los sitios web. Cada municipio presenta un diseño diferente, lo que complica la automatización del proceso de recolección de datos, ya que se debía desarrollar prácticamente un algoritmo específico para cada municipio, lo que incrementaría los tiempos de desarrollo del proyecto y perdería el objetivo de ser una extracción automatizada.

Finalmente, nos comunicamos con el Instituto Departamental de Cultura y Turismo de Cundinamarca (IDECUT) que es la entidad responsable de la promoción y organización de actividades culturales, solicitando un listado de los eventos programados en dichos municipios y así poder poblar la información en la base de datos.

Haciendo uso de la librería de Python FuzzyUzzy, se realiza la búsqueda de los eventos a partir de la coincidencia y similitud de las palabras que se ingresen, esto con el fin de identificar el tipo de evento que se realizan como el rango de fechas en el que se lleva a cabo y así limitar los rangos de análisis dentro del modelo.

En la Tabla 3 se presenta la estructura que se manejó para dicha información.

*Tabla 3 Estructura de Datos – Festividades (Autores).*

ID	ETIQUETA	DESCRIPCIÓN	TIPO	FORMATO
E001	Municipio	Código Contaduría General Nacional	DISCRETE	NUMBER
E002	Evento	Nombre del Municipio	DISCRETE	CHARACTER
E003	Fecha de Inicio	Nombre del Departamento	DISCRETE	CHARACTER
E004	Fecha de Finalización	Cantidad de Población por Municipio	DISCRETE	CHARACTER

## 5.3. COMPORTAMIENTO DE TRÁFICO INICIAL.

Una vez que se cuentan con las diferentes fuentes de información, se procede a unificarlas por medio de una base de datos estructurada Oracle, utilizando como llave primaria el Municipio, con el fin de garantizar la completitud de los datos.

Como paso a seguir se toma como set de entrenamiento el 80% de la información y se seleccionan las

variables que se tendrán en cuenta en la entrada de los modelos.

Tabla 4 Variables de entrada a los modelos (Autores).

ETIQUETA	DESCRIPCIÓN	TIPO	FORMATO
Fecha Diaria	Fecha diaria en la que se toma el tráfico	DISCRETE	NUMBER
Sitio	Estación Base donde se realiza la toma de datos	DISCRETE	CHARACTER
Payload	Datos transmitidos por Estación Base	DISCRETE	CHARACTER
LS Payload	Dos desviaciones estándar superior en Payload	DISCRETE	CHARACTER
LI Payload	Dos desviaciones estándar inferior en Payload	DISCRETE	CHARACTER
Festividad	Identificador para identificar si se celebró una festividad o no	DISCRETE	CHARACTER

Posteriormente, se procede a garantizar la calidad de los datos (Formatos y caracteres especiales) y así realizar una visualización que permita conocer el comportamiento del tráfico en el municipio seleccionado en los últimos años.

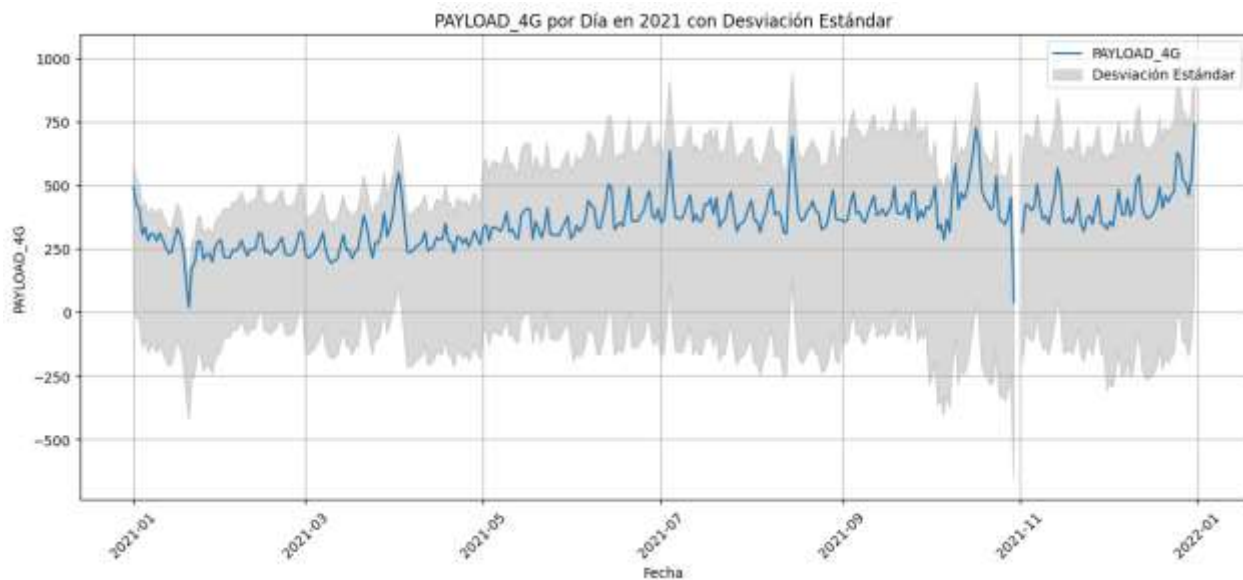


Ilustración 7 Comportamiento del tráfico año 2021. En línea azul el promedio diario y en gris la desviación estándar (Autores).

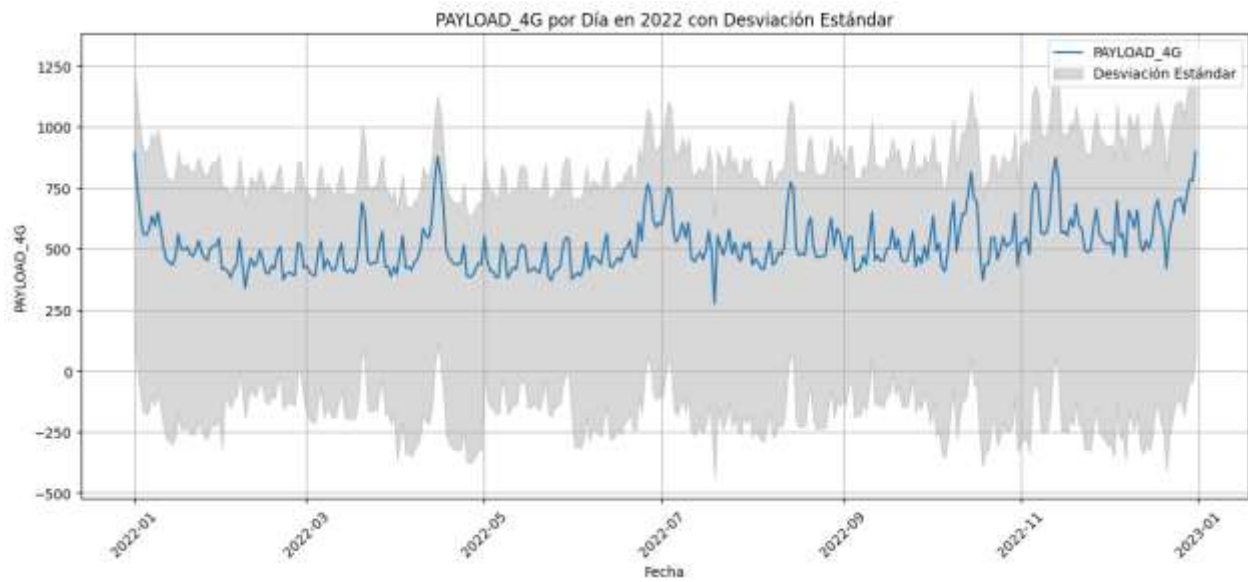


Ilustración 8 Comportamiento del tráfico año 2022. En línea azul el promedio diario y en gris la desviación estándar (Autores).

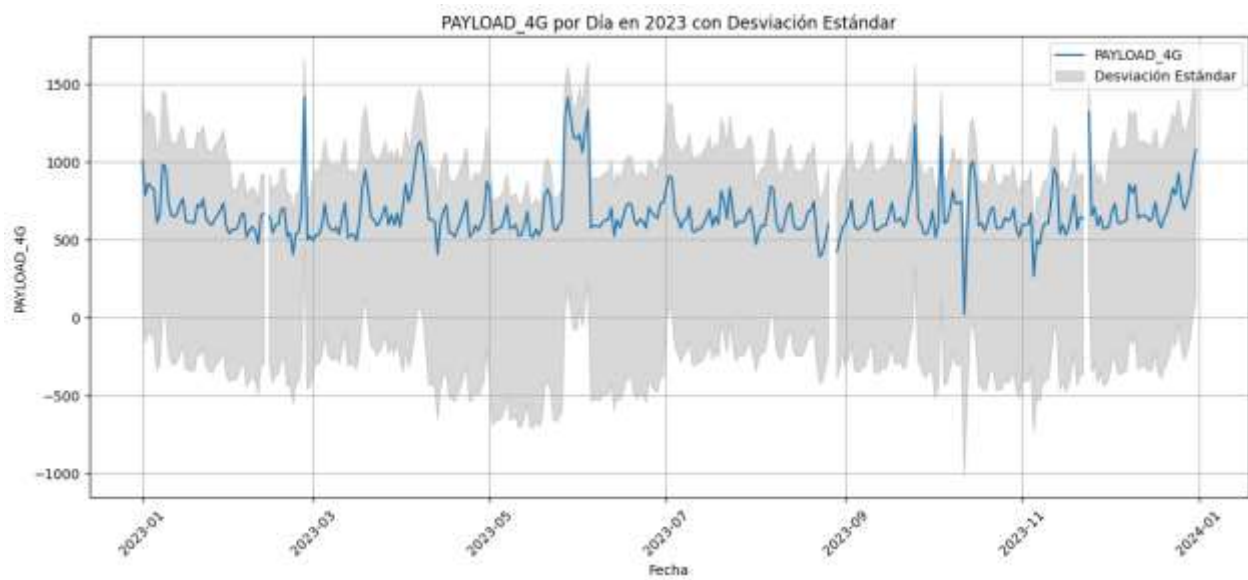


Ilustración 9 Comportamiento del tráfico año 2023. En línea azul el promedio diario y en gris la desviación estándar (Autores).

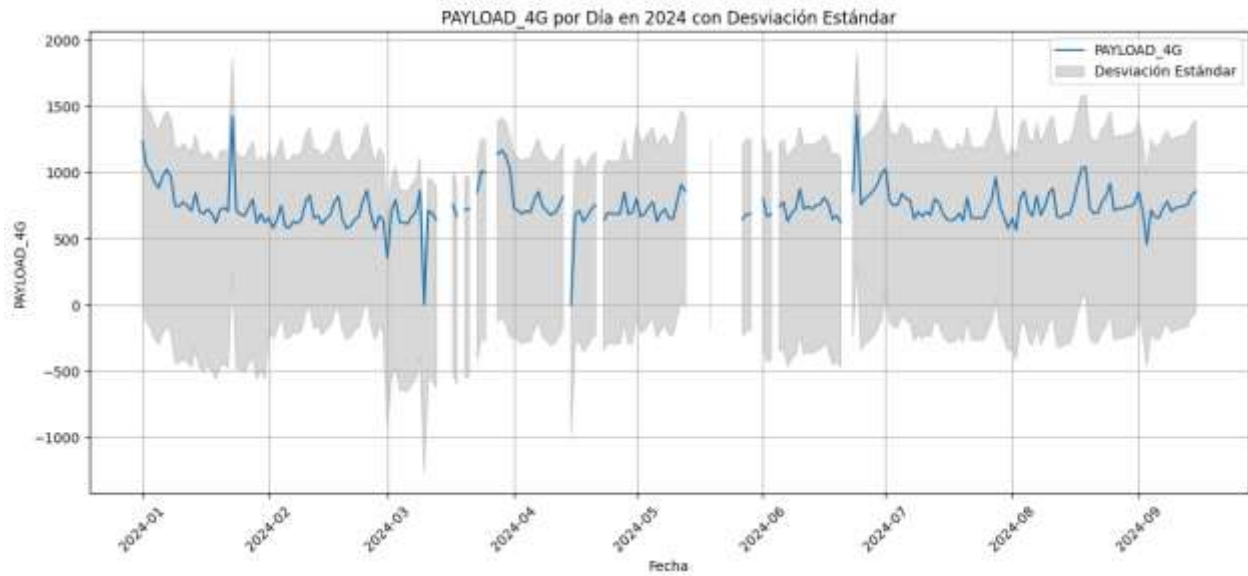


Ilustración 10 Comportamiento del tráfico año 2024. En línea azul el promedio diario y en gris la desviación estándar (Autores).

En las Ilustraciones del número 8 a la 10 se presenta el comportamiento del tráfico en los últimos cuatro años, donde observamos una tendencia creciente y picos de tráfico (línea azul) cuando existen eventos o festividades. De igual manera, la variabilidad del tráfico es superior (área gris más amplia) lo que podría estar relacionado con factores externos, como un aumento repentino de la demanda.

Como se evidencia en la Ilustración 10, existe información faltante en los transcurrido entre el mes de marzo y finales de junio del año 2024, esto debido a fallas técnicas e intermitencia que se presentaron en los dispositivos de red que capturan los datos. El proceso tiene una periodicidad de actualización diaria, lo que quiere decir que la información debe almacenarse con tiempos muy cortos de diferencia entre la ejecución y el almacenamiento, lo cual al momento de realizar el proyecto no fue viable recuperar el histórico ya que habían transcurrido varios meses y por lo cual no se cuenta con dicha información para periodos de tiempo específicos.

Cabe aclarar que por políticas de seguridad de la compañía se realiza el enmascaramiento de cada las estaciones base a analizar.

## 6. PREPARACIÓN DE LOS DATOS

Para garantizar que no exista información atípica dentro de los modelos y dando cumplimiento a la fase 3, se realiza una preparación de los datos capturados en pasos anteriores.

### 6.1. IMPUTACIÓN DE NULOS

Para llevar a cabo el análisis de la información, se realiza una preparación de datos previa donde se inicia con la imputación de los valores nulos por el valor de tráfico más próximo hacia adelante o hacia atrás de acuerdo con la fecha en la que se tomó la muestra, garantizando así la completitud de los datos a nivel diario (Ilustración 11).

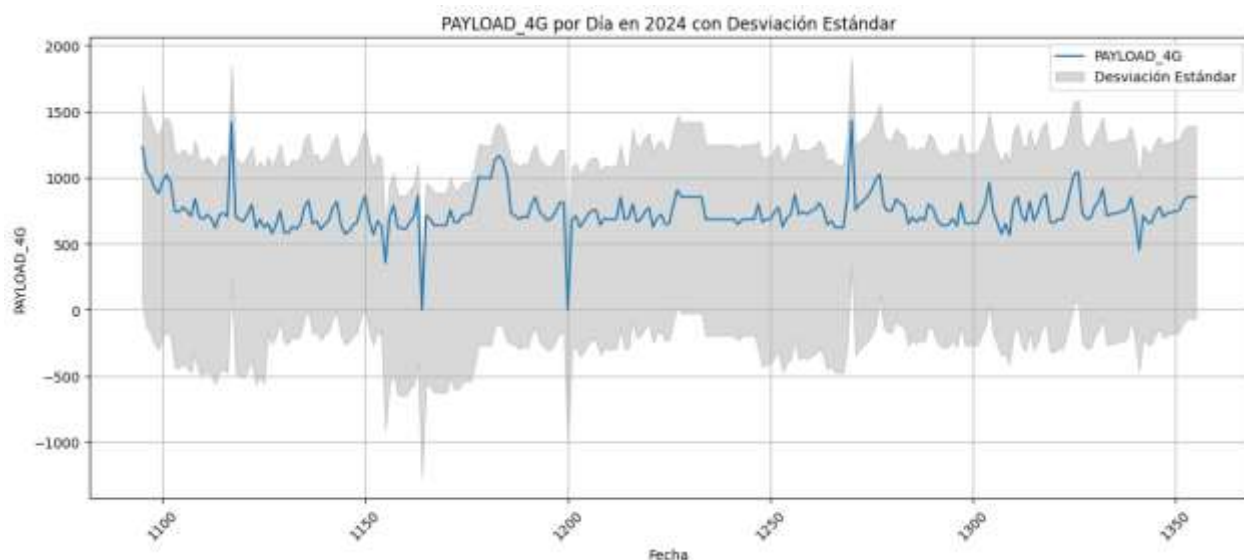


Ilustración 11 Serie de tiempo posterior a la imputación de nulos (Año 2024) (Autores).

### 6.2. SUAVIZADO DE DATOS O DETECCIÓN DE VALORES ATÍPICOS.

Dado que el objetivo no es eliminar los valores atípicos por completo, se utiliza la media para suavizarlos sin anularlos. Para llevar a cabo el tratamiento de estos datos, se hace uso del filtro de Hampel donde se utiliza un valor de  $k=7$  (ventana de 14 días) y un umbral  $t_0=1$ , lo que significa que cualquier punto que se desvíe más de una vez la desviación mediana absoluta será modificado con el valor de la media. Este método nos permitió limpiar la serie de datos sin afectar su estructura general, mejorando la calidad de las predicciones posteriores.

La Ilustración 12, presenta los resultados obtenidos en la Payload en lo transcurrido entre el año 2021 al 2024.

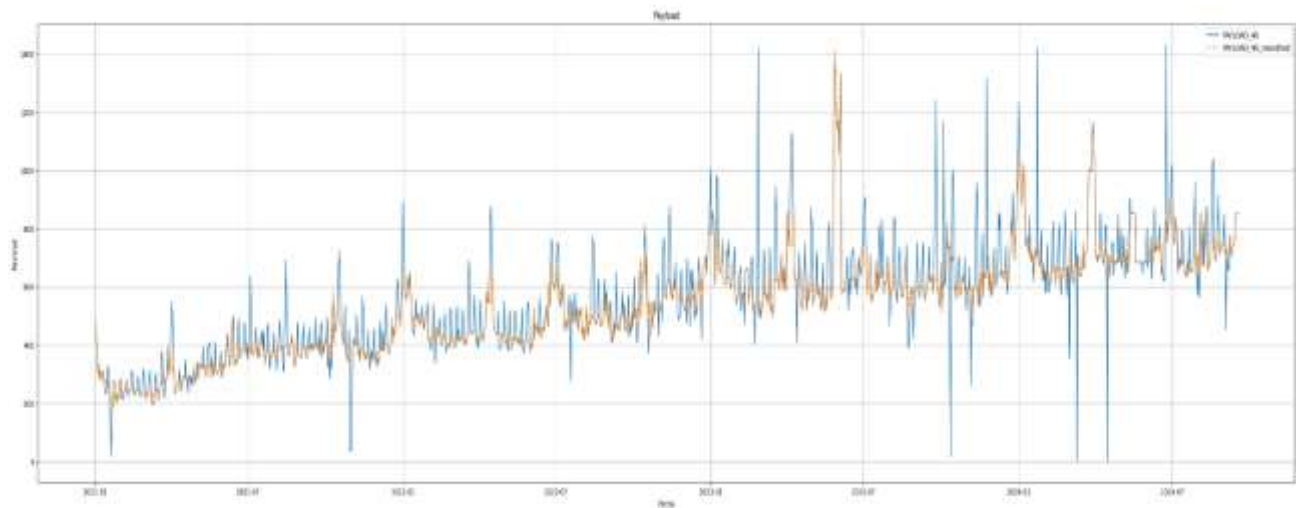


Ilustración 12 Comportamiento del Payload (Tráfico en línea azul) y posterior suavizado de datos (línea naranja) (Autores).

### 6.3. ANÁLISIS DEL TRÁFICO.

La ilustración 13 muestra la distribución multimodal, observando que el tráfico de uso normal de los usuarios oscila en un rango entre 200 y 800 bps. Mientras que existen casos ocasionales donde el tráfico es superior a este rango, lo cual permite concluir que existieron periodos de tiempos donde existió una gran demanda en el tráfico o anomalías en el comportamiento de la red y se puede llegar a designar a los eventos realizados en el municipio.

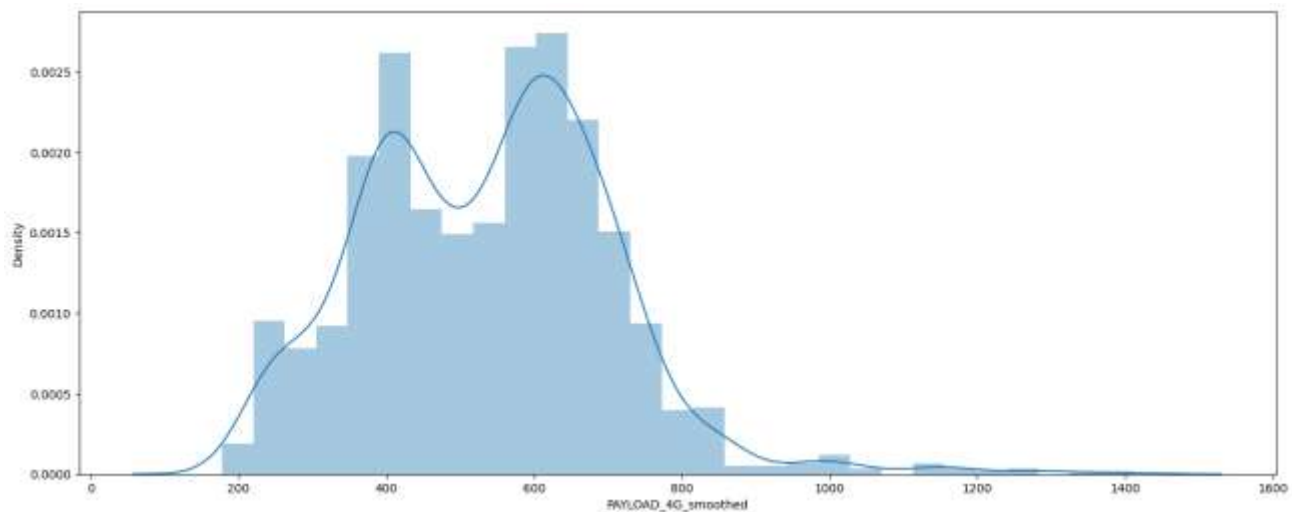
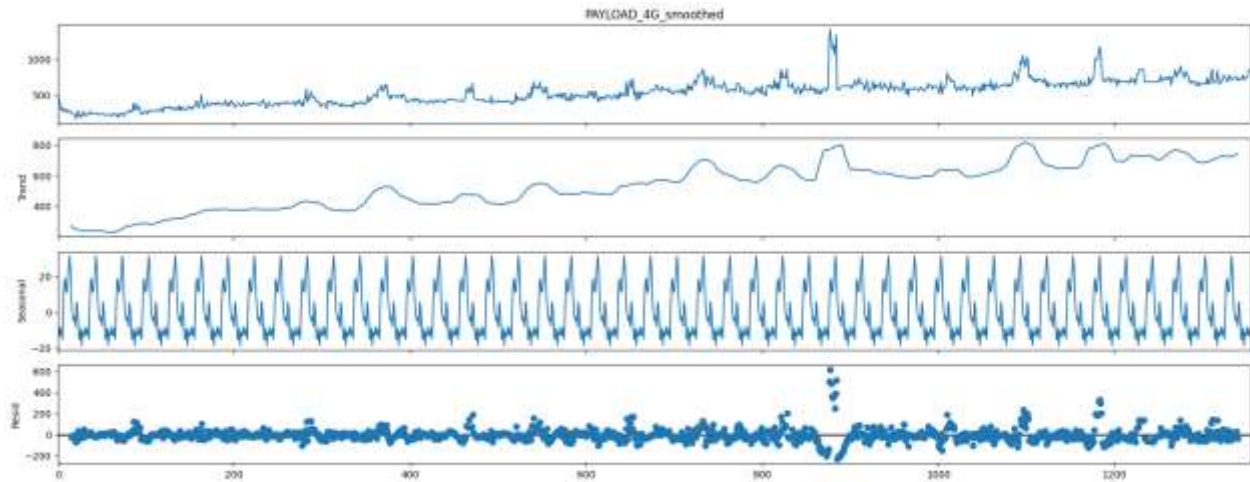


Ilustración 13 Distribución Modal del Tráfico (Autores).

Al realizar una descomposición estacional (Ilustración 14) de la información con el fin de analizar los componentes principales de tendencia, estacionalidad y residuos podemos observar que:

- **Serie original:** El tráfico presenta un comportamiento ascendente en los transcurrido entre el año 2021 y 2024, existen rangos de tiempos donde se presenta una variabilidad con respecto al comportamiento habitual.

- **Tendencia:** El tráfico de red ha crecido de forma exponencial y moderada, la tendencia muestra cambios suaves a lo largo del tiempo parece estabilizarse en periodos de tiempos específicos.
- **Estacionalidad:** El tráfico de red sigue un ciclo regular, observando un patrón estacional claro y repetitivo en los datos, lo que permite concluir que existen variaciones que son predecibles y se repiten en ciertos rangos de tiempo.
- **Residuos:** La mayor parte de los residuos son cercanos a cero, pero existen fluctuaciones que no siguen el patrón y que podrían indicar eventos anómalos en el tráfico.



*Ilustración 14 Descomposición Estacional del Tráfico (Autores).*

## 7. IMPLEMENTACIÓN Y EVALUACIÓN DE LOS MODELOS

Basado en lo desarrollado del capítulo anterior y siguiendo las fases de la metodología CRISP-DM Modelado, Evaluación y Verificación, se procede a realizar la implementación de los modelos seleccionados y observar el resultado de estos.

### 7.1. FACEBOOK PROPHET

Para el desarrollo de este modelo se hace uso de la variable Festividad y se indica que el evento inicia el mismo día y finaliza el día siguiente. Adicionalmente, la tabla 5 y 6 presentan las entradas y parámetros utilizados en el modelo, con el fin, de se incluye los días festivos y la estacionalidad mínima en la que se ejecuta el modelo es diaria.

*Tabla 5 Entradas Modelo Facebook Prophet (Autores).*

ENTRADA	DESCRIPCIÓN
DS	Corresponde a la fecha del día de la medición
y	Tráfico diario (Payload)
h	Flag que permite identificar si existe un evento

*Tabla 6 Parámetros Modelo Facebook Prophet (Autores)*

PARÁMETROS	VALOR
daily.seasonality	False
weekly.seasonality	True
yearly.seasonality	True
Changepoint.prior.scale	0.05

Una vez garantizado el conjunto de datos con las características mencionadas previamente, se procede a crear un nuevo conjunto con los mismos parámetros de festividades iniciales e incluimos en el pronóstico 365 días.

En la Ilustración 15 es la representación gráfica de la predicción del modelo donde observamos lo siguiente:

- **Línea Azul:** Hace referencia a la predicción del tráfico, la cual muestra una tendencia ascendente hasta 2023 después de este año se estabiliza.
- **Banda Azul Clara:** Corresponde al nivel de confianza que normalmente es del 80% al 95%.
- **Puntos Negros:** Valor real del Payload
- **Líneas Rojas:** Son los puntos de cambio detectados por el modelo Prophet, son puntos donde el modelo detecta un cambio de tendencia en el tráfico.

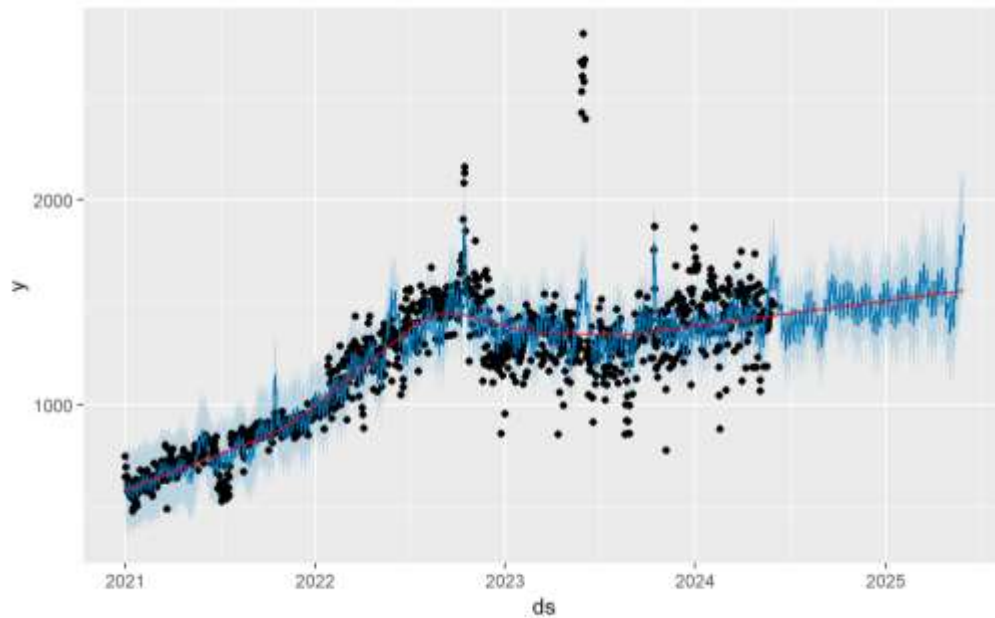


Ilustración 15 Predicción del Tráfico Modelo Prophet (Autores).

Al realizar la descomposición del tráfico por cada una de las temporalidades (Ilustración 16), obtenemos lo siguiente:

- **Trend:** Tendencia que muestra como el tráfico aumenta a lo largo del tiempo y luego del año 2023 se estabiliza
- **Holidays:** Impacto que generan los días festivos en el tráfico.
- **Weekly:** Presenta como el comportamiento de los últimos días de la semana es creciente, mientras que los primeros son los más bajos.
- **Yearly:** Este gráfico muestra que hay mayor tráfico los meses de Junio y Octubre, mientras que Enero y Marzo tienen menor tráfico.

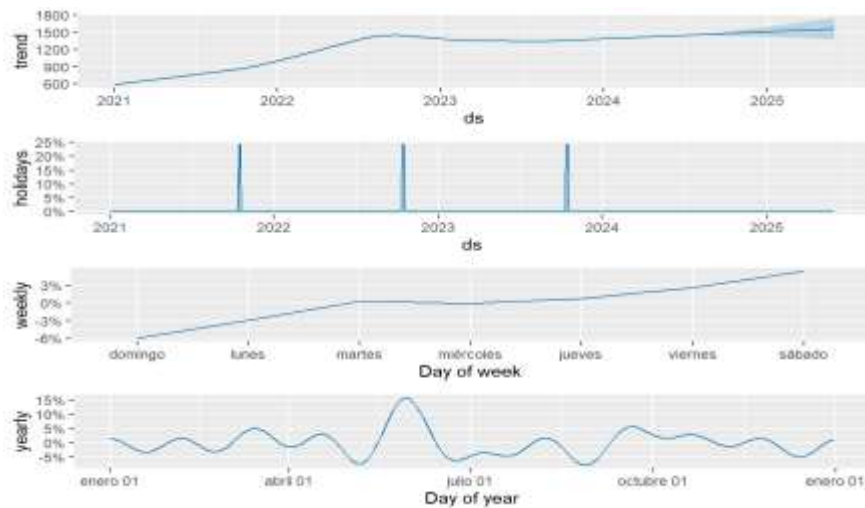


Ilustración 16 Descomposición del Tráfico - Modelo Prophet (Autores).

Si bien el modelo permite captar regresores en una serie de tiempo, el volumen de festivos repercute altamente a la hora de entrenar el modelo.

En primera instancia al modelo de generar un conjunto de datos de entrenamiento y prueba, la cantidad de festivos no es uniforme y por esta causa el modelo no puede explicar la variabilidad de los datos, obteniendo así un valor de R2 es negativo (-13%).

Si bien el modelo captura tendencias y patrones de tráfico con la incorporación de las festividades, existe un sobre ajuste que el modelo se adapta a lo que ya conoce, pero no generaliza los datos nuevos y existe una alta auto correlación de residuos.

## 7.2. RED NEURONAL ELMAN

La red Neuronal de Elman implementada usa las entradas originales y con rezagos (Tabla 7) está compuesta por 5 neuronas con una tasa de aprendizaje del 10% y 1000 iteraciones con la cual se realiza un análisis diario como semanal (Tabla 8). Para el análisis de este modelo el 75% de los datos corresponden al conjunto de entrenamiento, mientras que el 25% restante corresponde al conjunto de pruebas.

*Tabla 7 Entradas Modelo Elman (Autores).*

ENTRADA	DESCRIPCIÓN
Original	Corresponde a la fecha del día de la medición
Con retardo	Tráfico diario (Payload)

*Tabla 8 Parámetros Modelo Elman (Autores).*

PARÁMETROS	VALOR
Neuronas	5
Tasa de aprendizaje	0.1
Iteraciones	1000

En los resultados generados por los modelos se observa que los resultados de la temporalidad diaria tienen un rendimiento significativamente mejor con respecto a la temporalidad semanal, tanto en las métricas de errores y como en el valor de R2.

Adicional a esto ambos modelos se aplicó un suavizado de outliers por media de media móvil, esperando que el resultado mejorará significativamente pero no fue así. El modelo con temporalidad diaria, aunque con errores altos mejora la capacidad de explicar el modelo con el R2 en comparación del semanal.

En la Ilustración 17, se observa cómo se puede explicar la variabilidad de la variable Festividad en un 27% con respecto a los datos originales, pero aun así es un valor bajo para lograr predecir el tráfico en futuros eventos con un porcentaje de error mínimo.

De igual manera, el modelo no tiene la capacidad de capturar la variabilidad de los datos.

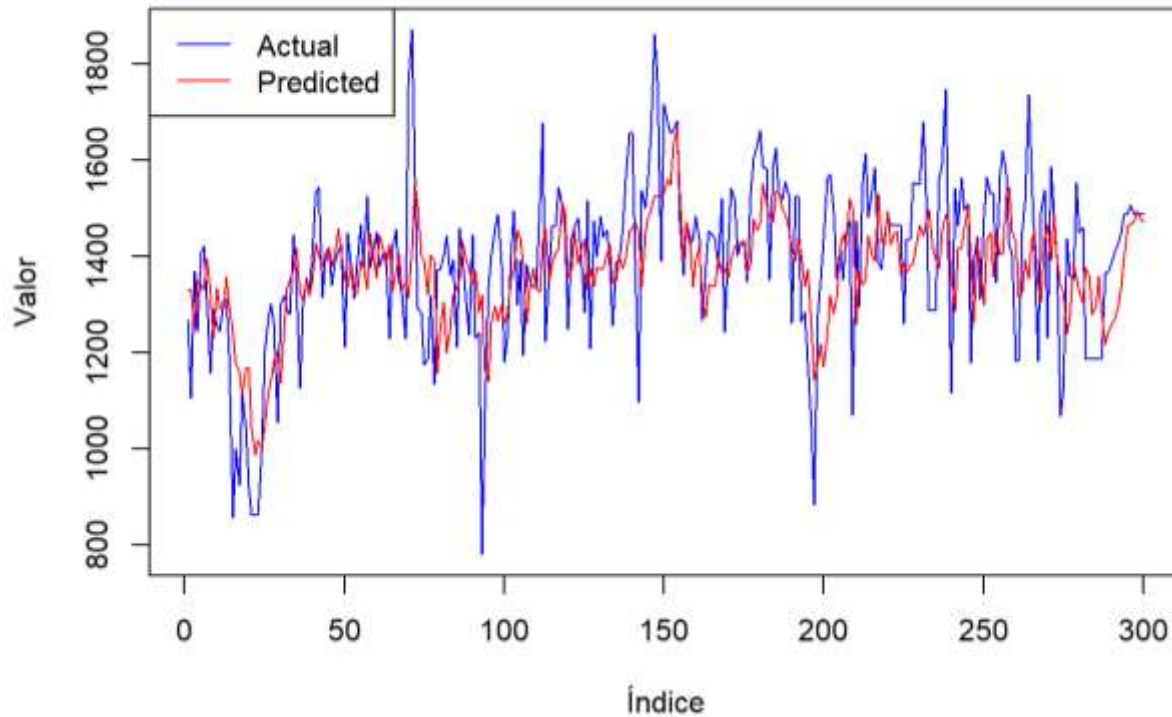


Ilustración 17 Comparación Tráfico Actual vs Predicción - Red Neuronal Elman (Autores).

### 7.3. RED NEURONAL JORDAN

Al igual que en el caso anterior la red neuronal usa las mismas entradas (Tabla 9), y los mismos parámetros (Tabla 10) los cuales está compuesta por 5 neuronas con una tasa de aprendizaje del 10% y 1000 iteraciones con la diferencia que la salida de una neurona es insumo de entrada para la siguiente.

Tabla 9 Entradas Modelo Jordan (Autores).

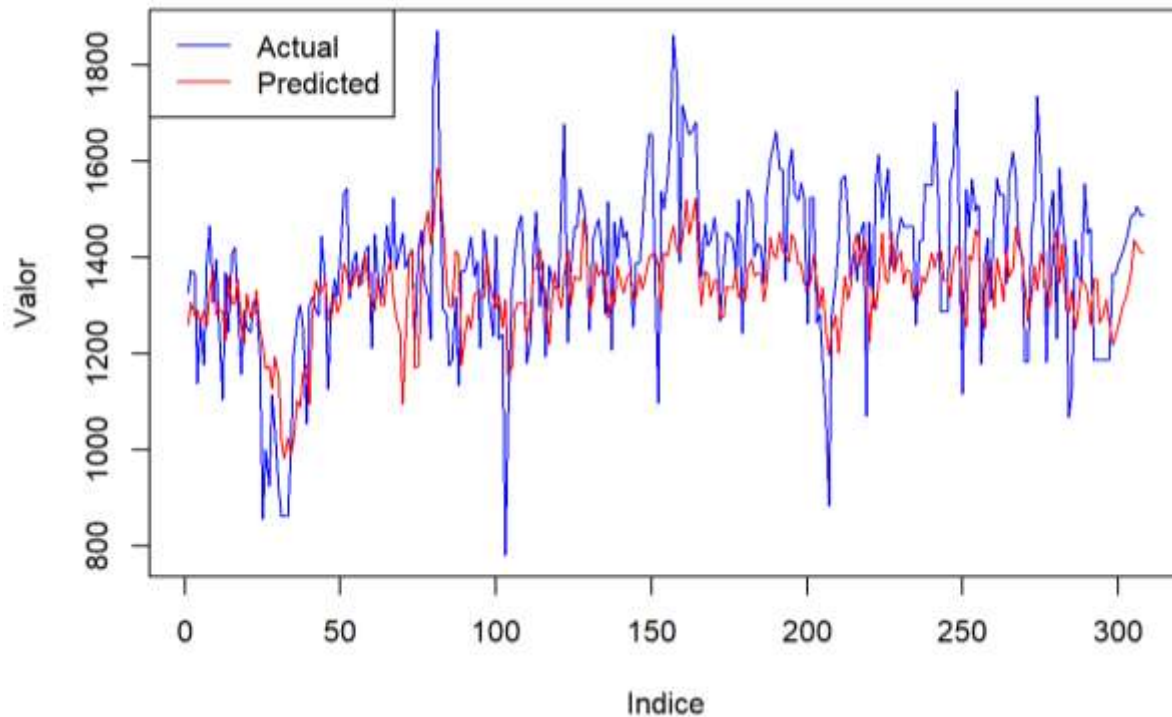
ENTRADA	DESCRIPCIÓN
Original	Serie Original
Con retardo	Serie con 12 retardos

Tabla 10 Parámetros Modelo Jordan (Autores).

PARÁMETROS	VALOR
Neuronas	5
Tasa de aprendizaje	0.1
Iteraciones	100

Al igual que en el caso anterior y como se observa en la Ilustración 18, la red neuronal Jordan presenta mejor rendimiento y precisión en la temporalidad diaria en comparación con la temporalidad semanal. Al

incluir la variable Festividad el modelo logra explicar la variabilidad en un 34% de la variabilidad, pero al igual que en la red neuronal Elman, el modelo tiene problemas para capturar la dinámica de los datos con métricas de error altas y baja capacidad explicativa.



*Ilustración 18 Comparación Tráfico Actual vs Predicción - Red Neuronal Jordan (Autores).*

#### 7.4. NEURALPROPHET

Al ser un modelo híbrido entre las redes neuronales y un modelo aditivo (Facebook Prophet) se inicia con la identificación de los hiperparámetros a utilizar para el desarrollo del modelo NeuralProphet.

El modelo Prophet nos provee un modelo para la estacionalidad, tendencia y puntos de cambio y sus parámetros son:

- **Changepoints\_range:** Controla la fracción de la serie donde se detectan los cambios de tendencia y el cual el valor por defecto es de 0.8.
- **N\_changepoints:** Número de puntos de cambio permitidos para la tendencia, el valor por defecto es 10
- **Seasonality\_mode:** Tipo de estacionalidad (aditiva o multiplicativa) y la selección por defecto es additive.
- **Yearly\_seasonality:** Controla la estacionalidad anual y la cual por defecto es True
- **Weekly\_seasonality:** Controla la estacionalidad semana y la cual por defecto es True
- **Daily\_seasonality:** Controla la estacionalidad diaria y la cual por defecto es False
- **Seasonality\_reg:** Coeficiente de regularización de la estacionalidad y su valor por defecto es 0
- **Trend\_reg:** Regularización de la tendencia y su valor por defecto es 0

- **Season\_global\_local:** Indica si la estacionalidad es compartida entre series (global) o específica de cada serie (local), la selección por es 'global'
- **Trend\_global\_local:** Indica si la tendencia es compartida entre series o específica de cada serie y la selección por defecto es 'global'
- **Changepoint\_prior\_scale:** Controla la sensibilidad del modelo a los puntos de cambio en la tendencia, en los casos de los valores más altos permiten detectar más cambios de tendencia y su valor por defecto es de '0.05'
- **MC samples:** Controla el número de muestras de Monte Carlo para la inferencia bayesiana y su valor por defecto es 0.
- **Interval\_width:** Define el ancho del intervalo de confianza para las predicciones y su valor es de 0.8
- **Uncertainty\_samples:** Número de muestras utilizadas para generar la incertidumbre de las predicciones y su valor por defecto es 1000

NeuralProphet añade un componente de redes neuronales a la metodología mencionada anteriormente y los cuales se explican a continuación:

- **N\_lags:** Número de rezagos usados para la predicción autoregresiva (AR). Esto agrega una estructura de red neuronal en la que se toman en cuenta valores anteriores de la serie para predecir los futuros. Valor por defecto es 0
- **N\_forecasts:** Cantidad de pasos futuros que el modelo debe predecir, su valor por defecto es 1
- **Epochs:** Número de épocas de entrenamiento. En cada época, el modelo ve los datos completos para ajustar sus parámetros y su valor por defecto es 40
- **Learning\_rate:** La tasa de aprendizaje que controla la velocidad con la que se ajustan los pesos del modelo durante el entrenamiento y su valor por defecto es 1.0
- **Batch\_size:** Tamaño del lote que se usa para actualizar los pesos en cada paso de entrenamiento, los tamaños más grandes hacen que el entrenamiento sea más estable, pero más lento y su valor por defecto es 64
- **Loss\_func:** La función de pérdida utilizada para ajustar el modelo, el valor predeterminado es 'Huber', ya que, es robusto frente a valores atípicos.
- **AR\_sparsity:** Controla la dispersión (sparsity) en la matriz de autoregresión, lo que promueve soluciones más simples con menos rezagos seleccionados, su valor predeterminado es 0
- **Normalize:** Si los datos deben ser normalizados antes de ser procesados por la red neuronal, normalmente la selección es auto.
- **Dropout:** Técnica utilizada para regularizar el modelo, establece una proporción de unidades neuronales que se deben "apagar" aleatoriamente en cada iteración para evitar el sobreajuste. Valor por defecto: 0.1

Una vez identificados los hiperparámetros que componen el modelo se realizan 195 pruebas con diferentes combinaciones y con un rezago de 6 días con el fin de identificar los más representativos al de aplicar el modelo.

Se usa como criterio de selección basado en R2 el cual mide la capacidad explicabilidad del modelo indicando que porcentaje de variabilidad es capturado por la predicción, en nuestro caso validamos que

fuera mayor a 70%. El otro parámetro Ljung-Box evalúa la autocorrelación de residuos el cual filtra que sea superior a 0.05, el cual nos indica que se comporta como ruido blanco, asegurando que no hay patrones no explicados. Con estas 2 métricas garantizamos que el modelo captura correctamente las tendencias sin sobreajuste, mejorando su capacidad de generalización. Como criterios de selección se tiene en cuenta que el valor de Ljung-box sea superior al 0.05, R2 tanto en Entrenamiento como en Prueba sea mayor a 60 con las siguientes entradas de la Tabla 11.

*Tabla 11 Entradas Modelo Neuralprophet (Autores).*

ENTRADAS	DESCRIPCIÓN
ds	M Fecha día es la fecha de la medición
y	Payload Tráfico diario

De acuerdo con lo mencionado anteriormente de las entradas y las pruebas los criterios de selección, los hiperparámetros seleccionados son:

- **inlags=6:** Define el número de "lags" o rezagos utilizados para la predicción autoregresiva en este caso se tomarán los últimos 6 días para predecir el siguiente valor.
- **n\_changepoints=2:** Este modelo va a detectar 2 puntos de cambio que permiten al modelo detectar diferencias en la tendencia.
- **trend\_reg=0.08:** Este parámetro regulariza la tendencia. Controla la suavidad de la tendencia global del modelo y penaliza en un menor valor los cambios bruscos en la tendencia, resultando en una predicción sin tanto suavizado a los datos de entrenamiento.
- **yearly\_seasonality=True:** Indica que el modelo debe capturar la estacionalidad anual.
- **weekly\_seasonality=True:** Activa la estacionalidad semanal, lo que significa que el modelo intenta capturar patrones que se repiten semanalmente. Esto es útil para datos que presentan variaciones regulares durante los días de la semana.
- **daily\_seasonality=True:** Indica que el modelo tiene en cuenta la estacionalidad diaria. Esto permite capturar patrones que ocurren a diario, como fluctuaciones horarias en ciertos tipos de datos.
- **n\_forecasts=60:** Especifica cuántas predicciones hacia el futuro se desean hacer. En este caso, el modelo predice los valores para los próximos 60 períodos (días). Este parámetro define la ventana de predicción.

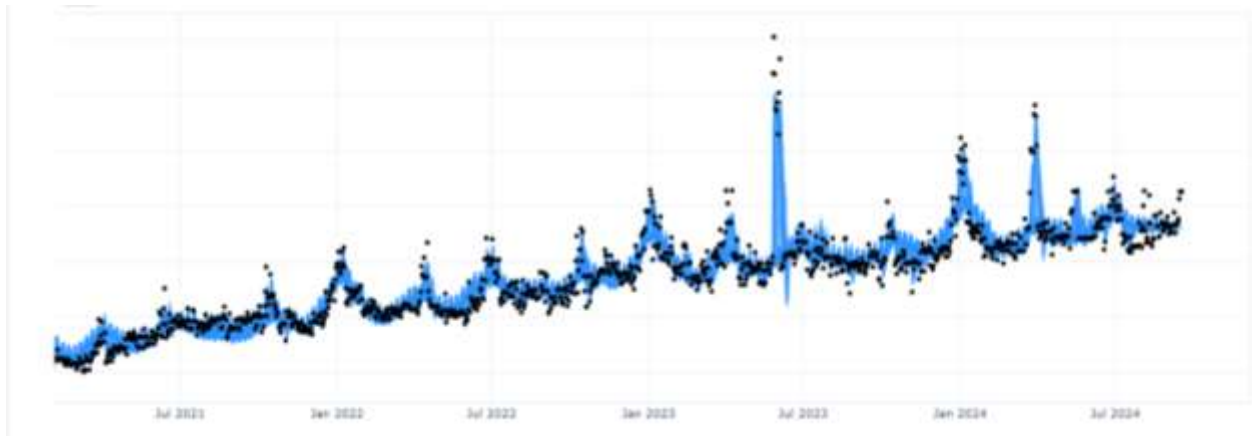


Ilustración 19 Comparación Tráfico Real vs Predicción del Modelo - NeuralProphet (Autores).

En la ilustración 19 se observa un comparativo de la predicción del tráfico (Línea Azul) con respecto a los valores reales (Puntos Negro), donde se identifica que el modelo captura de manera adecuada los valores y los cambios de estacionales que se presentan, aunque existen valores atípicos estos difieren de la predicción, ya que, su naturaleza corresponde al origen de los datos.

Para generar un caso de análisis más detallado se toma como muestra el mes de octubre, para este escenario se tomará el Error Porcentual Absoluto (EPA) y el cual se calcula con la siguiente formula:

$$(Ecuación 1) \text{ EPA} = \frac{\text{Valor Real} - \text{Predicción}}{\text{Valor Real}} * 100$$

En la siguiente ilustración se observa como el modelo (Línea Roja) tiene la capacidad de seguir la tendencia de los datos reales (Línea Azul) incluyendo los días festivos.

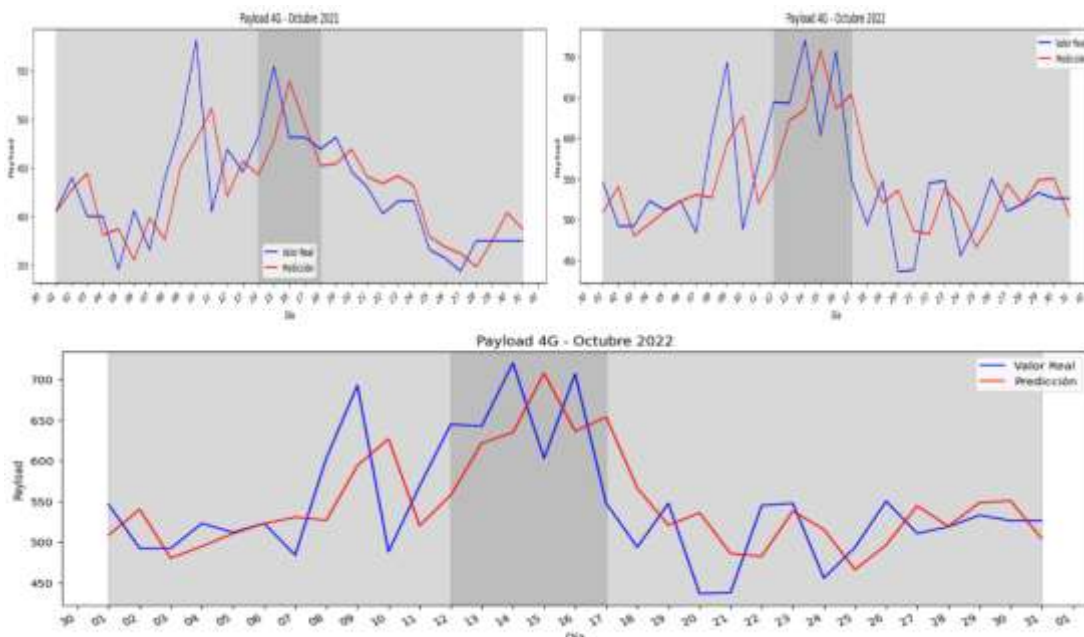
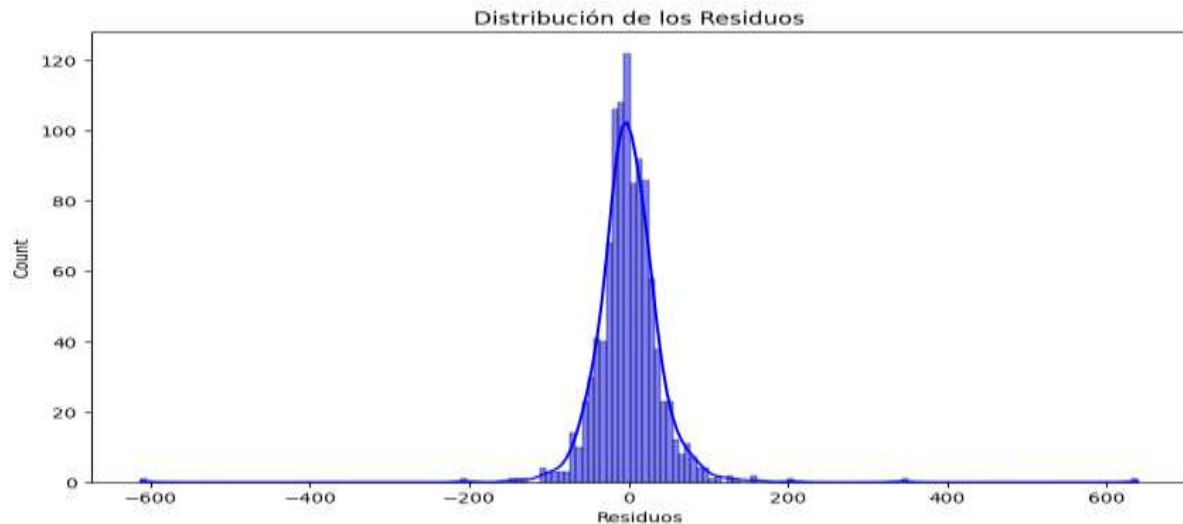


Ilustración 20 Análisis de Tráfico haciendo uso Modelo NeuralProphet (Autores).

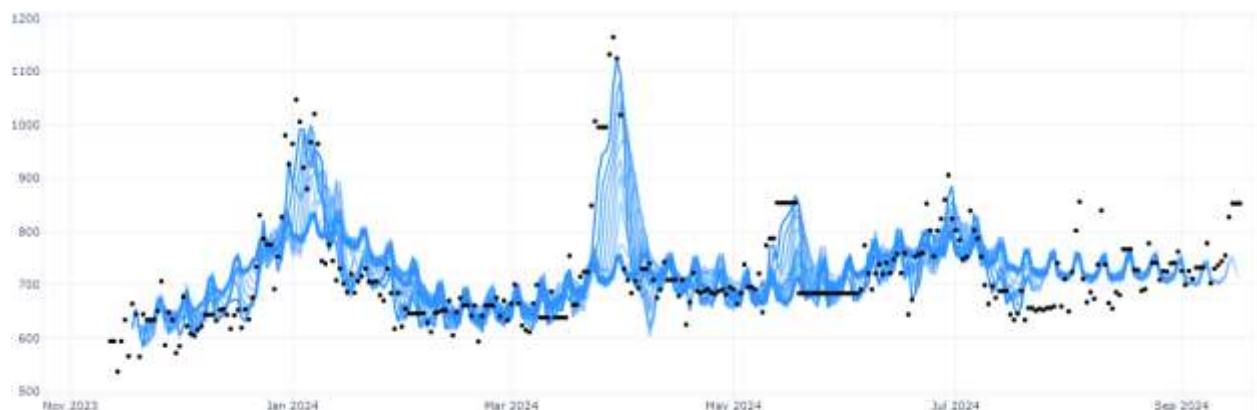
Con el fin de evaluar el resultado del modelo se realiza una distribución de residuos (Ilustración 21) evidenciando una distribución normal y con una autocorrelación baja, lo que indica que gran porcentaje del modelo esta ajustado correctamente.



*Ilustración 21 Distribución de Residuos - NeuralProphet (Autores).*

Para conocer la correlación de los residuos se aplica la prueba de Ljung-Box, puesto que el tráfico tiene comportamientos no estacionarios y variaciones abruptas. El valor obtenido en esta prueba es del 0.7682, lo que indica que no existe una autocorrelación significativa entre los residuos para los primeros 6 rezagos.

Al no existir una correlación entre los residuos, métricas de error bajas y aplicando el conjunto de pruebas, se observa que el modelo esta realiza una predicción aceptable con respecto a los datos reales (Ilustración 22). Captando la tendencia, los picos y valles de tráfico que se presentan en la red.



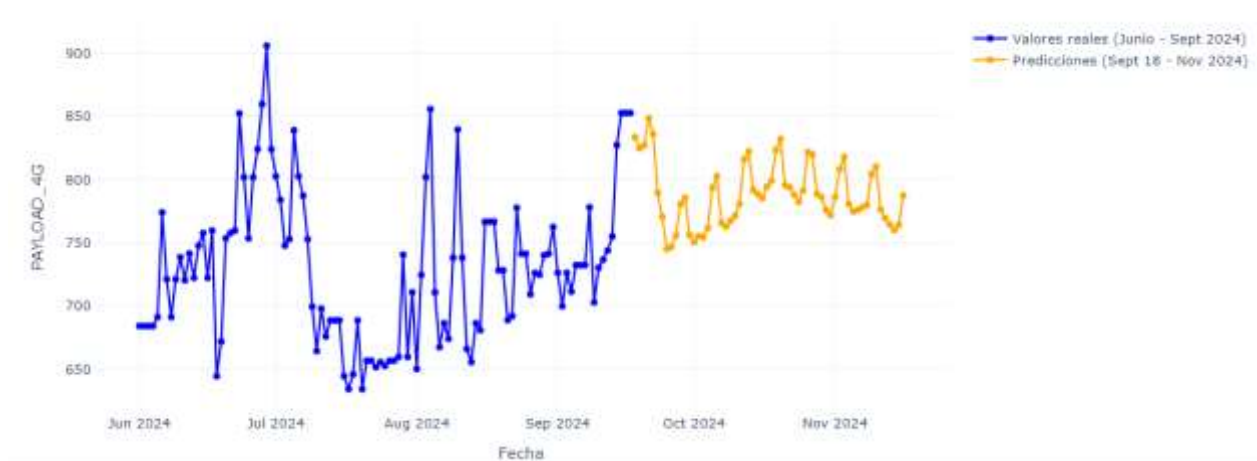
*Ilustración 22 Tráfico real versus predicción con conjunto de datos de prueba. (Autores).*

#### 7.4.1. EVALUACIÓN DEL MODELO

El modelo NeuralProphet ha mostrado ser eficaz para predecir el tráfico con una buena precisión tanto en el conjunto de entrenamiento como en el de test, con respecto a los modelos mencionados previamente. Por lo cual se implementa y evalúa, con el fin, de generar una predicción del tráfico en el periodo del mes de Octubre (10 al 14 del presente año) para el municipio seleccionado.

La tendencia que genera la predicción del modelo es estable y ligera al alza a futuro para el periodo de tiempo transcurrido entre septiembre a noviembre de 2024 (Ilustración 23), oscilando entre 820 y 890 bps. Lo que permite concluir que la red espera un aumento moderado en el tráfico de la red en corto plazo.

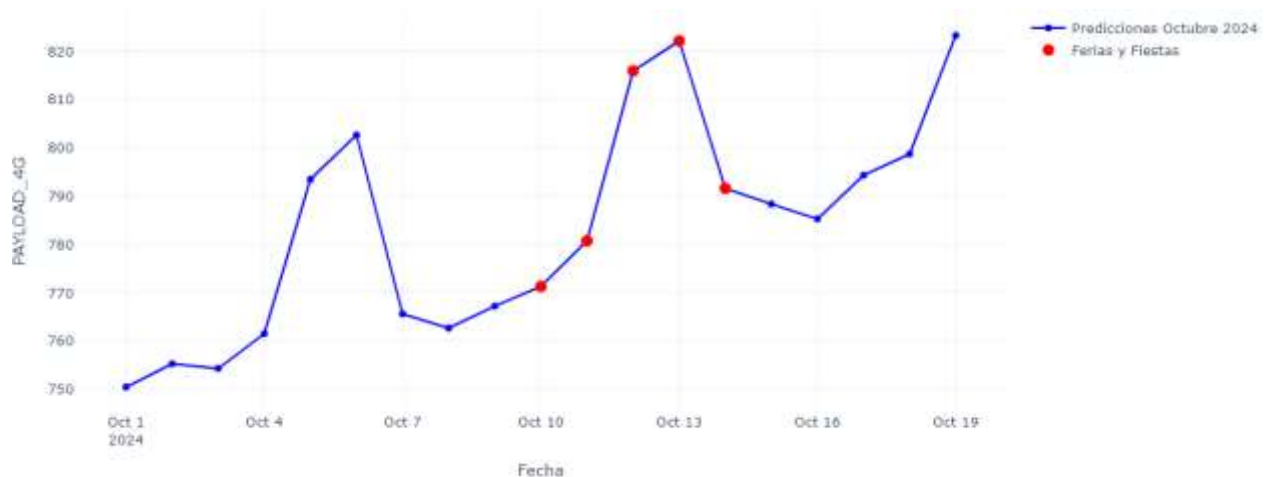
En la transición entre valores reales y predichos se observa una continuidad relativamente suave entre el tráfico real y el resultado del modelo, ya que, el modelo ha capturado bien los patrones de los datos históricos y los proyecta con cierta estabilidad hacia el futuro.



*Ilustración 23 Predicción de Tráfico septiembre 2024 - noviembre 2024 (Autores).*

Para el periodo del 10 al 14 de octubre la Ilustración 24 se muestra con los días festivos (marcadores rojos) coinciden con los picos más altos en la curva, lo que permite concluir que durante este periodo de tiempo hay un aumento significativo en el uso del Tráfico.

La coincidencia mencionada anteriormente puede tener un impacto positivo en el aumento del tráfico de datos.



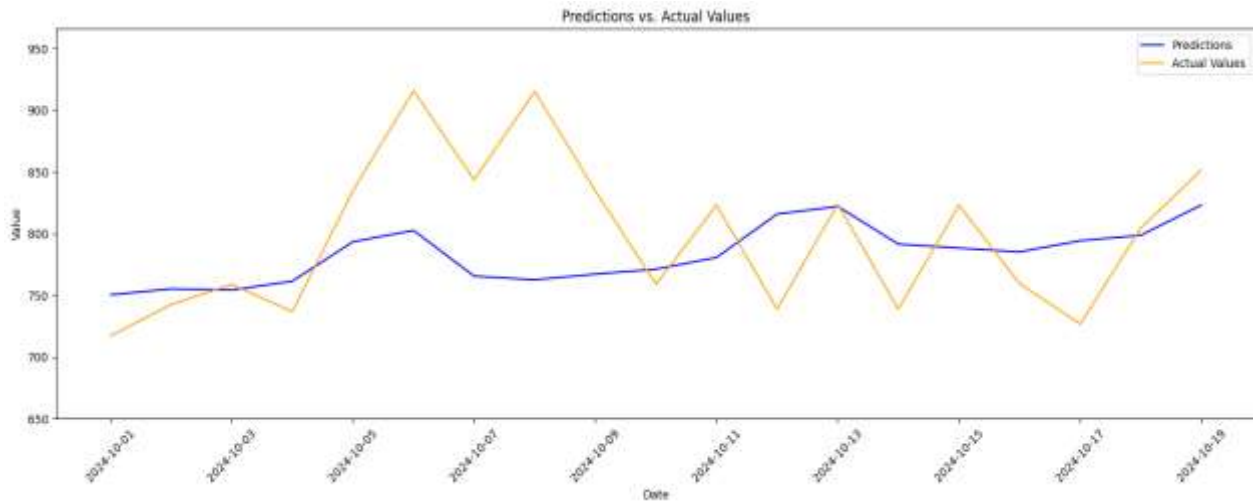
*Ilustración 24 Predicción de Tráfico para Festividades (Autores).*

Acotando a la predicción a los días feriados podemos identificar que la tendencia es creciente precisamente en los días de ferias y fiestas de dicho municipio, esperando un aumento en el tráfico exactamente los días sábado 12 y domingo 13 de octubre del 2024 en el municipio de Cundinamarca.

Aunque el R2 del modelo es del 75% en el conjunto de test tiene un margen de mejora, los errores son bajos (MAPE, MAE, RMSE) y las predicciones siguen una tendencia moderada y estable hacia el futuro, con menos fluctuaciones que en los meses previos. Por lo cual podemos deducir que el tráfico probablemente se estabilizará en los próximos meses con un crecimiento sostenido en el uso del tráfico en la red.

A continuación presentaremos la fase de “Verificación” según la metodología CRISP -DM ya que el modelo NeuralProphet logra capturar correctamente las tendencias sin sobreajustes , como muestra la ilustración 25, presenta la comparación entre la predicción del tráfico realizado por el modelo y el comportamiento real del tráfico en la festividad que se llevó a cabo entre el 10 al 14 de octubre. Se observa que la proyección predice una ocupación de red moderada y estable, lo cual permite prever la estabilización del tráfico en los próximos meses.

Aunque se observa una ligera diferencia en los picos de demanda tanto el modelo como el tráfico real, el modelo predice correctamente el aumento que presenta el tráfico desde el día que inicia la festividad (10 de octubre) y concuerda con los días de mayor ocupación 13 de octubre, posterior a esto, se observa como la predicción y el comportamiento de la red descienden y tiende a estabilizarse.



*Ilustración 25 Comparación Predicción de Tráfico vs Tráfico Real del 10 al 14 de octubre 2024 (Autores).*

El modelo NeuralProphet con el conjunto de datos de prueba, captura correctamente la tendencia y estabilización en el tráfico de red, manteniendo un margen de error bajo (MAPE de 5.64%) y sin dejar información sin modelar, como lo indica la ausencia de autocorrelación en los residuos. Aunque el R2 de 75% sugiere un margen para mejorar, las predicciones son suficientemente precisas para planificar la carga de la red.

La capacidad predictiva resulta valiosa para las empresas de Telecomunicaciones, ya que les permite gestionar proactivamente la infraestructura de red y optimizar los recursos para mejorar la experiencia del usuario, incluso en periodos de alta demanda.

## 8. CONCLUSIONES Y TRABAJOS FUTUROS

### 8.1. CONCLUSIONES

- ❖ Actualmente, el modelo permite proyectar la carga de la red móvil y así realizar una planeación y diseño de mejoras en busca de que el usuario no se vea impactado en su experiencia a la hora de usar sus datos móviles o realizar llamadas.
- ❖ La capacidad del modelo NeuralProphet para seguir las tendencias de los datos históricos y captar los efectos de los días festivos es un buen indicador de su calidad predictiva, aunque se pueden presentar alteraciones en el indicador por capturas erróneas desde la fuente de datos.
- ❖ Aunque en lo transcurrido del año se realizan diferentes eventos culturales en los municipios de Cundinamarca, son los eventos de alta relevancia y Ferias y Fiestas mayor impacto generan en la red por la migración de usuarios de otros municipios al que se realiza el evento.
- ❖ Algunos de los modelos analizados en este estudio presentan considerables limitaciones en su capacidad para capturar la variabilidad y las dependencias de los datos, particularmente debido a la influencia de los días festivos y la autocorrelación de residuos. Aunque la inclusión de variables regresoras proporciona algunas mejoras, persisten desafíos significativos en términos de precisión y capacidad explicativa.

### 8.2. TRABAJOS FUTUROS

El proyecto permite ser ampliado o tiene las siguientes oportunidades de mejora:

- ❖ Integrar como parámetros de entrada las características con las cuales se diseñaron e implementaron las Estaciones Base, puesto que, dependiendo de la cantidad de población y topología del Municipio pueden variar dichas características.
- ❖ Conocer e incluir análisis de sentimientos, que permita conocer identificar cual es el sentir del cliente y así evaluar acciones preventivas para mejorar las estrategias comerciales y de experiencia del usuario.
- ❖ Ampliar el análisis descriptivo de los datos con estadísticas que permitan conocer la cantidad de usuarios que viajan a dichas regiones es fechas específicas.

## BIBLIOGRAFÍA

- [1] ONU, "Las comunicaciones móviles y el desarrollo socioeconómico: Una perspectiva latinoamericana," 01 junio 2023. [En línea]. Disponible en: <https://www.un.org/es/chronicle/article/las-comunicaciones-moviles-y-el-desarrollo-socioeconomicouna-perspectiva-latinoamericana>.
- [2] R. E. Tiempo, "Saturación de redes es catalogada como un buen problema para las empresas telefónicas," *Portafolio*, p. 1, 18 agosto 2010.
- [3] D. N. d. Planeación, "Departamento Nacional de Planeación," [En línea]. Disponible en: <https://www.dnp.gov.co/programas/Desarrollo%20Digital/Paginas/Analitica-de-datos.aspx>. [Último acceso: 01 junio 2023].
- [4] C. D. R. D. COMUNICACIONES, "Postdata Más Allá de los Datos," 09 junio 2022. [En línea]. Disponible en: <https://postdata.gov.co/story/medicion-de-percepcion-de-calidad-de-los-servicios-de-telecomunicaciones-2021>.
- [5] MINTIC, "BOLETÍN TRIMESTRAL DE LAS TIC - PRIMER SEMESTRE DE 2023," *Información Estadísticas para las TIC*, Bogotá D.C., 2023.
- [6] S. Quevedo, "La Satisfacción del Cliente en la Industria de las Telecomunicaciones," *P&M*, Bogotá, 2022.
- [7] MINTIC, "Artículo 1," *Resolución 202 de 2010*, Bogotá, 2010.
- [8] S. d. I. y Comercio, "DECRETO NUMERO 0741 DE 1993," Capítulo I, Bogotá, 2023.
- [9] D. Luna, "Decreto 1370 de 2018," *Ministerio de Tecnologías de la Información y las Comunicaciones*, Bogotá, 2018.
- [10] M. d. T. d. I. I. y. I. Comunicaciones, "Preguntas Frecuentes," [En línea]. Disponible en: [https://www.mintic.gov.co/portal/inicio/Atencion-y-Servicio-a-la-Ciudadania/Preguntas-frecuentes/5237:Servicios-de-Telecomunicaciones#:~:text=¿En%20Colombia%20qué%20empresas%20prestan,y%20NOVATOR%20PARTNERS%20\(WOM\)](https://www.mintic.gov.co/portal/inicio/Atencion-y-Servicio-a-la-Ciudadania/Preguntas-frecuentes/5237:Servicios-de-Telecomunicaciones#:~:text=¿En%20Colombia%20qué%20empresas%20prestan,y%20NOVATOR%20PARTNERS%20(WOM).). [Último acceso: 01 junio 2023].
- [11] IBM, "¿Qué es la ciencia de datos?," *IBM*, [En línea]. Disponible en: <https://www.ibm.com/es/topics/data-science>. [Último acceso: 01 junio 2023].
- [12] F. Perez Ramírez, *Introducción a las Series de Tiempo*, Medellín: Sello Editorial, 2007.
- [13] Desarrollo, P. d. (20 de Diciembre de 2024). *Programa de la Nacióna Unidad para el Desarrollo*. Obtenido de <https://www.undp.org/es/sustainable-development-goals/industria-innovacion-infraestructura>
- [14] C. J. Benavides Rivera, "Predicción de comportamiento en tráfico de red LTE y ajuste de parametrización para maximizar performance de red," *Universidad de Chile*, Santiago de Chile, 2021.
- [15] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia y D. Anastasiu, "Stock Price Prediction Using News Sentiment Analysis," *IEEE Computer Society*, Estados Unidos, 2019.
- [16] M. A. Curipallo Martínez, "Detección de patrones de comportamiento de parámetros de RF en redes de comunicación móvil mediante mediciones de campo y técnicas de Machine Learning," *Escuela Politécnica Nacional*, Quito, 2022.

- [17] G. Gamús, "Análisis de sentimientos de datos de redes sociales usando técnicas de Machine Learning," *Universidad Autónoma de Barcelona*, Barcelona, 2021.
- [18] D. N. d. Planeación, "Departamento Nacional de Planeación," [En línea]. Disponible en: <https://www.dnp.gov.co/programas/Desarrollo%20Digital/Paginas/Analitica-de-datos.aspx>. [Último acceso: 07 febrero 2023].
- [19] Ministerio de Tecnologías de la Información y las Comunicaciones, "Ministerio de Tecnologías de la Información y las Comunicaciones," 01 marzo 2020. [En línea]. Disponible en: [https://www.mintic.gov.co/portal/715/articles-126335\\_recurso\\_1.pdf](https://www.mintic.gov.co/portal/715/articles-126335_recurso_1.pdf).
- [20] José, H., *Comunicaciones Móviles. Sistemas GSM, UMTS y LTE*, España: Grupo Editorial RA-MA, 2002.
- [21] C. A. Serra Jiménez y F. R. Marante Rizo, "Arquitectura general del sistema LTE," *Telemática*, vol. 12, no. 2, pp. 81–90, 2013. [En línea]. Disponible en: <https://revistatelematica.cujae.edu.cu/index.php/tele/article/view/106>.
- [22] 3rd Generation Partnership Project (3GPP), "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Overall description; Stage 2 (Release 15)," *3GPP TS 36.300*, 2021.
- [23] M. Sauter, *From GSM to LTE: An Introduction to Mobile Networks and Mobile Broadband*, Wiley, 2016.
- [24] ACENS, "Qué es Payload," 16 septiembre 2024. [En línea]. Disponible en: <https://ayuda.acens.com/hc/es/articles/360018220377--Qu%C3%A9-es-Payload#:~:text=En%20inform%C3%A1tica%20y%20telecomunicaciones%20es,facilitar%20la%20entrega%20del%20mensaje>.
- [25] A. J. Ribeiro, R. Mendes y M. do C. Duarte, "Requisitos para a ciência de dados: Analisando anúncios de vagas de emprego com mineração de texto," *RISTI - Revista Ibérica de Sistemas y Tecnologías de Información*, vol. 46, pp. 54-70, 2022. [En línea]. Disponible en: <https://www.risti.xyz/issues/risti46.pdf>.
- [26] R. Florez y J. Fernandez, *Redes Neuronales Artificiales*, La Coruña: Netbiblo, 2008.
- [27] IBM, "Natural Language Processing," 31 septiembre 2024. [En línea]. Disponible en: <https://www.ibm.com/mx-es/topics/natural-language-processing>.
- [28] Facebook Open Source, "Prophet," [En línea]. Disponible en: <https://facebook.github.io/prophet/>.
- [29] J. L. Elman, "Finding Structure in Time," *Cognitive Science*, vol. 14, no. 2, pp. 179-211, 1990.
- [30] M. I. Jordan, "Attractor Dynamics and Parallelism in a Connectionist Sequential Machine," *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, pp. 531-546, 1986.
- [31] O. Triebe, N. Laptev y R. Rajagopal, "NeuralProphet: Explainable Forecasting at Scale," *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021.
- [32] "Ley 1341 de 2009 - Por la cual se establece el marco regulatorio para el uso y desarrollo de las tecnologías de la información y la comunicación en Colombia," [En línea]. Disponible en: <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=36913>.
- [33] "Resolución 3066 de 2011 - Por la cual se definen los parámetros de calidad del servicio que deben cumplir los operadores de telecomunicaciones," [En línea]. Disponible en: <https://www.crcom.gov.co/es>.

- [34] "Ley 1450 de 2011 - Por la cual se expide el Plan Nacional de Desarrollo 2010-2014," [En línea]. Disponible en: <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=43101>.
- [35] "Ley 1978 de 2019 - Por la cual se establecen medidas para promover la inclusión digital en el país," [En línea]. Disponible en: <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=98210>.
- [36] "Resolución 421 de 2015 - Por la cual se modifica y complementa la regulación sobre calidad de servicio," [En línea]. Disponible en: <https://www.crcm.gov.co/es>.
- [37] "Resolución 124 de 2020 - Por la cual se establece el modelo de medición de calidad de servicio," [En línea]. Disponible en: <https://www.crcm.gov.co/es>.
- [38] "Resolución 8 de 2019 - Por la cual se definen los estándares de calidad que deben cumplir los operadores," [En línea]. Disponible en: <https://www.crcm.gov.co/es>.
- [39] "Plan Nacional de Desarrollo 2018-2022 - Pacto por Colombia, Pacto por la Equidad," [En línea]. Disponible en: <https://colaboracion.dnp.gov.co/CDT/Prensa/Resumen-PND2018-2022-final.pdf>.
- [40] A. Nokia, "LTE End to End System Part 1," Nokia Academy, 2014.
- [41] "Elman network vs Jordan network in RNN," [En línea]. Disponible en: [https://www.researchgate.net/figure/Elman-network-vs-Jordan-network-in-RNN\\_fig1\\_380062533](https://www.researchgate.net/figure/Elman-network-vs-Jordan-network-in-RNN_fig1_380062533).
- [42] "Block diagram of proposed hybrid model [En línea]. Disponible en: <https://www.mdpi.com/1996-1073/15/6/2158>.

## ANEXOS

**Anexo 1.** Detalle del modelo implementado Neuronal-Prophet.