

# **ESCUELA POLITÉCNICA NACIONAL**

## **FACULTAD DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA**

**ESTUDIO DEL COMPORTAMIENTO DE LAS MEDICIONES DE  
RADIO FRECUENCIA EN LA ZONA DE COBERTURA DE UNA  
CELDA CELULAR EN BASE A MEDICIONES DE CAMPO Y  
TÉCNICAS DE MACHINE LEARNING PARA ESTUDIOS DEL DOA**

**EVALUACIÓN DE UNA TÉCNICA DE MACHINE LEARNING PARA LA  
PREDICCIÓN DE LA DIRECCIÓN DE ARRIBO (DOA) UTILIZANDO  
MEDICIONES DE CAMPO Y RSTUDIO.**

**TRABAJO DE INTEGRACIÓN CURRICULAR PRESENTADO COMO  
REQUISITO PARA LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN  
TELECOMUNICACIONES**

**DENNIS ISMAEL LOMAS GARCÍA**

**([dennis.lomas@epn.edu.ec](mailto:dennis.lomas@epn.edu.ec))**

**DIRECTOR: PhD. PABLO ANÍBAL LUPERA MORILLO**

**([pablo.lupera@epn.edu.ec](mailto:pablo.lupera@epn.edu.ec))**

**DMQ, abril 2024**

## **CERTIFICACIONES**

Yo, Dennis Ismael Lomas García declaro que el trabajo de integración curricular aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

---

**Ismael Lomas**

Certifico que el presente trabajo de integración curricular fue desarrollado por Dennis Ismael Lomas García, bajo mi supervisión.

---

**PhD. PABLO ANÍBAL LUPERA**  
**DIRECTOR**

## **DECLARACIÓN DE AUTORÍA**

A través de la presente declaración, afirmamos que el trabajo de integración curricular aquí descrito, así como el (los) producto(s) resultante(s) del mismo, son públicos y estarán a disposición de la comunidad a través del repositorio institucional de la Escuela Politécnica Nacional; sin embargo, la titularidad de los derechos patrimoniales nos corresponde a los autores que hemos contribuido en el desarrollo del presente trabajo; observando para el efecto las disposiciones establecidas por el órgano competente en propiedad intelectual, la normativa interna y demás normas.

Dennis Ismael Lomas García

PhD. PABLO ANÍBAL LUPERA

## **DEDICATORIA**

Este trabajo esta dedicado primero a Dios y a todas las personas que han sido un soporte en mi vida, para ellos y de forma especial me dedico este trabajo porque, aunque parecía interminable, aunque no cobraba sentido muchas cosas no permití que se estancará y continúe dando lo mejor de mí.

## **AGRADECIMIENTO**

Agradezco inmensamente a Dios por todo su amor impartido en mi vida y por permitirme culminar esta etapa de mi vida. Doy gracias a mi madre Silvia que ha estado en innumerables noches acompañándome en todas mis actividades. A mi padre Rodolfo que siempre me mostro un apoyo, me supo guiar, hacer un buen hombre y siempre trato de corregir mis caminos, con su gran frase "Ya falta poco". Agradezco a mis hermanos los cuales han sido un apoyo incondicional, en toda área de mi vida, le doy gracias a Dios por mi familia la cual siempre fue ese pilar fundamental en mi vida para no caer.

De la misma manera agradezco a mi abuelita, que gracias a todos sus consejos y enseñanzas he podido avanzar. A la Srta. Jessica Tandazo agradezco por su manera de ayudarme en mis momentos más tristes y mis bajones donde encontré consuelo cuando parecía que no tenía salida. A mis amigos los cuales han sido una roca en la cual me he sabido arrimar cuando no he entendido algo.

Al PhD. Pablo Lupera por su guía a lo largo de todo este trabajo, le agradezco por su paciencia, constancia en su enseñanza y todo su conocimiento impartido. Doy gracias a todos mis profesores que a lo largo de esta carrera me han enseñado a ser una mejor persona y han buscado dar lo mejor de ellos.

## ÍNDICE DE CONTENIDO

CERTIFICACIONES .....	II
DECLARACIÓN DE AUTORÍA .....	III
DEDICATORIA .....	IV
AGRADECIMIENTO .....	V
ÍNDICE DE CONTENIDO.....	VI
RESUMEN .....	VIII
ABSTRACT.....	IX
1 INTRODUCCIÓN.....	1
Objetivo general.....	2
Objetivos específicos .....	2
Alcance.....	2
Marco teórico .....	3
1.1.1 DOA .....	3
1.1.2 Técnicas aplicadas para localizar el DoA.....	4
1.1.2.1 Redes neuronales profundas (DNN) .....	4
1.1.2.2 Accuracy Enhancement Technique using Oversized Lens-Loaded Cavity .....	5
1.1.3 Comunicaciones Móviles .....	5
1.1.4 LTE .....	7
1.1.4.1 Bandas de Frecuencias .....	8
1.1.4.2 Arquitectura LTE .....	9
1.1.4.3 IMS.....	10
1.1.5 Estación Base.....	10
1.1.6 Machine Learning .....	11
1.1.7 R y Rstudio .....	13
2 METODOLOGÍA.....	15
Herramienta de recolección de datos.....	15
2.1.1 Netmonitor pro .....	15
Etapa de recolección de datos .....	18
2.1.2 Planificación .....	18
2.1.3 Zonas y rutas utilizadas en la medición .....	19
2.1 Parámetros adicionales utilizados en la base de datos .....	22
2.1.1 DoA real y cálculo de distancias.....	22
Preprocesamiento de los datos.....	28

Implementación de las técnicas de Machine Learning .....	34
2.1.2    Regresión lineal .....	35
Machine learning en Rstudio y preparación de datos .....	36
2.1.3    Entrenamiento del modelo y evaluación .....	40
<b>3 RESULTADOS, CONCLUSIONES Y RECOMENDACIONES .....</b>	<b>44</b>
Resultados .....	44
3.1.1    Modelo 80-170 grados .....	46
3.1.2    Modelo de 65-80 grados.....	48
3.1.3    Modelo 0-65 grados .....	49
3.1.4    Modelo 170-250 grados .....	52
3.1.5    Modelo 250-305 grados .....	53
3.1.6    Modelo 305 a 360 grados .....	55
3.1.7    RESUMEN DE LOS MODELOS .....	58
Conclusiones .....	59
Recomendaciones .....	60
<b>4 REFERENCIAS BIBLIOGRÁFICAS.....</b>	<b>61</b>
<b>5 ANEXOS .....</b>	<b>63</b>

## RESUMEN

La tesis se centra en el desarrollo y evaluación de nuevos modelos para predecir los ángulos de dirección de arribo (DoA) en entornos reales de comunicaciones móviles utilizando mediciones de parámetros de radiofrecuencia y el método de regresión lineal como técnica de aprendizaje automático. Comienza con una exploración detallada del concepto de DoA y su importancia en aplicaciones inalámbricas, seguido por una descripción de la metodología utilizada, que incluye la recopilación de datos, el análisis del entorno de muestreo y las herramientas empleadas, con énfasis en Rstudio y librerías de Machine Learning relevantes. Los resultados obtenidos se presentan y analizan en profundidad, destacando la eficacia y utilidad de los modelos desarrollados para la estimación del DoA. Se examinan las ecuaciones resultantes y se discuten las variables implicadas en términos de mejoras potenciales en la precisión y eficacia de los sistemas de localización en entornos móviles.

**PALABRAS CLAVE:** DoA, Machine Learning, Regresión lineal, Modelo, Rstudio.



## **ABSTRACT**

The thesis focuses on the development and evaluation of new models to predict the angles of arrival direction (DoA) in real mobile communication environments using radiofrequency parameter measurements and linear regression as a machine learning technique. It begins with a detailed exploration of the DoA concept and its importance in wireless applications, followed by a description of the methodology used, which includes data collection, sampling environment analysis, and tools employed, with emphasis on Rstudio and relevant Machine Learning libraries. The obtained results are presented and analyzed in depth, highlighting the effectiveness and usefulness of the developed models for DoA estimation. The resulting equations are examined, and the variables involved are discussed in terms of potential improvements in the accuracy and effectiveness of location systems in mobile environments.

**KEYWORDS:** DoA, Machine Learning, Linear regression, Model, Rstudio.

# 1 INTRODUCCIÓN

En el ámbito de las comunicaciones móviles, la determinación precisa de la dirección de arribo (DoA) de una señal es fundamental para diversas aplicaciones, como la localización de dispositivos móviles, el seguimiento de objetos y la optimización de redes inalámbricas. Con el avance de la tecnología y el crecimiento exponencial de datos disponibles, se ha vuelto crucial emplear métodos inteligentes para modelar y predecir estos ángulos de arribo con alta precisión. En este contexto, el presente trabajo de investigación se enfoca en el desarrollo de modelos para la generación de ángulos de dirección de arribo mediante la inclusión de parámetros de entrada utilizando el método de regresión lineal en el marco del aprendizaje automático.

El primer capítulo de esta tesis aborda conceptos fundamentales que sientan las bases para comprender el objetivo y el alcance de esta investigación. Se comienza con una definición detallada del concepto de dirección de arribo (DoA), explorando sus aplicaciones en el contexto de las comunicaciones inalámbricas y la importancia de su precisión para diversas aplicaciones. Además, se discuten las diferentes técnicas para obtener el DoA, destacando la relevancia del uso de algoritmos de Machine Learning para mejorar la precisión de estas predicciones. Se ofrece una visión general de los tipos de aprendizaje automático, con especial énfasis en el método de regresión lineal, que se empleará en este estudio.

El segundo capítulo se centra en la metodología empleada para la recopilación de datos y el desarrollo de los modelos. Se describen en detalle el entorno donde se tomaron las muestras, las herramientas utilizadas para el procesamiento y análisis de datos, con énfasis en la plataforma Rstudio y las librerías específicas de Machine Learning empleadas en el estudio. Se profundiza en el significado y la selección de los parámetros incluidos en los modelos, detallando su relevancia y su contribución a la precisión de las predicciones de DoA.

En el tercer capítulo se presentan los resultados obtenidos tras la aplicación de los modelos desarrollados. Se describen los diferentes modelos generados, se analiza la eficacia de cada uno y se identifican las ecuaciones seleccionadas para su inclusión en Rstudio. Se examinan las gráficas generadas para validar la precisión de los modelos.

## **OBJETIVO GENERAL**

Desarrollar un modelo basado en una técnica de Machine Learning para la obtención del Direction of Arrival (DoA) utilizando datos recopilados a través de la aplicación Netmonitor Pro, con el propósito de evaluar su eficacia en la estimación del DoA.

## **OBJETIVOS ESPECÍFICOS**

1. Recopilar exhaustivamente datos de obtención del DoA mediante la aplicación Netmonitor, abarcando diversos escenarios y condiciones para lograr una representación completa.
2. Realizar un preprocesamiento riguroso sobre los datos recopilados, eliminando ruidos y asegurando la coherencia y calidad de los datos para el análisis subsiguiente.
3. Implementar y entrenar un modelo de aprendizaje automático supervisado, utilizando RStudio y las bibliotecas correspondientes.
4. Evaluar el desempeño del modelo de aprendizaje automático mediante métricas pertinentes como precisión, y error medio cuadrado.

## **ALCANCE**

El alcance de esta investigación se enfoca en el posible desarrollo de un modelo para la obtención del Direction of Arrival (DoA) a través de una técnica de Machine Learning en el entorno de RStudio. La comparación entre dos técnicas específicas no será abordada; en cambio, se busca profundizar en la evaluación de la eficacia de una técnica elegida en el contexto exclusivo del DoA.

La investigación estará dividida en tres etapas clave con el objetivo de lograr un análisis exhaustivo y robusto del proceso de obtención del DoA mediante el uso de una técnica de Machine Learning.

Primera Etapa: Se procederá con la recolección de datos para la obtención del DoA a partir de la aplicación Netmonitor Pro. Esta aplicación se selecciona debido a su capacidad de registrar datos pertinentes relacionados con las señales de red móvil y los parámetros de ubicación. Diversos escenarios de prueba se establecerán en ubicaciones y condiciones variadas para garantizar la representatividad de los datos recopilados. Utilizando los

sensores y antenas del dispositivo móvil, se registrarán datos de señales de red móvil, incluyendo la ubicación geográfica.

Segunda Etapa: La etapa siguiente involucra el procesamiento de los datos a través de RStudio y las bibliotecas de Machine Learning pertinentes. Se llevará a cabo el preprocesamiento de los datos con el fin de eliminar ruidos y asegurar la calidad y consistencia de la información. Se optará por una técnica específica de Machine Learning para la obtención del DoA. Esta técnica será implementada en RStudio utilizando las bibliotecas de Machine Learning supervisado correspondientes. Para el entrenamiento y la evaluación de los modelos, los datos se dividirán en conjuntos de entrenamiento y prueba.

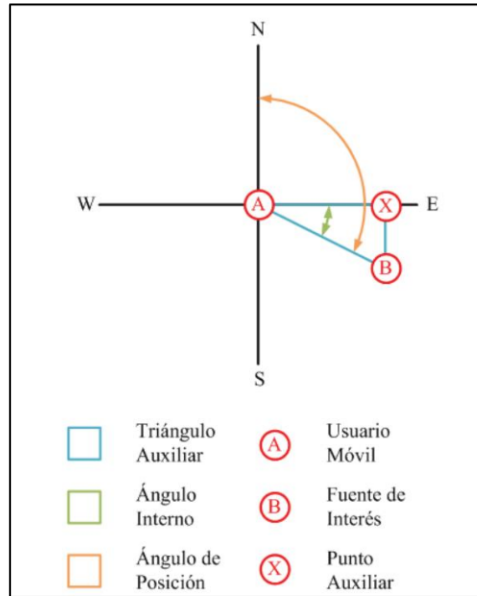
Tercera Etapa: La etapa final se centrará en el análisis de los resultados obtenidos. Se procederá al entrenamiento del modelo de Machine Learning utilizando los datos de entrenamiento, ajustando los parámetros para obtener la mejor eficacia posible. A continuación, se evaluará la eficacia de la técnica de Machine Learning seleccionada mediante el uso de los datos de prueba. Se medirán y compararán diversas métricas de rendimiento, incluyendo eficacia y error medio cuadrado. La comparación detallada de los resultados y el análisis de la técnica se llevarán a cabo considerando la calidad y las particularidades de los datos obtenidos a través de Netmonitor Pro.

## **MARCO TEÓRICO**

### **1.1.1 DOA**

La dirección de llegada, también conocida como Direction of Arrival (DOA), se refiere a la técnica empleada para estimar la fuente de origen de las señales inalámbricas que llegan a una antena receptora desde una antena emisora [1]. Este tipo de estimación adquiere relevancia en diversas aplicaciones, como la localización de objetos, detección de interferencias y comunicaciones inalámbricas, preferentemente aquellas de alta velocidad [2].

En el capítulo 2 de este trabajo se presenta un enfoque para calcular el ángulo de la dirección de arribo, abordando de manera detallada este proceso. La figura 1.1 muestra cómo se obtiene el ángulo de la dirección de arribo.



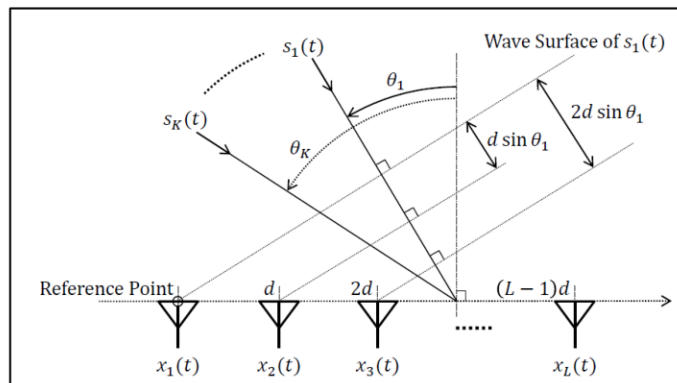
**Figura 1. 1.** Cálculo de la dirección de arriba mediante la utilización de mediciones de latitud y longitud [3].

## 1.1.2 TÉCNICAS APLICADAS PARA LOCALIZAR EL DOA

### 1.1.2.1 Redes neuronales profundas (DNN)

En el artículo "Performance Analysis of DOA Estimation of Two Targets Using Deep Learning," se ilustra en la figura 1.2 la aplicación de la técnica de Deep Learning (DDN) para la estimación de la dirección de llegada de señales inalámbricas. Durante el proceso de entrenamiento, se emplean diversas técnicas de aprendizaje, entre las que destacan el Aprendizaje por lotes y técnicas de optimización.

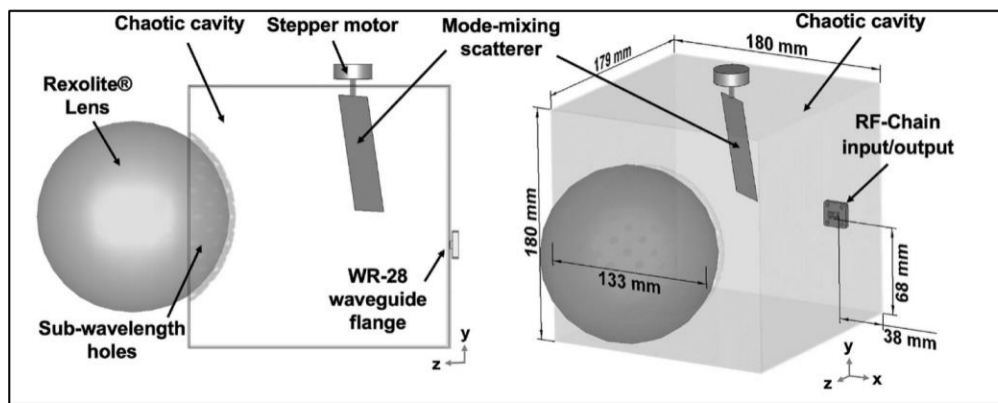
Los parámetros fundamentales para modelar este sistema incluyen la estructura de la antena, que abarca el número de elementos, la frecuencia de la señal, el SNR (relación señal a ruido), la cantidad de fuentes de señal, así como el número de capas intermedias y las unidades por capa intermedia utilizadas en la red neuronal profunda (DNN) [1].



**Figura 1. 2.** Arreglo lineal uniforme de L-elementos y ondas incidentes [1].

### 1.1.2.2 Accuracy Enhancement Technique using Oversized Lens-Loaded Cavity

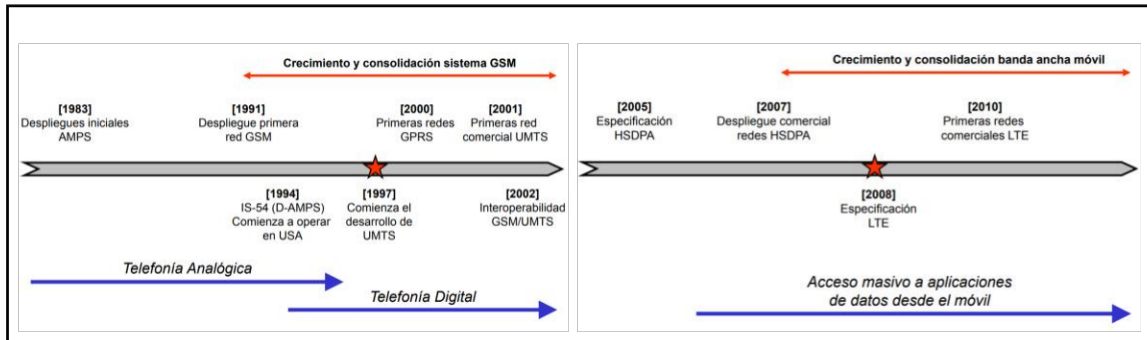
La metodología empleada en el artículo "Machine Learning-assisted direction-of-arrival accuracy enhancement technique using oversized lens-loaded cavity" se fundamenta en el uso de una antena de cavidad cargada con una lente sobredimensionada. Esta técnica incorpora la combinación de mecanismos de modos reconfigurables dentro de la antena con cavidad, como se ilustra en la figura 1.3 Posteriormente, se introduce un segundo paso que involucra una capa computacional donde se implementa el aprendizaje automático con el objetivo de optimizar la selección del estado óptimo de la antena de cavidad sobredimensionada [2].



**Figura 1. 3.** “Configuración geométrica de la apertura dinámica cargada con lente”[2]

### 1.1.3 COMUNICACIONES MÓVILES

El entendimiento previo sobre la tecnología abordada en este escrito es esencial y se remonta a la interpretación y evolución de los sistemas utilizados desde la antigüedad hasta el presente. La capacidad de realizar llamadas o llevar a cabo procesos de comunicación entre usuarios no ha sido posible sin la aparición de cada estructura celular a lo largo de los años. La figura siguiente evidencia cómo ha progresado a lo largo del tiempo este avance.

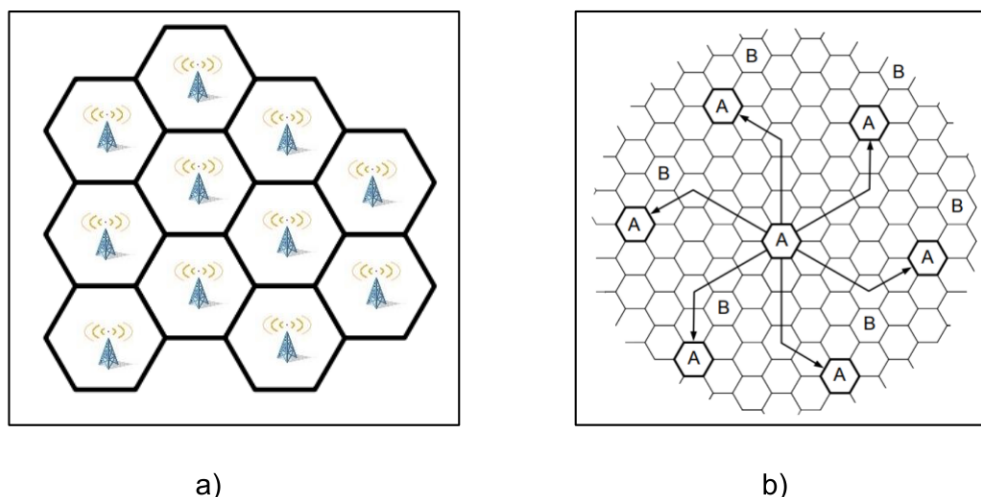


**Figura 1. 4.** Evolución de las comunicaciones móviles [4].

La gestión de estos sistemas de comunicación se ha llevado a cabo mediante la implementación de redes celulares, que posibilitan el intercambio de información entre diversas estaciones (BS - Base stations) [5]. Este proceso guarda similitud con el intercambio entre las células de un organismo vivo. De manera análoga, las células biológicas emiten mensajeros químicos que se desplazan a través del líquido extracelular hasta alcanzar una célula diana, conocida como célula receptora [6]. En los sistemas celulares, una estación base envía un mensaje a través de su medio de disipación (generalmente aire) hasta llegar a otra estación, la cual retransmite el mensaje al receptor final (usuario).

La transmisión de estos mensajes se lleva a cabo mediante el empleo de señales de radiofrecuencia, las cuales ocupan rangos específicos de frecuencias. Dado que las frecuencias no son infinitas, un grupo de usuarios no puede utilizarlas de manera constante. Es en este contexto donde surge el concepto de "reutilización de frecuencias". Para lograr la reutilización eficiente de las frecuencias, se han realizado estudios que indican el espaciamiento adecuado entre las estaciones base con el fin de evitar interferencias.

La Figura 1.5a) ilustra la estructura de un sistema celular, mientras que la Figura 1.5b) muestra cómo se realiza la reasignación (reúso) de frecuencias entre celdas vecinas. Al conjunto de celdas que comparten frecuencias únicas se les denomina "Cluster". Además, es importante destacar que las celdas adoptan una forma hexagonal debido a la mayor área de cobertura que proporcionan.



**Figura 1. 5. a) Sistema celular; b) reutilización de frecuencias [4].**

### 1.1.4 LTE

A lo largo de los años y la evolución de la tecnología celular se ha presentado a LTE (Long Term Evolution) como una solución para cubrir necesidades como: mayor velocidad para la carga y descarga de datos, reducción de costos al aplicar estándares con menor complejidad.

En la tabla 1.1 se muestra la descripción de cada tipo de tecnologías LTE con sus velocidades de subida y bajada.

**Tabla 1. 1. Velocidades de DL y UL para las diferentes tecnologías de LTE y LTE-A [7].**

<b>Estándar</b>	<b>Descripción</b>	<b>Velocidad DL (Mbps)</b>	<b>Velocidad UL (Mbps)</b>
LTE (Cat 3)	Long Term Evolution	100	50
LTE (Cat 4)	Long Term Evolution	150	50
LTE (Cat 5)	Long Term Evolution	300	75
LTE - A (Cat 6)	Long Term Evolution Advanced	300	50
LTE - A (Cat 7)	Long Term Evolution Advanced	300	150
LTE - A (Cat 8)	Long Term Evolution Advanced	1200	600



### 1.1.4.1 Bandas de Frecuencias

LTE posee un grupo de 25 bandas de frecuencia, de las cuales 17 son pares y 8 son impares, en la actualidad existe una coexistencia entre algunas tecnologías con LTE, estas bandas se muestran en las tablas 1.2 y 1.3.

**Tabla 1. 2.** Frecuencias en bandas pares de UL y DL para LTE [8] .

Banda de operación	Nombre según 3GPP	Frecuencia UL	Frecuencia DL
1	2100	1920 MHz - 1980 MHz	2110 MHz - 2170 MHz
2	1900	1850 MHz - 1910 MHz	1930 MHz - 1990 MHz
3	1800	1710 MHz - 1785 MHz	1805 MHz - 1880 MHz
4	1700/2100	1710 MHz - 1755 MHz	2110 MHz - 2155 MHz
5	850	824 MHz - 849 MHz	869 MHz - 894 MHz
6	800	830 MHz - 840 MHz	875 MHz - 850 MHz
7	2600	2500 MHz - 2570 MHz	2620 MHz - 2690 MHz
8	900	880 MHz - 915 MHz	925 MHz - 960 MHz
9	1700	1750 MHz - 1785 MHz	1845 MHz - 1880 MHz
10	1700/2100	1710 MHz - 1770 MHz	2110 MHz - 2170 MHz
11	1500	1427.9 MHz - 1452.9 MHz	1475.9 MHz - 1500.9 MHz
12	US 700	698 MHz - 716 MHz	728 MHz - 746 MHz
13	US 700	777 MHz - 787 MHz	746 MHz - 756 MHz
14	US 700	788 MHz - 798 MHz	758 MHz - 768 MHz
17	US 700	704 MHz - 716 MHz	734 MHz - 746 MHz
18	Japón 800	815 MHz - 830 MHz	860 MHz - 875 MHz
19	Japón 800	830 MHz - 845MHz	875 MHz - 890MHz

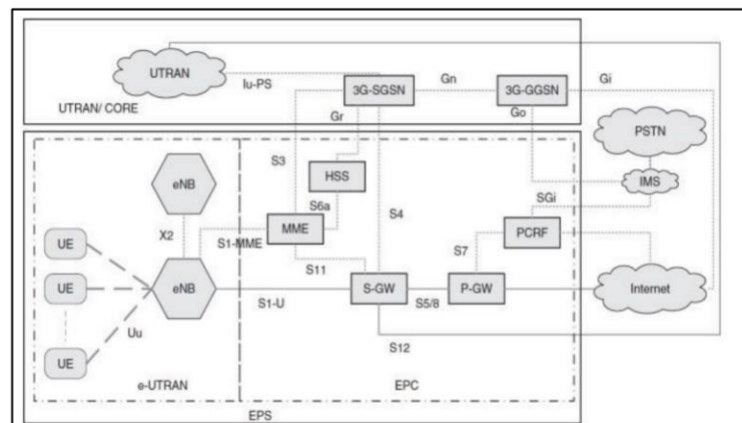
**Tabla 1. 3.** Frecuencias en bandas impares de UL y DL para LTE [8] .

Banda de operación	Nombre según 3GPP	Frecuencia UL y DL
33	UMTS TDD1	1900 MHz - 1920 MHz
34	UMTS TDD2	2010 MHz - 2025 MHz
35	US 1900 UL	1850 MHz - 1910 MHz
36	US 1900 DL	1930 MHz - 1990 MHz

37	US 1900	1910 MHz - 1930 MHz
38	2600	2570 MHz - 2620 MHz
39	UMTS TDD	1880 MHz - 1920 MHz
40	2300	2300 MHz - 2400 MHz

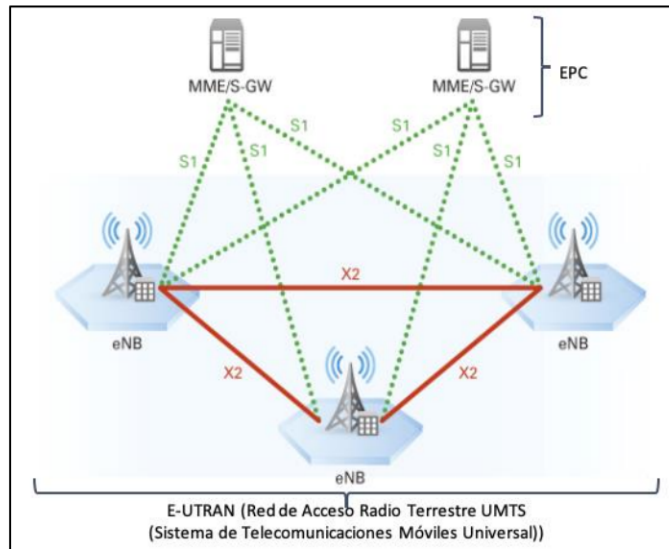
### 1.1.4.2 Arquitectura LTE

Se presenta una estructura compuesta principalmente por tres elementos clave: la red de acceso (E-UTRAN), la red de core (EPC), y IMS. Un cuarto elemento a considerar es el dispositivo utilizado por el usuario final. Tanto E-UTRAN como EPC proporcionan servicios diseñados para la transferencia eficiente de información a través de paquetes IP. La figura 1.6 ilustra los componentes dentro de la arquitectura de LTE, incluyendo aquellos necesarios para la interconexión con otras tecnologías, aunque no sean propios de esta especificación [9].



**Figura 1. 6.** Arquitectura completa de LTE [9].

La estructura física de la E-UTRAN está compuesta principalmente por evolved Node B (eNBs), los cuales proporcionan conectividad al usuario. La configuración de la EPC se presenta en la figura 1.7, mostrando dicha estructura.



**Figura 1. 7.** Red de acceso (E-UTRAN) [10].

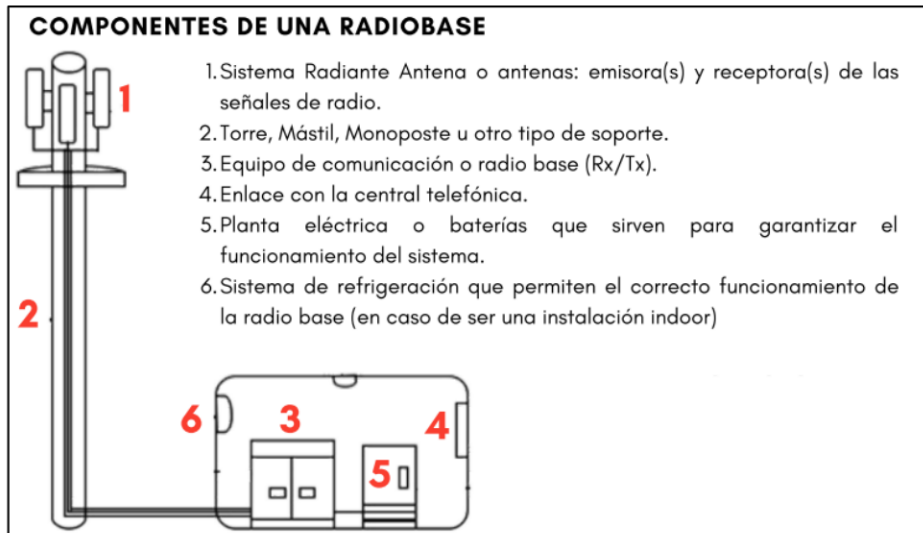
### 1.1.4.3 IMS

El Subsistema Multimedia sobre IP (IMS) es un mecanismo que facilita la provisión basada en el protocolo IP para la comunicación multimedia en la red LTE dirigida a los usuarios. La infraestructura implementada posibilita la operación de servicios como voz, video, SMS, entre otros, a través del protocolo IP. El protocolo de señalización especificado por 3GPP para LTE es SIP. La estructura se fundamenta en capas que abarcan transporte, control y aplicación [8].

### 1.1.5 ESTACIÓN BASE

Se define como una estación radioeléctrica fija que opera bajo el control directo desde un punto específico. Este control puede llevarse a cabo localmente o a través de líneas telefónicas o enlaces de radio. Su característica principal radica en su capacidad para funcionar como estación emisora o receptora de tráfico y señalización [11].

En la Figura 1.8 se presentan de manera general los componentes de una estación base.



**Figura 1. 8.** Componentes de una estación base [12].

### 1.1.6 MACHINE LEARNING

Comprender que los sistemas o aplicaciones "inteligentes" se basan en un conjunto de decisiones formuladas mediante condiciones "si ocurre" o "si no ocurre", previamente codificadas o programadas por una persona. Sin embargo, surge la pregunta de qué sucedería al intentar clasificar automóviles según sus colores. El proceso de programación sería sencillo al separar entre una cantidad pequeña de colores, pero si se quisiera ampliar esto a un gran número, la complejidad aumentaría, requiriendo más tiempo para llevar a cabo dicha tarea.

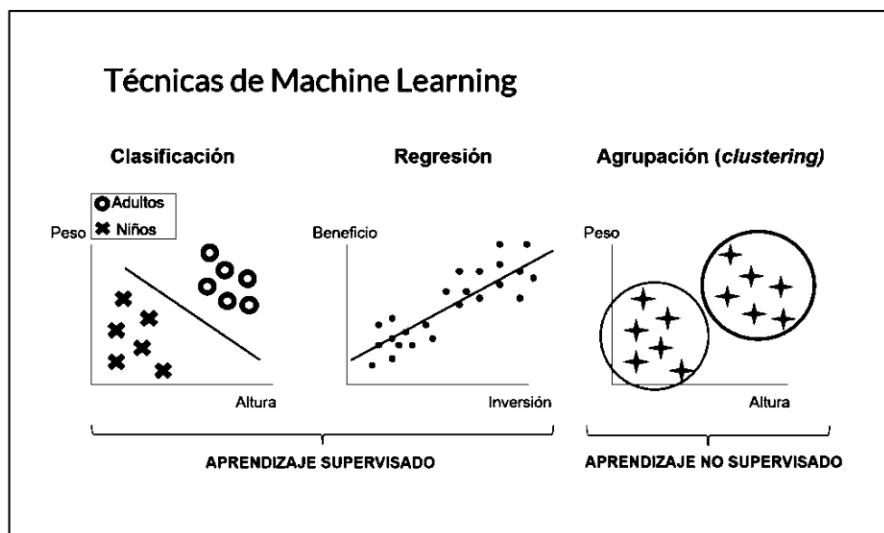
Es esencial comprender que el mundo está evolucionando hacia nuevas tecnologías gobernadas por inteligencia artificial, lo que nos lleva a preguntarnos qué tipo de aprendizaje o inteligencia artificial puede optimizar procesos, ya sean tecnológicos, médicos o aquellos que se llevan a cabo en la vida diaria de una persona.

Tipos de aprendizaje automático:

- **Aprendizaje supervisado:** Basado en el aprendizaje predictivo, maneja un conjunto de datos de entrada y salida, generando resultados a partir de los valores o características ingresadas.
  - **Clasificación:** Permite determinar si un grupo comparte ciertas características o rasgos similares a otros; un ejemplo es el reconocimiento facial.

- **Regresión:** Localiza un valor específico (numérico) en algún estudio realizado, como encontrar el precio de una computadora según ciertas características.
- **Aprendizaje no supervisado:** Se basa en el conocimiento de valores o características de entrada, sin contar con un valor específico o esperado de salida.
  - **Clustering:** También conocido como agrupamiento, crea conjuntos de datos basados en similitudes presentadas en los valores de entrada.
  - **Reducción de dimensionalidad:** Permite la eliminación de redundancia en las variables a analizar.
  - **Asociación:** Descubre patrones entre los datos ingresados.
  - **Detección de anomalías:** Realiza un análisis entre los datos de entrada y localiza instancias atípicas o inusuales.
  - **Generación de Datos:** Crea nuevos datos a partir de los ingresados para obtener información suficiente sobre un problema determinado [13].

En la figura 1.9 se presenta un ejemplo de las técnicas de Machine Learning más sencillas de estudiar.



**Figura 1. 9.** Técnicas de Machine Learning más sencillas [14].

Además de las técnicas ya mencionadas también existe una técnica conocida como aprendizaje por refuerzo la cual se basa en una interacción continua del modelo con un entorno dinámico. Esta se encuentra fundamentada en el aprendizaje mediante la

retroalimentación de recompensas, es decir mediante parámetros llamados agentes se realiza un control sobre si se está realizando bien o mal una tarea encomendada del aprendizaje. Este agente es el encargado de evaluar la tarea realizada y generar una recompensa positiva, negativa o neutra (en esta el agente no presenta un impacto significativo). La información proporcionada por el agente permite que el modelo mejore el desempeño, evitando acciones que conduzcan a resultados negativos [15].

### 1.1.7 R Y RSTUDIO

R se utiliza para llevar a cabo análisis estadísticos y generar gráficas correspondientes. RStudio, por otro lado, es el entorno de desarrollo integrado (IDE) que facilita la interacción con R mediante una interfaz gráfica amigable para el usuario. Este entorno posibilita la creación de diversos modelos estadísticos, desde los más simples, como la regresión lineal, hasta aquellos de mayor complejidad, como las redes neuronales [16].

Entre las principales capacidades que posee R se encuentra [17]:

- La automatización se logra de manera sencilla mediante la creación de scripts que realizan procesos como la lectura y operaciones de datos de forma automatizada.
- Tiene la capacidad de procesar prácticamente cualquier tipo de datos.
- Presenta funciones avanzadas de representación gráfica, posibilitando la creación de gráficos y dashboards para presentar resultados de manera visualmente atractiva.
- Es compatible con diversas plataformas.

La figura 1.10 muestra las partes principales de Rstudio.

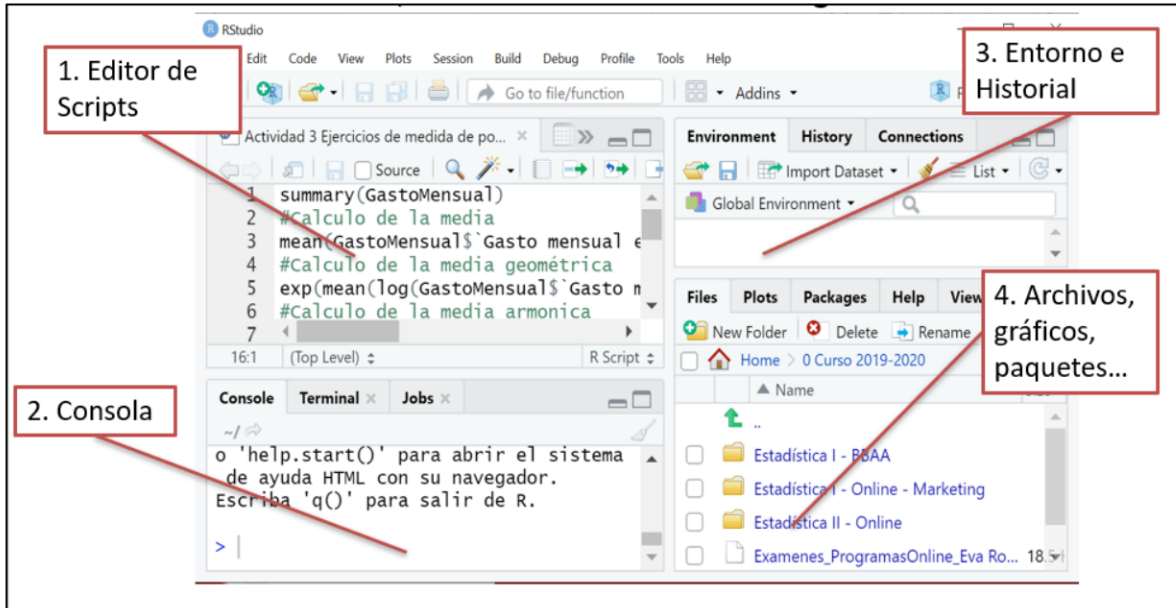


Figura 1. 10. Rstudio y sus partes en la interfaz gráfica [18].

## 2 METODOLOGÍA

### HERRAMIENTA DE RECOLECCIÓN DE DATOS

#### 2.1.1 NETMONITOR PRO

Net monitor es una aplicación diseñada para monitorear y analizar redes en dispositivos móviles, especialmente en teléfonos Android. Esta app permite realizar un análisis exhaustivo del tráfico de la red, la velocidad de conexión, el uso de datos e incluso evalúa la calidad de la señal Wi-Fi, resaltando estos parámetros en su interfaz principal. Además, ofrece gráficos detallados que se superponen en un mapa de Google Maps, mostrando la ruta seguida durante el análisis. Esta aplicación proporciona al usuario herramientas para mejorar y optimizar la conexión de red, buscando un mejor rendimiento y una experiencia más fluida [19].

Entre las características principales de Net Monitor Pro para la recolección de datos en este trabajo investigativo se encuentran:

*Nombre:* Net Monitor Pro

*Tamaño:* 23.96 MB

*Versión:* 1.81

*Formato de archivos:* CSV, KML

La función utilizada para recopilar datos en trabajos escritos es el análisis de tráfico, que permite examinar la información recolectada en intervalos de tiempo predeterminados, tomando muestras cada segundo. Aunque las evidencias de monitoreo (figura 2.1.a)) se obtienen en tiempo real, no son tan claras como las del análisis, donde se puede tener un registro de múltiples muestras tomadas según el tiempo solicitado por el usuario (figura 2.1.b) ). La figura 2.1. a) muestra que la opción de monitoreo solo permite verificar la celda de servicio actual del usuario y los niveles de potencia de diferentes tecnologías (LTE, GSM entre las más comunes).





a)



b)

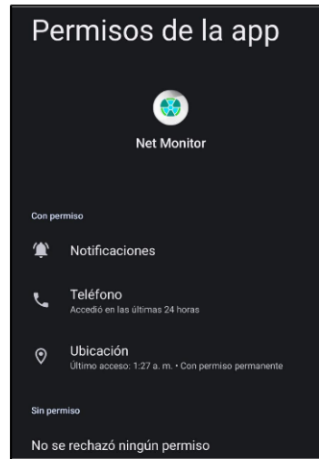
**Figura 2. 1.** NetMonitor Pro. a) función de monitoreo; b) Función de análisis  
Entre los principales parámetros que permite recolectar la aplicación se encuentran:

**Tabla 2. 1.** Principales parámetros generados por la aplicación Net monitor Pro [20].

Característica	Descripción
sim_state	Estado de la tarjeta SIM en el dispositivo.
net_type	Tipo de red actual (por ejemplo, LTE, UMTS, GSM).
data_state	Estado de la conexión de datos.
data_act	Actividad de la conexión de datos.
gsm_neighbors	Vecinos de red GSM.
umts_neighbors	Vecinos de red UMTS.
lte_neighbors	Vecinos de red LTE.
nr_neighbors	Vecinos de red NR (por ejemplo, 5G).
cdma_neighbors	Vecinos de red CDMA.
rsi_strongest	Intensidad de la señal más fuerte.
nstrong	Número de señales fuertes.
mcc	Código de país móvil.
mnc	Código de red móvil.
node_id_nid	ID de nodo o identificación de nodo.
cid_bid	ID de celda o ID de estación base.
psc_pci	Identificador de código de servicio primario o Identificador de código de identidad de célula.
rsi	Indicador de la intensidad de la señal.
rsrq	Relación señal-ruido cuadrada del canal.
rssnr	Relación señal-ruido.
ta	Tiempo de avance.
gps	Estado del sistema de posicionamiento global.
accuracy	Precisión de la ubicación GPS.

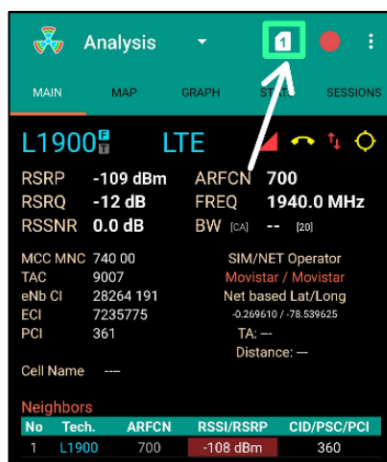
La forma principal de recolectar los datos dentro de la aplicación se presenta:

1. Garantizar los permisos concedidos a la aplicación para permitir la captura de datos sobre cualquier otra aplicación (ver figura 2.2).

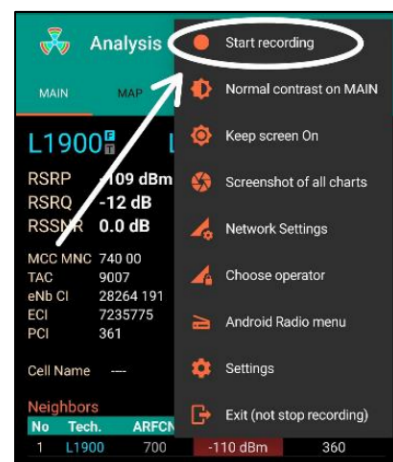


**Figura 2. 2** Permisos de la aplicación.

2. En Net Monitor, es crucial verificar la selección de la tarjeta SIM que se usará para las pruebas (en caso de dispositivos de doble SIM), como se ilustra en la figura 2.3 a). Para iniciar las pruebas, se puede presionar el icono rojo que se muestra en la figura 2.3 a) o seleccionar la opción 'start recording' en el menú de los tres puntos (ver figura 2.3 b)).



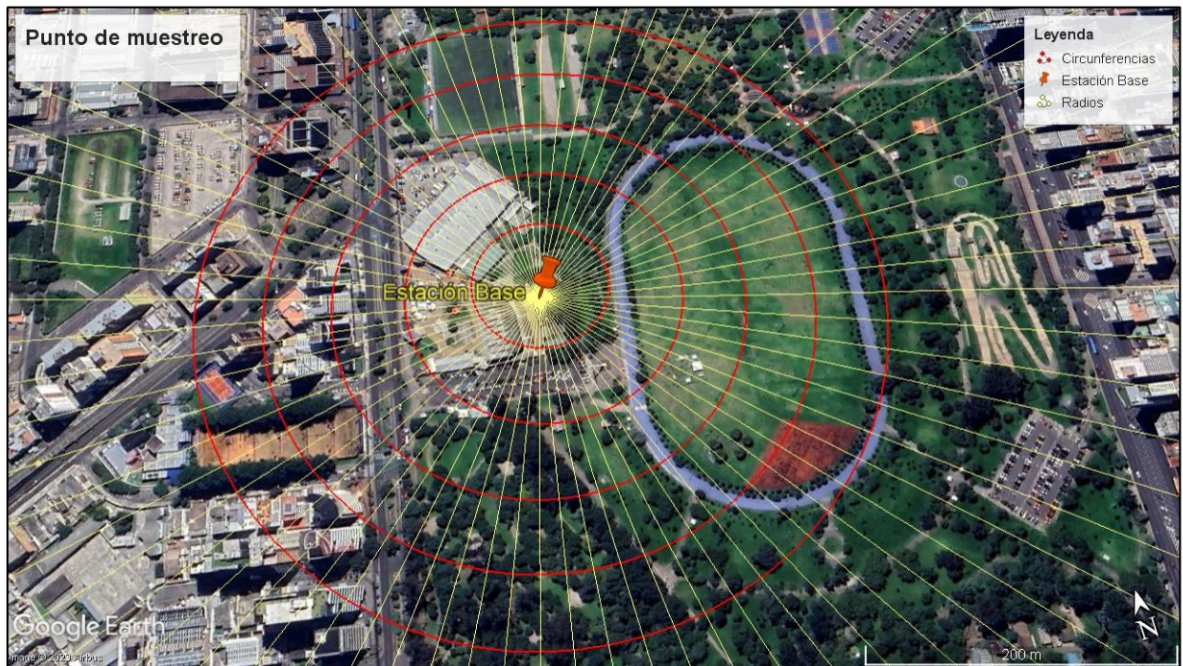
a)



b)

**Figura 2. 3. a)** Selección de la Sim para la prueba. **b)** Comenzar la grabación y toma de muestras dentro de la aplicación. Tomadas de NetMonitor.

3. Seguir el rastro generado en Google Earth, tal como se indica en la figura 2.4.



**Figura 2. 4.** Mapa del rastro para las mediciones realizadas en Google earth. Tomada de Google Earth.

## ETAPA DE RECOLECCIÓN DE DATOS

La recolección de datos facilita la recopilación de información relevante y necesaria para cumplir con el propósito de tomar muestras que se utilizarán en el modelo descrito en este documento.

Se pudo tomar los datos siguiendo los siguientes lineamientos:

1. **Etapas de planificación:** En este apartado se debe verificar los datos que se utilizarán para proponer un modelo de obtención del DoA mediante parámetros de radiofrecuencia. También se establecieron objetivos claros y se seleccionó un método adecuado de recolección.
2. **Etapas de Captura de datos:** Realización de mediciones de parámetros de radiofrecuencia y análisis de la red correspondiente.

### 2.1.2 PLANIFICACIÓN

Objetivo de la recolección de datos

Identificar el lugar idóneo para la recolección de datos, capturando todos los parámetros de radiofrecuencia necesarios para proponer un modelo de obtención del DoA. Además, se verificará la toma de muestras adecuadas en el sector seleccionado de la red.

En este momento se realizó todo un esquema de las actividades a desarrollarse y los recursos que fueron necesarios para la toma de las muestras, de esa manera también se ideó en qué manera se realizó el desplazamiento en el espacio a realizarse, esto se puede evidenciar en la figura 2.4.

Los principales parámetros recolectados que permitieron un análisis del ángulo del DoA:

- lte\_neighbors
- rssi
- rssi strongest
- rsrq
- rssnr
- latitude
- longitude
- altitude
- bearing

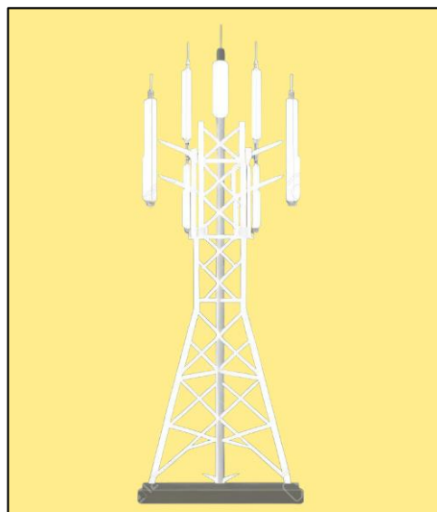
### **2.1.3 ZONAS Y RUTAS UTILIZADAS EN LA MEDICIÓN**

Las mediciones se realizaron en la ciudad de Quito, eligiendo un lugar con amplia visibilidad para llevar a cabo la toma de datos a pie. De esta manera, se logró obtener las muestras utilizando un dispositivo móvil previamente configurado con la aplicación seleccionada para la toma de muestras; el móvil utilizado fue un Xiaomi Note 12S con Android 14. En la figura 2.2.1 se observa la antena base desde la cual se llevaron a cabo las mediciones.



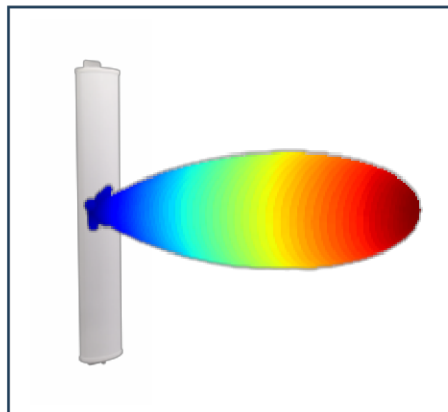
**Figura 2. 5.** Estación base para la toma de mediciones.

La estación base representada en la figura anterior hace referencia a aquella mostrada en la figura 2.6. En el capítulo 3, se presentará una explicación detallada sobre el uso de cinco modelos, cada uno de los cuales se construirá según la cantidad de antenas indicadas. De este modo, se podrá especificar cómo los lóbulos de radiación influyen en la captura de información por las muestras, manteniendo un patrón único para cada antena que sigue irradiando potencia hacia sus respectivos sectores.



**Figura 2. 6.** Estación base utilizada para la toma de muestras.

La estación presenta las siguientes características:



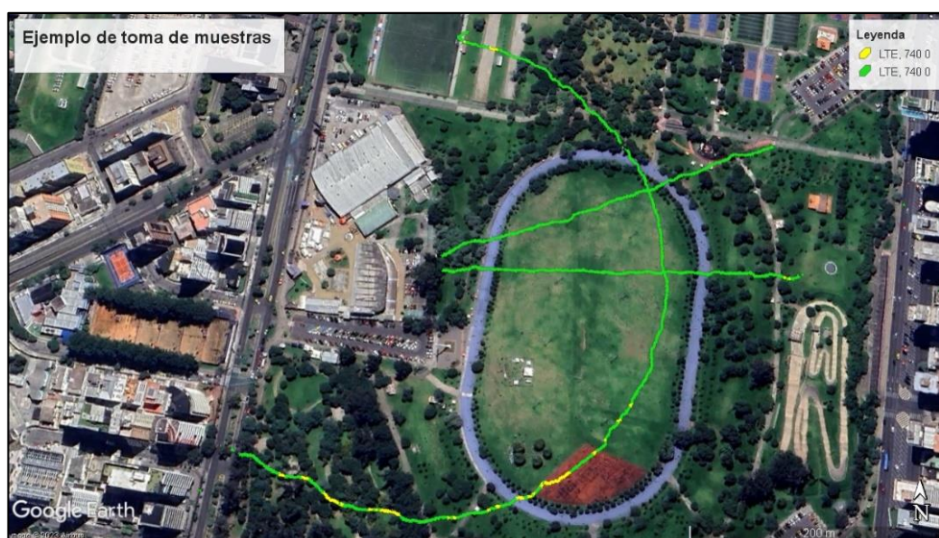
**Figura 2. 7.** Lóbulo de radiación estimado según percepción de cada antena colocada en la estación base.

Altura de la estación base: Aproximadamente 35 metros.

Sectorización: Se estima, según la apreciación visual, que la estación cuenta con 5 sectores, dado el número de antenas dispuestas en la figura 2.6.

Apuntamiento: En la figura 2.7, se realiza una estimación del lóbulo de radiación para cada antena base, evidenciando su carácter direccional.

Las rutas seguidas se presentan en la figura 2.4 Las muestras se recolectaron de manera radial cada 5° y en círculos concéntricos en radios de 50 m, 100 m y 150 m. Se muestra un ejemplo de la toma de muestras en la figura 2.8 capturada a través de la aplicación Net Monitor Pro.



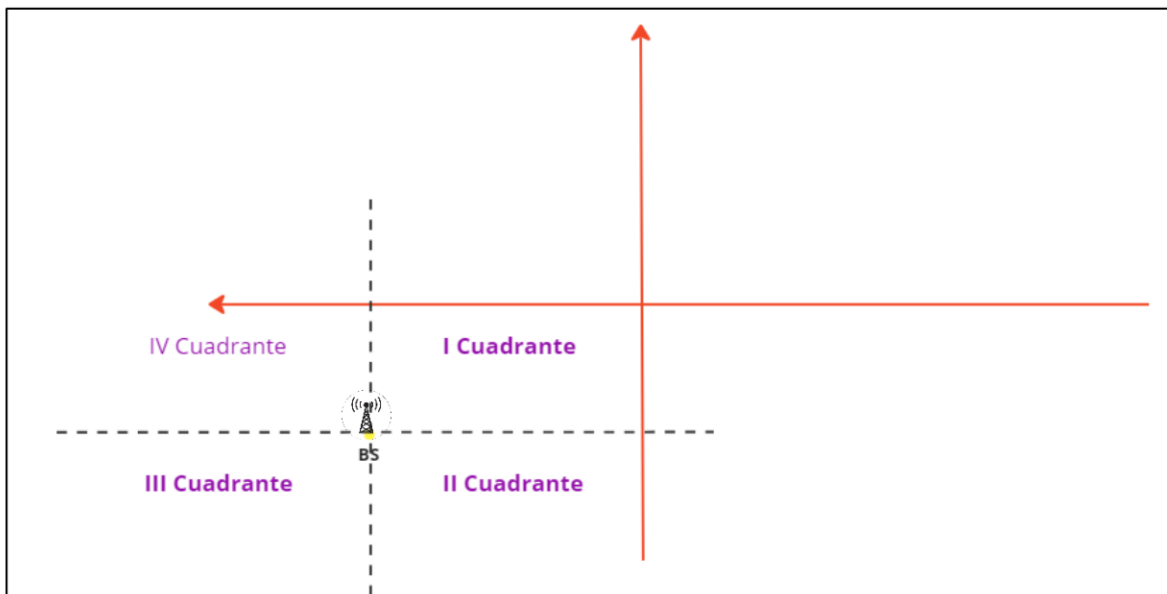
**Figura 2. 8.** Ejemplo de toma de muestras en la aplicación Google Earth.

Las muestras fueron tomadas mediante el cronograma mostrado en la tabla 5.1 mostrada en el Anexo I.

## 2.1 PARÁMETROS ADICIONALES UTILIZADOS EN LA BASE DE DATOS

### 2.1.1 DOA REAL Y CÁLCULO DE DISTANCIAS

Los parámetros adicionales implementados en la base de datos proveniente de la aplicación NetMonitor Pro incluyeron una columna para registrar el ángulo que indica la dirección de arribo, referenciado desde el punto cardinal norte. Esta medida se ajusta a la ubicación de la estación base para facilitar las mediciones en sentido horario. La principal dificultad encontrada radicó en calcular los ángulos a partir de la latitud y longitud de la estación base, así como determinar la muestra a ser tomada en cada medición. Para abordar estas incógnitas, se llevó a cabo un análisis de los cuatro cuadrantes formados alrededor de la estación base, tal como se ilustra en la figura 2.9, este análisis se basó en el artículo “Desarrollo de Métodos de Seguimiento para un Sistema de Antena Inteligente basado en GPS” [3].

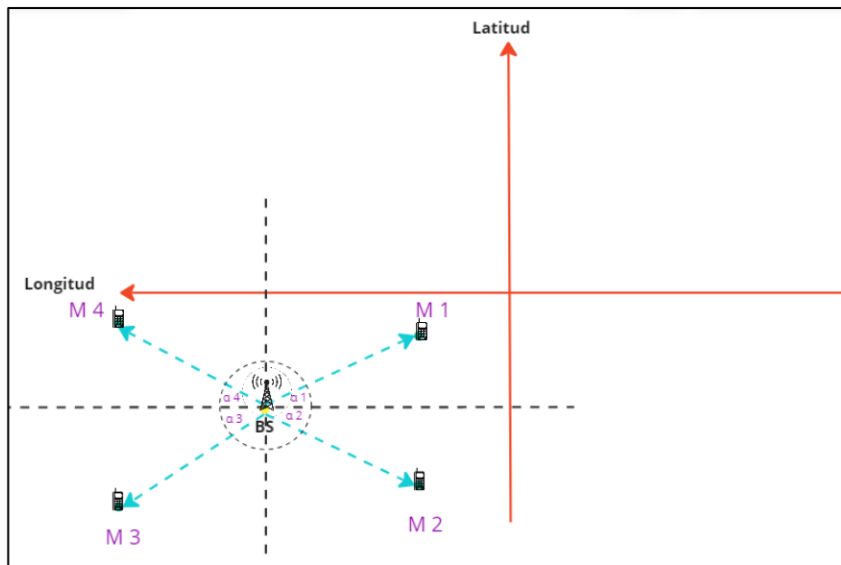


**Figura 2. 9.** Sistema de referencia colocado sobre la estación base analizada.

Nota: los cuadrantes se identifican siguiendo la disposición presentada en la figura 2.9, considerando que los ángulos se registran en sentido horario. La estación base se posiciona en el sistema de referencia marcado en rojo en el III cuadrante, debido a que su

ubicación geográfica exhibe una latitud al sur de la línea ecuatorial y una longitud al oeste del meridiano de Greenwich.

El análisis se dividió en cuatro casos, representados en la figura 2.10, que ilustra la distribución de las muestras en cada cuadrante, siguiendo el orden previamente mencionado en sentido horario. El valor del ángulo  $\theta$  (utilizado para el cálculo de la dirección de arribo) se determina según las especificaciones de los casos mencionados en la tabla 2.2.



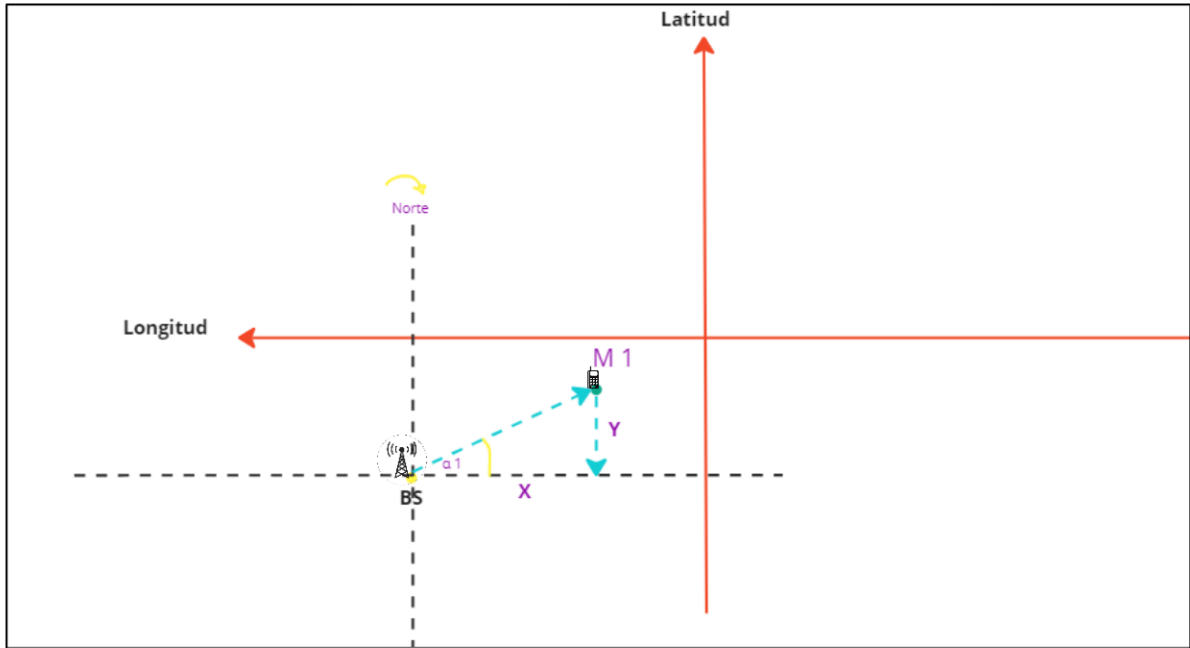
**Figura 2. 10.** Casos de estudio para los ángulos del DoA.

**Tabla 2. 2.** Consideraciones a tener en cuenta para el análisis del ángulo de la dirección de arribo real.

Caso	Consideraciones	Valor de $\theta$
1	$BS_{LAT} < M_{LAT} \wedge BS_{LON} < M_{LON}$	$\theta = 90^\circ - \alpha$
2	$BS_{LAT} > M_{LAT} \wedge BS_{LON} < M_{LON}$	$\theta = 90^\circ + \alpha$
3	$BS_{LAT} > M_{LAT} \wedge BS_{LON} > M_{LON}$	$\theta = 270 - \alpha$
4	$BS_{LAT} < M_{LAT} \wedge BS_{LON} > M_{LON}$	$\theta = 270 + \alpha$

Los valores de los ángulos  $\alpha$  (figura 2.10) en cada cuadrante se calculan utilizando las funciones trigonométricas de los ángulos formados en los triángulos rectángulos correspondientes. Esta metodología se visualiza en la figura 2.11, la cual demuestra cómo realizar estos cálculos utilizando las distancias derivadas de la latitud y longitud de la estación base, junto con el punto de muestra.





**Figura 2. 11.** Triángulo rectángulo utilizado para poder obtener el valor del ángulo  $\alpha$ .

El ángulo  $\alpha$  se obtiene mediante la utilización de la función tangente:

$$\tan(\alpha) = \frac{Y}{X} = \frac{Dist(\Delta LAT)}{Dist(\Delta LON)}$$

**Ecuación 2. 1.** Valor del ángulo comprendido por la  $\Delta LAT$  y  $\Delta LON$

Los valores de Y y X se calculan mediante la fórmula de Haversine, utilizada para determinar la distancia entre dos ubicaciones basadas en sus respectivas latitudes y longitudes. Es importante tener en cuenta que, para ubicaciones situadas al sur de la línea ecuatorial, se emplea un signo negativo para la latitud, y de manera similar, para los meridianos al oeste del meridiano de Greenwich, se usa un signo negativo para la longitud. Regresando a la explicación, la fórmula de Haversine se representa en la Ecuación 2.2 y presenta una variación para calcular la distancia entre las variaciones de latitud y longitud.

$d(Muestra, BS)$

$$= 2R \text{sen}^{-1} \left( \sqrt{\text{sen}^2 \left( \frac{Muestra_{LAT} - BS_{LAT}}{2} \right) + \cos(Muestra_{LAT}) \cos(BS_{LAT}) \text{sen}^2 \left( \frac{Muestra_{LON} - BS_{LON}}{2} \right)} \right)$$

**Ecuación 2. 2.** Ecuación de la distancia entre dos ubicaciones de Haversine [21].

Es importante tener en cuenta que los ángulos de latitud y longitud deben expresarse en radianes, y el valor de R corresponde al radio de la Tierra, aproximadamente 6378 kilómetros. La Ecuación 2.2 presenta una variante para calcular la distancia entre  $\Delta LAT$  y  $\Delta LON$ , reflejada en las Ecuaciones 2.3 y 2.4.

$$d(\Delta LAT) = 2R \text{sen}^{-1} \left( \sqrt{\text{sen}^2 \left( \frac{Muestra_{LAT} - BS_{LAT}}{2} \right)} \right)$$

$$d(\Delta LAT) = 2R \text{sen}^{-1} \left( \left| \text{sen} \left( \frac{Muestra_{LAT} - BS_{LAT}}{2} \right) \right| \right)$$

**Ecuación 2. 3.** Ecuación para obtener la  $\Delta LAT$ .

$$d(\Delta LON) = 2R \text{sen}^{-1} \left( \left| \text{sen} \left( \frac{Muestra_{LON} - BS_{LON}}{2} \right) \right| \right)$$

**Ecuación 2. 4.** Ecuación para obtener la  $\Delta LON$ .

Sustituyendo las Ecuaciones 2.3 y 2.4 en la Ecuación 2.1 se tiene la ecuación para encontrar el ángulo  $\alpha$ .

$$\tan(\alpha) = \frac{2R \text{sen}^{-1} \left( \left| \text{sen} \left( \frac{Muestra_{LAT} - BS_{LAT}}{2} \right) \right| \right)}{2R \text{sen}^{-1} \left( \left| \text{sen} \left( \frac{Muestra_{LON} - BS_{LON}}{2} \right) \right| \right)}$$

$$\tan(\alpha) = \frac{\text{sen}^{-1} \left( \left| \text{sen} \left( \frac{Muestra_{LAT} - BS_{LAT}}{2} \right) \right| \right)}{\text{sen}^{-1} \left( \left| \text{sen} \left( \frac{Muestra_{LON} - BS_{LON}}{2} \right) \right| \right)}$$

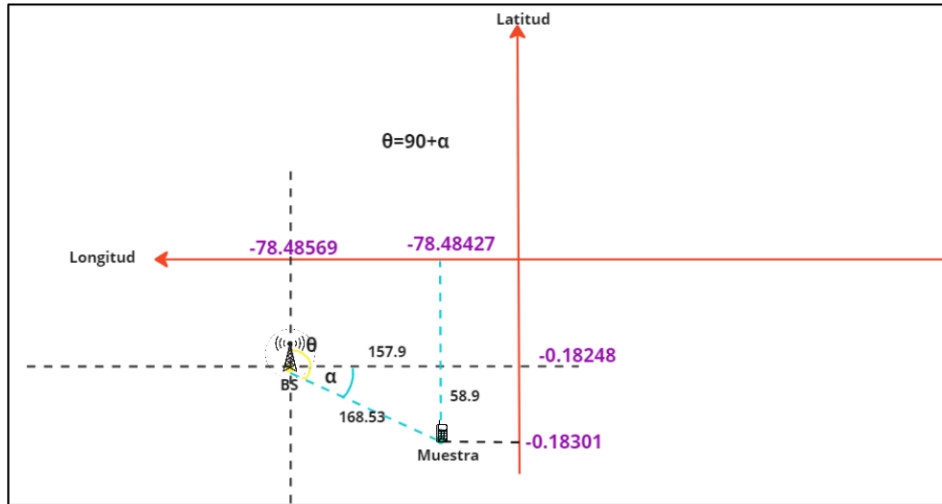
**Ecuación 2. 5.** Ángulo entre la  $\Delta LAT$  y  $\Delta LON$

Un ejemplo práctico del cálculo del ángulo de arribo (DoA) y la distancia entre una muestra y la estación base se ilustra con las siguientes ubicaciones:

Muestra: Latitud -0.18301; Longitud -78.48427

Estación Base (BS): Latitud -0.18248; Longitud -78.48569

Las variaciones en la latitud y longitud se determinan utilizando las Ecuaciones 2.3 y 2.4, obteniendo los valores correspondientes que se presentan en la figura 2.12.

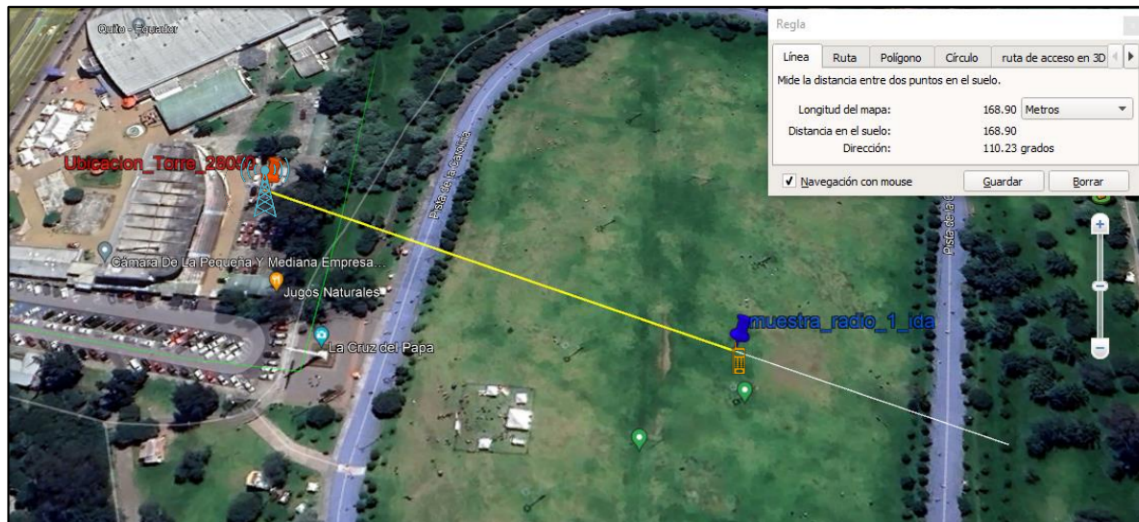


**Figura 2. 12.** Ejemplo para la obtención del DOA.

Ya que solo se está calculando el ángulo  $\alpha$ , el ángulo de llegada (DoA) se determina observando que corresponde al caso 2 de la tabla 2.2, lo que indica que se debe agregar un incremento de  $90^\circ$ .

$$\theta = 90 + \tan^{-1}\left(\frac{58.9}{157.9}\right) = 110.456^\circ$$

La figura 2.13 muestra el valor del ángulo de llegada (DoA) obtenido a través de la aplicación Google Earth.



**Figura 2. 13.** DoA real obtenido mediante la aplicación Google Earth.

El error porcentual registrado es del 0.205 %, lo que indica que el valor obtenido puede considerarse confiable para llevar a cabo el análisis correspondiente.

El cálculo del DoA se ha llevado a cabo de manera manual para determinar el error porcentual. Dado el volumen considerable de datos, realizar todos los cálculos de esta manera resultaría en una pérdida de tiempo considerable. Por esta razón, los datos adicionales fueron calculados mediante el uso de Microsoft Excel. En este programa, se generaron múltiples columnas adicionales, evitando así la necesidad de utilizar una única ecuación para el DoA, como se explicó anteriormente.

En la siguiente tabla se presentan los parámetros adicionales utilizados. La columna "Cuadrante" muestra la selección del cuadrante según lo especificado en la tabla 2.3. Las columnas "d(Variación\_latitud)" y "d(Variación\_Longitud)" indican las distancias entre las variaciones de latitud y longitud. La columna "Ángulo sin cuadrante" calcula el ángulo según la ecuación 2.1, mientras que la columna DoA2 representa el ángulo de la dirección de arribo, tal como se decidió en la tabla 2.3.

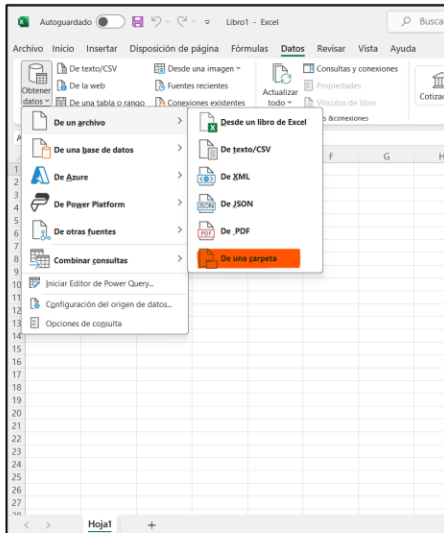
**Tabla 2. 3.** Parámetros adicionales para la localización del ángulo de la dirección de arribo.

cuadrante	d(Variacion_latitud)	d(Variacion_Longitud)	Angulo sin cuadrante	DoA2
Tercer Cuadrante	86,63735162	53,90333061	58,11128442	211,8887156
Tercer Cuadrante	87,42140974	51,55121363	59,47272668	210,5272733
Tercer Cuadrante	88,00945173	50,76715316	60,02203901	209,977961
Tercer Cuadrante	102,7103347	24,5014765	76,58288566	193,4171143
Tercer Cuadrante	102,3183163	24,10948069	76,74113186	193,2588681
Tercer Cuadrante	100,7502497	23,52145414	76,85893291	193,1410671
Tercer Cuadrante	98,59406658	12,34879531	82,86094432	187,1390557

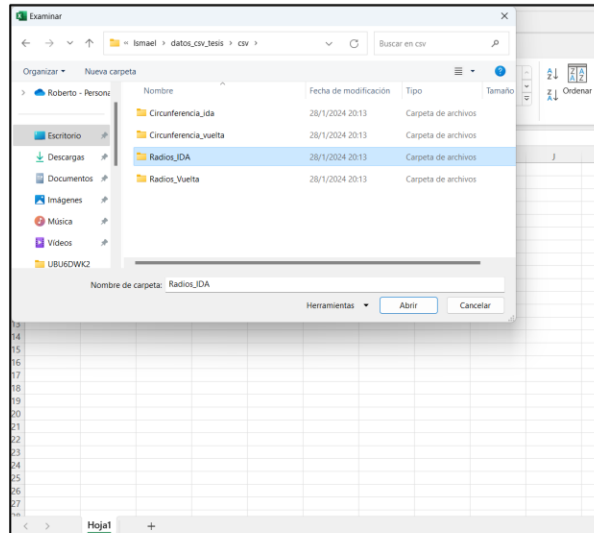
El cálculo de la distancia entre la estación base y la muestra adquirida por la aplicación NetMonitor Pro se realiza utilizando la ecuación 2.2. Un ejemplo de este cálculo se obtiene al emplear la tabla previamente mencionada. Considerando las columnas descritas en el punto anterior, se aplica la fórmula proporcionada en la ecuación 2.6 utilizando Microsoft Excel.

## PREPROCESAMIENTO DE LOS DATOS

Las muestras recopiladas se presentaron en formato CSV, el cual se transformó en valores interpretables mediante la conversión al formato XLSX. Los pasos a seguir se detallan en la figura 14a) y 14b).



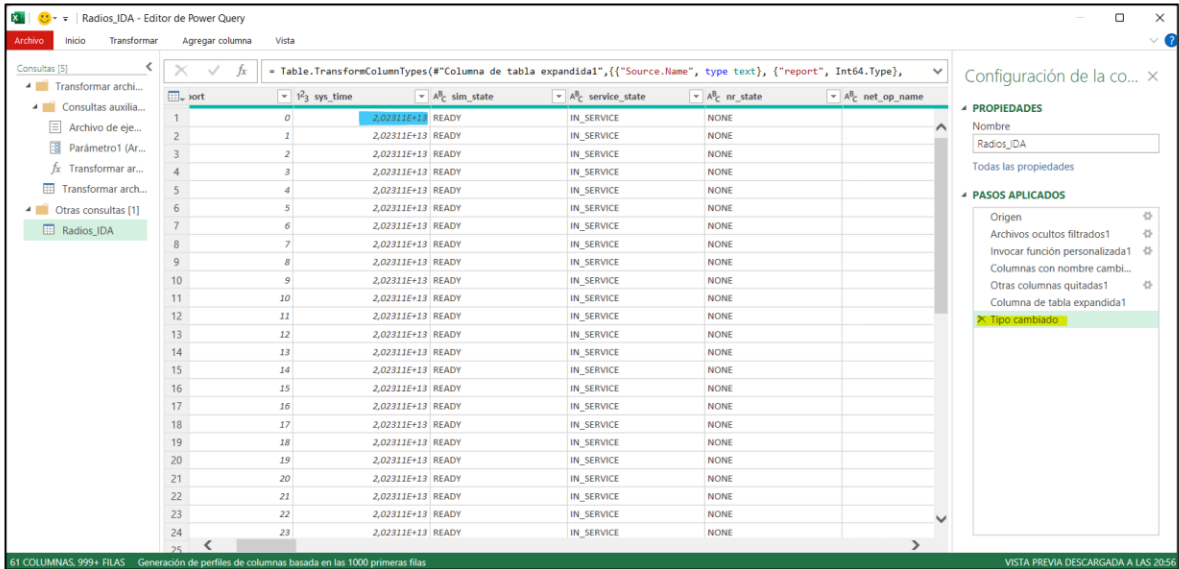
a)



b)

**Figura 2. 14.** Paso de formato de CSV a XSLX. a) Selección desde una carpeta; b) localización de la carpeta.

Una vez seleccionada la carpeta, es posible llevar a cabo los procedimientos indicados en la figura 2.15. En este paso, se opta por transformar los datos mediante una combinación, permitiendo así obtener el conjunto completo de muestras recopiladas. En el ejemplo proporcionado, se visualizan todos los radios en dirección radial desde la estación base hasta el punto donde se produce un handover entre estaciones base. Es imperativo eliminar la opción marcada en amarillo para evitar obtener valores no deseados, como se destaca en color celeste.



**Figura 2. 15.** Selección de datos a variables sin notación científica.

Una vez que se han obtenido los valores en forma de tablas de Excel, se procede a llevar a cabo la eliminación de variables que no serán relevantes para este estudio en particular. La tabla resultante se presenta en la figura 2.16, reflejando los valores obtenidos tras realizar un análisis de estos datos mediante la herramienta de análisis, en este caso, RStudio. Se profundizará en el análisis de estos datos en etapas posteriores.

rsi_strongest	rsi	rsrq	lat	long	altitude	speed	bearing
2147483647	-70	-12	-0.18247666666666668	-78.485511666666667	2807.1	0	139.44
2147483647	-70	-12	-0.18247666666666668	-78.485511666666667	2807.1	0	139.44
2147483647	-67	-11	-0.18247666666666668	-78.485511666666667	2807.1	0	139.44
2147483647	-67	-11	-0.18247666666666668	-78.485511666666667	2807.1	0	139.44
2147483647	-70	-13	-0.18250166666666667	-78.485473333333333	2814.6	0.306774	80.15
2147483647	-70	-13	-0.18250166666666667	-78.485468333333333	2814.5	0.364423	106.53
2147483647	-79	-14	-0.18251	-78.485465	2814.2	0.401483	116.12
2147483647	-79	-14	-0.182515	-78.48546	2813.9	0.443176	125.83
2147483647	-79	-14	-0.18252	-78.485456666666668	2813.3	0.499795	111.65
2147483647	-77	-10	-0.18252166666666667	-78.48545	2813.2	0.549723	109.79
2147483647	-77	-10	-0.18252333333333332	-78.485445	2813	0.612519	101.22
-91	-81	-11	-0.18252833333333335	-78.485441666666667	2812.6	0.592445	26.89
-91	-81	-11	-0.18253333333333333	-78.485436666666666	2812.4	0.60274	118.95
-91	-81	-11	-0.18253666666666665	-78.485428333333333	2812.3	0.662962	106.79
2147483647	-79	-14	-0.18254	-78.485423333333333	2812	0.622814	100.14
2147483647	-79	-14	-0.182545	-78.48542	2811.7	0.599137	115.08
2147483647	-79	-14	-0.18255	-78.485411666666668	2811.6	0.622814	131.6
-84	-79	-14	-0.182555	-78.485401666666666	2812	0.661933	103.57
-84	-79	-14	-0.18256	-78.485396666666667	2812	0.631049	122.14
2147483647	-82	-16	-0.18256166666666665	-78.485390000000001	2811.9	0.623843	119.02
2147483647	-81	-12	-0.18256833333333333	-78.485383333333333	2811.5	0.622814	125.26
2147483647	-82	-11	-0.18256833333333333	-78.485375	2811.4	0.674801	109.73
2147483647	-82	-11	-0.18257333333333334	-78.485366666666668	2811.1	0.698993	109.16
2147483647	-83	-15	-0.18257666666666666	-78.485355	2810.7	0.687154	111.99
2147483647	-83	-15	-0.18257833333333334	-78.485345000000001	2810.2	0.7412	95.68

**Figura 2. 16.** Tabla obtenida para el análisis del modelo para la obtención del ángulo de la dirección de arribo

Con base en los valores recopilados y presentados en la tabla de la figura 2.16, se lleva a cabo un tratamiento inicial de los datos. En este proceso, se procede a eliminar aquellos datos que presentan inconsistencias, específicamente valores extremadamente grandes

que fueron registrados por la herramienta de recolección de datos (NetMonitor Pro). Para cada variable, se identifican y eliminan los valores problemáticos, como se ilustra en la figura 2.17. Aquí, los valores que deben ser eliminados se resaltan en color rosado para facilitar su identificación y exclusión del conjunto de datos. Este paso es esencial para asegurar la calidad y la integridad de los datos antes de avanzar en análisis posteriores.

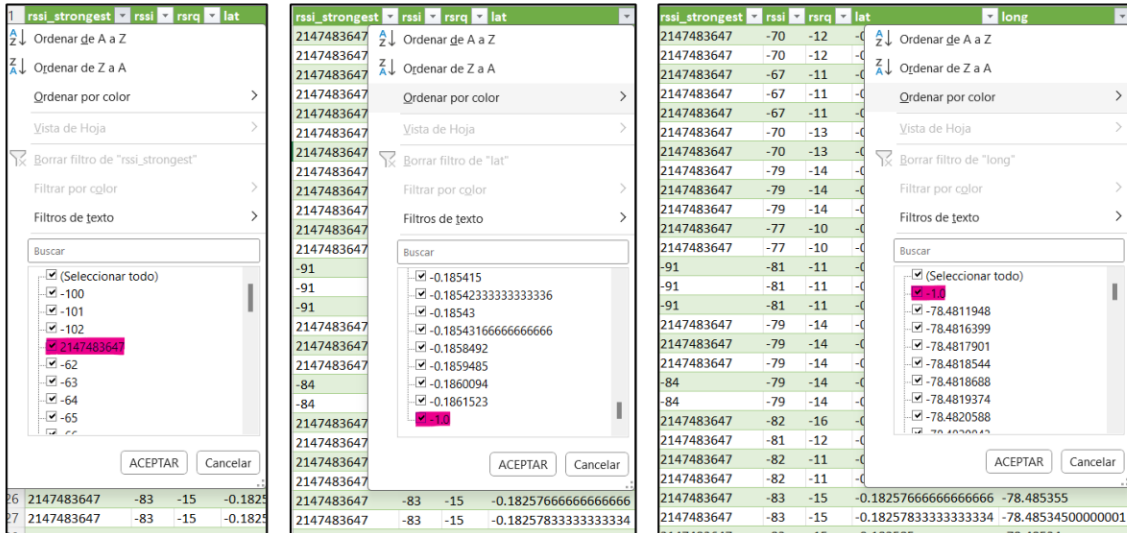


Figura 2. 17. Valores con inconsistencias a eliminarse en excel.

Después de la eliminación de los datos inconsistentes, se procedió a la expansión de las variables utilizadas en el presente estudio. Entre estas variables, se incluyó el ángulo de la dirección de arriba, obtenido mediante la agregación de columnas adicionales para identificar el cuadrante en el cual se encuentra la muestra con respecto a la estación base, como se ilustra en la figura 2.10. La inclusión de estas variables adicionales se detalla en la figura 2.18.

lat	long	cuadrante	d(Variacion_latitud)	d(Variacion_Longi)	Angulo_sin_cuadr	DoA2	Distancia
-0,182528333	-78,48544167	Cuarto Cuadrante	5,6835954	18,62111218	16,97340505	106,973405	19,46918271
-0,182533333	-78,48543667	Cuarto Cuadrante	6,272209167	19,20929871	18,08286615	108,0828661	20,20736907
-0,182536667	-78,48542833	Cuarto Cuadrante	6,664228151	20,18914973	18,26751883	108,2675188	21,26061391
-0,182555	-78,48540167	Cuarto Cuadrante	8,82066139	23,32539287	20,71446427	110,7144643	24,93728007
-0,18256	-78,48539667	Cuarto Cuadrante	9,408045842	23,913392	21,47574978	111,4757498	25,6973065
-0,182598333	-78,48531833	Cuarto Cuadrante	13,91720058	33,1260133	22,78869085	112,7886909	35,93064356
-0,182601667	-78,485315	Cuarto Cuadrante	14,30868704	33,51814029	23,11733661	113,1173366	36,44439835
-0,182618333	-78,48521667	Cuarto Cuadrante	16,26889657	45,08274128	19,84286617	109,8428662	47,92817953
-0,182618333	-78,48520667	Cuarto Cuadrante	16,26889657	46,2588265	19,37640657	109,3764066	49,03606694
-0,182618333	-78,48519667	Cuarto Cuadrante	16,26889657	47,43497753	18,9305625	108,9305625	50,14712279
-0,182623333	-78,48516667	Cuarto Cuadrante	16,8569196	50,96318635	18,3024901	108,3024901	53,67841144
-0,18263	-78,48515833	Cuarto Cuadrante	17,64103702	51,94322149	18,75859391	108,7585939	54,8568519
-0,182635	-78,48514667	Cuarto Cuadrante	18,22904486	53,31533806	18,87606891	108,8760689	56,3453024
-0,182638333	-78,48513667	Cuarto Cuadrante	18,62111218	54,49136168	18,86662897	108,866629	57,58492919
-0,18264	-78,48513333	Cuarto Cuadrante	18,8174312	54,88340319	18,9248914	108,9248914	58,01942275

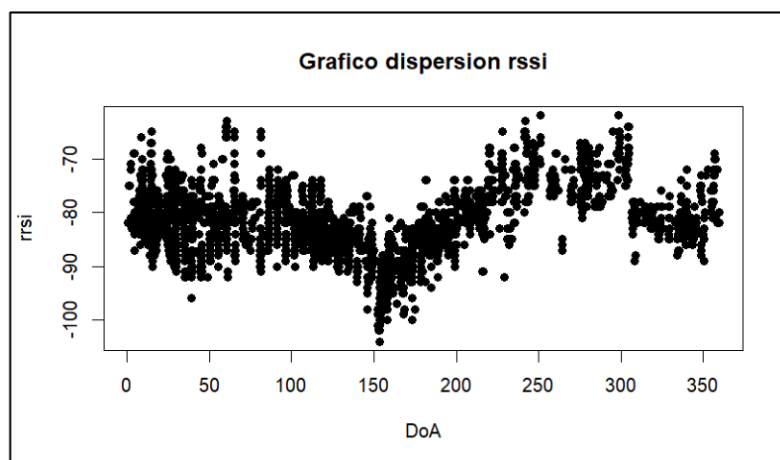
Figura 2. 18. Ángulo de la dirección de arriba y la distancia obtenida entre la estación base y cada muestra captada.

Los cálculos para determinar la distancia entre la estación base y las muestras se llevaron a cabo empleando la ecuación 2.2. Sin embargo, en la implementación en Excel, se observa una variación en la formulación de la ecuación, ya que implica el uso de ángulos en radianes y la aplicación de funciones trigonométricas en dicha unidad angular. Debido a esto, es necesario realizar un preproceso antes de ingresar las ecuaciones en la hoja de cálculo, tal como se ejemplifica en la ecuación 2.6.

$$\begin{aligned}
 &1000 * 6738.4 * \cos^{-1} \left( \cos(\text{Radianes}(90 - \text{latitud})) \right. \\
 &\quad * \cos \left( (\text{Radianes}(90 + 0.18248)) \right) \\
 &\quad + \sin(\text{Radianes}(90 - \text{latitud})) \\
 &\quad \left. * \sin \left( (\text{Radianes}(90 + 0.18248)) \right) * \cos(\text{Radianes}(\text{longitud} + 78.4856)) \right)
 \end{aligned}$$

**Ecuación 2. 6.** Ecuación obtenida en Excel para obtener la distancia (latitud y longitud hace referencia a la muestra ejemplo).

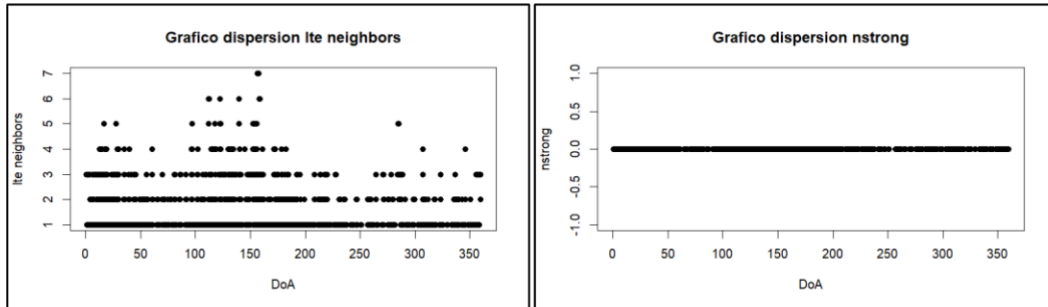
Antes de iniciar el proceso para desarrollar el modelo, es esencial comprender el análisis realizado en RStudio. Este análisis proporcionó información crítica para la selección de variables que se incorporarán en el modelo destinado a la localización del ángulo de Direction of Arrival (DoA). La obtención de variables de radiofrecuencia para la modelación se llevó a cabo mediante la exploración del Received Signal Strength Indicator (RSSI) en función del DoA, como se presenta en la figura 2.19. Aunque no se evidencia una función específica debido a la gran cantidad de datos ingresados, se observa un patrón general que justifica la elección de esta variable como elemento esencial para el modelo. En el Anexo 1 se presenta en la figura 5.1 una tendencia “lineal” en rangos para el rssi en función del DoA.



**Figura 2. 19.** Gráfico de dispersión de rssi en función del DoA.

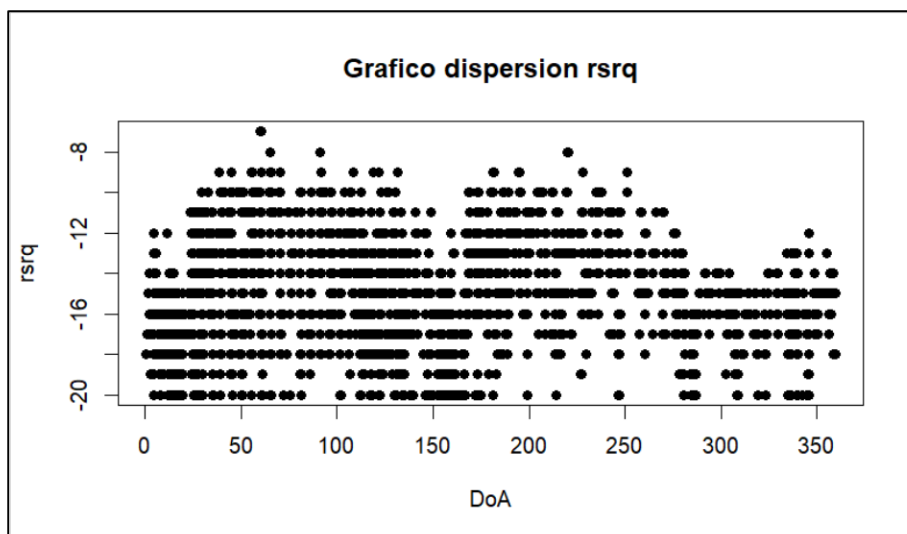


Las imágenes presentadas en la Figura 2.20 revelan la presencia de variables que carecían de relevancia para su inclusión en los modelos desarrollados posteriormente. A modo de ilustración, se han considerado las gráficas de dispersión correspondientes a las variables "lte neighbors" y "nstrong", evidenciando así su falta de aportación significativa en el análisis llevado a cabo.

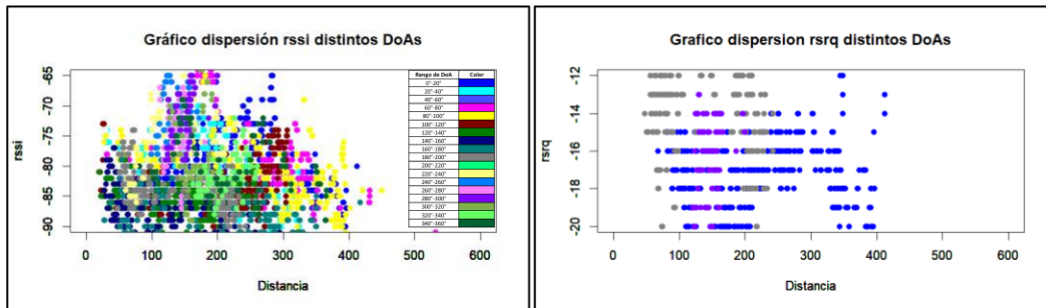


**Figura 2. 20.** Gráficos de dispersión de lte neighbors y nstrong en función del DoA.

La Figura 2.21 detalla la existencia de un patrón sutil entre "rsrq" y "DoA", lo cual ha permitido considerar este patrón en el análisis para el modelo supervisado de Machine Learning. Esta relación se visualiza de manera más clara en las gráficas presentadas en la Figura 2.22 a) y 2.22 b). En estas representaciones gráficas, se emplea una codificación de colores para ilustrar cómo "rsrq" y "rssi" se relacionan en función de la distancia para diversos ángulos de llegada ("DoAs"). El rango de "DoAs" utilizado es de 20° entre cada muestra en las gráficas. La parte b) de la Figura 2.22 consolida tres gráficos en uno para facilitar la comprensión del patrón observado.



**Figura 2. 21.** Dispersión de rsrq en función del DoA.

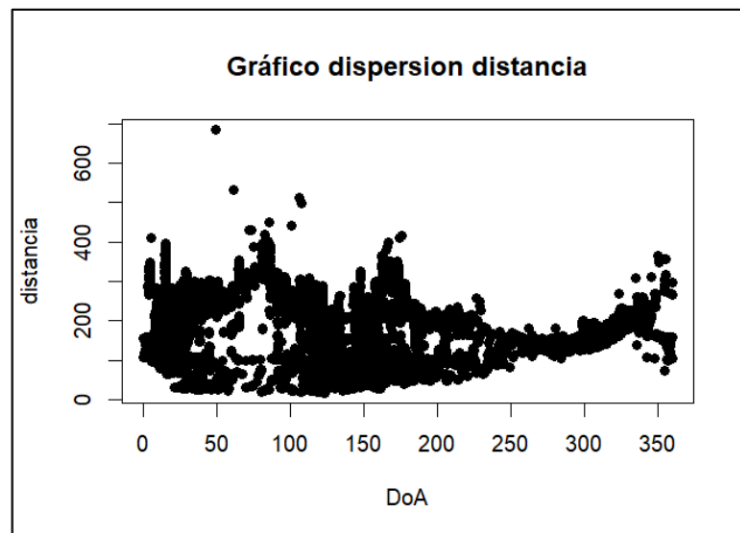


a)

b)

**Figura 2. 22.** a) dispersión de rssi para distintos DoAs en función de la distancia; b) Grafico de dispersión de rsrq para 3 rangos de DoAs en función de la Distancia.

La variable agregada que proporciona información sobre las mediciones realizadas se presenta en la Figura 2.23, en la cual se observa una tendencia polinómica en rangos establecidos presentados en el capítulo 3 de DoA para la distancia.



**Figura 2. 23.** Dispersión entre la distancia en función del DoA.

# IMPLEMENTACIÓN DE LAS TÉCNICAS DE MACHINE LEARNING

En el contexto teórico, se exploraron las técnicas de aprendizaje automático (Machine Learning), incluyendo tanto las no supervisadas como las supervisadas. Esta sección, se enfoca en el análisis de la técnica supervisada, centrándose de manera específica en el método de regresión.

Antes de abordar el análisis de la técnica de aprendizaje por regresión, se describen las fases empleadas en la técnica de Machine Learning de tipo SUPERVISADO. Las etapas fundamentales comprenden:

- **Recolección:** Además de adquirir los datos de entrada y salida (denominados en este estudio como datos de salida, representando el ángulo de la dirección de arribo calculado en la sección 2.3) y su correspondiente etiquetado.
- **Preprocesamiento:** Se lleva a cabo una limpieza y, de ser necesario, normalización de datos, incluyendo la eliminación de valores atípicos (abordado en la sección 2.4). Se menciona también la transformación de variables; en la sección anterior, fue necesario convertir el tipo de variable para obtener valores numéricos, lo cual posibilitó la generación de gráficas para el análisis ya presentado.
- **Entrenamiento:** En esta fase, se realiza el aprendizaje de la neurona a través de la preparación de datos de entrada y la asignación de pesos a cada valor. Las fórmulas específicas utilizadas en la técnica de regresión serán mencionadas en la siguiente sección.
- **Prueba:** La evaluación del modelo obtenido se efectúa con datos no empleados en el proceso de aprendizaje, con el fin de verificar su capacidad de generalización.
- **Evaluación:** Para determinar la idoneidad del modelo, se compara la información de entrada y salida utilizando métricas de rendimiento, tales como el error medio cuadrático y el coeficiente de determinación.
- **Ajuste:** Como paso final, se realiza un ajuste al modelo, si es necesario, con el propósito de minimizar los valores del error medio cuadrático y refinar el modelo con la máxima precisión posible [22].

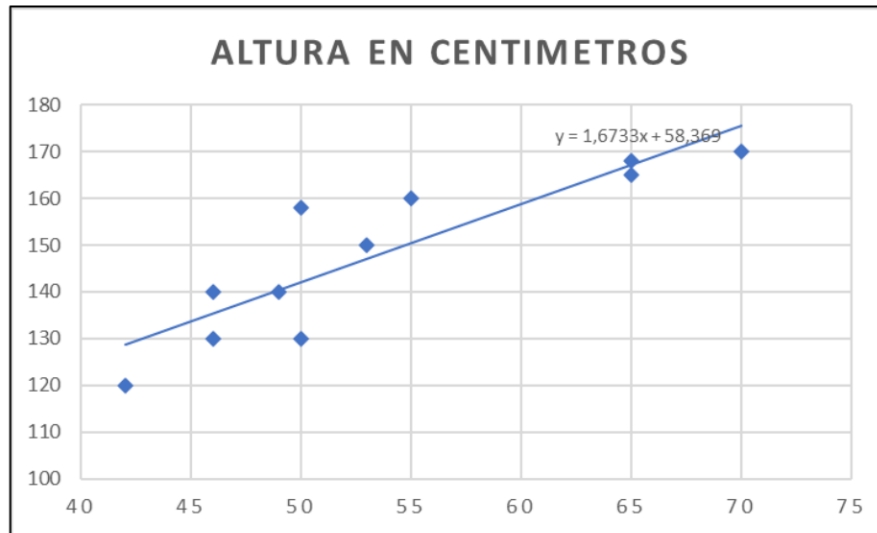
## 2.1.2 REGRESIÓN LINEAL

Este enfoque es reconocido como un modelo estadístico en el cual se busca identificar un patrón que vincule una o varias variables de entrada y, a partir de ellas, prever un valor [22]. Para una comprensión más clara, se ilustra con el ejemplo de estimar la altura de una persona en función de diversas características, como peso, edad y sexo. Estas características pueden o no proporcionar información relevante sobre el parámetro deseado. Se podría incluso considerar una cuarta característica, como el nombre de la persona. El ejemplo específico utilizado se basa en la Tabla 2.4, que presenta datos sobre la altura de un grupo de individuos según su edad y peso.

**Tabla 2. 4.** Ejemplo de regresión.

Peso en Kg	Altura en centímetros	Edad en años
70	170	18
65	168	17
50	158	16
55	160	15
65	165	14
53	150	13
49	140	12
50	130	11
46	140	10
46	130	9
42	120	8

Los datos tabulados se representan gráficamente mediante un diagrama de dispersión, en este caso, se examinó la relación entre la altura y el peso. Aunque la altura está influida por diversos factores, el peso puede ser un indicador relevante. En la Figura 2.24, se observa una clara tendencia y un patrón al relacionar la altura con el peso. La gráfica incluye una función de regresión lineal que tiene un punto de intersección en el eje "y" (altura) y una pendiente que multiplica los pesos para estimar una posible altura. En Rstudio, este punto de intersección se conoce como el intercepto o parámetro inicial para cálculos posteriores. Aunque el ejemplo analizado considera un parámetro en función de una variable, en Rstudio se puede realizar un análisis con varios factores, es decir, múltiples variables predictivas para deducir un patrón. Más adelante, se explorará cómo en Rstudio se pueden visualizar los pesos asignados a cada predictor para la estimación de un parámetro.



**Figura 2. 24.** Altura en función del peso

El análisis de la técnica de aprendizaje supervisado, en congruencia con el ejemplo ilustrativo previo, se enfocará en la técnica de regresión implementada a través de Rstudio. Se busca obtener un modelo capaz de estimar valores similares a la dirección de arriba, lo que motiva la aplicación de una variante de la técnica de regresión lineal, como se mostró en el ejemplo anterior. En el cálculo de este modelo, se emplea una regresión lineal con covariables del tipo polinómicos, en ciertos casos (debido a las características específicas de la estación base), aunque es pertinente destacar que en algunos modelos se aplicaron regresiones "lineales". Aunque se menciona "regresión lineal" debido a su representación gráfica por una función de línea recta, esta función se denomina así por la utilización de dos parámetros: la variable dependiente (en este ejemplo, la altura) y la variable independiente (el peso). Dado que se emplearon cuatro parámetros de entrada, también podría hablarse de una regresión lineal o una función en varias variables.

## **MACHINE LEARNING EN RSTUDIO Y PREPARACIÓN DE DATOS**

Rstudio ofrece herramientas no solo para análisis estadísticos, sino también para modelar sistemas mediante el aprendizaje automático. Esto posibilita la obtención de distintos tipos de variables basadas en pesos obtenidos a partir de los datos utilizados en el entrenamiento. Entre los paquetes empleados en Rstudio para estos modelos se encuentran:

- Caret: Esta herramienta, cuyo nombre es una abreviatura de clasificación y entrenamiento de regresión, proporciona una interfaz unificada diseñada para simplificar el proceso de construcción y evaluación de modelos predictivos [23].
- Tidymodels: Este paquete se centra en ofrecer una estructura coherente para el modelado, trabajando con librerías que permiten el preprocesamiento de los datos a utilizar, como `recipes` [24].

Una vez conocidos los paquetes que serán utilizados en la creación de los diferentes modelos es necesario saber cuál es el proceso que se debe seguir para poder conseguir estructuras predictivas. El escenario que se tendrá en cuenta es el utilizado para poder conseguir el ángulo de la dirección de arriba, el cual será comparado con el ángulo localizado en la sección 2.3.

Previo a la utilización de un modelo de aprendizaje supervisado (regresión lineal) se debe tener datos limpios para evitar tener redundancias o incoherencias para poder tener un modelo con mayor precisión. Este proceso ya fue descrito en la sección 2.4, dentro de este apartado se pudo visualizar todo el desarrollo realizado sobre los datos obtenidos de las muestras generadas por NetMonitor Pro sobre Microsoft Excel. Así mismo se debe realizar un procesamiento de datos dentro de Rstudio el cual será mencionado más adelante.

La comprensión de todas las funciones empleadas para encontrar y desarrollar los diversos modelos utilizados, así como el entendimiento de cada fase llevada a cabo, será ilustrada mediante un ejemplo detallado. En este caso, se expondrán las librerías utilizadas y el proceso de generación de cada una.

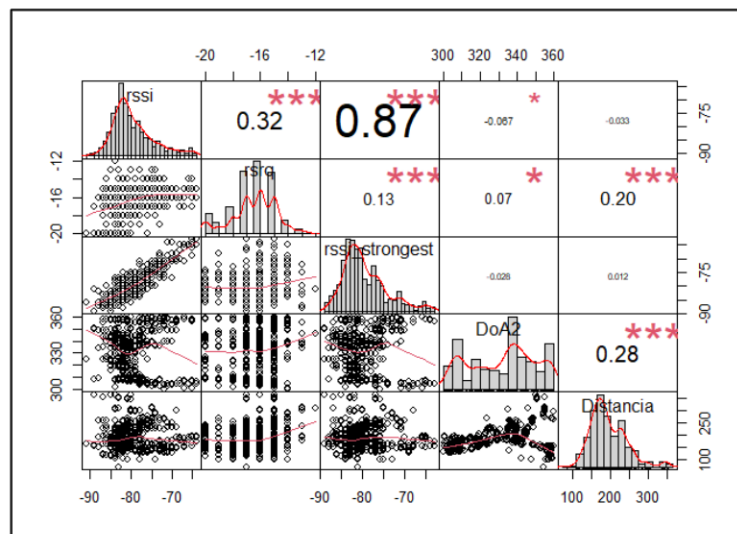
La carga de datos preprocesados desde Excel se realizó a través de la librería `"openxlsx"`, la cual facilita la importación de matrices en forma de tabla y su conversión en conjuntos de datos dentro del entorno de Rstudio. Los datos obtenidos se almacenan en una variable, cuya cabecera se muestra en la figura 2.25, utilizando la función `"head"` para visualizarla en la consola. En dicha imagen, se muestran únicamente las variables pertinentes para el modelo, resultado del filtrado realizado con la librería `"select"`.

```
> head(data_doa_mod_train)
  rssi rsrq rssi_strongest DoA2 Distancia
1  -78  -19             -75 347.8429 103.4264
2  -78  -19             -75 347.8429 103.4264
3  -74  -18             -73 347.8429 103.4264
4  -78  -16             -76 347.8429 103.4264
5  -79  -18             -81 335.7048 137.1768
6  -77  -16             -80 342.8361 162.0672
```

**Figura 2. 25.** Cabecera del conjunto de datos con las variables que ingresarán al entrenamiento.

El siguiente paso consistió en la conversión de los datos a valores numéricos, con el fin de manipularlos y obtener información de manera más lineal. Las variables que experimentaron este cambio fueron: "rssi", "rssi strongest" y "rsrq", ya que estos datos se obtuvieron directamente de la aplicación NetMonitor Pro. En cuanto al tipo de dato de la distancia entre la estación base y cada muestra, no se modificó, dado que este dato se calculó a partir de la ubicación entre la estación base y la muestra tomada.

Posteriormente, se generó una matriz de correlación para examinar la relación entre cada variable y su homóloga. Para esto, se utilizó la función "**chart.Correlation**" la cual generó la matriz de correlación junto con su respectivo histograma de las variables a utilizar dentro del modelo, así como su relación con el parámetro de salida (ángulo de la dirección de arriba). La cuarta columna de la siguiente figura muestra la cantidad de información que posee el DoA en función de las variables. De esta manera, se puede observar que la variable "Distancia" es la que presenta una mayor cantidad de información sobre el ángulo en comparación con el resto de las entradas.



**Figura 2. 26.** Matriz de correlación entre las variables de entrada y la de salida.

La matriz de correlación proporcionó información clave para el entrenamiento de la neurona. Antes de iniciar el proceso de entrenamiento, fue necesario preparar los datos, tarea que se llevó a cabo mediante la función **"recipe"**. Las siguientes líneas de código ilustran cómo se realizó esta preparación. La receta creada muestra a "DoA2" como un data frame, junto con el resto de los parámetros que contribuyen a determinar la dirección de arriba. En la segunda sentencia, se observa cómo se convierten estos parámetros en variables numéricas para su utilización en el modelo [25].

```
DoA_receta1<-recipe(DoA2~.,data=data_doa_mod_train) %>% #para que se ejecute
secuencialmente

step_zv(all_numeric(),-all_outcomes())%>%

step_mutate(distcic=runif(nrow(data_doa_mod_train),0,1))%>%

step_mutate(cic=as.factor(if_else(distcic>0.8,"Norm","Rec")))%>%

step_select(-distcic)%>%

prep()
```

Los datos procesados en la receta solo pueden ser obtenidos como valores utilizables mediante la función "juice". Esta función permite extraer los datos generados en la receta y presentarlos como un data frame. La figura 2.27 ilustra cómo se presentan los datos dentro de la receta obtenida.

DoA_receta1	list [10] (S3: recipe)	List of length 10
var_info	list [5 x 4] (S3: tbl_df, tbl, data.fra	A tibble with 5 rows and 4 columns
term_info	list [6 x 4] (S3: tbl_df, tbl, data.fra	A tibble with 6 rows and 4 columns
steps	list [4]	List of length 4
template	list [5742 x 6] (S3: tbl_df, tbl, data	A tibble with 5742 rows and 6 columns
levels	list [6]	List of length 6
retained	logical [1]	TRUE
requirements	list [1]	List of length 1
tr_info	list [1 x 2] (S3: data.frame)	A data.frame with 1 row and 2 columns
orig_lvls	list [5]	List of length 5
last_term_info	list [7 x 6] (S3: grouped_df, tbl_df	A tibble with 7 rows and 6 columns

**Figura 2. 27.** Datos obtenidos de la receta creada para preparar los datos para el modelo.



Los valores obtenidos mediante la función "juice" son los que se incorporaron al modelo. Sin embargo, antes de introducir todos los datos, se crearon dos particiones de forma aleatoria. Estas particiones permitieron obtener un conjunto de datos que la neurona utilizará para aprender, mientras que el resto se utilizará para evaluar el modelo. Esta división se realizó asignando el 80% de los datos para el entrenamiento y el 20% restante para la evaluación del modelo.

### 2.1.3 ENTRENAMIENTO DEL MODELO Y EVALUACIÓN

El modelo fue entrenado utilizando el método de regresión, el cual se detalló previamente en la sección "**Regresión lineal**". La función "lm" (regresión lineal simple) se utilizó para generar el modelo mencionado. Este modelo estadístico se basa en la ecuación 2.7, que muestra una variable de salida Y en función de una suma entre el intercepto (valor de la ordenada en el origen, similar al ejemplo de la figura 2.24) y las variables independientes, que en este caso son: "rssi", "rssi strongest", "rsrq" y la distancia.

$$Y_i = \beta_0 + \beta_1 rssi_i + \beta_2 rssi\_strongest_i + \beta_3 rsrq_i + \beta_4 Distancia_i$$

**Ecuación 2. 7.** Función de regresión lineal [26].

Las variables utilizadas en la ecuación mostrada son las siguientes:  $Y_i$ , que representa el ángulo de la dirección de arribo predicho;  $\beta_0$ , que es el valor del intercepto; y,  $\beta_1, \beta_2, \beta_3$ , que representan los pesos estimados para cada parámetro del modelo que influyen en la obtención del DoA. Este concepto se ilustra más claramente en la figura 2.28, la cual presenta un resumen de los parámetros utilizados, incluyendo el intercepto, los pesos asignados a cada variable y algunas medidas estadísticas relevantes. Además, en la parte inferior de la figura se muestra el error residual estándar, el cual indica la discrepancia entre los valores predichos por el modelo y los valores reales. Para que el modelo genere predicciones cercanas a la realidad, este indicador debe tender a ser lo más bajo posible.

```

Call:
lm(formula = DoA2 ~ ., data = training_base_doa)

Residuals:
    Min       1Q   Median       3Q      Max
-157.701  -68.911   -8.475   66.254  257.172

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  312.74486   18.80489   16.631 < 2e-16 ***
rssi          0.29889    0.50276    0.595  0.55220
rsrq         -0.80483    0.67568   -1.191  0.23366
rssi_strongest 1.44107    0.45350    3.178  0.00149 **
Distancia    -0.22048    0.01568  -14.063 < 2e-16 ***
cicRec       -4.40829    3.37924   -1.305  0.19212
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 90.86 on 4588 degrees of freedom
Multiple R-squared:  0.05957, Adjusted R-squared:  0.05854
F-statistic: 58.12 on 5 and 4588 DF, p-value: < 2.2e-16

```

**Figura 2. 28.** Resumen del modelo entrenado.

La evaluación del modelo se llevó a cabo utilizando la función "predict", la cual permite introducir los valores separados para realizar las pruebas, correspondientes al 20% de la partición mencionada en el apartado anterior. Una vez obtenidos los valores predichos, se procedió a evaluar el rendimiento del modelo utilizando las siguientes métricas:

- La raíz del error cuadrático medio (RMSE, por sus siglas en inglés): este indicador es crucial para evaluar el rendimiento del modelo, especialmente en el contexto de la regresión. Cuanto menor sea el valor del RMSE, mejores serán los resultados del modelo. La ecuación 2.8 muestra cómo calcular este indicador para cada valor obtenido [27].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{X}_i - X_i)^2}{n}}$$

**Ecuación 2. 8.** Raíz del error cuadrático medio [27].

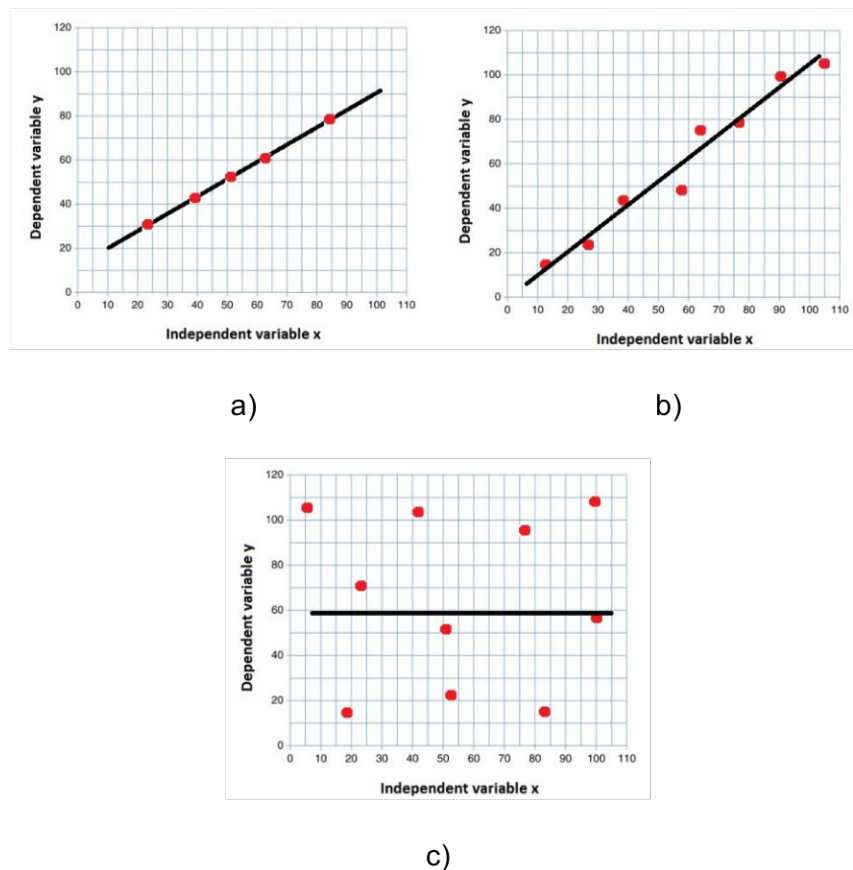
Donde:

$\hat{X}_i$ : valores predichos

$X_i$ : valores reales

n: número de muestras

- El coeficiente de determinación ( $R^2$ ), conocido como la medida que proporciona información sobre el ajuste de un modelo, es fundamental en el análisis. Cuanto más cercano sea el valor obtenido a la unidad, mejor será el modelo. En la figura 2.29 se muestran los valores generados por un modelo, mientras que en negro se representan los valores reales. En (a), se muestra un valor de  $R^2 = 1$ , en (b) un valor de  $R^2 < 1$ , y en (c) un valor de  $R^2 = 0$ [28].



**Figura 2. 29.** Valores de un modelo comparado con los valores reales. a)  $R^2$  igual a 1; b) Valor del  $R^2$  menor a 1; c) valor de  $R^2$  igual a 0 [28].

La información generada por estos dos indicadores se obtiene en Rstudio mediante las funciones "**rmse**" y "**R2**" respectivamente. Un ejemplo de los valores obtenidos se ilustra en la figura 3.1. El análisis de estos resultados se presenta en el capítulo 3.

Todo el procedimiento para obtener las gráficas, el preprocesamiento de los datos capturados por NetMonitor Pro, el entrenamiento de la neurona y la evaluación del modelo se detalla en el Anexo 1.

Por otro lado, las bibliotecas utilizadas se muestran en la tabla 2.5.

**Tabla 2. 5.** Librerías utilizadas en la creación y evaluación del modelo obtenido [29].

<b>Librerías</b>	<b>Especificación</b>
library("tidyverse")	Conjunto de paquetes para poder manipular, importar, explorar y visualizar datos.
library("lattice")	Permite la creación de gráficos
library("caret")	Crea modelos predictivos
library("skimr")	Creación de datos estadísticos resumidos
library("devtools")	Manipulación ágil de los datos, como la carga de ellos
library("openxlsx")	Permite la lectura, escritura y edición de archivos creados en Excel
library("readxl")	Importa las hojas de excel a data frames
library("dplyr")	Permite el manejo de ficheros
library("recipes")	Creación de recetas
library("psych")	Objetos de estadística descriptiva
library("PerformanceAnalytics")	Permite analizar el rendimiento de un modelo
library("Metrics")	Estimación de un modelo

# 3 RESULTADOS, CONCLUSIONES Y RECOMENDACIONES

## RESULTADOS

Antes de generar las ecuaciones de cada modelo, es crucial comprender que el área donde se llevaron a cabo las mediciones exhibe una sectorización estimada, como se ilustra en la figura 3.1. La sectorización real, representada en la figura 3.2, delineó la segmentación de la estación base a partir de mediciones recopiladas por la aplicación Cellmapper, la cual se ha construido a partir de datos provenientes de diversos usuarios.



**Figura 3. 1.** Sectorización estimada para localizar las ecuaciones de los modelos.

En la figura 3.2 se exhiben los distintos sectores, cada uno identificado por un color único. Para el análisis, se llevó a cabo una comparación entre la sectorización estimada y la sectorización real, revelando la similitud entre ambos. Con este conocimiento, se procede a intentar obtener un modelo general para la localización del ángulo de la dirección de arribo.



**Figura 3. 2.** Sectorización de la estación base tomada para el análisis. Tomada de la aplicación Cellmapper.

Una vez asimilado el procedimiento de entrenamiento, entrega de resultados y verificación de la funcionalidad del modelo obtenido, se evidencia inicialmente su rendimiento deficiente debido a la sectorización de 5 antenas en la estación base. La sectorización, detallada en la sección de zonas y rutas, confirma que la variación en la transitabilidad de usuarios en diferentes áreas ha generado celdas con diferentes antenas para abordar dicha congestión.

Los primeros resultados del modelo general exhiben un elevado valor de error medio cuadrático. El RMSE obtenido para cada modelo general fluctúa en rangos muy similares, todos ellos elevados, lo cual sugiere que aún no se ha alcanzado un modelo eficiente para obtener un DoA en cualquier ubicación. Esta discrepancia se visualiza claramente en la figura 3.3.

modelo	RMSE	R2
1 modDoA1	90.96388	0.05774683
2 modDoA2	90.44433	0.06844051
3 modDoA3	90.49521	0.06876501

**Figura 3. 3.** Valores de RMSE y R2 para cada modelo general obtenido.

Los siguientes modelos descritos son los que se evaluaron en la figura 3.3:

lm(DoA2~., data=training\_base\_doa): Este modelo utiliza todas las variables de forma lineal para predecir el ángulo de la dirección de arribo.

lm(DoA2~rssi+Distancia+rssi:Distancia, data=training\_base\_doa): En este modelo se emplean los parámetros rssi y distancia con diversas combinaciones.

lm(DoA2~l(((Distancia)rsrq+rssi\*rssi\_strongest-rssi)) + rsrq + l(rssi^5) + rssi\_strongest + l(Distancia^3+rsrq+rssi^3), data=training\_base\_doa): Este modelo utiliza una función de tipo polinómica.

Una vez examinado un modelo general y constatado que no sería factible obtenerlo debido a la configuración de la estación base y la ubicación donde se realizaron las muestras, se sugirió la creación de varios modelos alternativos. Estos modelos fueron generados mediante un filtro que selecciona valores de ángulos con una diferencia menor a 10° entre las muestras obtenidas por el modelo y los valores reales de la dirección de arribo.

### 3.1.1 MODELO 80-170 GRADOS

El primer modelo obtenido a través de este análisis abarca el rango de 80° a 170°. La figura 3.4 presenta la cabecera de los datos producidos por el modelo lineal, resaltando las variaciones inferiores a 10°.

```
> head(Comparacion_Filtrado)
```

	DoA2	predic_modDoA1	predic_modDoA2	predic_modDoA3	Delta_DoA1	Delta_DoA2	Delta_DoA3
8	106.7661	115.4095	115.3149	117.0954	8.643464	8.5488672	10.329357
9	106.7200	115.1074	114.8575	116.5926	8.387369	8.1375160	9.872630
10	106.1789	112.1246	105.5045	101.8433	5.945697	0.6744249	4.335605
34	112.8760	122.6622	125.7908	136.5530	9.786160	12.9147470	23.676984
36	112.8797	121.9178	124.5010	134.2389	9.038091	11.6212980	21.359190
38	112.6083	118.0148	121.1528	131.7792	5.406495	8.5444972	19.170821

**Figura 3.4.** Filtrado de los valores obtenidos con los reales para el DoA.

La ecuación generada por el modelo para el rango entre 80° y 170° se visualiza a través del resumen proporcionado por el modelo, el cual se presenta en la figura 3.5.

```

Call:
lm(formula = DoA2 ~ ., data = training_base_doa)

Residuals:
    Min       1Q   Median       3Q      Max
-162.699  -81.675   -4.936    60.764   250.290

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  244.19740    12.94368   18.866 < 2e-16 ***
rssi         -0.53929     0.35554   -1.517  0.129
rsrq         -0.80461     0.49061   -1.640  0.101
rssi_strongest  1.61790     0.30947    5.228 1.75e-07 ***
Distancia    -0.15890     0.01158  -13.719 < 2e-16 ***
cicRec       -0.90582     2.40124   -0.377  0.706
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 93.33 on 9375 degrees of freedom
Multiple R-squared:  0.03094,    Adjusted R-squared:  0.03042
F-statistic: 59.86 on 5 and 9375 DF,  p-value: < 2.2e-16

```

**Figura 3. 5.** Resumen obtenido para el modelo obtenido para un rango entre 80° a 170°.

Observando el modelo se puede determinar la ecuación 3.1

$$\begin{aligned}
 DoA_{j_{80^{\circ}-170^{\circ}}} = & 244.19740 - 0.53929 * rssi_j - 0.80461 * rsrq_j + 1.61790 * rssi_{strongest_j} \\
 & - 0.15890 * Distancia_j
 \end{aligned}$$

**Ecuación 3. 1.** Ángulos de la dirección de arribo estimados para un rango entre 80°-170°

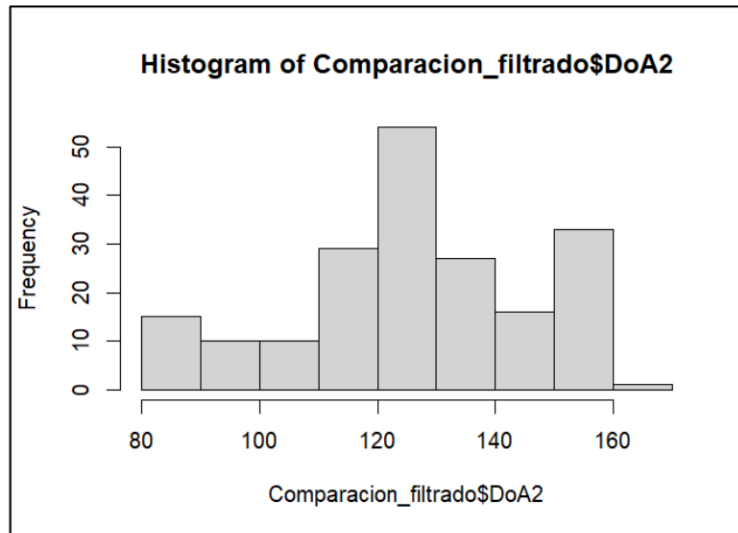
El modelo presenta el rendimiento que se muestra en la figura 3.6.

modelo	RMSE	R2
1 modDoA1	16.6771	0.540114

**Figura 3. 6.** Rendimiento del modelo entre 80°-170° para la localización del ángulo del DoA.

Debido al volumen de datos utilizado para el entrenamiento y las pruebas, el error calculado (RMSE) exhibe un valor estimado que se considera relativamente confiable, en función de los parámetros ingresados en el modelo. En la figura 3.7 se puede visualizar la concentración de los datos obtenidos por el modelo.





**Figura 3. 7.** Histograma de los resultados obtenidos para el rango entre 80°-170°.

### 3.1.2 MODELO DE 65-80 GRADOS

De manera análoga al rango anterior, se llevó a cabo un procedimiento similar para un rango entre 65° y 80°. En este caso, se observa una ecuación similar a la 3.1, en la que todas las variables están incluidas, pero los pesos de cada variable varían. Como se explicó en el capítulo 2, sección 2.5, es posible que estas ecuaciones muestren una tendencia lineal, pero con pendientes diferentes. Esta determinación se realizó al observar cada parámetro que se presenta en el resumen del modelo (figura 3.8).

Los resultados observados en la figura 3.8 permitieron conocer la ecuación del modelo presentado en la ecuación 3.2.

```
Call:
lm(formula = DoA2 ~ ., data = training_base_doa)

Residuals:
    Min       1Q   Median       3Q      Max
-6.5440 -2.7133 -0.0805  1.6686 11.1083

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  48.770943   4.740662  10.288 < 2e-16 ***
rssi        -0.476991   0.146674  -3.252  0.00138 **
rsrq         0.103709   0.169665   0.611  0.54184
rssi_strongest 0.123624   0.088778   1.393  0.16557
Distancia   -0.017958   0.003874  -4.636 7.02e-06 ***
cicRec      -0.595969   0.698688  -0.853  0.39486
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.694 on 171 degrees of freedom
Multiple R-squared:  0.1779,    Adjusted R-squared:  0.1538
F-statistic:  7.4 on 5 and 171 DF,  p-value: 2.632e-06
```

**Figura 3. 8.** Resumen obtenido para el modelo obtenido para un rango entre 65° a 80°.

$$DoA_{j_{65^{\circ}-80^{\circ}}} = 48.770943 - 0.476991 * rssi_j + 0.103709 * rsrq_j + 0.123624 * rssi_{strongest_j} - 0.017958 * Distancia_j$$

**Ecuación 3. 2.** Ángulos de la dirección de arribo estimados para un rango entre 65°-80°

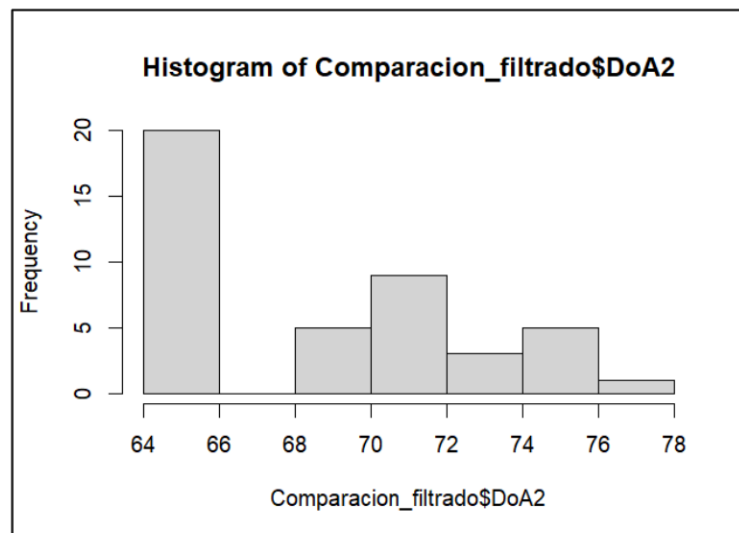
En la figura 3.9 se ilustra el rendimiento del modelo, mientras que en la figura 3.10 se presenta el histograma correspondiente. En este último, se observa un hueco en el ángulo de 67°. Esta anomalía se generó debido a la ubicación de donde se tomaron las muestras; en esa zona, había un área boscosa que no pudo ser atravesada durante el muestreo.

modelo	RMSE	R2
1 modDoA1	3.636324	0.1545067

**Figura 3. 9.** Rendimiento del modelo entre 65°-80° para la localización del ángulo del DoA.

El rendimiento se presenta con una tendencia lineal, la cual se considera muy confiable debido a la cantidad limitada de muestras utilizadas.

De este modo, se puede afirmar que existe el mismo modelo entre 65° y 170°, con la única diferencia de que en el rango de 65° a 80° tiene una “pendiente” diferente.

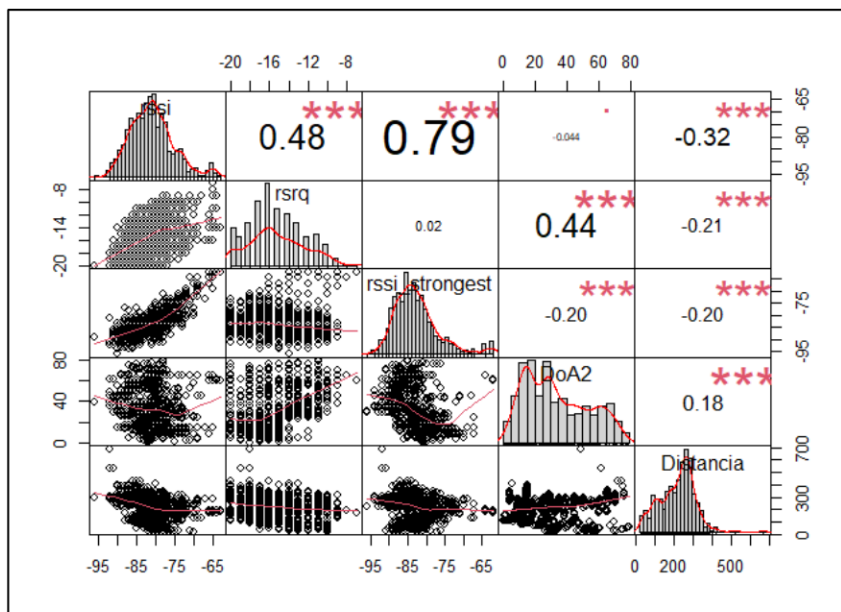


**Figura 3. 10.** Histograma de los resultados obtenidos para el rango entre 65°-80°.

### 3.1.3 MODELO 0-65 GRADOS

El tercer modelo propuesto abarca un intervalo de 0° a 65°, los cuales fueron generados utilizando una ecuación distinta a las previamente analizadas. Esta ecuación se basó en la matriz de correlación mostrada en la figura 3.11, donde se destaca que la variable rsrq

contiene la mayor cantidad de información. Sin embargo, debido a la naturaleza de las muestras tomadas, no se observa una tendencia lineal clara. Por lo tanto, se optó por emplear ecuaciones de tipo polinómico para evitar la linealidad mencionada. El proceso de ajuste para determinar estas ecuaciones fue similar al descrito en la tabla 3.1, que muestra el error residual estándar asociado a cada parámetro incluido en el modelo. Como se explicó en el capítulo de metodología, este error debería tender a cero; para lograr esto, se evaluaron diferentes combinaciones de variables según la matriz de correlación. Los valores obtenidos en esta matriz proporcionaron una visión más clara del modelo a generar, lo que permitió la inclusión progresiva de posibles modelos en la tabla 3.1.



**Figura 3. 11.** Matriz de correlación para un rango de 0°-65°.

**Tabla 3. 1.** Estimación de parámetros incluidos en los modelos y su RSE.

Coeficiente	Estimación	Residual standard error (debe tener tendencia a cero)
rrsi	-0,13463	19,84
Distancia	0,039578	19,54
rrsi_strongest	-0,62105	19,49
rsrq	3,1683	17,79
rrsi^20	4,99E-39	19,77
Distancia^3+rsrq		16,82
Distancia^3+rsrq+rrsi		16,31
Distancia^3+rsrq+rrsi^3		16,2
Distancia^3+rsrq+rrsi^3+rrsi_strongest		16,07

El parámetro "**distancia**" desempeña un papel significativo en la estimación del ángulo de llegada (DoA), lo cual se refleja en la generación de una ecuación interna que incluye todos los parámetros, con una ponderación negativa para rssi. El modelo resultante se presenta en la ecuación 3.3, que incluye el valor de intercepto y el peso asociado a cada parámetro, como se muestra en el resumen del modelo en la figura 3.12. Además, en la figura 3.13 se muestra el rendimiento de esta ecuación del modelo obtenido.

```
Call:
lm(formula = DoA2 ~ I(((Distancia) * rsrq * rssi * rssi_strongest -
  rssi)) + rsrq + I(rssi^5) + rssi_strongest + I(Distancia^3 +
  rsrq + rssi^3), data = training_base_doa)

Residuals:
    Min       1Q   Median       3Q      Max
-48.041 -10.798  -2.392   8.730  49.739

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.296e+02  9.806e+00  13.218 < 2e-16 ***
I(((Distancia) * rsrq * rssi * rssi_strongest - rssi)) -3.740e-07  9.483e-08  -3.944 8.39e-05 ***
rsrq              5.463e+00  2.298e-01  23.771 < 2e-16 ***
I(rssi^5)        -6.681e-09  8.087e-10  -8.261 3.18e-16 ***
rssi_strongest   5.629e-01  1.206e-01   4.666 3.35e-06 ***
I(Distancia^3 + rsrq + rssi^3)  7.667e-08  5.525e-08   1.388  0.165
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.91 on 1470 degrees of freedom
Multiple R-squared:  0.3599,    Adjusted R-squared:  0.3577
F-statistic: 165.3 on 5 and 1470 DF,  p-value: < 2.2e-16
```

**Figura 3. 12.** Resumen obtenido para el modelo obtenido para un rango entre 0° a 65°.

modelo	RMSE	R2
1 modDoA1	15.77966	0.3666952

**Figura 3. 13.** Rendimiento del modelo entre 0°- 65° para la localización del ángulo del DoA.

$$DoA_{j_{0^{\circ}-65^{\circ}}} = 129.6 - 3.74 \cdot 10^{-7} * (Distancia_j * rsrq_j * rssi_j * rssi_{strongest_j} - rssi_j) + 5.463 * rsrq_j - 6.68 \cdot 10^{-9} * rssi_j^5 + 0.5629 * rssi_{strongest_j} + 7.667 \cdot 10^{-8} * (Distancia_j^3 + rsrq_j + rssi_j^3)$$

**Ecuación 3. 3.** Ángulos de la dirección de arribo estimados para un rango entre 0°- 65°

En la figura 3.4 se representa el histograma generado, el cual revela una mayor concentración de datos en el rango de 5 a 60°. Esta distribución se explica por la presencia de calles en los extremos de 0 a 5 y de 60 a 65°, donde no fue posible obtener muestras precisas para su inclusión en el modelo.

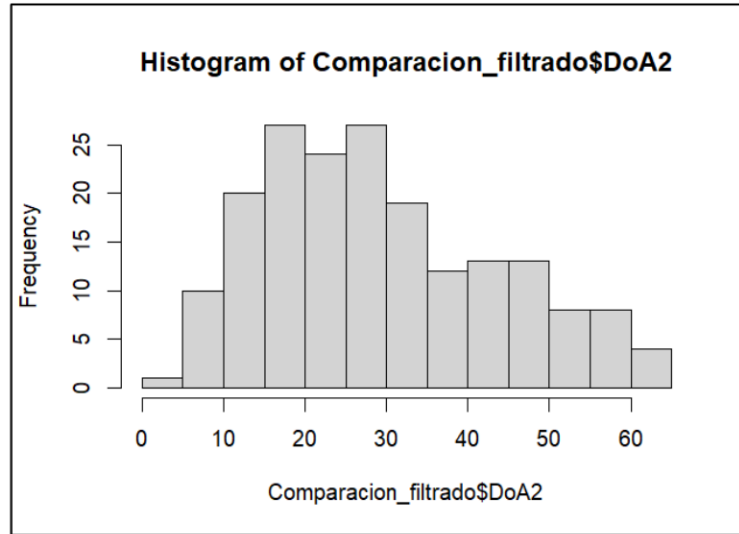


Figura 3. 14. Histograma de los resultados obtenidos para el rango entre 0°- 65°.

### 3.1.4 MODELO 170-250 GRADOS

Este modelo presenta una ecuación similar a la encontrada en el modelo anterior, pero con valores de intercepto y pendientes diferentes, lo cual se refleja en su rendimiento mostrado en la figura 3.16. Por otro lado, el resumen del modelo se exhibe en la figura 3.15. Los valores proporcionados por esta última figura permiten obtener la ecuación 3.4.

```
Call:
lm(formula = DoA2 ~ I(((Distancia) * rsrq * rssi * rssi_strongest -
  rssi)) + rsrq + I(rssi^5) + rssi_strongest + I(Distancia^3 +
  rsrq + rssi^3), data = training_base_doa)

Residuals:
    Min       1Q   Median       3Q      Max
-34.146  -7.885  -1.001   6.483  60.084

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.260e+02  9.976e+00  32.676 < 2e-16 ***
I(((Distancia) * rsrq * rssi * rssi_strongest - rssi)) -1.065e-06  1.291e-07  -8.250 5.71e-16 ***
rsrq              9.136e-02  2.603e-01   0.351  0.726
I(rssi^5)         1.135e-08  6.493e-10  17.475 < 2e-16 ***
rssi_strongest    1.094e+00  1.182e-01   9.260 < 2e-16 ***
I(Distancia^3 + rsrq + rssi^3) -6.841e-07  1.194e-07  -5.732 1.36e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.04 on 885 degrees of freedom
Multiple R-squared:  0.7098,    Adjusted R-squared:  0.7082
F-statistic:  433 on 5 and 885 DF,  p-value: < 2.2e-16
```

Figura 3. 15. Resumen obtenido para el modelo obtenido para un rango entre 170° a 250°.

modelo	RMSE	R2
1 modDoA1	13.1589	0.713693

Figura 3. 16. Rendimiento del modelo entre 170°- 250° para la localización del ángulo del DoA.

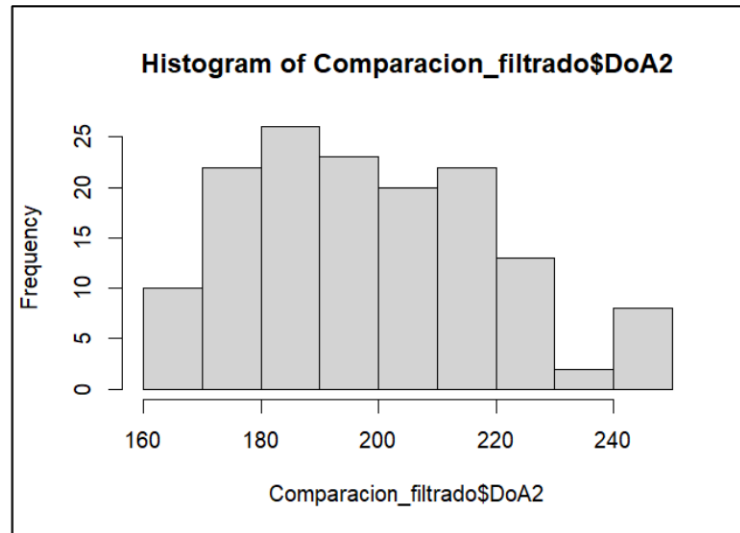
$$DoA_{j_{170^{\circ}-250^{\circ}}} = 326 - 1.065 \cdot 10^{-6} * (Distancia_j * rsrq_j * rssi_j * rssi_{strongest_j} - rssi_j) \\ + 0.09136 * rsrq_j + 1.135 \cdot 10^{-8} * rssi_j^5 + 1.09 * rssi_{strongest_j} - 6.841 \\ \cdot 10^{-7} * (Distancia_j^3 + rsrq_j + rssi_j^3)$$

**Ecuación 3. 4.** Ángulos de la dirección de arribo estimados para un rango entre 170°-250°

En la figura 3.17 se evidencian obstáculos entre los ángulos de 130° a 140°, lo que resulta en un histograma de la figura 3.18 con una menor cantidad de datos en dicho rango.



**Figura 3. 17.** Obstáculos entre 130° a 140°



**Figura 3. 18.** Histograma de los resultados obtenidos para el rango entre 170°- 250°.

### 3.1.5 MODELO 250-305 GRADOS

La ecuación 3.5, detallada en el resumen del modelo mostrado en la figura 3.19, presenta una estructura similar a los modelos previos, pero con valores de intercepto y pesos

ajustados. La figura 3.20 exhibe el rendimiento de esta ecuación, con un RMSE de 10.71, indicando su fiabilidad. El histograma representado en la figura 3.22 refleja las dificultades encontradas durante la toma de muestras, particularmente en áreas con calles e intersecciones, lo que concuerda con la distribución de datos observada en la figura 3.21.

```
Call:
lm(formula = DoA2 ~ I(((Distancia) * rsrq * rssi * rssi_strongest -
  rssi)) + rsrq + I(rssi^5) + rssi_strongest + I(Distancia^3 +
  rsrq + rssi^3), data = training_base_doa)

Residuals:
    Min       1Q   Median       3Q      Max
-30.7148  -6.4848  -0.2279   6.1426  28.0825

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.146e+02  4.121e+01  2.780  0.00578 **
I(((Distancia) * rsrq * rssi * rssi_strongest - rssi)) 7.454e-06  1.735e-06  4.297  2.33e-05 ***
rsrq        -8.910e+00  1.503e+00 -5.926  8.36e-09 ***
I(rssi^5)    -4.078e-09  2.653e-09 -1.537  0.12535
rssi_strongest -1.023e+00  4.237e-01 -2.415  0.01632 *
I(Distancia^3 + rsrq + rssi^3) 1.228e-05  2.128e-06  5.770  1.95e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.14 on 305 degrees of freedom
Multiple R-squared:  0.4233,    Adjusted R-squared:  0.4138
F-statistic: 44.77 on 5 and 305 DF,  p-value: < 2.2e-16
```

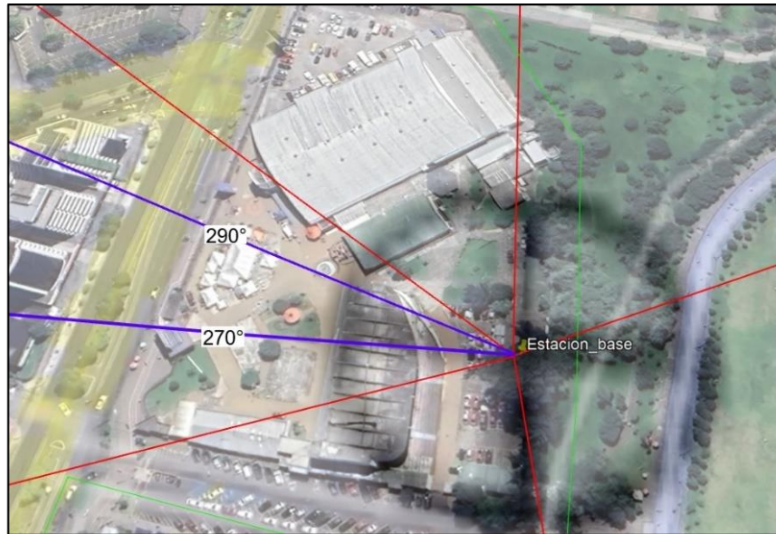
**Figura 3. 19.** Resumen obtenido para el modelo obtenido para un rango entre 250° a 305°.

$$DoA_{j_{250^{\circ}-305^{\circ}}} = 114.6 + 7.454 \cdot 10^{-6} * (Distancia_j * rsrq_j * rssi_j * rssi_{strongest_j} - rssi_j) - 8.91 * rsrq_j - 4.078 \cdot 10^{-9} * rssi_j^5 - 1.023 * rssi_{strongest_j} + 1.228 \cdot 10^{-5} * (Distancia_j^3 + rsrq_j + rssi_j^3)$$

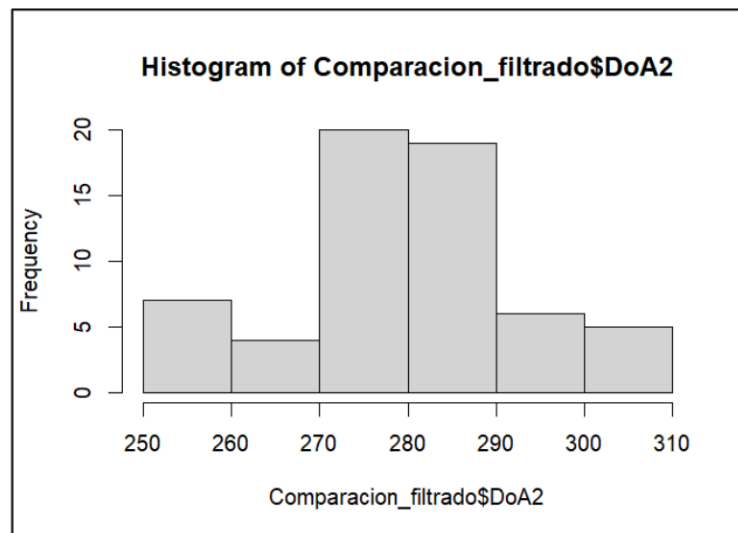
**Ecuación 3. 5.** Ángulos de la dirección de arribo estimados para un rango entre 250°-305°

modelo	RMSE	R2
1 modDoA1	10.71045	0.4707027

**Figura 3. 20.** Rendimiento del modelo entre 250°- 305° para la localización del ángulo del DoA.



**Figura 3. 21.** Apertura entre 270° a 290°



**Figura 3. 22.** Histograma de los resultados obtenidos para el rango entre 250°- 305°.

### 3.1.6 MODELO 305 A 360 GRADOS

La ecuación 3.6, generada por el modelo, adopta un enfoque distinto debido a la alta densidad de personas en el sector, como se aprecia visualmente. Esta ecuación fue evaluada según los criterios expuestos en la tabla 3.1, priorizando un menor valor de RSE. La figura 3.23 presenta un resumen detallado del modelo, mostrando los valores de intercepto y los pesos asignados a cada variable, derivados de la matriz de correlación analizada en la figura 3.24. El rendimiento del modelo se ilustra en la figura 3.25. El histograma de la figura 3.27 refleja las características del terreno de pruebas, donde las áreas cercanas a 305° se ven afectadas por la presencia de calles, mientras que la mayor



concentración de ángulos se observa en el rango de 335 a 360, como se evidencia en la figura 3.26.

```
Call:
lm(formula = DoA2 ~ I(Distancia^2) + I(rssi_strongest^20) + I(rssi^3),
    data = training_base_doa)

Residuals:
    Min       1Q   Median       3Q      Max
-25.065 -10.711  -0.253   8.343  39.861

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.217e+02  6.184e+00  52.014 < 2e-16 ***
I(Distancia^2)  3.416e-04  3.432e-05   9.953 < 2e-16 ***
I(rssi_strongest^20) 1.802e-38  6.859e-39   2.628  0.00899 **
I(rssi^3)       1.540e-05  1.335e-05   1.154  0.24939
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.68 on 342 degrees of freedom
Multiple R-squared:  0.242,    Adjusted R-squared:  0.2354
F-statistic: 36.4 on 3 and 342 DF,  p-value: < 2.2e-16
```

Figura 3. 23. Resumen obtenido para el modelo obtenido para un rango entre 305° a 360°.

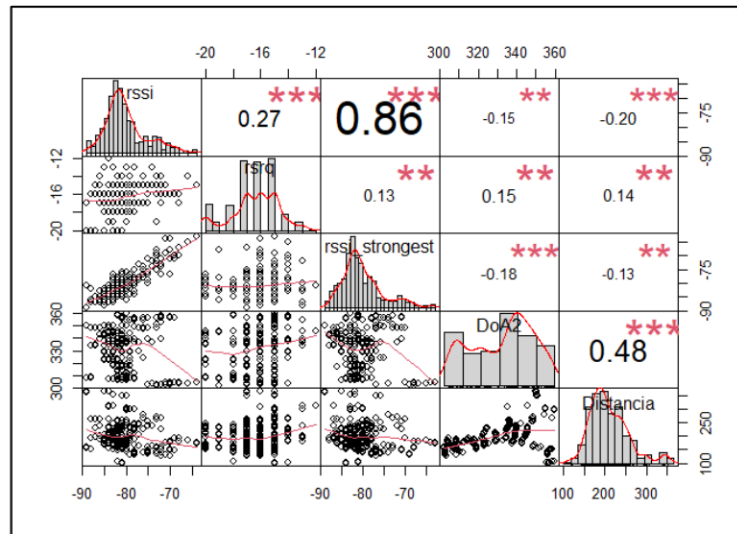


Figura 3. 24. Matriz de correlación para un rango de 305°-360°.

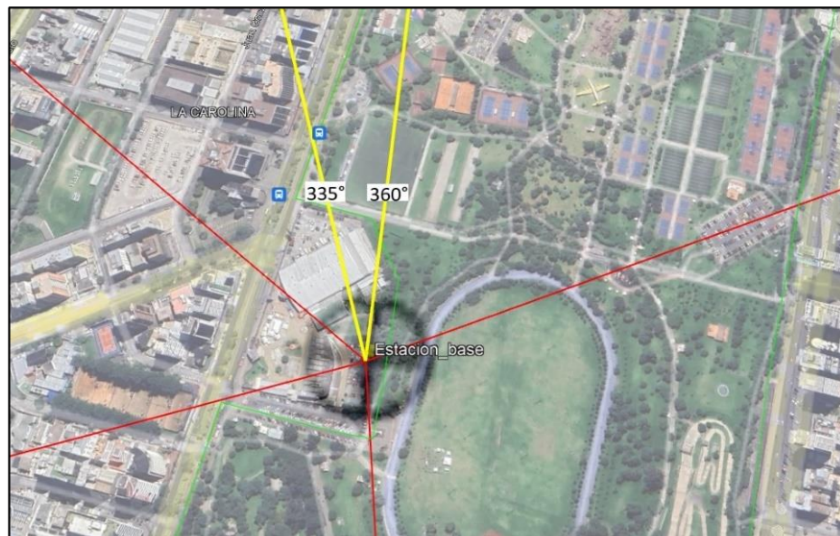
$$DoA_{j_{305^{\circ}-360^{\circ}}} = 321.7 + 3.416 \cdot 10^{-4} * Distancia_j^2 + 1.802 \cdot 10^{-38} * rssi_{strongest_j}^{20} + 1.54 \cdot 10^{-5} * rssi_j^3$$

Ecuación 3. 6. Ángulos de la dirección de arribo estimados para un rango entre 305°-360°

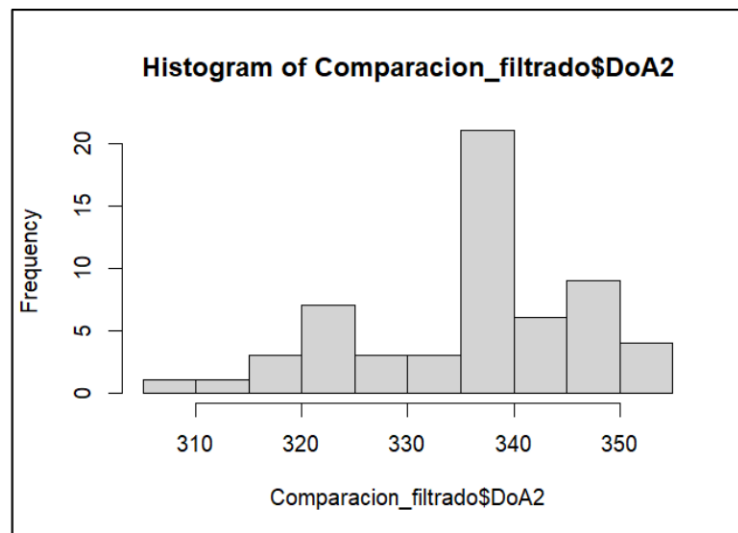
Las variables seleccionadas se basan en el hecho de que la distancia aporta la mayor cantidad de información para predecir el DoA. Además, se elevó el rssi strongest a un exponente alto para capturar información en las áreas más relevantes del modelo.

modelo	RMSE	R2
1 modDoA1	14.54602	0.261918

**Figura 3. 25.** Rendimiento del modelo entre 305°- 360° para la localización del ángulo del DoA.



**Figura 3. 26.** Apertura entre 335° a 360°



**Figura 3. 27.** Histograma de los resultados obtenidos para el rango entre 305°- 360°.

### 3.1.7 RESUMEN DE LOS MODELOS

Los modelos mostrados en los anteriores apartados se muestran en la tabla resumen 3.2.

**Tabla 3. 2.** Resumen de los modelos.

Rangos	Ecuación	Evaluación (RMSE)
0°-65°	$DoA_{j_{0^{\circ}-65^{\circ}}} = 129.6 - 3.74 \cdot 10^{-7}$ $* \left( Distancia_j * rsrq_j * rssi_j * rssi_{strongest_j} \right.$ $\left. - rssi_j \right) + 5.463 * rsrq_j - 6.68 \cdot 10^{-9} * rssi_j^5$ $+ 0.5629 * rssi_{strongest_j} + 7.667 \cdot 10^{-8}$ $* \left( Distancia_j^3 + rsrq_j + rssi_j^3 \right)$	15.78
65°-80°	$DoA_{j_{65^{\circ}-80^{\circ}}} = 48.770943 - 0.476991 * rssi_j + 0.103709 * rsrq_j$ $+ 0.123624 * rssi_{strongest_j} - 0.017958$ $* Distancia_j$	3.63
80°-170°	$DoA_{j_{80^{\circ}-170^{\circ}}} = 244.19740 - 0.53929 * rssi_j - 0.80461 * rsrq_j$ $+ 1.61790 * rssi_{strongest_j} - 0.15890$ $* Distancia_j$	16.68
170°-250°	$DoA_{j_{170^{\circ}-250^{\circ}}} = 326 - 1.065 \cdot 10^{-6}$ $* \left( Distancia_j * rsrq_j * rssi_j * rssi_{strongest_j} \right.$ $\left. - rssi_j \right) + 0.09136 * rsrq_j + 1.135 \cdot 10^{-8}$ $* rssi_j^5 + 1.09 * rssi_{strongest_j} - 6.841 \cdot 10^{-7}$ $* \left( Distancia_j^3 + rsrq_j + rssi_j^3 \right)$	13.15
250°-305°	$DoA_{j_{250^{\circ}-305^{\circ}}} = 114.6 + 7.454 \cdot 10^{-6}$ $* \left( Distancia_j * rsrq_j * rssi_j * rssi_{strongest_j} \right.$ $\left. - rssi_j \right) - 8.91 * rsrq_j - 4.078 \cdot 10^{-9} * rssi_j^5$ $- 1.023 * rssi_{strongest_j} + 1.228 \cdot 10^{-5}$ $* \left( Distancia_j^3 + rsrq_j + rssi_j^3 \right)$	10.71
305°-360°	$DoA_{j_{305^{\circ}-360^{\circ}}} = 321.7 + 3.416 \cdot 10^{-4} * Distancia_j^2 + 1.802$ $\cdot 10^{-38} * rssi_{strongest_j}^{20} + 1.54 \cdot 10^{-5} * rssi_j^3$	14.54

## CONCLUSIONES

Se logró el desarrollo de cinco modelos basados en técnicas de regresión lineal para la estimación del Direction of Arrival (DoA). Estos modelos representan herramientas útiles y precisas para la predicción de los ángulos de llegada de las señales en entornos de comunicaciones móviles.

La recopilación exhaustiva de datos a través de la aplicación Netmonitor Pro permitió obtener un conjunto de datos completo y representativo que abarca diversos escenarios y condiciones. El preprocesamiento riguroso garantizó la coherencia y calidad de los datos, preparándolos para el análisis.

Se implementaron y entrenaron con éxito los modelos de regresión lineal utilizando RStudio y las bibliotecas correspondientes. Esto demuestra la eficacia de la regresión lineal en la creación de modelos predictivos en entornos de comunicaciones móviles, donde se evidenció que el parámetro que más información brindó fue el  $r^2$  dado que este permite evaluar la calidad de la señal de referencia.

El desempeño de cada uno de los modelos se evaluó meticulosamente utilizando métricas relevantes como la precisión y el error medio cuadrado, el cual generó un valor mínimo de 3.53 y un valor máximo de 16.68, generando este error elevado debido a la vegetación que se cubría en la zona de toma de muestras. Los resultados demuestran la efectividad de los modelos en la estimación del DoA, identificando áreas de mejora para optimizar su rendimiento.

Este trabajo contribuye al avance de la tecnología de comunicaciones móviles al proporcionar modelos de regresión lineal para la estimación del DoA. Estos modelos tienen el potencial de mejorar el rendimiento y la eficiencia de los sistemas de comunicación inalámbrica, y su aplicabilidad y posibles mejoras pueden ser exploradas en investigaciones adicionales.

## RECOMENDACIONES

Realizar investigaciones adicionales en entornos variados para obtener una comprensión más completa de los factores que influyen en la precisión del Direction of Arrival (DoA), especialmente en áreas urbanas densamente pobladas y en entornos con terreno irregular.

Explorar técnicas de Machine Learning alternativas, además de la regresión lineal, para determinar si otras metodologías pueden mejorar aún más la precisión de la predicción del DoA en entornos móviles.

Investigar el potencial de la fusión de datos de múltiples fuentes, como datos de sensores adicionales y técnicas de posicionamiento por satélite, para mejorar la precisión de la localización del DoA en condiciones desafiantes.

Explorar la posibilidad de realizar pruebas en áreas más amplias con una mayor densidad de estaciones base para capturar una gama más diversa de escenarios y condiciones, lo que podría mejorar la precisión de los modelos de predicción del DoA.

Considerar la utilización de equipos especiales como drones para realizar mapeos aéreos y superar obstáculos como cruces de calles y terrenos irregulares, lo que permitiría una toma de muestras más efectiva y completa para la evaluación del DoA.

## 4 REFERENCIAS BIBLIOGRÁFICAS

- [1] Y. Kase, T. Nishimura, T. Ohgane, Y. Ogawa, D. Kitayama, y Y. Kishiyama, "Performance Analysis of DOA Estimation of Two Targets Using Deep Learning", en *2019 22nd International Symposium on Wireless Personal Multimedia Communications (WPMC)*, nov. 2019, pp. 1–6. doi: 10.1109/WPMC48795.2019.9096165.
- [2] M. A. B. Abbasi *et al.*, "Machine learning-assisted direction-of-arrival accuracy enhancement technique using oversized lens-loaded cavity", *IET Microw. Antennas Propag.*, vol. 16, núm. 6, pp. 305–315, 2022, doi: 10.1049/mia2.12257.
- [3] F. Cuesta y R. Jativa, "Desarrollo de Métodos de Seguimiento para un Sistema de Antena Inteligente basado en GPS", *Av. En Cienc. E Ing.*, vol. 5, dic. 2013, doi: 10.18272/aci.v5i2.141.
- [4] R. A. Calvo, "Tema 3 – Dimensionado de sistemas celulares".
- [5] "Comunicaciones móviles y sociedad | Telos". Consultado: el 27 de enero de 2024. [En línea]. Disponible en: <https://telos.fundaciontelefonica.com/archivo/numero084/comunicaciones-moviles-y-sociedad/>
- [6] "¿Cómo se comunican las células? El complejo lenguaje biomolecular", Explorer BioGen. Consultado: el 27 de enero de 2024. [En línea]. Disponible en: <https://explorerbiogen.wordpress.com/2016/05/31/comunicacion-celular/>
- [7] M. Martínez, "Categorías LTE: Velocidades de descarga de cada una de ellas", MovilZona. Consultado: el 10 de enero de 2024. [En línea]. Disponible en: <https://www.movilzona.es/2020/05/27/tipos-categorias-lte/>
- [8] "09751.pdf". Consultado: el 10 de enero de 2024. [En línea]. Disponible en: <https://dspace.uazuay.edu.ec/bitstream/datos/2563/1/09751.pdf>
- [9] C. Calle y M. S. Jiménez, "Estudio y Análisis Técnico Comparativo entre las Tecnologías LTE y LTE Advanced".
- [10] A. Londoño, W.-Y. Campo-Muñoz, y W. D. Jimenez, "Evaluación del desempeño de una red LTE mediante parámetros de QoS", *Rev. Espac.*, vol. 40, núm. 18, jun. 2019, Consultado: el 11 de enero de 2024. [En línea]. Disponible en: <https://www.revistaespacios.com/a19v40n18/19401808.html>
- [11] "2. Fundamento teórico".
- [12] "BoletinEstadistico-May2020-SMA-CoberturaInfraestructura.pdf". Consultado: el 16 de enero de 2024. [En línea]. Disponible en: <https://www.arcotel.gob.ec/wp-content/uploads/2015/01/BoletinEstadistico-May2020-SMA-CoberturaInfraestructura.pdf>
- [13] A. C. Müller y S. Guido, *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media, Inc., 2016.
- [14] "Modelos de Machine Learning", OpenWebinars.net. Consultado: el 28 de enero de 2024. [En línea]. Disponible en: <https://openwebinars.net/blog/modelos-de-machine-learning/>
- [15] R. S. Sutton y A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 1998.
- [16] Rafa, "Qué es R y RStudio? [curso de R Studio] Rafa González Gouveia", Rafa González Gouveia. Consultado: el 16 de enero de 2024. [En línea]. Disponible en: <https://gonzalezgouveia.com/que-es-r-y-rstudio/>
- [17] "Qué es R y por qué utilizarlo", OpenWebinars.net. Consultado: el 16 de enero de 2024. [En línea]. Disponible en: <https://openwebinars.net/blog/que-es-r-y-por-que-utilizarlo/>
- [18] "Tema 6a. Primeros pasos en r-studio - Estadisticaparatodos.com". Consultado: el 16 de enero de 2024. [En línea]. Disponible en: <https://estadisticaparatodos.com/primeros-pasos-en-r-studio/>
- [19] "Descarga de la aplicación NetMonitor Pro 2024 - Gratis - 9Apps". Consultado: el 14 de febrero de 2024. [En línea]. Disponible en: [https://www.9apps.com/es/android-apps/ru-v\\_a\\_v-netmonitorpro/](https://www.9apps.com/es/android-apps/ru-v_a_v-netmonitorpro/)

- [20] "ITU-T Recommendation database", ITU. Consultado: el 23 de enero de 2024. [En línea]. Disponible en: <https://www.itu.int/itu-t/recommendations/rec.aspx?rec=12831&lang=es>
- [21] "Geographic Information Science and Systems, 4th Edition | Wiley", Wiley.com. Consultado: el 28 de enero de 2024. [En línea]. Disponible en: <https://www.wiley.com/en-us/Geographic+Information+Science+and+Systems%2C+4th+Edition-p-9781119128458>
- [22] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc., 2019.
- [23] M. Kuhn y K. Johnson, *Applied Predictive Modeling*. New York, NY: Springer New York, 2013. doi: 10.1007/978-1-4614-6849-3.
- [24] "tidymodels - Aprende". Consultado: el 6 de febrero de 2024. [En línea]. Disponible en: <https://www.tidymodels.org/learn/>
- [25] "recipe function - RDocumentation". Consultado: el 7 de febrero de 2024. [En línea]. Disponible en: <https://www.rdocumentation.org/packages/recipes/versions/1.0.9/topics/recipe>
- [26] F. Hernández, O. Usuga, y M. Mazo, *1 Regresión lineal simple | Modelos de Regresión con R*. Consultado: el 7 de febrero de 2024. [En línea]. Disponible en: [https://fhernanb.github.io/libro\\_regresion/rls.html](https://fhernanb.github.io/libro_regresion/rls.html)
- [27] "Oracle® Fusion Cloud EPM Trabajo con Planning". Consultado: el 7 de febrero de 2024. [En línea]. Disponible en: [https://docs.oracle.com/cloud/help/es/pbcs\\_common/PFUSU/insights\\_metrics\\_RMSE.htm#PFUSU-GUID-FD9381A1-81E1-4F6D-8EC4-82A6CE2A6E74](https://docs.oracle.com/cloud/help/es/pbcs_common/PFUSU/insights_metrics_RMSE.htm#PFUSU-GUID-FD9381A1-81E1-4F6D-8EC4-82A6CE2A6E74)
- [28] "Aritmética, Matemáticas y Estadística - Kit de Habilidades Académicas". Consultado: el 7 de febrero de 2024. [En línea]. Disponible en: <https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/regression-and-correlation/coefficient-of-determination-r-squared.html>
- [29] "rseek.org - rstats search engine". Consultado: el 8 de febrero de 2024. [En línea]. Disponible en: <https://rseek.org/>

## **5 ANEXOS**