

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
ESTIMACIÓN DEL DOA EN BASE A MEDICIONES DE CAMPO Y
UNA TÉCNICA DE MACHINE LEARNING**

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

LENIN ALEXANDER GUACHALÁ ANDRANGO

(lenin.guachala@epn.edu.ec)

DIRECTOR: ING. PABLO ANÍBAL LUPERA MORILLO, PHD.

(pablo.lupera@epn.edu.ec)

DMQ, abril 2024

CERTIFICACIONES

Yo, Lenin Alexander Guachalá Andrango 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.

Lenin Alexander Guachalá Andrango

Certifico que el presente trabajo de integración curricular fue desarrollado por Lenin Alexander Guachalá Andrango, bajo mi supervisión.

Ing. Pablo Aníbal Lupera Morillo, PhD.
DIRECTOR

DECLARACIÓN DE AUTORÍA

A través de la presente declaración, afirmamos que el trabajo de integración curricular aquí descrito, así como el producto resultante 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.

Lenin Alexander Guachalá Andrango

Ing. Pablo Aníbal Lupera Morillo, PhD.

DEDICATORIA

A mis amados padres, Gonzalo y Martha, por su amor incondicional, confianza, apoyo constante, enseñanzas y valores que aportan en cada etapa de mi vida, son mi fuente de inspiración y me demuestran que no existen sueños imposibles. A mis queridas mascotas, que son mis compañeros fieles y mi consuelo en los momentos más difíciles, brindándome alegrías y momentos compartidos que permanecerán conmigo. Esto es para ustedes, que son pilares fundamentales de mi vida y mi éxito.

AGRADECIMIENTO

A mi querida Tiffany Suárez, te agradezco por acompañarme en estas instancias de mi vida, tu amor, paciencia y apoyo incondicional han sido fundamentales para este logro, gracias por estar en momentos de felicidad y tristeza, cada palabra de aliento a superarme y creer en mis capacidades, eres mi inspiración para mejorar de manera personal y profesional, siempre tendrás mi apoyo incondicional.

Agradezco de manera especial a mis padres y mis abuelos, que me han enseñado a no rendirme en ninguna circunstancia de la vida, sin ellos no habría llegado hasta aquí. A todos mis amigos dentro y fuera de la universidad, que me brindaron su apoyo e influyeron para continuar con mi etapa universitaria.

También quiero agradecer a cada uno de los profesores que se convirtieron en amigos y mentores, brindándome su apoyo y sabiduría para mi crecimiento personal y profesional, son fuente de inspiración y motivación dentro y fuera de las aulas.

Finalmente, quiero expresar mi agradecimiento y gratitud hacia el profesor Pablo Lupera, por brindarme sus conocimientos, guiarme y soportar mis dudas a lo largo de este trabajo de integración curricular.

ÍNDICE DE CONTENIDO

CERTIFICACIONES	I
DECLARACIÓN DE AUTORÍA	II
DEDICATORIA	III
AGRADECIMIENTO	IV
ÍNDICE DE CONTENIDO.....	V
RESUMEN	VIII
ABSTRACT.....	IX
1 INTRODUCCIÓN.....	1
1.1 Objetivo general	1
1.2 Objetivos específicos	1
1.3 Alcance	2
1.4 Marco teórico	2
1.4.1 DoA (Dirección de Arribo)	2
1.4.1.1 Definición	2
1.4.2 Aplicaciones y técnicas utilizadas para estimación del DoA.....	3
1.4.2.1 Método de antenas inteligentes (Smart Antennas).....	3
1.4.2.2 Método de formación de Haz (Beamforming).....	3
1.4.2.3 Método MUSIC	3
1.4.2.4 Método ESPRIT	4
1.4.3 Red celular LTE.....	4
1.4.3.1 Definición	4
1.4.3.2 Banda de frecuencias	4
1.4.3.3 Elementos de la red LTE	5
1.4.3.4 Sectorización.....	6
1.4.3.5 Potencia de transmisión	7
1.4.3.6 Azimut	7
1.4.3.7 Ángulo de la antena.....	8
1.4.4 Machine Learning	8
1.4.4.1 Técnicas de Machine Learning.....	8
1.4.4.2 Descripción de etapa de recolección, preprocesamiento, entrenamiento, prueba, evaluación y ajuste.....	9
1.4.4.3 Método de aprendizaje supervisado y algoritmo de regresión lineal	10
1.4.5 Infraestructura de la estación base.....	11
1.4.5.1 Altura de la antena.....	11
1.4.5.2 Número de antenas	11

1.4.5.3	Apuntamiento de las antenas (ángulo de elevación).....	12
1.4.5.4	Tipo de zona en mediciones.....	12
1.4.6	RStudio.....	13
1.4.6.1	Capacidades.....	13
2	METODOLOGÍA.....	13
2.1	Herramientas de recolección.....	14
2.1.1	Google Maps.....	14
2.1.2	Google Earth Pro.....	14
2.1.3	Wikiloc.....	14
2.1.4	NetMonitor Pro.....	15
2.1.5	CellMapper.....	16
2.2	Método de recolección de datos.....	16
2.3	Recolección de datos.....	19
2.3.1	Descripción del proceso y etiquetado de medición radial y de circunferencias.....	20
2.3.1.1	Radiales.....	20
2.3.1.2	Circunferencias.....	21
2.3.1.3	Cronograma de la etapa de recolección de datos.....	22
2.4	Preprocesamiento de Datos.....	22
2.4.1	Obtención de archivos “.csv”.....	23
2.4.2	Clasificación y generación de Archivos de excel de mediciones obtenidas	23
2.4.3	Atributos adicionales.....	23
2.4.3.1	Distancia entre BS y MS.....	24
2.4.3.2	Ángulo de direccionamiento (DoA).....	26
2.4.4	Filtrado de atributos erróneos en medición.....	31
2.5	Implementación De la técnica de Machine Learning en RStudio.....	31
2.5.1	Ingreso y configuración de atributos en archivo (dataset).....	32
2.5.1.1	Eliminación de atributos y cambio de tipo de atributo.....	32
2.5.2	Análisis de tendencia entre parámetros de radiofrecuencia y el DoA.....	33
2.5.2.1	Análisis sin rangos del DoA.....	34
2.5.2.2	Análisis con rangos del DoA.....	38
2.5.3	Implementación de una técnica de Machine Learning.....	40
2.5.3.1	Creación de dataset y observación de histogramas en atributos.....	40
2.5.3.2	Matriz de correlación.....	40
2.5.3.3	Implementación de la receta.....	41
2.5.3.4	Generación de modelo.....	41

2.5.3.5	Ajuste del modelo	41
2.5.3.6	Evaluación del modelo.....	42
2.5.3.7	Comparación de modelos.....	42
2.5.4	Implementación de los modelos del DoA por rangos	42
3	RESULTADOS, CONCLUSIONES Y RECOMENDACIONES	44
3.1	Resultados	44
3.1.1	Implementación de una técnica de Machine Learning	44
3.1.1.1	Matriz de correlación obtenida	44
3.1.1.2	Implementación de la receta	45
3.1.1.3	Generación, ajuste y rendimiento de los modelos obtenidos	46
3.1.1.4	Comparación de modelos mediante análisis estadístico.....	47
3.1.2	Implementación de modelos del DoA por rangos.....	48
3.1.2.1	Matriz de correlación obtenida	48
3.1.2.2	Implementación de las recetas.....	50
3.1.2.3	Generación, ajustes y evaluación de los modelos obtenidos.....	51
3.1.2.4	Comparación de modelos mediante un análisis estadístico.....	56
3.1.3	Resumen de resultados	58
3.2	Conclusiones	58
3.3	Recomendaciones	60
4	REFERENCIAS BIBLIOGRÁFICAS.....	60

RESUMEN

El presente estudio permite indagar la aplicación de una técnica de Machine Learning con el objetivo de predecir la dirección de arribo (DoA) dentro de las redes celulares basándose en mediciones de campo y parámetros de radiofrecuencia. Se elaboran los correspondientes modelos de predicción haciendo uso de los datos recolectados de una estación base específica, con el correspondiente preprocesamiento de datos para la identificación y detección de patrones que permiten realizar una estimación del DoA. Los resultados que se obtienen indican que los modelos desarrollados durante el proceso de aprendizaje automático se diferencian de los métodos tradicionales haciendo uso de una aplicación para medición de parámetros de radiofrecuencia, ofreciendo herramientas valiosas para la ingeniería en Telecomunicaciones en el ámbito de la optimización de las redes celulares. La metodología empleada y los resultados del estudio contribuirían en gran manera al avance y mejoramiento de la cobertura y calidad de servicio en una red celular LTE, sobresaliendo mediante la aplicación de técnicas de Machine Learning en el análisis de parámetros de radiofrecuencia.

PALABRAS CLAVE: DoA, Machine Learning, red celular, Estación móvil (MS), Estación base (BS), handover.

ABSTRACT

The present study allows us to investigate the application of a Machine Learning technique with the objective of predicting the direction of arrival (DoA) within cellular networks based on field measurements and radio frequency parameters. The corresponding prediction models are developed from the data collected from a specific base station, with the corresponding data preprocessing for the identification and detection of patterns that allow an estimation of the DoA. The results obtained indicate that the models developed during the machine learning process differ from traditional methods by using an application for measuring radio frequency parameters, offering valuable tools for Telecommunications Engineering in the field of communications optimization. cellular networks. The methodology used and the results of the study would contribute greatly to the advancement and improvement of coverage and quality of service in an LTE cellular network, excelling through the application of Machine Learning techniques in the analysis of radio frequency parameters.

KEYWORDS: DoA, Machine Learning, cellular network, Mobile Station (MS), Base Station (BS), handover.

1 INTRODUCCIÓN

El presente estudio aborda el problema de la estimación de la Dirección de Arribo (DoA) en redes celulares mediante la aplicación de una técnica de Machine Learning, elaborando mediciones para la obtención de parámetros de radiofrecuencia como por ejemplo RSSI STRONGEST, RSSNR y el RSSI [1], [2], [3]. Este estudio se enfoca en la implementación de un método innovador que combina mediciones reales de campo elaborando un análisis de datos, buscando superar las limitaciones y usos de los métodos convencionales en la estimación del DoA a través de una técnica de regresión lineal.

La búsqueda de una detección precisa del DoA en base a predicciones mediante el uso de aprendizaje automático presenta una forma actual de obtener patrones, elaborar clasificaciones, predicción de atributos específicos, automatización de procesos y entre otros beneficios que pueden brindar a la red celular con el fin de mejorar la localización de fuentes de señales, sistemas de radar y comunicación móvil [4], [5], [6].

1.1 OBJETIVO GENERAL

Desarrollar un modelo de obtención de la dirección de arribo (DoA) utilizando una técnica de Machine Learning haciendo uso de datos obtenidos en una estación base específica con la aplicación NetMonitor Pro.

1.2 OBJETIVOS ESPECÍFICOS

Para el cumplimiento del objetivo general se tienen los siguientes objetivos específicos.

1. Obtener mediciones de parámetros de radiofrecuencia en base a un patrón determinado para realizar mediciones considerando rutas existentes dentro de la cobertura de la estación base.
2. Clasificar y filtrar los datos en base a parámetros útiles de radiofrecuencia.
3. Analizar los datos obtenidos mediante una técnica Machine Learning supervisada específica.
4. Evaluar el rendimiento del modelo, comparando el DoA estimado respecto a las posiciones reales del dispositivo móvil.

1.3 ALCANCE

En el alcance se define el estudio para la estimación del DoA en base a parámetros de radiofrecuencia definidos para una red celular de una estación base determinada mediante la recolección de datos con un patrón de medición definido. En la recolección de los datos se hará uso de una aplicación para registrar parámetros, tales como: RSSI STRONGEST, RSRQ, RSSI, RSSNR, PCI, CI y EARFCN. En la primera fase del estudio del DoA se hará la recolección de los datos de radiofrecuencia de una estación base en diferentes condiciones. En la segunda fase de estudio del DoA se realizará el procesamiento y análisis de los datos recolectados en base al uso de una técnica de Machine Learning determinada, como primer paso se clasifican los datos que se desean emplear y se definen las variables de entrada que se usarán en el modelo seleccionado de Machine Learning, el modelo se basará en el uso de un lenguaje de programación a través del software denominado “R Studio”, con esto se desea obtener el modelo adecuado que permita obtener un patrón y determinar el DoA. Como tercera fase se tiene el análisis de los resultados y la evaluación del modelo de aprendizaje automático (Machine Learning) escogido.

1.4 MARCO TEÓRICO

1.4.1 DOA (DIRECCIÓN DE ARRIBO)

1.4.1.1 Definición

La dirección de arribo o también denominado “DoA” como se observa en la **Figura 1.1**, hace referencia a la dirección del ángulo de incidencia desde la cual una señal electromagnética o de otra forma de onda es captada mediante un dispositivo, sensor o antena. La determinación del DoA es esencial para diversas aplicaciones como localización de fuentes sonoras, sistemas de localización y sistemas de comunicación inalámbricos avanzados [2], [7].

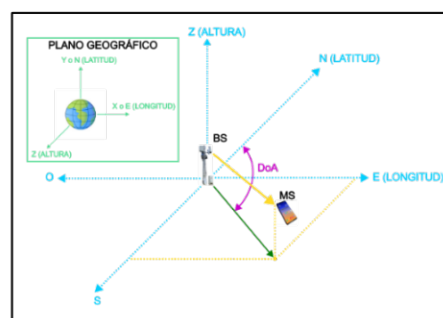


Figura 1.1. Dirección de arribo basado en ángulo de azimut y cartografía [7], [8], [9], [10].

1.4.2 APLICACIONES Y TÉCNICAS UTILIZADAS PARA ESTIMACIÓN DEL DOA

Las técnicas utilizadas para la estimación del DoA y que se describen brevemente más adelante, son de fundamental importancia en diversas áreas donde se necesita determinar la dirección desde la cual proviene una señal, entre las aplicaciones y propósitos que se usan se tiene:

- **Sistemas de comunicación inalámbricos:** Mejoramiento en la calidad de la comunicación con un uso eficiente del espectro y reducción de interferencia, permitiendo adaptación de antenas en transmisión y recepción de señales en direcciones específicas [2].
- **Radars:** Determina la dirección desde la cual un objetivo refleja la señal electromagnética, de esta manera permite el rastreo de objetos móviles [11], [12].
- **Sonido y acústica:** Determinación de fuentes acústicas como submarinos o monitorización de vida marina [13].
- **Antenas inteligentes MIMO:** facilita y permite la adaptación de patrones de radiación en antenas para transmisión de señales en direcciones específicas permitiendo un aumento de capacidad y confiabilidad en la comunicación [2], [3].

1.4.2.1 Método de antenas inteligentes (Smart Antennas)

Las antenas inteligentes utilizan arrays o arreglos de antenas que permiten una capacidad de adaptación mejorando de esta manera la resolución espacial y la precisión en la estimación del DoA [7], [14].

1.4.2.2 Método de formación de Haz (Beamforming)

Utilizan arrays o arreglos de antenas que permite formar haces de señales y determinar la dirección de llegada dominante acorde al haz del lóbulo de radiación, de esta manera permite estimar el DoA [1], [7].

1.4.2.3 Método MUSIC

El método MUSIC se basa en un análisis sobre la matriz de covarianza de las señales captadas con el propósito de discernir las direcciones desde las cuales provienen las fuentes de señales de onda. Este proceso de análisis permite la separación de las fuentes en el dominio de la frecuencia, lo que permite obtener una estimación precisa del DoA [1], [4], [7].

1.4.2.4 Método ESPRIT

El método ESPRIT hace uso de la invarianza rotacional dentro de un subespacio a través de dos sub-arreglos de antenas desplazados en movimiento de traslación para estimar los parámetros de señales, incluyendo el DoA [4], [7].

1.4.3 RED CELULAR LTE

1.4.3.1 Definición

La red LTE con sus siglas en inglés (Long-Term Evolution) es una tecnología de comunicación inalámbrica de banda ancha que permite la transmisión de datos a alta velocidad en las redes móviles. LTE es la evolución de la tecnología móvil 3G y está diseñada con el objetivo de ofrecer mayor velocidad de datos, reducción o minoración en la latencia y una mayor eficiencia espectral [15], [16].

La arquitectura de una red LTE como se muestra en la **Figura 1.2**, consta de dos partes esenciales: núcleo de la red y red de acceso, además consta de elementos como estaciones base (eNodeB), que se conectan a través de una red de acceso al núcleo EPC (Evolved Packet Core), el cual está relacionado al núcleo de red y la red de acceso que se encuentra interconectada entre estaciones base. La interfaz radioeléctrica utiliza técnicas avanzadas de modulación y acceso al medio para lograr un rendimiento óptimo en el sistema [15], [16].

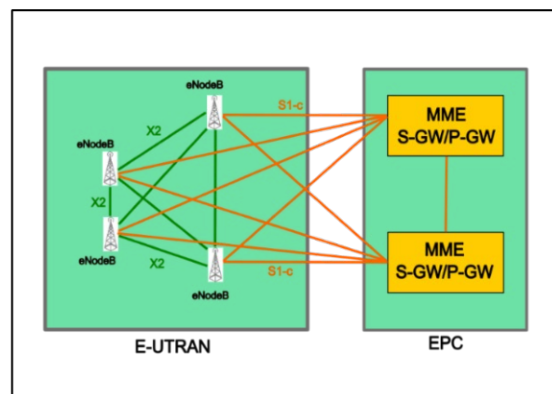


Figura 1.2. Arquitectura de la red celular LTE [15], [16].

1.4.3.2 Banda de frecuencias

La banda de frecuencias en una red celular LTE se refiere al conjunto específico de frecuencias asignadas para la transmisión de datos inalámbricos utilizando la tecnología LTE. El espectro de frecuencias se divide en diferentes bandas para satisfacer las necesidades de capacidad, cobertura y eficiencia espectral. La asignación de la banda de frecuencias puede variar según las regulaciones y políticas regionales [17].

Las bandas de frecuencia asignadas para cada uno de los diferentes operadores como se observa en la **Figura 1.3** son [17]:

- Conecel (Claro): 1700/2100 MHz (banda 4)
- Otecel (Movistar): 1900 MHz (banda 2)
- CNT: 1700/2100 MHz (banda 4) y 700 MHz (banda 28)

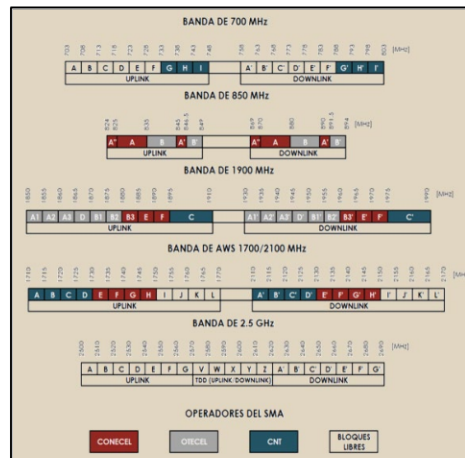


Figura 1.3. Banda de frecuencias en Ecuador tomado de Boletín estadístico Arcotel [17].

1.4.3.3 Elementos de la red LTE

Los elementos de una red celular LTE que se pueden observar en la **Figura 1.4** son componentes esenciales que trabajan en conjunto para proporcionar conectividad inalámbrica de alta velocidad y baja latencia, entre los elementos que conforman la estructura de la red celular LTE se tiene [15], [16]:

- **Estación base (eNodeB):** Elemento encargado de la transmisión y recepción de datos hacia y desde los dispositivos móviles.
- **Dispositivo de usuario o estación móvil (UE):** Dispositivo capaz de hacer uso de la red LTE, como teléfonos y tabletas.
- **E-UTRAN (Evolved Universal Terrestrial Radio Access Network):** Es la parte de la red LTE que incluye todas las estaciones base (eNodeB). Es la interfaz de radio evolucionada que proporciona la conectividad inalámbrica entre los UEs y las estaciones base.
- **EPC (Evolved Packet Core):** Corresponde al núcleo de la red LTE y se encuentra conformado de varios elementos, incluidos el MME (Mobility Management Entity), el SGW (Serving Gateway), y el PGW (PDN Gateway). Estos elementos son

responsables de la gestión de la movilidad, la gestión de sesiones y la conexión a la red externa.

- **MME (Mobility Management Entity):** Permite gestionar la señalización de movilidad, como el registro y desregistro del UE en la red, la autenticación y la gestión de la ubicación.
- **SGW (Serving Gateway):** Responsable de la gestión de la movilidad a nivel local y de la conmutación de paquetes de datos entre el UE y la red central.
- **PGW (PDN Gateway):** Permite conectar la red LTE con la red externa, como el Internet. Gestiona la asignación de direcciones IP y enruta los datos entre la red LTE y la red externa.
- **BACKHAUL:** Esta parte de la infraestructura permite conectar las estaciones base (eNodeB) con el núcleo de la red (EPC) y, a menudo, se utiliza para transportar grandes volúmenes de datos entre ellos [16].

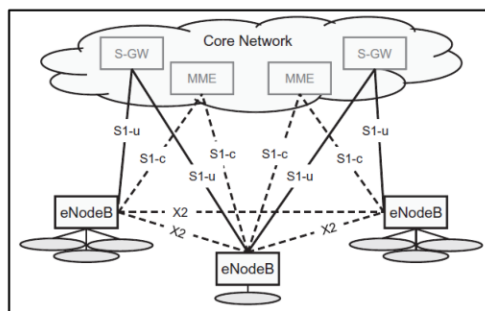


Figura 1.4. Elementos de la red LTE tomado de LTE/ LTE-Advanced for mobile Broadband [15], [16].

1.4.3.4 Sectorización

El concepto de sectorización como ejemplo de la **Figura 1.5**, se basa en la división de una celda de red móvil en sectores más pequeños, esto tiene como ventajas una mejora en la eficiencia y capacidad de la red [2], [3], [16]. Cada sector se representa con una antena dentro de la celda sectorizada, entre los beneficios se tiene:

- Mayor capacidad de la red
- Mejora en la zona de cobertura
- Reducción de interferencias

Se debe recalcar que al momento de aplicar el concepto de sectorización también existen desventajas que deben ser consideradas, entre las más importantes se tiene:

- Aumento de costos en equipos
- Complejidad de gestión
- Interferencias internas entre sectores

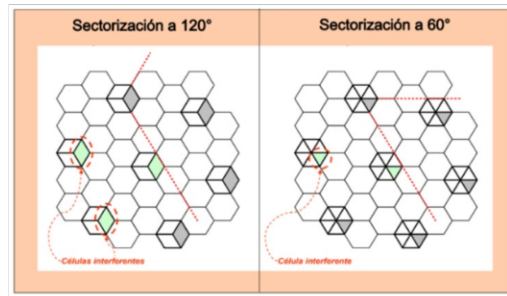


Figura 1.5. Sectorización en una red celular LTE tomado de *Dimensionado de sistemas celulares* [2], [3], [18].

1.4.3.5 Potencia de transmisión

La potencia de transmisión se refiere a la cantidad de energía radiada por un dispositivo de transmisión, como un teléfono móvil o una estación base, para comunicarse a través de la red. La gestión de la potencia es crucial para garantizar una comunicación eficiente y confiable [2], [3], [15], [16].

1.4.3.6 Azimut

El concepto de Azimut se define como una medida angular en el plano horizontal como se evidencia en la **Figura 1.6**, generalmente expresada en grados, que indica la dirección de un punto en relación con el norte geográfico. En términos más simples, el azimut representa la orientación angular de un objeto o ubicación con respecto a un punto de referencia [8], [9].

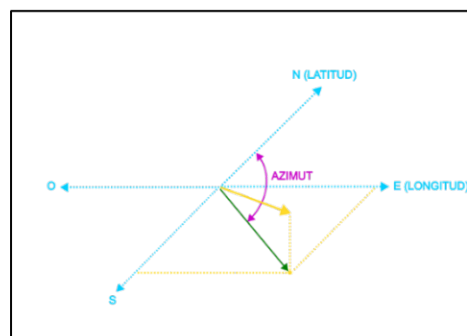


Figura 1.6. Representación del Azimut basado en cartografía [8], [9], [19].

1.4.3.7 Ángulo de la antena

En el concepto de antenas de comunicación inalámbrica, existen dos ángulos fundamentales, ángulo de elevación y el ángulo de azimut observados en la **Figura 1.7**. Estos ángulos son cruciales para determinar la dirección en la que una antena debe apuntar para establecer una conexión de enlace efectiva.

- **Ángulo de elevación:** Este ángulo mide la inclinación vertical de la antena con respecto al plano horizontal. Es esencial para determinar la altura a la que se debe colocar una antena para establecer una conexión efectiva entre dos puntos.
- **Ángulo de azimut:** Este ángulo mide la dirección horizontal de la antena con respecto al norte magnético. Indica la orientación en grados hacia la cual la antena debe apuntar para alcanzar su objetivo [8], [9], [19].

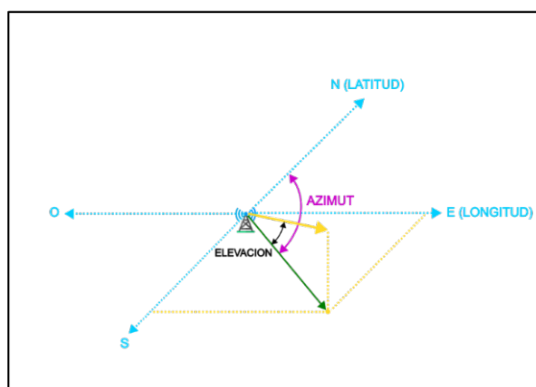


Figura 1.7. Representación de ángulos físicos en antenas.

1.4.4 MACHINE LEARNING

Machine Learning es el concepto de estudio que brinda a un computador la capacidad de ser programado de manera no explícita para que pueda aprender, determinar y predecir información o patrones en base a datos proporcionados [5].

1.4.4.1 Técnicas de Machine Learning

Existen diferentes técnicas de Machine Learning, las cuales se adaptan acorde a los problemas que se desean solventar, existe una gran cantidad de aplicaciones de estas técnicas como detección de la voz, predicción de diagnósticos médicos, clasificación de datos, etc.

Dentro de estas técnicas existen tres diferentes tipos, las cuales son:

- **Aprendizaje supervisado:** En este tipo de aprendizaje, los datos que se ingresan para el entrenamiento poseen etiquetas, es decir que el algoritmo es entrenado con

las entradas de datos y las respuestas o salidas esperadas a las que se nombra con etiquetas.

- Aprendizaje no supervisado: En este tipo de aprendizaje, los datos que se ingresan para el entrenamiento no poseen ninguna etiqueta, es decir que no se saben de qué tipo son los datos, de esta manera el algoritmo buscará de manera autónoma patrones que permitan interpretar la tendencia que poseen los datos ingresados.
- Aprendizaje por refuerzo: En este tipo de aprendizaje el algoritmo es entrenado en base a la observación de acciones que se realizan con los datos, la información de entrada es la retroalimentación que obtiene en base a las decisiones tomadas que pueden ser juzgadas como correctas o incorrectas, y de esta manera genera políticas para aprender cual es la acción que más le beneficia al conjunto de datos [5], [6], [20].

1.4.4.2 Descripción de etapa de recolección, preprocesamiento, entrenamiento, prueba, evaluación y ajuste

Existen etapas necesarias dentro del proceso de aprendizaje automático, a continuación, se describen cada una de ellas para la aplicación de técnicas de Machine Learning [21]:

- Recolección: En la etapa de recolección de información o datos, se procede a obtener la información de datos necesaria en cuanto al estudio que se elabora, en este caso, se recolectan mediciones de parámetros de radiofrecuencia en un dispositivo móvil (MS) bajo diferentes lineamiento descritos en los métodos de recolección, además, se consideran atributos adicionales para el uso de aprendizaje supervisado denominados etiquetas, tales como distancia y DoA.
- Preprocesamiento: Para la etapa de preprocesamiento se busca que el conjunto de datos sea depurado, es decir, que todos los datos erróneos o que pueden causar conflicto en la etapa de entrenamiento sean eliminados, garantizando resultados al hacer uso de una técnica de Machine Learning.
- Entrenamiento: dentro de esta etapa de entrenamiento, el modelo generado en base a la técnica de Machine Learning escogida es entrenado mediante el conjunto de datos de entrenamiento, el cual se conforma por atributos de entrada y salida con el objetivo de detectar un patrón y realizar una predicción.
- Prueba: En la etapa de prueba se busca el poder evaluar el modelo generado mediante el conjunto de datos de prueba, esto quiere decir que los datos ingresados al modelo son distintos a los de entrenamiento, esto permite ver resultados de

patrones en base a un conjunto desconocido, con el objetivo de predecir una salida específica en base a lo aprendido en la etapa de entramiento.

- Evaluación: La etapa de evaluación es la más importante dentro del uso de la técnica de Machine Learning, esta etapa permite analizar el rendimiento del modelo generado, esto quiere decir que nos permite obtener mediciones estadísticas, como el error cuadrático, que permiten discernir si el modelo es funcional, además permite comparar los atributos adicionales calculados con la estimación de los atributos de salida del modelo, comprobando así su eficiencia en la predicción.
- Ajuste: La etapa de ajuste es crucial, esta etapa permite mejorar el modelo generado en caso de que el rendimiento de este no sea satisfactorio, brindando la opción de modificar, añadir o eliminar atributos que aporten con el aumento del rendimiento del modelo [21].

1.4.4.3 Método de aprendizaje supervisado y algoritmo de regresión lineal

Dentro del método de aprendizaje supervisado existen diferentes clasificaciones, en este estudio nos centraremos en el uso del algoritmo de regresión lineal.

- Regresión Lineal: El algoritmo de regresión lineal permite realizar una estimación del atributo de salida deseado, su funcionamiento se basa en establecer relaciones entre el atributo de salida o variable independiente y los atributos de entrada o variables dependientes [5], [6], [20].
- Atributo de salida: Variable independiente de interés a obtener, en base a los atributos de entrada, esta es obtenida mediante el algoritmo aplicado [6], [20].
- Atributos de entrada: Variables independientes que son obtenidas del preprocesamiento de los datos y de los atributos adicionales que se añaden dentro de este proceso, estas pueden o no poseer una relación con el modelo que se desee implementar [6], [20].
- Slope e intercept: el Slope e intercept son coeficientes obtenidos de la minimización de la suma de la diferencia al cuadrado de la distancia entre los puntos de los datos y la línea de regresión [6].
- Hiperparámetros: Son atributos adicionales externos, obtenidos y configurados de manera manual antes de implementar el modelo [6], [20].

1.4.5 INFRAESTRUCTURA DE LA ESTACIÓN BASE

Para este estudio se debe tomar en cuenta la infraestructura de la estación base seleccionada que se divide en diferentes aspectos que permiten identificar sus características físicas, como altura de la antena, número de antenas, apuntamiento de la antena, tipo de zona para las mediciones y entre otros aspectos que pueden ser considerados como se muestra en la **Figura 1.10**.

1.4.5.1 Altura de la antena

La altura de la antena es un parámetro fundamental al momento de posicionar la estación base dentro de una zona geográfica determinada, este parámetro permite que la cobertura y el alcance de la señal sea mayor conforme la altura aumenta esto permitiendo que pueda existir conexión de enlaces y que la zona de Fresnel sea adecuada para cada uno de ellos.

En este caso la altura de la antena es de aproximadamente 9 metros respecto al suelo, como se observa en la **Figura 1.8**.



Figura 1.8. Medición de altura en la estación base.

1.4.5.2 Número de antenas

El número de antenas hace referencia a la sectorización que se usa para poder radiar la señal hacia un área determinada, de tal manera que se puede aumentar la capacidad de la red móvil y para cubrir de una mejor manera la zona a la cual está destinado, como desventaja se tiene que se aumentan los costos de equipos para su implementación. En este caso el número de antenas que posee la estación base es de 2, como se observa en la **Figura 1.9**.

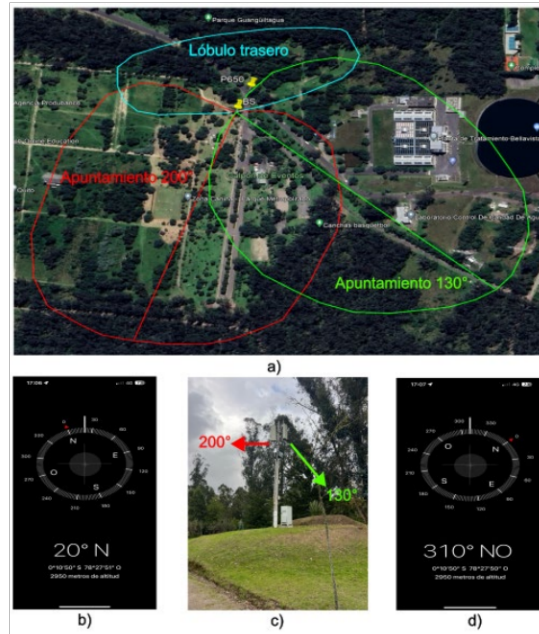


Figura 1.9. a) Zona de cobertura y apuntamiento de las antenas b) Apuntamiento de antena 1 c) Estación base a 0° respecto al norte geográfico d) Apuntamiento de antena 2.

1.4.5.3 Apuntamiento de las antenas (ángulo de elevación)

El ángulo de elevación direcciona la antena de manera vertical respecto al plano horizontal, de tal manera que permite transmitir la señal de red hacia todos los dispositivos móviles permitiendo conexión de enlaces de manera eficiente acorde a la zona geográfica. En este caso el apuntamiento de las antenas es de aproximadamente 40° .

1.4.5.4 Tipo de zona en mediciones

Las mediciones de los parámetros de radiofrecuencia pueden variar conforme al tipo de zona geográfica, entre los factores que pueden generar una variación de las mediciones debido al tipo de zona están el clima, vegetación, relieve geográfico, hidrografía, entre otros. En este caso de estudio se tiene una zona geográfica de tipo montañosa con densidad de vegetación arbórea alta y un clima tropical húmedo.

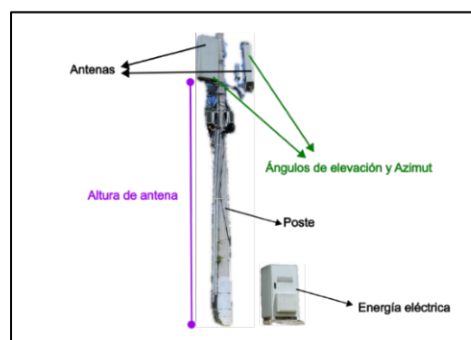


Figura 1.10. Representación básica de infraestructura en una estación base.

1.4.6 RSTUDIO

RStudio es un entorno de desarrollo diseñado para facilitar la programación en el lenguaje R, usado principalmente en el ámbito de estadística y análisis de datos. Esta herramienta “Open Source” utilizada en ciencia de datos, ofrece características robustas que hacen que la exploración, manipulación y modificación de datos, generación de reportes sobre rendimiento y estadística, visualización de información y uso de librerías faciliten su uso y accesibilidad al momento de aplicar técnicas de Machine Learning [22].

1.4.6.1 Capacidades

Entre las capacidades que más resaltan dentro de este software se tiene [22]:

- **Consola interactiva de R:** RStudio ofrece una consola interactiva de interfaz GUI que facilita la ejecución de comandos de R y la visualización inmediata de resultados. Esto permite la exploración de datos y el uso de código de manera didáctica.
- **Editor de código:** Posee un editor de código integrado que proporciona corrección de sintaxis, autocompletado y otras funciones que mejoran la eficiencia de programación, facilitando la elaboración de códigos en el lenguaje de R.
- **Historial y administrador de entornos:** La herramienta mantiene un historial de comandos ejecutados y proporciona administración de entornos que permite el manejo de variables y objetos en GUI de trabajo de RStudio.
- **Generación de informes dinámicos:** RStudio permite la creación de informes de manera dinámica utilizando R Markdown, que permite combinar texto, código y resultados en un solo documento.
- **Visualización interactiva:** Con capacidades integradas para visualización de datos, RStudio permite la creación de gráficos interactivos, facilitando la interpretación de resultados y facilitando el análisis mediante uso de librerías [22].

2 METODOLOGÍA

En esta sección se aborda el procedimiento para la obtención de la información, preprocesamiento de esta, uso de una técnica de Machine Learning, generación de modelo para regresión lineal, comparaciones de gráficos y análisis de estos en función del rendimiento del modelo del DoA para su correspondiente ajuste.

2.1 HERRAMIENTAS DE RECOLECCIÓN

2.1.1 GOOGLE MAPS

La herramienta Google Maps permite realizar el seguimiento de las rutas establecidas para el correspondiente proceso de recolección de datos, su tamaño de instalación es de 307 MB.

Entre los parámetros que permite recolectar se encuentran:

- Seguimiento de rutas mediante posicionamiento GPS.
- Posiciones exactas en formato de latitud y longitud mediante waypoints.

Google Maps permite importar archivos “KMZ”, de tal manera que es compatible con diferentes herramientas que serán necesarias para poder realizar el proceso de mediciones [23].

2.1.2 GOOGLE EARTH PRO

La herramienta Google Earth Pro permite realizar el establecimiento de las rutas mediante la creación de circunferencias, líneas, rutas y entre otras opciones mediante ángulos respecto al norte geográfico, coordenadas geográficas o distancias con el objetivo de cumplir el correspondiente proceso de recolección de datos, su tamaño de instalación es de 188 MB.

Entre los parámetros que permite recolectar se encuentran:

- Seguimiento de rutas mediante posicionamiento GPS.
- Posiciones exactas en formato de latitud y longitud mediante waypoints.
- Inserción de marcadores mediante diferentes formatos de coordenadas.

Google Maps permite importar archivos “KMZ” y posee compatibilidad con diferentes herramientas que serán necesarias para poder realizar el proceso de mediciones [24].

2.1.3 WIKILOC

La herramienta Wikiloc permite realizar el registro de rutas y posiciones mediante marcadores sin la necesidad de tener acceso a internet para monitoreo de posibles zonas que pueden ser medidas durante el proceso de recolección de datos, su tamaño de instalación es de 78 MB.

Entre los parámetros que permite recolectar se encuentran:

- Archivo “KMZ” con la correspondiente ruta y puntos de referencia guardados durante una sesión de uso.
- Posiciones exactas en formato de latitud y longitud mediante waypoints.
- Inserción de marcadores sin necesidad de acceso a internet.

Wikiloc permite exportar archivos “KMZ” y posee compatibilidad con diferentes herramientas de navegación que serán necesarias para poder realizar el proceso de mediciones con el fin de determinar y cubrir rutas dentro de una zona específica [25].

2.1.4 NETMONITOR PRO

La herramienta NetMonitor Pro permite realizar la obtención de las mediciones de los parámetros de radiofrecuencia para el correspondiente proceso de recolección de datos, su tamaño de instalación es de 295 MB.

Entre los parámetros que permite recolectar se encuentran [26]:

- Tipo de tecnología: Permite conocer el tipo de tecnología con la cual opera la celda celular al momento de realizar la recolección de datos.
- RSSI: Indica el nivel de potencia de la señal recibida en un dispositivo, su valor es medido en (dBm).
- RSSI STRONGEST: Indica el nivel de potencia de la señal recibida más fuerte en un dispositivo, su valor es medido en (dBm).
- RSRQ: Indica la calidad de las señales pilotos que son recibidas, su valor es medido en (dB).
- RSSNR: Indica la relación de la potencia señal a ruido respecto a la señal de referencia, su valor es medido en (dB).
- MNC: Indica el código de red móvil de la operadora correspondiente.
- LAC: Indica el código de localización de área con el fin de poder identificar la zona de ubicación.
- RNC: Permite gestionar los recursos de radio de la red celular y sus diferentes conjuntos de nodos acorde a su tipo de tecnología.
- ID: identificador de un dispositivo, servicio o elemento de una red celular.

- CID: Indica el ID de la celda a la que se encuentra conectado el dispositivo celular.
- PCI: Indica el ID físico de la celda a la que se encuentra conectado el dispositivo celular.
- TA: Tiempo de avance en el que un móvil inicia una transmisión mediante un enlace ascendente para compensar el retardo existente debido a la propagación.
- Latitud y longitud: Permite adquirir el punto de localización del dispositivo móvil de manera instantánea [26].

NetMonitor Pro permite obtener mayor número de parámetros dado que es una versión de pago, además se puede descargar archivos en formato “CSV”, “KML”, “JSON” [27].

2.1.5 CELLMAPPER

La herramienta CellMapper permite identificar el handover existente mediante notificaciones de sonido y proporciona visualización de conexión de enlace entre BS y MS mediante un mapa de navegación o notificación permanente en la barra de notificaciones del dispositivo móvil [28].

2.2 MÉTODO DE RECOLECCIÓN DE DATOS

El método propuesto para el proceso de mediciones de los parámetros de radiofrecuencia en la zona de interés (o etapa de recolección de datos) se basa en la toma de datos mediante la división de la zona de cobertura de la estación base en radios y circunferencias como se muestra en la **Figura 2.3**; es decir, se secciona la zona de cobertura seleccionada, de tal manera que los datos que se desean obtener estén conforme a los siguientes lineamientos de medición a seguir:

- Radio de medición cada 5°, se debe tomar en cuenta que se realizan dos mediciones, una desde la BS hasta el punto de ejecución del handover y otra desde el handover hacia la BS.
- Circunferencia de medición cada 50 metros respecto de la estación base.

La elección de la estación base se basa en la accesibilidad existente dentro de la zona escogida para el correspondiente proceso de análisis, de esta manera, se busca obtener la mejor zona con el fin de poder realizar el mayor número de mediciones de forma adecuada. De esta manera se escoge la estación base señalada en la **Figura 2.1** dado que es la zona más adecuada y accesible de manera peatonal.



Figura 2.1. Recorrido de estaciones base con Wikiloc vinculado a Google Earth Pro.

Se realiza una inspección detallada como se visualiza en la **Figura 2.2** para conocer de manera real la zona de cobertura de la estación base escogida, de esta manera, se podrá seccionar la zona descrita en base al método de recolección de datos escogido.



Figura 2.2. Recorrido de cobertura para estación base en Wikiloc vinculado a Google Earth Pro.

Dadas estas condiciones para la elaboración de las mediciones y la inspección realizada a la estación base, se procede a hacer uso del software Google Earth Pro, Google Maps y Wikiloc.

Se debe tomar en cuenta que existe compatibilidad entre las diferentes herramientas descritas, además, se debe pasar la información obtenida mediante Wikiloc hacia Google Earth Pro y de este generar un archivo KMZ para poder usarlo en Google Maps, esto permitirá guiar la elaboración de las correspondientes mediciones de tipo radial y de circunferencias como se muestra en la **Figura 2.3**.

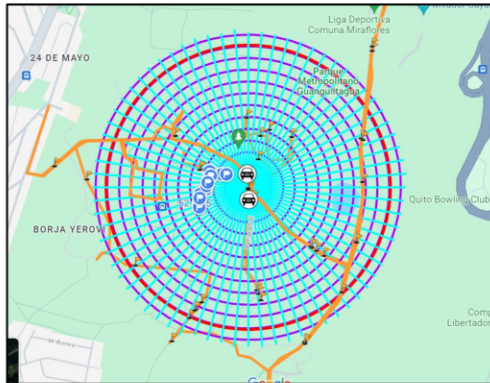


Figura 2.3. Mapa con trazado de radios y circunferencias elaboradas en Google Earth Pro para la correspondiente recolección de datos mediante Google Maps.

Una vez elaborado este trabajo de inspección y de haber identificado zonas adecuadas para las respectivas mediciones, se hace un cálculo de la cantidad de las mediciones totales a obtener al finalizar el proceso de adquisición de datos.

Para el caso de las mediciones radiales su valor es fijo ya que acorde a la inspección elaborada, en su mayoría es posible obtener mediciones a través de los obstáculos, dado que se mide cada 5° y la zona a cubrir es de 360° , entonces se obtienen 72 radios, por lo tanto, se tendrán 144 mediciones radiales, 72 de “ida” y 72 de “vuelta”.

Para el caso de las mediciones en circunferencia (alrededor de la BS) su valor varía dependiendo del alcance de la BS y de la accesibilidad que existe dentro de cada una de las trayectorias en forma de circunferencia observada a través de la inspección elaborada. En este caso, este tipo de mediciones se ven restringidas debido al tipo de zona escogida dado que existen obstáculos como quebradillas, zonas boscosas inaccesibles, densidad de vegetación alta con difícil acceso, trayectos irregulares, zonas restringidas para tratamiento de agua potable, entre otros.

Dadas estas condiciones se establecen 6 circunferencias trazadas en Google Maps como se observa en la **Figura 2.4**, por lo tanto, se debe tener alrededor de 12 mediciones de circunferencia teniendo la posibilidad de dividir dichas mediciones en tramos debido a las restricciones descritas anteriormente, siendo así, 6 en sentido “horario” y 6 en “antihorario”.

Dado estos factores se tendría de manera parcial un total de 156 mediciones por elaborar.

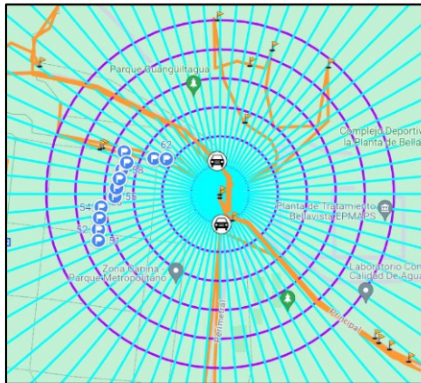


Figura 2.4. Mapa con rutas finales propuestas de radios y circunferencias para las mediciones en Google Maps.

2.3 RECOLECCIÓN DE DATOS

Una vez planteada la forma en la que se elaboran las mediciones de los parámetros de RF, se procede a obtener cada una de ellas haciendo uso de las herramientas Google Maps y Netmonitor Pro, cada uno cumpliendo con su función mediante un dispositivo Android, en este caso se hace uso específicamente del equipo Samsung Galaxy Note 9 Plus para el proceso de medición, cabe recalcar que este dispositivo será de utilidad tanto para la parte de geoposicionamiento como para la toma de muestras en cada uno de los radios y circunferencias propuestas.

Los parámetros obtenidos durante el proceso de recolección de datos son:

Tabla 2.1. Parámetros de Radiofrecuencia obtenidos en aplicación NetMonitor Pro.

Categoría	Parámetros
Información de la red celular	Source.Name, net_op_name, net_op_code, roaming, net_type, tech, mcc, mnc, mnc_master, band, arfcn, bw, bwlist
Estados del servicio de conexión	sim_state, service_state, nr_state, call_state, data_state, data_act
Datos de servicio en red celular	data_rx, data_tx, thp_rx, thp_tx
Señales y Mediciones de Radiofrecuencia	rss_i_strongest, nstrong, rssi, rsrq, rssi_ev, ecio_ev, rssnr, nrssrsrp, nrssrsrq, nrsssinr, nrscsirsrp, nrscsirsrq, nrscsisinr, slev, ta
Celdas vecinas cercanas de diferentes tecnologías celulares	gsm_neighbors, umts_neighbors, lte_neighbors, nr_neighbors, cdma_neighbors
Información de ubicación de la estación móvil	gps, accuracy, lat, long, altitude, speed, bearing
Detalles específicos de la red celular	lac_tac_sid, long_cid, node_id_nid, cid_bid_psc_pci, nrtac, nrnci, nrpci, nrarfcn, sys_time, Report

2.3.1 DESCRIPCIÓN DEL PROCESO Y ETIQUETADO DE MEDICIÓN RADIAL Y DE CIRCUNFERENCIAS

De manera similar se hará uso de un software de apoyo denominado CellMapper, esto permitirá detectar el cambio de handover mediante notificaciones con audio al momento de realizar las mediciones de la manera descrita anteriormente, ida y vuelta, horario y antihorario cada una respectivamente a su forma radial y de circunferencias.

Para las correspondientes mediciones se sigue el siguiente procedimiento descrito a continuación:

- Se debe tener abierto las aplicaciones descritas, Google Maps, NetMonitor Pro y CellMapper.

2.3.1.1 Radiales

- Para el caso de la medición radial, se elabora un recorrido previo para posicionarse dentro de la línea radial trazada en el software Google Maps y buscando la zona de handover existente dentro del radio seleccionado.
- Una vez ubicado el punto de handover, se procede a hacer uso de banderas de referencias, la primera colocada en el punto radial y de handover, otra aproximadamente en el punto intermedio entre la MS-BS y para finalizar si existe línea de vista se hace uso de la propia BS, si no, se coloca la tercera bandera lo más cercana a la BS para generar línea de vista dado que la zona de medición no es regular.
- Para el caso de la medición radial, comienzan de la forma denominada “vuelta”; es decir, la medición se realiza desde la zona de handover hacia la BS, esto haciendo uso del software NetMonitor Pro y presionando la opción grabar.
- Se procede a seguir mediante Google Maps de la manera más exacta posible la línea radial escogida, esto hasta llegar hacia la BS para finalizar la medición “vuelta” y parar mediante el botón stop.
- Se guarda la medición “vuelta” bajo la nomenclatura escogida como rXXvuelta, donde XX indica el número de medición que está relacionado al radio escogido dentro del software Google Maps, esto facilitará la etapa de preprocesamiento de datos.

- De manera similar, se inician las mediciones denominadas “ida” o desde la BS hacia la zona de handover del MS, haciendo uso de NetMonitor Pro y presionando la opción grabar.
- Se procede a seguir mediante Google Maps de la manera más exacta posible la línea radial escogida, esto hasta llegar hacia la bandera de referencia del handover del MS para finalizar la medición “ida” y parar mediante el botón stop.
- Se guarda la medición “ida” bajo la nomenclatura escogida como rXXida, donde XX indica el número de medición que está relacionado al radio escogido dentro del software Google Maps, esto facilitará la etapa de preprocesamiento de datos.

2.3.1.2 Circunferencias

- Para el caso de medición en circunferencias, comienzan de la forma denominada “antihorario”, esto haciendo uso del software NetMonitor Pro y presionando la opción grabar, en esta etapa no se considera el uso de CellMapper.
- Se procede a seguir mediante Google Maps de la manera más exacta posible la circunferencia escogida, esto hasta cubrir los 360° considerando condiciones de existencia de obstáculos para finalizar la medición “antihorario” y parar mediante el botón stop.
- Se guarda la medición “antihorario” bajo la nomenclatura escogida como cXXah, donde XX indica el número de medición que está relacionado a la circunferencia escogida dentro del software Google Maps, esto facilitará la etapa de preprocesamiento de datos.
- De manera similar, se comienza de la forma denominada “horario”, esto haciendo uso del software NetMonitor Pro y presionando la opción grabar, en esta etapa no se considera el uso de CellMapper.
- Se procede a seguir mediante Google Maps de la manera más exacta posible la circunferencia escogida, lo cual se realiza hasta cubrir los 360° considerando condiciones de existencia de obstáculos para finalizar la medición “horario” y parar mediante el botón stop.
- Se guarda la medición “horario” bajo la nomenclatura escogida como cXXh, donde XX indica el número de medición que está relacionado a la circunferencia escogida dentro del software Google Maps, esto facilitará la etapa de preprocesamiento de datos.

2.3.1.3 Cronograma de la etapa de recolección de datos

Una vez ejecutado el proceso de medición se obtienen en cada recorrido radial y circular un total de 160 mediciones finales, esto debido a que se dividieron en tramos las circunferencias por los obstáculos dentro de las trayectorias.

Las rutas en la zona escogida y donde se realizó y culminó la recolección de datos se presentaron en la **Figura 2.4**.

Se elaboró un cronograma para realizar todos los recorridos establecidos y para tener una cantidad adecuada de datos que permitan registrar la variación temporal de los parámetros de radiofrecuencia en la zona de cobertura de la BS. La etapa de recolección de datos se desarrolló de acuerdo con el siguiente cronograma.

Tabla 2.2. Cronograma de etapa de recolección de datos.

Fecha	Tareas específicas
17/10/2023	Se realizaron las mediciones de los radios 1, 2, 3, 4, 5, 6, 7, 8 ida y vuelta
18/10/2023	Se realizaron las mediciones de los radios 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 ida y vuelta
21/10/2023	Se realizaron las mediciones de los radios 26, 27, 28, 29, 30, 31, 32, 33, 34, 35 ida y vuelta
23/10/2023	Se realizaron las mediciones de los radios 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 ida y vuelta
24/10/2023	Se realizaron las mediciones de los radios 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72 ida y vuelta
11/11/2023	Se realizaron las mediciones de circunferencias 1, 2, 3, 4, 5, 6 en los sentidos horario y antihorario

2.4 PREPROCESAMIENTO DE DATOS

Culminado el proceso de medición o adquisición de datos se procede a la etapa de preprocesamiento, en el cual se hará uso del software Excel y NetMonitor Pro para las diferentes acciones que se elaborarán con el fin de obtener una base de datos conjunta que pueda ser usada en el proceso de modelamiento en RStudio.

2.4.1 OBTENCIÓN DE ARCHIVOS “.CSV”

Como primer punto dentro del preprocesamiento de datos se necesita obtener cada una de las mediciones elaboradas previamente; esto quiere decir que, mediante NetMonitor Pro se obtienen los archivos con el formato respectivo compatible con Excel.

NetMonitor Pro nos permite generar tres tipos de archivos diferentes acorde a nuestras mediciones elaboradas, para nuestro caso de interés se hará uso del archivo de extensión “.csv”, el cual puede ser editado mediante el software Excel, cabe recalcar que se genera un archivo por cada sesión guardada dentro de NetMonitor Pro, tanto de extensión “.csv” como “.zip” dentro de la carpeta que almacena las sesiones en el dispositivo celular como se observa en la **Figura 2.5**.

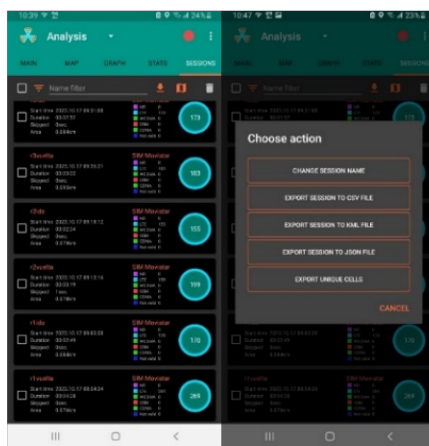


Figura 2.5. NetMonitor Pro con las sesiones obtenidas del proceso de medición y exportación a “.csv”.

Una vez generados todos los archivos de mediciones, se procede a transferir toda la recolección de datos a la computadora designada para realizar el preprocesamiento y estudio respectivo del DoA.

2.4.2 CLASIFICACIÓN Y GENERACIÓN DE ARCHIVOS DE EXCEL DE MEDICIONES OBTENIDAS

En esta etapa se describe el proceso de clasificación y generación de archivos en Excel de las mediciones obtenidas, el detalle de cada procedimiento se puede ser observado dentro del ANEXO I.

2.4.3 ATRIBUTOS ADICIONALES

El cálculo de atributos adicionales nos ayuda en dos temas fundamentales dentro de nuestro estudio: el primero es poder determinar de manera real los valores del DoA, y el

segundo, es poder entrenar al modelo con el objetivo de observar si se logran obtener mejores resultados gracias a la inclusión de parámetros externos a las mediciones obtenidas mediante el software NetMonitor Pro.

2.4.3.1 Distancia entre BS y MS

El atributo de la distancia entre la BS y MS contribuye a conocer el valor exacto y real de la posición de la medición en cada muestra respecto a la estación base, ya que pudieron haber existido desviaciones al realizar las mediciones.

Los datos previstos para el cálculo de la distancia corresponden a las posiciones geográficas obtenidas mediante el software NetMonitor Pro, estos serán usados para conseguir el atributo deseado.

En base a una búsqueda de cómo resolver este problema, el método propuesto para el cálculo de la distancia es la fórmula de Haversine.

La fórmula de Haversine se utiliza para calcular la distancia entre dos puntos en la superficie de una esfera, como puede ser la Tierra. La fórmula se usa en sistemas de geolocalización y cálculos de distancia en navegación mediante valores de latitud y longitud [29], [30].

Su fórmula se puede definir de diferentes maneras debido a la existencia de identidades trigonométricas, en este estudio está definida de la siguiente forma:

$$D = 1000 * R * \arccos (\text{Sen} (Lat BS) * \text{Sen} (Lat MS) + \text{Cos} (Lat BS) * \text{Cos} (Lat MS) * \text{Cos} (Long BS - Long MS)) \quad (2.1)$$

Donde:

- *D*: Distancia calculada en metros.
- *R*: Radio de la tierra en kilómetros.
- *Lat BS*: Valor de la latitud de la estación base.
- *Long BS*: Valor de la longitud de la estación base.
- *Lat MS*: Valor de la latitud de la estación móvil.
- *Long MS*: Valor de la longitud de la estación móvil.

Se debe tener en cuenta que, para el cálculo de la distancia, los valores de latitud y longitud deben estar en radianes, de otra manera, el cálculo será erróneo.

La expresión para el uso como fórmula en Excel se la proporciona de la siguiente manera:

$$= 1000 * 6371 * ACOS((SENO(RADIANES(BJ2)) * SEÑO(RADIANES(BL2))) + (COS(RADIANES(BJ2)) * COS(RADIANES(BL2)) * COS(RADIANES(BK2 - BM2)))) \quad (2.2)$$

Además, se puede observar el significado de cada celda asignada en la **Figura 2.6** mediante las etiquetas de cada atributo obtenido durante el preprocesamiento.

=1000*6371*ACOS((SENO(RADIANES(BJ2))*SEÑO(RADIANES(BL2)))+(COS(RADIANES(BJ2))*COS(RADIANES(BL2))*COS(RADIANES(BK2-BM2))))									
	BJ	BK	BL	BM	BN	BO	BP	BQ	BR
	latitud BS	longitud BS	latitud MS	longitud MS	Distancia_m	DoA	DoAR	d_Distancia_m	d_lte_neighbors
	-0,180445	-78,464067	-0,18043952	-78,46407401	BM2))))	308,0161755	308	14,85166757	0
	-0,180445	-78,464067	-0,18030538	-78,46409539	15,84282118	348,506309	348,5	0,986184259	0

Figura 2.6. Implementación de la fórmula de Haversine para el cálculo de la distancia en Excel.

Como ejemplo se ha escogido la fila 558 de la hoja "MEDICIONES IDA FINAL", en el cual se realiza el cálculo correspondiente de la distancia y se obtiene un resultado de 30.39 m de manera real.

	thp_tx	latitud BS	longitud BS	latitud MS	ongitud M	distancia k	Dlat	Dlong	DoAP	DoA	DoAR
556	0	-0,180445	-78,464067	-0,18018235	-78,463994	0,03031173	4,58413E-06	1,2736E-06	15,5264005	74,4735995	344,473599
557	0	-0,180445	-78,464067	-0,18018252	-78,463993	0,03032347	4,58116E-06	1,2912E-06	15,7404187	74,2595813	344,259581
558	0	-0,180445	-78,464067	-0,18018161	-78,4639938	0,03039709	4,59705E-06	1,2772E-06	15,5272914	74,4727086	344,472709
559	0	-0,180445	-78,464067	-0,18018186	-78,4639924	0,03041443	4,59268E-06	1,3029E-06	15,837997	74,162003	344,162003
560	9	-0,180445	-78,464067	-0,18018026	-78,463992	0,0305951	4,62061E-06	1,3083E-06	15,8091689	74,1908311	344,190831
561	0	-0,180445	-78,464067	-0,18017894	-78,4639928	0,03071476	4,64365E-06	1,2957E-06	15,590879	74,409121	344,409121
562	0	-0,180445	-78,464067	-0,18017489	-78,4639918	0,03117653	4,71433E-06	1,3121E-06	15,5534629	74,4465371	344,446537

Figura 2.7. Cálculo de la distancia real para una muestra de la fila 558.

Ahora se procede a verificar si es correcto de manera empírica haciendo uso del software Google Earth Pro, este permite hacer uso de herramientas tanto para medición de distancias entre punto como para determinar el ángulo de barrido respecto al norte geográfico.

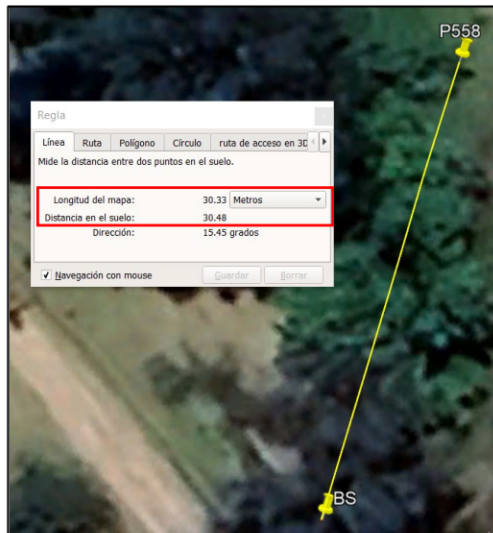


Figura 2.8. Cálculo de la distancia empírica en Google Earth Pro para una muestra de la fila 558.

Como se puede observar en la **Figura 2.8**, la distancia obtenida mediante la colocación de los respectivos puntos geográficos en el mapa posee un valor de 30.33 m, esto varía en un factor pequeño, debido a que al momento de medir no se pueden seleccionar los puntos de manera exacta, dado que el software no posee esa opción, pero se comprueba que la fórmula de Haversine es correcta y permite obtener la distancia entre dos puntos geográficos cualquiera.

Como observación para el atributo de distancia, de manera empírica y bajo el recorrido realizado en las mediciones, la distancia máxima de la estación base no sobrepasa los 1000 metros de cobertura; por ende, es un factor para tener en cuenta para posibles análisis e inconvenientes que pueden presentarse para la obtención del modelo de estimación del DoA.

2.4.3.2 Ángulo de direccionamiento (DoA)

En este estudio el atributo adicional calculado en el dataset del ángulo de direccionamiento o denominado DoA contribuye en conocer el valor exacto y real de la dirección desde la cual una onda o señal llega a un sensor o conjunto de sensores. Este concepto es fundamental en el procesamiento de señales para determinar la ubicación de la fuente de señal que radia una onda [1], [2], [3], [4].

Los datos previstos para el cálculo del DoA corresponden a las posiciones geográficas obtenidas mediante el software NetMonitor Pro, estos serán usados para conseguir el atributo deseado.

En base a una búsqueda de cómo resolver este problema, el método propuesto para el cálculo del atributo DoA se basa en el uso de coordenadas geográficas aplicando una analogía respecto a las coordenadas cartesianas, esto debido a que la estación base (BS) posee coordenadas estáticas, mientras que las coordenadas obtenidas mediante NetMonitor Pro, y que corresponden a la MS, varían, de esta manera los pares ordenados geográficos se podrán trabajar como si fueran pares ordenados cartesianos [9], [10], [29].

Un punto clave para calcular el atributo es que, dado que las mediciones se elaboran dentro del Ecuador y el país se encuentra en el cuadrante 3 correspondiente al plano sur-oste, donde sus latitudes y longitudes son escalares negativos, para simplificar los cálculos se asume que los datos recolectados están dentro del cuadrante mencionado, ya que no existen distancias lejanas cerca de este cuadrante.

Posteriormente a realizar la analogía se encuentran las expresiones generales de las variaciones de latitudes y longitudes en cada uno de los cuadrantes auxiliares respecto a un eje auxiliar colocado en la estación base, esto permite encontrar de manera sencilla los ángulos auxiliares para obtener la expresión general del DoA en cada uno de los cuadrantes del eje auxiliar.

Bajo todas las condiciones mencionadas, los parámetros descritos se pueden observar en la **Figura 2.9**.

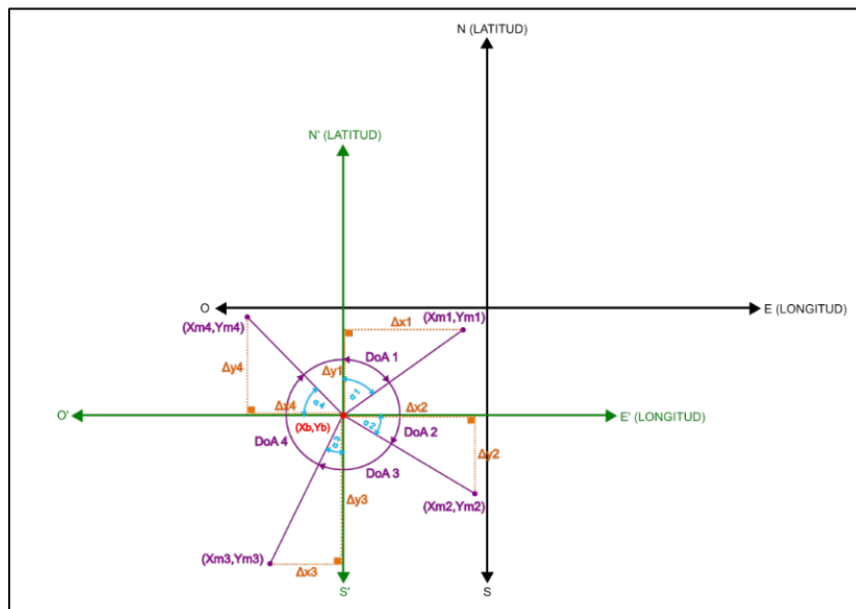


Figura 2.9. Parámetros y esquema para la determinación del DoA.

De esta manera se obtienen las siguientes expresiones para cada uno de los deltas en el cuadrante auxiliar respectivo:

1. Cuadrante auxiliar 1

- $\Delta_{X1} = -X_b + X_{m1}, X_{m1} \geq X_b$ (2.3)

- $\Delta_{Y1} = -Y_b + Y_{m1}, Y_{m1} \geq Y_b$ (2.4)

2. Cuadrante auxiliar 2

- $\Delta_{X2} = -X_b + X_{m2}, X_{m2} \geq X_b$ (2.5)

- $\Delta_{Y2} = -Y_{m2} + Y_b, Y_{m2} \leq Y_b$ (2.6)

3. Cuadrante auxiliar 3

- $\Delta_{X3} = -X_{m3} + X_b, X_{m3} \leq X_b$ (2.7)

- $\Delta_{Y3} = -Y_{m3} + Y_b, Y_{m3} \leq Y_b$ (2.8)

4. Cuadrante auxiliar 4

- $\Delta_{X4} = -X_{m4} + X_b, X_{m4} \leq X_b$ (2.9)

- $\Delta_{Y4} = -Y_b + Y_{m4}, Y_{m4} \geq Y_b$ (2.10)

Ahora para el caso de los ángulos auxiliares se obtienen las siguientes expresiones de cada cuadrante auxiliar usando razones trigonométricas:

1. Cuadrante auxiliar 1

- $\alpha_1 = \tan^{-1} \left(\frac{\Delta_{X1}}{\Delta_{Y1}} \right)$ (2.11)

2. Cuadrante auxiliar 2

- $\alpha_2 = \tan^{-1} \left(\frac{\Delta_{Y2}}{\Delta_{X2}} \right)$ (2.12)

3. Cuadrante auxiliar 3

- $\alpha_3 = \tan^{-1} \left(\frac{\Delta_{X3}}{\Delta_{Y3}} \right)$ (2.13)

4. Cuadrante auxiliar 4

- $\alpha_4 = \tan^{-1} \left(\frac{\Delta_{Y4}}{\Delta_{X4}} \right)$ (2.14)

Una vez obtenidos todos los ángulos auxiliares necesarios para el cálculo del DoA en cada cuadrante, se obtienen las fórmulas finales sumando ángulos implícitos que se pueden

observar en la **Figura 2.9** para cada cuadrante auxiliar que cubre al cuadrante geográfico 3.

De esta manera se obtienen las fórmulas que permiten calcular el DoA, las cuales son:

$$\bullet \quad DoA_1 = \alpha_1 \quad (2.15)$$

$$\bullet \quad DoA_2 = \alpha_2 + 90^\circ \quad (2.16)$$

$$\bullet \quad DoA_3 = \alpha_3 + 180^\circ \quad (2.17)$$

$$\bullet \quad DoA_4 = \alpha_4 + 270^\circ \quad (2.18)$$

Donde:

- $DoA_{1,2,3,4}$: Ángulo de direccionamiento DoA en cada cuadrante auxiliar dentro del cuadrante 3.
- X_b : Valor de la longitud de la estación base.
- Y_b : Valor de la latitud de la estación base.
- $X_{m1,2,3,\dots,n}$: valor de la longitud de cualquier zona dentro del cuadrante 3.
- $Y_{m1,2,3,\dots,n}$: valor de la latitud de cualquier zona dentro del cuadrante 3.
- $\Delta_{X1,2,3,\dots,n}$: valor del delta de longitud de cualquier zona dentro del cuadrante 3.
- $\Delta_{Y1,2,3,\dots,n}$: valor del delta de latitud de cualquier zona dentro del cuadrante 3.

Se debe tener en cuenta que para el cálculo del ángulo de direccionamiento, los valores de latitud y longitud deben estar en radianes, de otra manera, el cálculo será erróneo, además, estas fórmulas se obtuvieron de manera propia con el fin de simplificar cálculos.

La expresión usada en Excel se la presenta a continuación:

$$= SI(Y(BM2 >= BK2; BL2 >= BJ2); GRADOS(ATAN((-RADIANES(BK2) + RADIANES(BM2))/(-RADIANES(BJ2) + RADIANES(BL2)))); SI(Y(BM2 >= BK2; BL2 <= BJ2); 90 + GRADOS(ATAN((-RADIANES(BL2) + RADIANES(BJ2))/(-RADIANES(BK2) + RADIANES(BM2)))); SI(Y(BM2 <= BK2; BL2 <= BJ2); 180 + GRADOS(ATAN((-RADIANES(BM2) + RADIANES(BK2))/(-RADIANES(BL2) + RADIANES(BJ2)))); SI(Y(BM2 <= BK2; BL2 >= BJ2); 270 + GRADOS(ATAN((-RADIANES(BJ2) + RADIANES(BL2))/(-RADIANES(BM2) + RADIANES(BK2)))); 0))) \quad (2.19)$$

Además, se puede observar el significado de cada celda mediante las etiquetas de cada atributo obtenido durante el preprocesamiento.

RADIANES(BL2))/(-RADIANES(BM2)+RADIANES(BK2)))));0))))					
BJ	BK	BL	BM	BN	BO
latitud BS	longitud BS	latitud MS	longitud MS	Distancia_m	DoA
-0,180445	-78,464067	-0,18043952	-78,46407401	0,991153608	BK2)))));0))))
-0,180445	-78,464067	-0,18030538	-78,46409539	15,84282118	348,506309

Figura 2.10. Implementación de la fórmula para el cálculo del DoA en Excel.

Como ejemplo se ha escogido la fila 650 de la hoja “MEDICIONES IDA FINAL”, en la cual se calcula el DoA y se obtiene un resultado de 26.13° de manera real.

	latitud BS	longitud BS	latitud MS	longitud MS	Distancia_m	DoA
649	-0,180445	-78,464067	-0,18009421	-78,4638977	43,31120597	25,76311527
650	-0,180445	-78,464067	-0,18008381	-78,46388985	44,73296438	26,12619955
651	-0,180445	-78,464067	-0,18007616	-78,46388529	45,72009436	26,22728426
652	-0,180445	-78,464067	-0,18007001	-78,46388283	46,45451	26,15718253

Figura 2.11. Cálculo del DoA real para la muestra de la fila 650.

Ahora se procede a verificar si el cálculo del DoA real es correcto, de manera empírica, haciendo uso del software Google Earth Pro, este nos permite hacer uso de herramientas tanto para medición de distancias entre el punto como para determinar el ángulo de barrido respecto al norte geográfico.

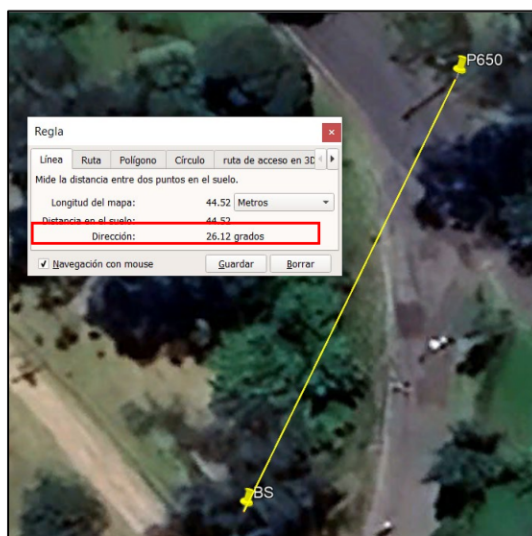


Figura 2.12. Cálculo del DoA empírico o estimado para una muestra de la fila 650.

Como se puede observar en la **Figura 2.12**, el DoA obtenido mediante la colocación de los respectivos puntos geográficos en el mapa y midiendo el ángulo respecto al norte geográfico, resultando en un valor de 26.12° , esto varía en un factor pequeño debido a que al momento de medir no se pueden seleccionar los puntos de manera exacta, dado que el software no posee esa acción, pero se comprueba que la fórmula propuesta es correcta y permite obtener el DoA entre dos puntos geográficos cualquiera.

2.4.4 FILTRADO DE ATRIBUTOS ERRÓNEOS EN MEDICIÓN

Como última parte dentro del preprocesamiento se tiene la etapa de filtrado de valores extraños, erróneos o innecesarios de los atributos en las mediciones. Este proceso se debe efectuar, como por ejemplo, debido a que nuestro estudio está dentro de la tecnología 4G, se debe mencionar que NetMonitor Pro al momento de estar tomando muestras registra diferentes tipos de tecnologías y al momento que no existe ninguna conexión con ninguna estación base, entonces el software arroja valores repetitivos y sin coherencia de uso.

Debido a esto, se deben filtrar los atributos que contienen valores que indican una desconexión o pérdida de la señal, generalmente el valor de este error está dado por el número 2147483647, esto permitirá eliminar variaciones en los atributos que podrían alterar desfavorablemente los análisis.

Después de esto, se sigue con el proceso de importación de archivos clasificados a Excel descrito anteriormente, filtrando cada uno de los atributos que son de mayor interés y eliminando valores que presentan algún error.

El orden de los parámetros que son filtrados para obtener los datos preprocesados de manera final es:

- Se filtra primero rssi, rsrq y rssnr de manera simultánea.
- Se filtra rssi_strongest como último atributo dado que es la mejor configuración para eliminar la menor cantidad de datos en nuestro archivo.

Una vez culminados los procesos descritos anteriormente se obtienen los archivos preprocesados y listos para ser usados en una técnica de Machine Learning.

2.5 IMPLEMENTACIÓN DE LA TÉCNICA DE MACHINE LEARNING EN RSTUDIO

En esta sección se indicará el proceso de obtención de los modelos para la predicción o estimación del DoA mediante el uso de archivos preprocesados, se debe tener en cuenta

que se mostrará el proceso aplicado para uno de los cuatro archivos obtenidos, con el fin de entender la forma en que se aplicaron cada una de las etapas.

2.5.1 INGRESO Y CONFIGURACIÓN DE ATRIBUTOS EN ARCHIVO (DATASET)

Una vez determinado el archivo a usar dentro de RStudio, en este caso “MEDICIONES UNIDO”, se procede a cargarlo haciendo uso de programación en R, donde se ingresa como parámetro, la dirección del archivo en la computadora. El comando utilizado para cargar y visualizar el archivo “.xlsx” es el siguiente:

Código 1. Función para lectura de dataset.

```
#importacion de dataset
data1<-read_xlsx("C:\\Users\\Lenin\\Desktop\\9NO
SEMESTRE\\TIC\\MEDICIONES\\MEDICIONES ARCHIVOS M2
MODIFICADO3D\\MEDICIONES_UNIDO_FINAL3D.xlsx")
data1
```

- La función “read_xlsx()” permite cargar diferentes tipos de archivos, pueden ser xlsx, csv u otros tipos de formatos que sean compatibles con RStudio.

2.5.1.1 Eliminación de atributos y cambio de tipo de atributo

Realizada una inspección de los atributos más útiles dentro de nuestro estudio, se procede a eliminar atributos que no poseen información relevante dentro del dataset, el comando implementado es el siguiente:

Código 2. Función para eliminación de atributos y visualización de dataset.

```
#eliminacion de atributos en dataset, sirve para todos los archivos
dataf<-select(data1,-Source.Name,-report,-sys_time,-sim_state,-service_state,-
nr_state,-net_op_name,-net_op_code
, -roaming,-net_type,-call_state,-data_state,-data_act,-data_rx,-data_tx,-
gsm_neighbors
, -umts_neighbors,-nr_neighbors,-cdma_neighbors,-nstrong,-tech,-mcc,-mnc,-
mnc_master,-lac_tac_sid
, -long_cid,-node_id_nid,-cid_bid,-psc_pci,-nrtac,-nrnci,-nrpci,-nrarfcn,-rssi_ev,-
ecio_ev
```

```

, -nrssrsrp, -nrssrsrq, -nrsssinr, -nrscsirsrp, -nrscsirsrq, -nrscsisinr, -slev, -ta, -gps, -
accuracy
, -bearing, -band, -arfcn, -bw, -bwlist, -thp_rx, -thp_tx)
dataf
#creacion de otro dataset para poder modificar y probar gráficas y modelos
data_doa <- dataf
data_doa

```

- La función “select()” permite escoger solo las variables del dataset que se desean conservar o eliminar, además de primera instancia se le define el dataset a ser modificado.

Una vez realizada la eliminación, se procede a cambiar el tipo de atributo, es decir, se emplea el siguiente comando para cambiar de tipo char a numérico:

Código 3. Función para cambio de tipo de atributos y visualización de dataset.

```

#cambio de atributo caracter a numerico
data_doa$rssi <- as.numeric(data_doa$rssi)
data_doa$rsrq <- as.numeric(data_doa$rsrq)
data_doa$rssnr <- as.numeric(data_doa$rssnr)
data_doa$rssi_strongest <- as.numeric(data_doa$rssi_strongest)
head(data_doa)

```

- La función “as.numeric()” permite realizar el cambio de una variable de tipo char a numérico y se la vuelve a almacenar dentro del mismo dataset.

Estos cambios permitirán trabajar posteriormente con la técnica de Machine Learning sin que existan conflictos de lectura de datos.

2.5.2 ANÁLISIS DE TENDENCIA ENTRE PARÁMETROS DE RADIOFRECUENCIA Y EL DOA

En esta sección se obtendrán ideas del comportamiento de los parámetros o atributos en función del DoA.

2.5.2.1 Análisis sin rangos del DoA

Para poder plantear ideas del modelo de la técnica de Machine Learning que se desea implementar, se debe hacer uso de gráficas para buscar patrones que puedan orientar el uso de las variables dentro del modelo.

Como observación se tiene que el radio del lóbulo de radiación de la estación base no supera los 600 metros, esto permite hacer un filtrado de datos en el dataset, de tal manera que permite analizar las variables en función del DoA, este proceso está dado por:

Código 4. Análisis de atributos en función del DoA.

```
#filtrado de delta distancia y distancia conocido que distancia maxima es aprox 600 m
data_d<-filter(data_doa,d_Distancia_m>=-1000 & d_Distancia_m<=1000)
data_d1<-filter(data_d,Distancia_m>=-1000 & Distancia_m<=1000)
data_d2<-data_d1
#data_d2 es una copia para las graficas y modelos

#graficas
#grafica doa vs parametros RF
#doa vs rssi
describe(data_d2$DoA)
plot(data_d2$DoA,data_d2$rssi,main="Gráfico de
dispersión",xlab="DoA",ylab="rssi",pch=19)

#grafica doa vs deltas parametros RF
#doa vs delta rssi
plot(data_d2$DoA,data_d2$d_rssi,main="Gráfico de
dispersión",xlab="DoA",ylab="Delta rssi",pch=19)
```

- La función “filter(data_doa,d_Distancia_m>=-1000 & d_Distancia_m<=1000)” permite realizar el filtrado de la distancia y el delta de distancia con el fin de poder observar las gráficas de manera detallada y sin ningún error de visualización.
- La función “plot(data_d2\$DoA,data_d2\$rssi,main="Gráfico de dispersión",xlab="DoA",ylab="rssi",pch=19)” permite generar gráficos de los atributos de entrada en función del DoA, de manera análoga se emplea este proceso para los atributos restantes.

Se debe considerar que el código completo de esta sección se encuentra en el ANEXO II.

El uso del código del ANEXO II proporciona las siguientes gráficas.

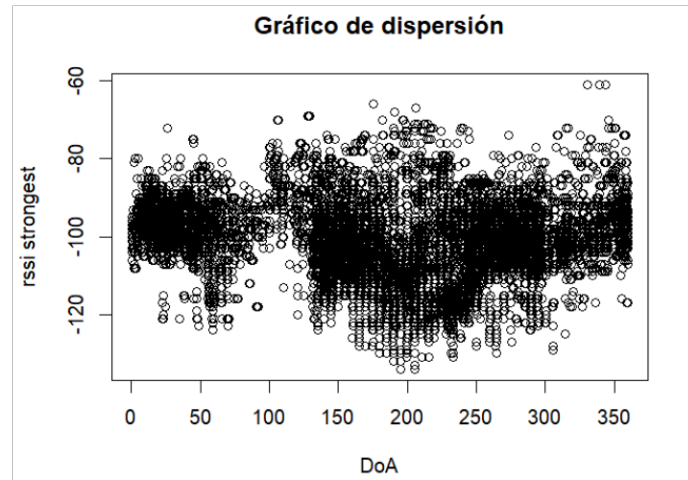


Figura 2.13. RSSI STRONGEST en función del DoA.

Como se observa en la **Figura 2.13**, los datos proporcionan información dado que alrededor de 220° los valores del “RSSI STRONGEST” disminuyen considerablemente debido a que en esa región del DoA apuntan las antenas instaladas en la BS permitiendo tener una mayor distancia de enlace, dado que estos atributos son inversamente proporcionales, además brinda una visualización de las mediciones del “RSSI STRONGEST” y su variación en cada valor del DoA.

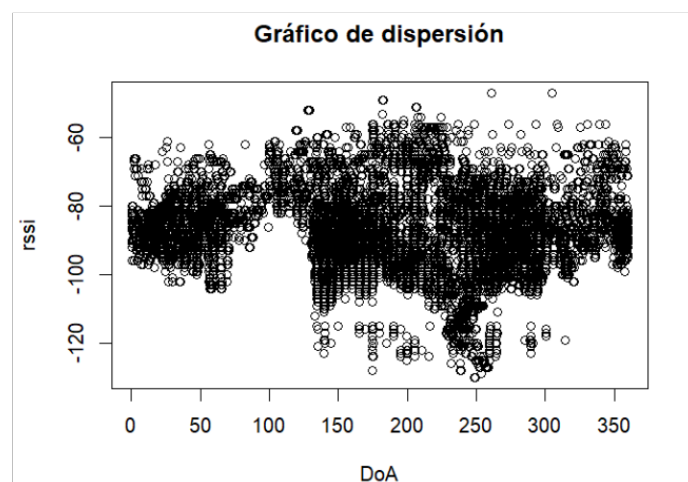


Figura 2.14. RSSI en función del DoA.

Como se observa en la **Figura 2.14**, los datos se separan unos de otros por grupos en donde se tiene que se concentran valores aproximadamente entre 0° a 50°, de 125° a 220°, de 240° a 300° y de 300° a 360°, esto permite obtener una idea de que las mediciones se encuentran agrupadas por valores cercanos del DoA, además en el rango de 150° a 220°

los valores del “RSSI” disminuyen considerablemente debido a que en esa región del DoA apuntan las antenas instaladas en la BS permitiendo una mayor distancia de enlace con la MS.

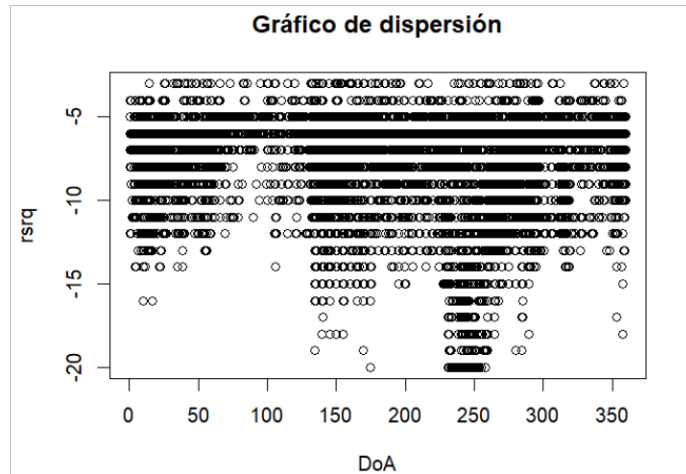


Figura 2.15. RSRQ en función del DoA.

Como se observa en la **Figura 2.15**, los datos de manera homóloga se separan unos de otros por grupos, esto permite obtener una idea de que las mediciones se encuentran agrupadas por valores cercanos del DoA, por otro lado se observa que los valores del “RSRQ” disminuyen dentro de un rango de 150° a 250°, indicando una mayor pérdida del enlace, medido a una distancia mayor respecto de las demás zonas debido al apuntamiento directo de las antenas dentro del rango indicado.

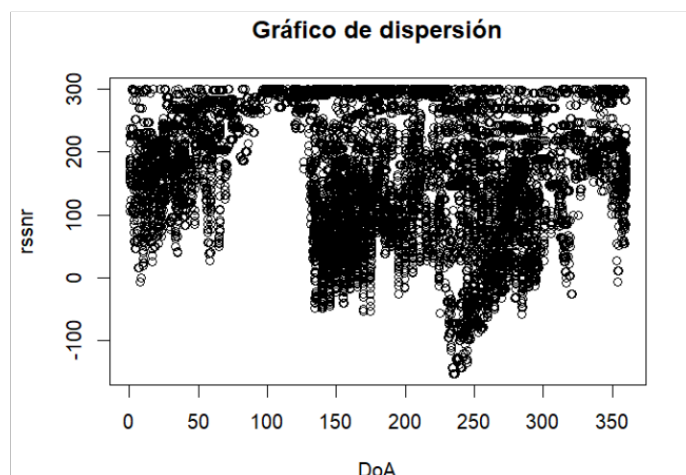


Figura 2.16. RSSNR en función del DoA.

Como se observa en la **Figura 2.16**, los datos de manera homóloga se separan unos de otros por grupo mediante el “RSSNR”, similar a la distribución del atributo “RSSI” y “RSRQ”, por tanto, se encuentran agrupadas en valores cercanos del DoA, además permite ver las

variaciones de la relación señal a ruido detectado por la estación móvil indicando que alrededor de los 250° disminuyen debido a que en esa región del DoA apuntan las antenas instaladas en la BS permitiendo una mayor distancia de enlace con la MS.

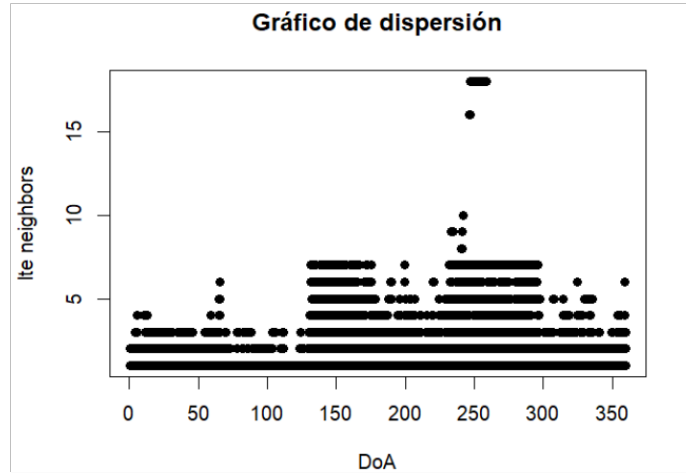


Figura 2.17. LTE NEIGHBORS en función del DoA.

Como se observa, en la **Figura 2.17**, la concentración de estaciones bases cercanas conforme varía el valor del DoA son entre un rango de 5, 7 y hasta 20, la zona con mayor cantidad de celdas vecinas se encuentra dentro de una región del DoA entre 240° a 250°, esto es debido a que en la región indicada no existe apuntamiento de las antenas de la estación base en estudio, por ende existe un proceso de handover entre estaciones vecinas cercanas.

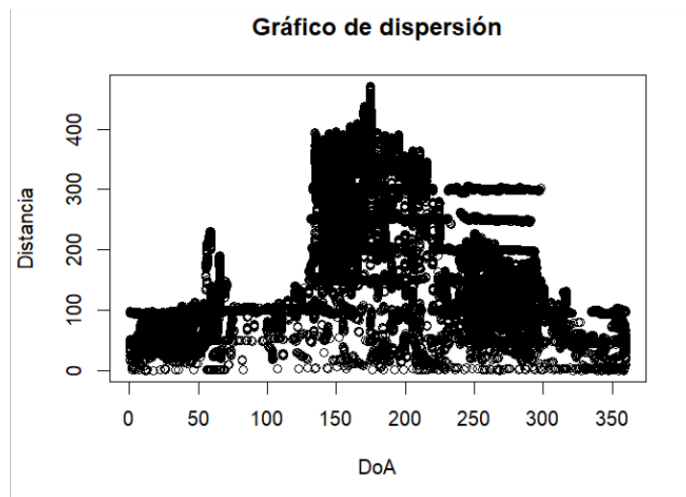


Figura 2.18. Distancia en función del DoA.

Como se observa en la **Figura 2.18**, los datos se separan unos de otros por grupos mediante la distancia, en donde se observa que se concentran valores aproximadamente entre 0° a 70° con 250 metros, de 100° a 220° con 500 metros, de 240° a 320° con 300

metros y de 320° a 360° con 100 metros, esto permite obtener una idea de que las mediciones se encuentran agrupadas por valores cercanos del DoA, por otro lado se visualiza que alrededor de los 180° en el DoA, las distancias aumentan considerablemente debido a que en esa dirección se encuentra la mayor concentración del lóbulo de radiación de la estación base escogida.

2.5.2.2 Análisis con rangos del DoA

Se debe considerar que el radio del lóbulo de radiación de la estación base no supera los 600 metros, además en este caso, se filtra al DoA mediante rangos, de 0° a 360° en pasos de 10° con una tolerancia de 20° respecto a su valor de DoA asignado, este proceso está dado por:

Código 5. Análisis de atributos en función de la distancia mediante rangos del DoA.

```
#creacion de grafico distancia vs rssi para diferentes DoAs
#encuentra el numero maximo de la distancia para poder graficar de manera correcta
m_dist<-max(data_d2$Distancia_m)
max_rssi<-max(data_d2$rssi)
min_rssi<-min(data_d2$rssi)
#0°=360°
data_doa10<-filter(data_d2,(DoA>=350 & DoAR<=360) | (DoA>=0 & DoAR<=10))
describe(data_doa10$Distancia_m)
plot(data_doa10$Distancia_m,data_doa10$rssi,xlim=c(0,m_dist),ylim=c(min_rssi,max_rssi),main="Gráfico de dispersión",xlab="Distancia",ylab="rssi",pch=19)
#axis(1, at = 0:m_dist)
#10°
par(new=TRUE)
data_doa11<-filter(data_d2,DoA>=0 & DoA<=20)
describe(data_doa11$Distancia_m)
plot(data_doa11$Distancia_m,data_doa11$rssi,xlim=c(0,m_dist),ylim=c(min_rssi,max_rssi),main="Gráfico de dispersión",col="blue",xlab="Distancia",ylab="rssi",pch=19,axes=FALSE)
```

- La función “m_dist<-max(data_d2\$Distancia_m)” permite obtener el máximo valor de la distancia para el uso de ejes en el gráfico.
- La función “max_rssi<-max(data_d2\$rssi)” permite obtener el máximo valor del atributo “RSSI” para el uso de ejes en el gráfico.

- La función “`min_rssi<-min(data_d2$rssi)`” permite obtener el mínimo valor del atributo “RSSI” para el uso de ejes en el gráfico.
- La función “`data_doa10<-filter(data_d2,(DoA>=350 & DoAR<=360) | (DoA>=0 & DoAR<=10))`” permite filtrar el DoA al valor de 0° en un rango de 350° a 10°.
- La función “`plot(data_doa10$Distancia_m,data_doa10$rssi, ...)`” permite generar gráficos de los atributos de entrada en función de la distancia mediante rangos del DoA, de manera análoga se emplea este proceso para los atributos restantes.

Se debe considerar que el código completo de esta sección se encuentra en el ANEXO III.

El uso del código del ANEXO III proporciona la siguiente gráfica, el resto de estas serán proporcionadas dentro de la sección mencionada.

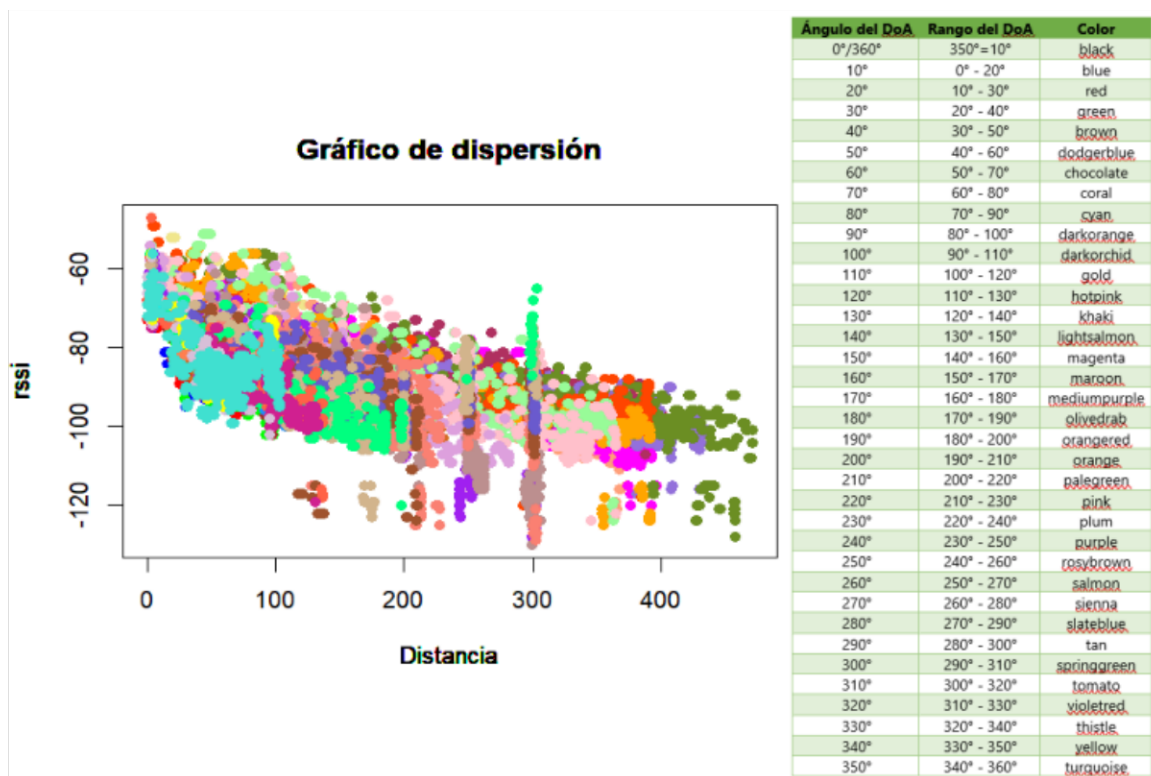


Figura 2.19. RSSI en función de la distancia para diferentes rangos del DoA.

La **Figura 2.19** esta descrita mediante la **Tabla 5.1**, descrita en el ANEXO III, que permite identificar mediante colores cada uno de sus rangos del DoA.

Como se observa en la **Figura 2.19**, los datos mediante esta tendencia de comportamiento proporcionan información interesante, se puede evidenciar que dependiendo del rango del DoA, este adquiere una función que la representa de diferente manera mediante la

“Distancia” y el “RSSI”, esto permite deducir que pueden plantearse modelos por rangos para la estimación del DoA.

Además, dentro de la **Figura 2.19** se puede observar que para una misma distancia pueden existir diferentes valores del DoA (lo cual es obvio), pero se debe mencionar que esto se ve también diferenciado por las características del lóbulo de radiación de la estación base en la que se obtuvieron las mediciones.

2.5.3 IMPLEMENTACIÓN DE UNA TÉCNICA DE MACHINE LEARNING

Dentro de esta sección se mostrará la forma de implementación de la técnica de Machine Learning y la generación de modelos que permitan considerar los parámetros de radiofrecuencia para la predicción o estimación del DoA.

Nota: Lo relacionado a este aspecto se encuentra anexado debido a la extensión del documento.

2.5.3.1 Creación de dataset y observación de histogramas en atributos

Para proceder con la implementación de la técnica, se requiere hacer uso de un dataset que pueda ser modificado en el proceso de la creación de modelos, se escogen atributos a ser insertados dentro de la técnica de Machine Learning, en este caso se definen como rssi_strongest, rssi, rsrq, rssnr, Distancia_m y DoA, además, se requiere una observación previa de la distribución de los atributos mediante el uso de histogramas.

El proceso descrito previamente que contiene el código y las gráficas resultantes puede ser observado en el ANEXO IV del **Código 3**.

2.5.3.2 Matriz de correlación

La matriz de correlación permite obtener los coeficientes de correlación, los cuales indican si puede existir una correlación lineal entre atributos, el coeficiente de relación más usado es el denominado “Rangos de Spearman” [31], el cual tiene tres valores fundamentales, “1” muestra una correlación positiva perfecta, “0” muestra que no existe correlación entre atributos y “-1” que muestra una correlación negativa perfecta [31], [32], como se muestra en la **Figura 2.20**.

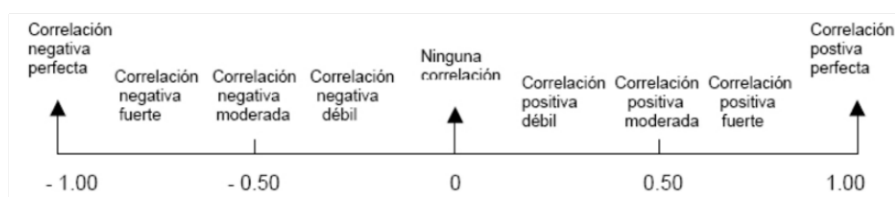


Figura 2.20. Rangos de correlación de Spearman [31].

El detalle de obtención de la matriz de correlación y de cada parámetro que contiene puede ser observado en el ANEXO IV del **Código 4**.

2.5.3.3 Implementación de la receta

La implementación de la receta permite especificar como preprocesar los datos, incluyendo la eliminación de atributos con varianza cero, mutación (creación) de nuevas variables y selección de variables específicas, de esta manera prepara la receta y extrae los datos para el correspondiente entrenamiento, culminando con la división del dataset en subconjuntos de entrenamiento y prueba.

El detalle de cada parámetro que contiene la implementación de la receta puede ser observado en el ANEXO IV del **Código 5**.

2.5.3.4 Generación de modelo

La generación del modelo permite la predicción del DoA basado en el conjunto de datos para el entrenamiento, muestra los coeficientes del modelo, que representan la intercepción y las pendientes que son estimadas para cada uno de los atributos de entrada, además de devolver un dataset con los valores del DoA predicho o estimado con el fin de poder realizar un análisis posterior de los resultados.

El detalle de cada parámetro que contiene la generación de la receta puede ser observado en el ANEXO IV del **Código 6**.

2.5.3.5 Ajuste del modelo

Dentro del ajuste del modelo se realiza un proceso en el cual se busca optimizar o ajustar de tal manera que dicho modelo posea una mayor correlación entre los atributos de entrada con el atributo de salida o DoA, de esta manera permitiendo que el modelo mejore su precisión en la predicción del DoA.

Estos ajustes se realizan en base al uso de la matriz de correlación que es obtenida, esta matriz permite observar la correlación de cada uno de los atributos entre sí, para el ajuste se inicia con el coeficiente de correlación más alto del atributo que se relacione al DoA, una vez obtenido esto, se empiezan a realizar operaciones entre atributos de entrada guiándose en los coeficientes de correlación y observando cual modelo permite aumentar el la precisión del modelo.

Se debe tomar en cuenta que para la reducción de código se generan varios modelos y se corren cada uno mediante las mismas líneas de código de predicción que arroja resultados

diferentes dependiendo del modelo, el detalle de cada parámetro que contiene la generación de la receta puede ser observado en el ANEXO IV del **Código 7**.

2.5.3.6 Evaluación del modelo

En la evaluación del modelo se hace uso de mediciones estadísticas que permiten determinar la precisión de cada uno de los modelos, su uso es frecuente para conocer la exactitud de los modelos obtenidos en base a los valores de las predicciones y los valores reales.

Dentro de este estudio se usan dos indicadores que permiten conocer la precisión de cada modelo, estos son [33]:

- RMSE: La raíz del error medio cuadrático permite medir la magnitud de los errores entre la predicción de valores de un atributo y los valores reales del atributo, indicando la precisión de un modelo.
- R2: Permite calcular el coeficiente de determinación que muestra la proporción de variabilidad de los datos en base al modelo propuesto, indica la relación entre atributos de entrada con el atributo de salida u objetivo a través del modelo.

Se hace uso de estos parámetros de evaluación estadística con el fin de poder comparar la precisión de cada modelo y de esa manera determinar su eficacia dentro de este estudio, el detalle de la evaluación de cada parámetro dentro de este apartado puede ser observado en el ANEXO IV del **Código 8**.

2.5.3.7 Comparación de modelos

Como último proceso dentro de este apartado se tiene la correspondiente comparación de los modelos obtenidos durante la implementación de una técnica de Machine Learning, esto permite realizar una comparativa de cómo está operando cada modelo dentro del conjunto de datos de entrenamiento y prueba mediante el uso de los atributos de radiofrecuencia.

El detalle de la implementación de este apartado puede ser observado en el ANEXO IV del **Código 9**.

2.5.4 IMPLEMENTACIÓN DE LOS MODELOS DEL DOA POR RANGOS

Debido a que los modelos obtenidos previamente no poseen una precisión alta como se observará en la sección de resultados, se procede a plantear la elaboración de modelos del DoA mediante rangos como se pudo observar en la generación del análisis de

tendencia para parámetros de radiofrecuencia en función del DoA, de manera resumida se obtendrá los 3 mejores modelos para cada rango designado.

Los rangos definidos para la implementación de los siguientes modelos a obtener están dados por:

- Rango 1: 0° a 150°
- Rango 2: 150° a 220°
- Rango 3: 220° a 300°
- Rango 4: 300° a 360°

A continuación se procede a mostrar los cambios existentes de manera general que permiten elaborar la implementación de una técnica de Machine Learning a través de cada una de las etapas descritas previamente.

- Creación de dataset: Se crea un nuevo dataset para cada rango con el fin de poder ser filtrado y se define el filtro del rango del DoA.
- Matriz de correlación: Se obtiene la matriz de correlación definida en base al filtro del rango, esto varía conforme a cada tipo de rango definido.
- Implementación de la receta: Dentro de la creación de la receta para cada rango varía el dataset filtrado para cada rango designado.
- Generación del modelo: Se genera el primer modelo para cada rango correspondiente haciendo uso de todos los atributos.
- Ajuste del modelo: Se ajusta el modelo en base a la matriz de correlación para cada rango.
- Evaluación del modelo: Se almacena la precisión de cada modelo ajustado dentro del rango definido.
- Comparación de modelos por rangos: Se comparan los 3 modelos obtenidos en un rango determinado del DoA con mayor precisión .
- Comparación de predicciones y valores reales del DoA: se crean tablas para observación y análisis de error porcentual que servirá para evaluaciones de modelos por rangos.

El detalle de la implementación de este apartado puede ser observado en el ANEXO IV del Código 10.

3 RESULTADOS, CONCLUSIONES Y RECOMENDACIONES

3.1 RESULTADOS

A continuación se procede a exponer los resultados obtenidos en el estudio del DoA, de esta manera, se indicarán los resultados gráficos, cálculos de parámetros estadísticos, tablas de modelos y su correspondiente evaluación, entre otra información que permite obtener la mejor forma de realizar el estudio correspondiente.

3.1.1 IMPLEMENTACIÓN DE UNA TÉCNICA DE MACHINE LEARNING

3.1.1.1 Matriz de correlación obtenida

```
> cor(data_doa_train)#matriz de correlacion mediante tabla
      rssi_strongest  rssi  rsrq  rssnr Distancia_m DoA
rssi_strongest  1.0000000 0.68375344 0.21729450 0.5520074 -0.57499626 -0.09386000
rssi            0.6837534 1.00000000 0.63508628 0.8647983 -0.65118366 -0.07245234
rsrq            0.2172945 0.63508628 1.00000000 0.6406609 -0.42164920 -0.06967686
rssnr          0.5520074 0.86479829 0.64066087 1.0000000 -0.70469843 -0.13371864
Distancia_m    -0.5749963 -0.65118366 -0.42164920 -0.7046984 1.00000000 -0.01075611
DoA            -0.0938600 -0.07245234 -0.06967686 -0.1337186 -0.01075611 1.00000000
```

Figura 3.1. Matriz de correlación entre atributos.

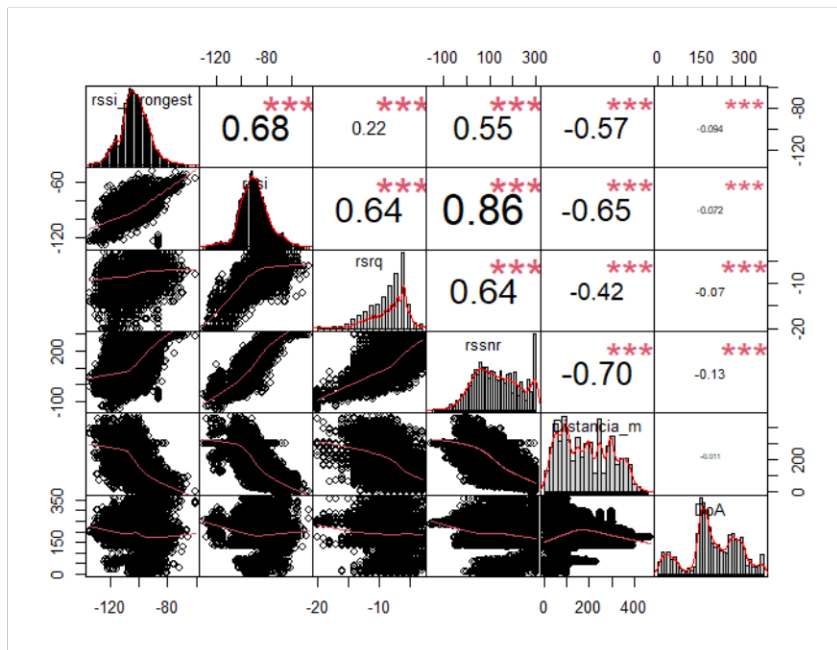


Figura 3.2. Matriz de correlación entre atributos con inclusión de histogramas.

Como se puede observar en la **Figura 3.1** y **Figura 3.2**, la obtención de la matriz de correlación permite observar los coeficientes de relación que existen entre cada atributo, como se puede visualizar, existe una gran relación entre los parámetros de radiofrecuencia que se obtienen en el proceso de medición de datos, además se observa que el atributo adicional calculado de la distancia también posee una relación negativa que indica que el aumento de un atributo está acompañado de una disminución del otro. Adicionalmente, permite ver de manera simplificada el histograma de cada atributo para conocer la concentración de datos y su frecuencia de ocurrencia en cada uno de sus valores correspondientes.

3.1.1.2 Implementación de la receta

```
> data_doa_receta

— Recipe —

— Inputs
Number of variables by role
outcome: 1
predictor: 5

— Training information
Training data contained 19402 data points and no incomplete rows.

— Operations
• Zero variance filter removed: <none> | Trained
• Variable mutation for: ~runif(nrow(data_doa_train), 0, 1) | Trained
• Variable mutation for: ~as.factor(if_else(distcic > 0.8, "Norm", "Rec")) | Trained
• Variables selected: rssi_strongest, rssi, rsrq, rssnr, Distancia_m, DoA, cic | Trained
```

Figura 3.3. Creación y detalle de la receta.

En la **Figura 3.3**, se puede observar que la receta se define con 5 atributos predictores y un atributo de salida o resultado, los datos de entrenamiento son de 19402 muestras, además se puede visualizar la forma en la que se aplicará el modelo mediante la receta configurada a través de diferentes operaciones de mutación.

3.1.1.3 Generación, ajuste y rendimiento de los modelos obtenidos

```

> mod_doa
Call:
lm(formula = DoA ~ ., data = training_base_doa)

Coefficients:
(Intercept)  rssi_strongest      rssi      rsrq      rssnr  Distancia_m      cicRec
 303.3204    -1.7740      2.4636    -1.2920    -0.3946    -0.1923    -1.5770

a)

> mod_doa2
Call:
lm(formula = DoA ~ rssi + Distancia_m + rssi:Distancia_m, data = training_base_doa)

Coefficients:
(Intercept)      rssi  Distancia_m  rssi:Distancia_m
 89.190414    -1.248113    0.057072    0.001425

b)

> mod_doa3
Call:
lm(formula = DoA ~ rssi_strongest:rssi_strongest + rssi:rssi_strongest +
  rsrq:rssi_strongest + Distancia_m:rssi:rsrq, data = training_base_doa)

Coefficients:
(Intercept)      rssi_strongest  rssi_strongest:rssi  rssi_strongest:rsrq
 77.4193852    -0.6074709    0.0023248    0.0547136
rssi:rsrq:Distancia_m
 -0.0001331

c)

```

Figura 3.4. a) Coeficientes del modelo 1 ajustado b) Coeficientes del modelo 2 ajustado c) Coeficientes del modelo 3 ajustado.

```

> rendimiento_mod_doa
  modelo  RMSE  R2
1 mod_doa1 82.59306 0.05556369
2 mod_doa2 84.33422 0.01547522
3 mod_doa3 83.60083 0.03261260

```

Figura 3.5. Rendimiento estadístico de cada modelo implementado.

Tabla 3.1. Rendimiento de los modelos para la predicción del DoA.

Modelo	Expresión	RMSE	R2
mod_doa1	rssi_strongest+rssi+rsrq+rssnr+Distancia_m	82.59306	0.05556369
mod_doa2	rssi+Distancia_m+rssi:Distancia_m	84.33422	0.01547522
mod_doa3	rssi_strongest:rssi_strongest+rssi:rssi_strongest+rsrq:rssi_strongest+Distancia_m:rssi:rsrq	83.60083	0.03261260

Como se puede observar mediante la **Figura 3.4**, **Figura 3.5** y la **Tabla 3.1**, mediante el RMSE de cada modelo implementado, se puede observar que no existe una mejora significativa dentro de la predicción y los valores reales del DoA, además, mediante el R2 se muestra que el porcentaje de variabilidad del DoA que es explicado por cada modelo no es adecuado, por lo tanto, se puede inferir que estos modelos no son precisos de la forma en que se aplican diferentes expresiones que permiten ajustar modelos en base a la matriz de correlación. En resumen se tiene que hasta el momento ninguno de los modelos tiene

una predicción alta y por ende no son precisos, dentro de esto, se plantea la posibilidad de emplear nuevas metodologías para la generación de los modelos, técnicas más complejas y atributos predictores.

Lo recomendable dentro de estos análisis estadísticos se basa en que el RMSE sea el más bajo posible para indicar que existe una precisión adecuada y el R2 el más alto posible para indicar que existe una variabilidad que explica el modelo mediante los atributos predictores.

3.1.1.4 Comparación de modelos mediante análisis estadístico

```
> modelos
# A tibble: 3,880 × 12
  rssi_strongest rssi rsrq rssnr Distancia_m DoA DoA_m1 DoA_m2 DoA_m3 delta_m1 delta_m2 delta_m3
  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 -77 -63 -7 296 0.991 308. 175. 168. 165. 133. 140. 143.
2 -105 -84 -5 244 20.2 351. 187. 193. 189. 164. 158. 162.
3 -104 -84 -5 218 24.6 356. 195. 192. 188. 161. 163. 168.
4 -102 -90 -6 166 30.3 359. 197. 199. 192. 162. 160. 167.
5 -100 -86 -6 156 34.7 0.532 207. 194. 189. 206. 194. 188.
6 -100 -86 -6 156 36.8 0.448 206. 194. 188. 206. 194. 188.
7 -97 -91 -8 156 38.0 0.134 193. 200. 196. 193. 200. 196.
8 -95 -90 -8 132 42.8 358. 199. 198. 192. 159. 160. 166.
9 -95 -92 -8 150 43.4 358. 186. 201. 193. 171. 157. 165.
10 -90 -91 -10 78 52.2 359. 209. 199. 194. 149. 160. 165.
# i 3,870 more rows
# i Use `print(n = ...)` to see more rows
```

Figura 3.6. Dataset para visualización de variaciones entre predicciones y valores reales del DoA para cada modelo implementado.

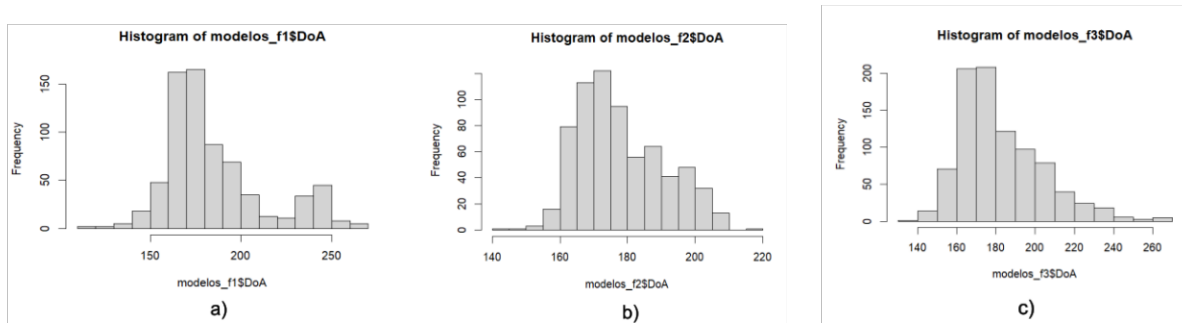


Figura 3.7. a) Histograma del modelo mediante predicción y valor real del DoA filtrado 1
 b) Histograma del modelo mediante predicción y valor real del DoA filtrado 2 c)
 Histograma del modelo mediante predicción y valor real del DoA filtrado 3.

Como se puede observar en la **Figura 3.6** y **Figura 3.7**, mediante el dataset se puede visualizar el valor real del DoA del conjunto de prueba y su comparación con las predicciones de cada modelo ajustado, además, se añade la variación que existe entre la predicción y su correspondiente valor real del DoA, esto permite visualizar que tan preciso es el modelo, como se puede ver, los valores predichos del DoA no están cercanos a los reales, por lo cual se hace uso de las variaciones, que nos permiten realizar las gráficas de los histogramas de cada modelo filtrando aquellas variaciones que son menores a 20°, de

esta manera nos permite visualizar la concentración de los datos del DoA y la frecuencia con la que se repiten.

Observando las concentraciones de los datos en los histogramas, vemos que donde poseen un menor error entre las predicciones y los valores reales del DoA se encuentran en un rango de entre 150° a 220° aproximadamente, por lo cual, de primera instancia mediante gráficas se obtienen rangos de 0° a 150° , 150° a 220° , 220° a 300° y de 300° a 360° , esto nos lleva al resultado de que se puede plantear la implementación de una técnica de Machine Learning por rangos con el objetivo de poder visualizar distribuciones normales, ya que dentro de la estadística son más fáciles de predecir en base a eventos o datos debido a su concentración dentro del pico de la función normal.

Esto permite corroborar que se puede plantear una metodología en la cual se pueda usar un modelo dentro de este rango, y aplicar el mismo concepto para los rangos restantes.

3.1.2 IMPLEMENTACIÓN DE MODELOS DEL DOA POR RANGOS

3.1.2.1 Matriz de correlación obtenida

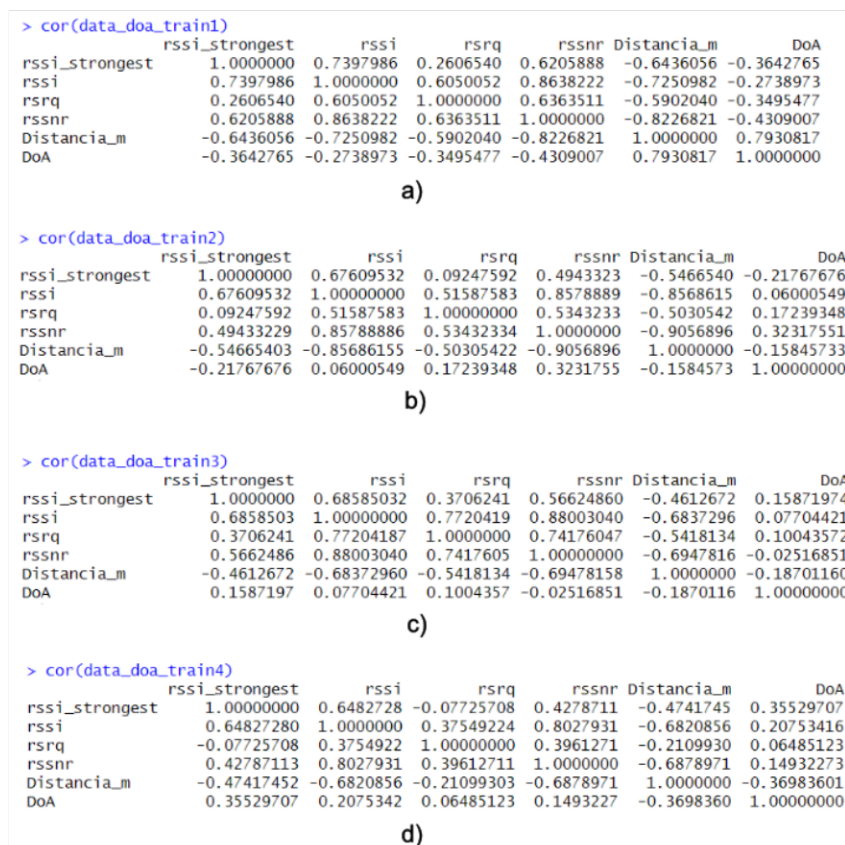


Figura 3.8. a) Matriz de correlación del rango de 0° a 150° b) Matriz de correlación del rango de 150° a 220° c) Matriz de correlación del rango de 220° a 300° d) Matriz de correlación del rango de 300° a 360° .

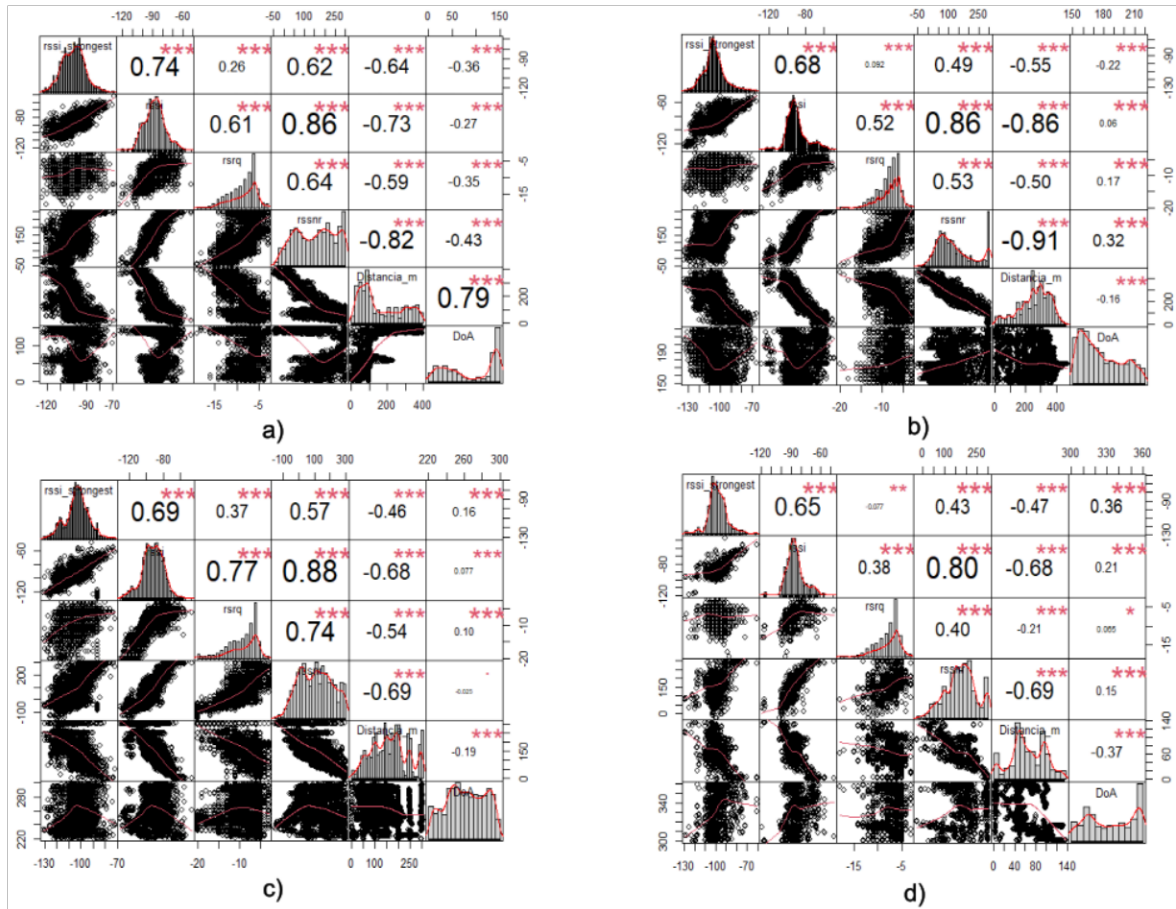


Figura 3.9. a) Matriz de correlación e histogramas del rango de 0° a 150° b) Matriz de correlación e histogramas del rango de 150° a 220° c) Matriz de correlación e histogramas del rango de 220° a 300° d) Matriz de correlación e histogramas del rango de 300° a 360°.

La obtención de las matrices de correlación correspondiente a cada rango como se puede observar en la **Figura 3.8** y **Figura 3.9** permite evidenciar que dependiendo del valor del rango del DoA, el coeficiente de correlación de cada atributo respecto a otro varía, por otra parte se observa que los coeficientes dentro de los rangos definidos aportan e influyen en el DoA debido al indicador de los tres puntos dentro de cada coeficiente, además, los atributos de radiofrecuencia poseen un coeficiente de correlación positivo y de gran importancia de relación entre estos. Por otro lado, se observa que los coeficientes de correlación positiva predominan dentro de la matriz de correlación, indicando que si un atributo aumenta el correspondiente aumenta de la misma manera, mientras que los atributos adicionales poseen coeficiente de correlación negativa respecto a los atributos de radiofrecuencia, indicando que si un atributo aumenta el correspondiente disminuye de la misma manera, y estos influyen con los atributos de radiofrecuencia en gran medida, esto

permite decidir que atributos pueden ser los más óptimos para ser considerados en el modelo.

3.1.2.2 Implementación de las recetas

<pre>> data_doa_receta1 — Recipe — — Inputs Number of variables by role outcome: 1 predictor: 5 — Training information Training data contained 5815 data points and no incomplete rows. — Operations • Zero variance filter removed: <none> Trained • Variable mutation for: ~runif(nrow(data_doa_train1), 0, 1) Trained • Variable mutation for: ~as.factor(if_else(distcic > 0.8, "Norm", "Rec")) Trained • Variables selected: rssi_strongest, rssi, rsrq, rssnr, Distancia_m, DoA, cic Trained</pre> <p style="text-align: center;">a)</p>	<pre>> data_doa_receta2 — Recipe — — Inputs Number of variables by role outcome: 1 predictor: 5 — Training information Training data contained 6516 data points and no incomplete rows. — Operations • Zero variance filter removed: <none> Trained • Variable mutation for: ~runif(nrow(data_doa_train2), 0, 1) Trained • Variable mutation for: ~as.factor(if_else(distcic > 0.8, "Norm", "Rec")) Trained • Variables selected: rssi_strongest, rssi, rsrq, rssnr, Distancia_m, DoA, cic Trained</pre> <p style="text-align: center;">b)</p>
<pre>> data_doa_receta3 — Recipe — — Inputs Number of variables by role outcome: 1 predictor: 5 — Training information Training data contained 5625 data points and no incomplete rows. — Operations • Zero variance filter removed: <none> Trained • Variable mutation for: ~runif(nrow(data_doa_train3), 0, 1) Trained • Variable mutation for: ~as.factor(if_else(distcic > 0.8, "Norm", "Rec")) Trained • Variables selected: rssi_strongest, rssi, rsrq, rssnr, Distancia_m, DoA, cic Trained</pre> <p style="text-align: center;">c)</p>	<pre>> data_doa_receta4 — Recipe — — Inputs Number of variables by role outcome: 1 predictor: 5 — Training information Training data contained 1446 data points and no incomplete rows. — Operations • Zero variance filter removed: <none> Trained • Variable mutation for: ~runif(nrow(data_doa_train4), 0, 1) Trained • Variable mutation for: ~as.factor(if_else(distcic > 0.8, "Norm", "Rec")) Trained • Variables selected: rssi_strongest, rssi, rsrq, rssnr, Distancia_m, DoA, cic Trained</pre> <p style="text-align: center;">d)</p>

Figura 3.10. a) Creación y detalle de la receta en el rango de 0° a 150° b) Creación y detalle de la receta en el rango de 150° a 220° c) Creación y detalle de la receta en el rango de 220° a 300° d) Creación y detalle de la receta en el rango de 300° a 360°.

Mediante la **Figura 3.10**, se puede observar que las recetas se definen mediante 5 atributos predictores y un atributo de salida o resultado, los datos de entrenamiento son de 5815, 6516, 5625 y 1446 muestras respectivamente en cada rango, además, se puede visualizar la forma en la que se aplicarán los modelos mediante las recetas configuradas a través de diferentes operaciones de mutación para insertar factores con niveles denominados “Norm” y “Rec”.

3.1.2.3 Generación, ajustes y evaluación de los modelos obtenidos

```

> mod_doal
Call:
lm(formula = DoA ~ ., data = training_base_doal)

Coefficients:
(Intercept)  rssi_strongest      rssi      rsrq      rssnr  Distancia_m      cicRec
 175.8891    -0.2393      2.7084    -0.4612     0.1601     0.6169     0.7710

a)

> mod_doal
Call:
lm(formula = DoA ~ (Distancia_m + rssi) * rssnr:rsrq + rssnr:Distancia_m,
    data = training_base_doal)

Coefficients:
(Intercept)      Distancia_m      rssi      rssnr:rsrq
 1.012e+02      4.742e-01      1.239e+00      -8.727e-02
Distancia_m:rssnr  Distancia_m:rssnr:rsrq
 1.336e-03      -1.810e-05      -1.063e-03

b)

> mod_doal
Call:
lm(formula = DoA ~ (Distancia_m + rssi) * rssnr * rsrq * rssi_strongest *
    rssi, data = training_base_doal)

Coefficients:
(Intercept)      Distancia_m      rssi      rssnr
 -3.487e+03      1.720e+01      1.656e+01      1.580e-01
rssi_strongest  rsrq      Distancia_m:rssnr
 -3.833e+01      -5.489e+01      -9.448e-02
Distancia_m:rsrq  Distancia_m:rssi      rssi:rssnr
 6.047e-01      6.226e-01      1.793e-01
rssi:rssi_strongest  rssi:rsrq      rssi:rssnr:rsrq
 -3.750e-01      -1.119e+00      8.462e-03
rsrq:rssi_strongest  Distancia_m:rssi_strongest
 -3.946e-01      1.606e-01      1.772e-03
Distancia_m:rssnr:rsrq  rssi:rssi_strongest
 -4.345e-03      1.756e-01      -1.055e-02
Distancia_m:rssnr:rssi_strongest  rssi:rsnr
 -9.401e-04      rssi:rssnr:rsrq
 5.446e-03      8.462e-03
rssnr:rsrq:rssi_strongest  rssi:rssnr:rssi_strongest
 7.038e-03      1.772e-03
Distancia_m:rssi:rsrq  rssi:rssi_strongest:rsrq
 7.501e-03      -1.028e-03
Distancia_m:rssnr:rsrq:rssi_strongest  Distancia_m:rssi:rssi_strongest
 -4.126e-05      1.678e-03
Distancia_m:rssi:rssnr:rsrq  rssi:rssnr:rsrq:rssi_strongest
 -5.053e-05      9.921e-05
Distancia_m:rssi:rsrq:rssi_strongest  Distancia_m:rssi:rssnr:rssi_strongest
 7.174e-05      -1.019e-05
      Distancia_m:rssi:rssnr:rsrq:rssi_strongest
      -4.969e-07
  
```

Figura 3.11. a) Coeficientes del modelo 1 del rango de 0° a 150° b) Coeficientes del modelo 2 del rango de 0° a 150° c) Coeficientes del modelo 3 del rango de 0° a 150°.

```

> rendimiento_mod_doal2
  modelo  RMSE  R2
1 mod_doal 20.79867 0.8477548
2 mod_doal 19.41548 0.8673992
3 mod_doal 17.93162 0.8870832
  
```

Figura 3.12. Rendimiento estadístico de cada modelo implementado para el rango de 0° a 150°.

Tabla 3.2. Rendimiento de los modelos para predicción del DoA en rango de 0° a 150°.

Modelo	Expresión	RMSE	R2
mod_doal1	rssi_strongest+rssi+rsrq+rssnr+Distancia_m	20.79867	0.8477548
mod_doal1	(Distancia_m+rssi)*rssnr:rsrq+rssnr:Distancia_m	19.41548	0.8673992
mod_doal1	(Distancia_m+rssi)*rssnr*rsrq*rssi_strongest*rssi	17.93162	0.8870832

Como se visualiza en la **Figura 3.11**, **Figura 3.12** y la **Tabla 3.2** correspondiente a los modelos ajustados, mediante el RMSE de cada modelo ajustado para el rango de 0° a 150° , se puede observar que existe una mejora significativa dentro de la predicción y los valores reales del DoA mediante el uso de rangos debido a que el valor del RMSE disminuye. Además, mediante el R^2 se muestra que la variabilidad del DoA que es explicado por cada modelo es muy adecuado y mejora acorde a cada ajuste, por lo tanto, se puede concluir que estos modelos son más precisos que los iniciales, ya que se aplican diferentes expresiones que permiten ajustar modelos en base a la matriz de correlación y mediante la delimitación de un rango del DoA analizado previamente. En resumen, se tiene que el mejor modelo del rango de 0° a 150° tiene una precisión alta y una alta variabilidad indicando la relación entre los atributos de entrada y salida a través del modelo, además, se tiene que los atributos que predominan en la influencia de la predicción del DoA son la distancia y el RSSI.

```

> mod_doa2
Call:
lm(formula = DoA ~ ., data = training_base_doa2)

Coefficients:
(Intercept)  rssi_strongest      rssi      rsrq      rssnr  Distancia_m  cicRec
-14.40785    -0.67708    -0.70998    -0.39264    0.26439    0.07218    0.89603

a)

> mod_doa2
Call:
lm(formula = DoA ~ rssnr * rssi * rssi_strongest, data = training_base_doa2)

Coefficients:
(Intercept)          rssnr          rssi          rssi_strongest
-5.565e+02      3.341e+00     -6.442e+00     -5.066e+00
 rssnr:rssi      rssnr:rssi_strongest      rssi:rssi_strongest  rssnr:rssi:rssi_strongest
 3.452e-02      2.655e-02      -4.328e-02      2.919e-04

b)

> mod_doa2
Call:
lm(formula = DoA ~ rssnr * rssi * rssi_strongest * rsrq * Distancia_m,
data = training_base_doa2)

Coefficients:
(Intercept)          rssnr          rssi          rssi_strongest
 3.737e+03     -1.285e+01     4.690e+01     3.982e+01
          rsrq          Distancia_m
 3.102e+01     -8.279e+00
 rssnr:rssi          rssnr:rssi_strongest
-1.688e-01     -1.491e-01
 rssi:rssi_strongest  rssnr:rsrq
 5.114e-01     -3.809e-01
 rssi:rsrq          rssi_strongest:rsrq
 9.973e-01     3.899e-01
 rssnr:Distancia_m  rssi:Distancia_m
 3.018e-02     -1.086e-01
 rssi_strongest:Distancia_m  rsrq:Distancia_m
-8.821e-02     9.473e-02
 rssnr:rssi:rssi_strongest  rssnr:rssi:rsrq
-1.923e-03     -7.271e-03
 rssnr:rssi_strongest:rsrq  rssi:rssi_strongest:rsrq
-4.607e-03     1.126e-02
 rssnr:rssi:Distancia_m  rssnr:rssi_strongest:Distancia_m
 3.965e-04     3.334e-04
 rssi:rssi_strongest:Distancia_m  rssnr:rsrq:Distancia_m
-1.146e-03     1.338e-03
 rssi:rsrq:Distancia_m  rssi_strongest:rsrq:Distancia_m
-7.669e-04     8.044e-04
 rssnr:rssi:rssi_strongest:rsrq  rssnr:rssi:rssi_strongest:Distancia_m
-8.528e-05     4.449e-06
 rssnr:rssi:rsrq:Distancia_m  rssnr:rssi_strongest:rsrq:Distancia_m
 2.500e-05     1.419e-05
 rssi:rssi_strongest:rsrq:Distancia_m  rssnr:rssi:rssi_strongest:rsrq:Distancia_m
-1.034e-05     2.693e-07

c)

```

Figura 3.13. a) Coeficientes del modelo 1 del rango de 150° a 220° b) Coeficientes del modelo 2 del rango de 150° a 220° c) Coeficientes del modelo 3 del rango de 150° a 220° .

```

> rendimiento_mod_doa2
  modelo      RMSE      R2
1 mod_doa2 15.50206 0.3680900
2 mod_doa2 14.61214 0.4386456
3 mod_doa2 13.11457 0.5479492

```

Figura 3.14. Rendimiento estadístico de cada modelo implementado para rango de 150° a 220°.

Tabla 3.3. Rendimiento de los modelos para predicción del DoA en rango de 150° a 220°.

Modelo	Expresión	RMSE	R2
mod_doa2	rss_i_strongest+rss_i+rsrq+rssnr+Distancia_m	15.50206	0.3680900
mod_doa2	rssnr*rss_i*rss_i_strongest	14.61214	0.4386456
mod_doa2	rssnr*rss_i*rss_i_strongest*rsrq*Distancia_m	13.11457	0.5479492

Como se puede observar mediante la **Figura 3.13** y **Figura 3.14** y la **Tabla 3.3** correspondiente a los modelos ajustados, mediante el RMSE de cada modelo ajustado para el rango de 150° a 220°, se puede observar que existe una mejora significativa dentro de la predicción y los valores reales del DoA mediante el uso de rangos debido a que el valor del RMSE disminuye. Además, mediante el R2 se muestra que la variabilidad del DoA para este rango es cercano a la mitad, por lo tanto, se obtiene que estos modelos son precisos en comparación a los iniciales de la forma en que se aplican diferentes expresiones que permiten ajustar modelos en base a la matriz de correlación y mediante la delimitación de un rango del DoA analizado previamente. En resumen, se tiene que el mejor modelo del rango de 150° a 220° tiene una alta precisión y una variabilidad aceptable indicando que la relación entre los atributos de entrada y salida es medianamente aceptable a través del modelo, cabe recalcar que dentro de este rango su precisión es la más alta para el DoA, además, los atributos que predominan en la influencia de la predicción del DoA son todos los atributos predictores.

```

> mod_doa3
Call:
lm(formula = DoA ~ ., data = training_base_doa3)

Coefficients:
(Intercept)  rssi_strongest      rssi      rsrq      rssnr  Distancia_m      cicRec
 355.88932    0.52815    -0.02931    1.73298    -0.11815    -0.09023    -0.08919

a)

> mod_doa3
Call:
lm(formula = DoA ~ (Distancia_m + rssnr) * rssi * rssi_strongest +
  rssnr * rsrq, data = training_base_doa3)

Coefficients:
(Intercept)          Distancia_m          rssnr
 5.823e+02    -1.747e+00    -1.955e+00
rssi          rssi_strongest          rsrq
 4.816e+00    3.460e+00    5.979e-01
Distancia_m:rssi          rssnr:rssi          Distancia_m:rssi_strongest
 -2.742e-02    -2.742e-02    -2.034e-02
rssnr:rssi_strongest          rssi:rssi_strongest          rssnr:rsrq
 -1.781e-02    5.428e-02    -5.002e-03
Distancia_m:rssi:rssi_strongest          rssnr:rssi:rssi_strongest
 -3.137e-04    -2.664e-04

b)

> mod_doa4
Call:
lm(formula = DoA ~ (rssi_strongest + Distancia_m) * rssnr * rsrq *
  rssi, data = training_base_doa4)

Coefficients:
(Intercept)          rssi_strongest          Distancia_m
 3.270e+03    3.144e+01    2.980e+00
rssi          rsrq          rssi
 -1.121e+01    1.781e+02    2.821e+01
rssi_strongest:rssi          Distancia_m:rssi          rssi_strongest:rsrq
 -1.232e-01    -1.973e-02    1.997e+00
Distancia_m:rsrq          rssnr:rsrq          rssi_strongest:rssi
 2.208e-01    -7.245e-01    3.129e-01
Distancia_m:rssi          rssnr:rssi          rsrq:rssi
 4.485e-02    -1.096e-01    1.774e+00
rssi_strongest:rssnr:rsrq          Distancia_m:rssi:rsrq          rssi_strongest:rssi:rsrq
 -8.322e-03    -1.089e-03    -1.257e-03
Distancia_m:rssi:rssi          rssi_strongest:rsrq:rssi          Distancia_m:rsrq:rssi
 -2.905e-04    2.028e-02    2.969e-03
rssnr:rsrq:rssi          rssi_strongest:rssi:rsrq:rssi          Distancia_m:rssi:rsrq:rssi
 -7.172e-03    -8.540e-05    -1.791e-05

c)

```

Figura 3.15. a) Coeficientes del modelo 1 del rango de 220° a 300° b) Coeficientes del modelo 2 del rango de 220° a 300° c) Coeficientes del modelo 3 del rango de 220° a 300°.

```

> rendimiento_mod_doa32
  modelo  RMSE  R2
1 mod_doa3 19.26547 0.1873852
2 mod_doa3 17.50235 0.3265683
3 mod_doa3 17.26673 0.3432788

```

Figura 3.16. Rendimiento estadístico de cada modelo implementado para rango de 220° a 300°.

Tabla 3.4. Rendimiento de los modelos para predicción del DoA en rango de 220° a 300°.

Modelo	Expresión	RMSE	R2
mod_doa 3	rssi_strongest+rssi+rsrq+rssnr+Distancia_m	19.2654	0.187385
mod_doa 3	(Distancia_m+rssnr)*rssi*rssi_strongest+rssnr*rsrq	17.5023	0.326568
mod_doa 3	(Distancia_m+rssnr)*rssi*rssi_strongest+rssnr*rsrq*Distancia_m*rssi*rssi_strongest	17.2667	0.343278

Como se puede observar mediante la **Figura 3.15**, **Figura 3.16** y la **Tabla 3.4** correspondiente a los modelos ajustados, mediante el RMSE de cada modelo ajustado para el rango de 220° a 300° , se puede observar que existe una mejora significativa dentro de la predicción y los valores reales del DoA mediante el uso de rangos debido a que el valor del RMSE disminuye. Además, mediante el R^2 se muestra que la variabilidad del DoA para este rango es cercano a la mitad, por lo tanto se obtiene que estos modelos son precisos de la forma en que se aplican diferentes expresiones que permiten ajustar modelos en base a la matriz de correlación y mediante la delimitación de un rango del DoA analizado previamente. En resumen, se tiene que el mejor modelo del rango de 220° a 300° tiene una alta precisión y una variabilidad aceptable indicando que la relación entre los atributos de entrada y salida es medianamente baja pero aceptable a través del modelo pero no todos los atributos aportan información en la predicción del DoA, además se tiene que los atributos que predominan en la influencia de la predicción del DoA son la distancia y el RSSNR.

```

> mod_doa4
Call:
lm(formula = DoA ~ ., data = training_base_doa4)

Coefficients:
(Intercept)  rssi_strongest      rssi      rsrq      rssnr  Distancia_m      cicRec
  407.78148    0.99010    -0.73308    1.59736   -0.05084   -0.29041   -0.93600

a)

> mod_doa4
Call:
lm(formula = DoA ~ (rssi_strongest + Distancia_m) * rssnr * rsrq,
  data = training_base_doa4)

Coefficients:
(Intercept)      rssi_strongest      Distancia_m      rssnr
  7.842e+02      4.083e+00      -6.314e-01      -1.549e+00
rsrq      rssi_strongest:rssnr      Distancia_m:rssnr      rssi_strongest:rsrq
  2.850e+01      -1.364e-02      3.532e-03      3.123e-01
Distancia_m:rsrq      rssnr:rsrq      rssi_strongest:rssnr:rsrq      Distancia_m:rssnr:rsrq
  8.677e-03      -1.151e-01      -1.211e-03      1.834e-04

b)

> mod_doa4
Call:
lm(formula = DoA ~ (rssi_strongest + Distancia_m) * rssnr * rsrq *
  rssi, data = training_base_doa4)

Coefficients:
(Intercept)      rssi_strongest      Distancia_m      rssnr
  3.270e+03      3.144e+01      2.980e+00      -1.121e+01
rsrq      rssi      rssi_strongest:rssnr      Distancia_m:rssnr
 -1.121e+01      1.781e+02      2.821e+01      -1.973e-02
rssi_strongest:rssnr      Distancia_m:rssnr:rsrq      rssi_strongest:rsrq      rssi_strongest:rssnr:rsrq
 -1.232e-01      -1.973e-02      1.997e+00      -7.245e-01
Distancia_m:rsrq      rssnr:rsrq      rssi_strongest:rssi      Distancia_m:rssi
  2.208e-01      -7.245e-01      3.129e-01      -1.096e-01
Distancia_m:rssi      rssnr:rssi      rsrq:rssi      rssi_strongest:rssi
  4.485e-02      -1.096e-01      1.774e+00      -1.089e-03
rssi_strongest:rssnr:rsrq      Distancia_m:rssnr:rsrq      rssi_strongest:rssnr:rssi      Distancia_m:rssi
 -8.322e-03      -1.089e-03      -1.257e-03      2.028e-02
Distancia_m:rssnr:rssi      rssi_strongest:rssi      Distancia_m:rssi      rssi_strongest:rssi
 -2.905e-04      2.028e-02      2.969e-03      -7.172e-03
rssi_strongest:rssi      rssi_strongest:rssi      Distancia_m:rssi      rssi_strongest:rssi
 -7.172e-03      -8.540e-05      -1.791e-05

c)

```

Figura 3.17. a) Coeficientes del modelo 1 del rango de 300° a 360° b) Coeficientes del modelo 2 del rango de 300° a 360° c) Coeficientes del modelo 3 del rango de 300° a 360° .

```
> rendimiento_mod_doa4
  modelo      RMSE      R2
1 mod_doa4 17.81874 0.1838969
2 mod_doa4 15.05957 0.4176007
3 mod_doa4 14.41402 0.4710272
```

Figura 3.18. Rendimiento estadístico de cada modelo implementado para el rango de 300° a 360°.

Tabla 3.5. Rendimiento de modelos para predicción del DoA en rango de 300° a 360°.

Modelo	Expresión	RMSE	R2
mod_doa4	rss_i_strongest+rss_i+rsrq+rssnr+Distancia_m	17.81874	0.1838969
mod_doa4	(rss_i_strongest+Distancia_m)*rssnr*rsrq	15.05957	0.41761007
mod_doa4	(rss_i_strongest+Distancia_m)*rssnr*rsrq*rss_i	14.41402	0.4710272

Como se puede observar mediante la **Figura 3.17**, **Figura 3.18** y la **Tabla 3.5** correspondiente a los modelos ajustados, mediante el RMSE de cada modelo ajustado para el rango de 300° a 360°, se puede observar que existe una mejora significativa dentro de la predicción y los valores reales del DoA mediante el uso de rangos debido a que el valor del RMSE disminuye. Además, mediante el R2 se muestra que la variabilidad del DoA para este rango es cercano a la mitad, por lo tanto se obtiene de que estos modelos son precisos ya que se aplican diferentes expresiones que permiten ajustar los modelos en base a la matriz de correlación y mediante la delimitación de un rango del DoA analizado previamente. En resumen, se tiene que el mejor modelo del rango de 300 a 360° tiene una predicción con una alta precisión y una variabilidad aceptable indicando que la relación entre los atributos de entrada y salida es medianamente aceptable a través del modelo, además, los atributos que predominan en la influencia de la predicción del DoA son la distancia y el RSSI STRONGEST.

3.1.2.4 Comparación de modelos mediante un análisis estadístico

```
> rendimiento_mod_doa_r
  modelo      RMSE      R2
1 mod_doa_rango1 17.93162 0.8870832
2 mod_doa_rango2 13.11457 0.5479492
3 mod_doa_rango3 17.26673 0.3432788
4 mod_doa_rango4 14.41402 0.4710272
```

Figura 3.19. Comparación del rendimiento estadístico para cada rango implementado.

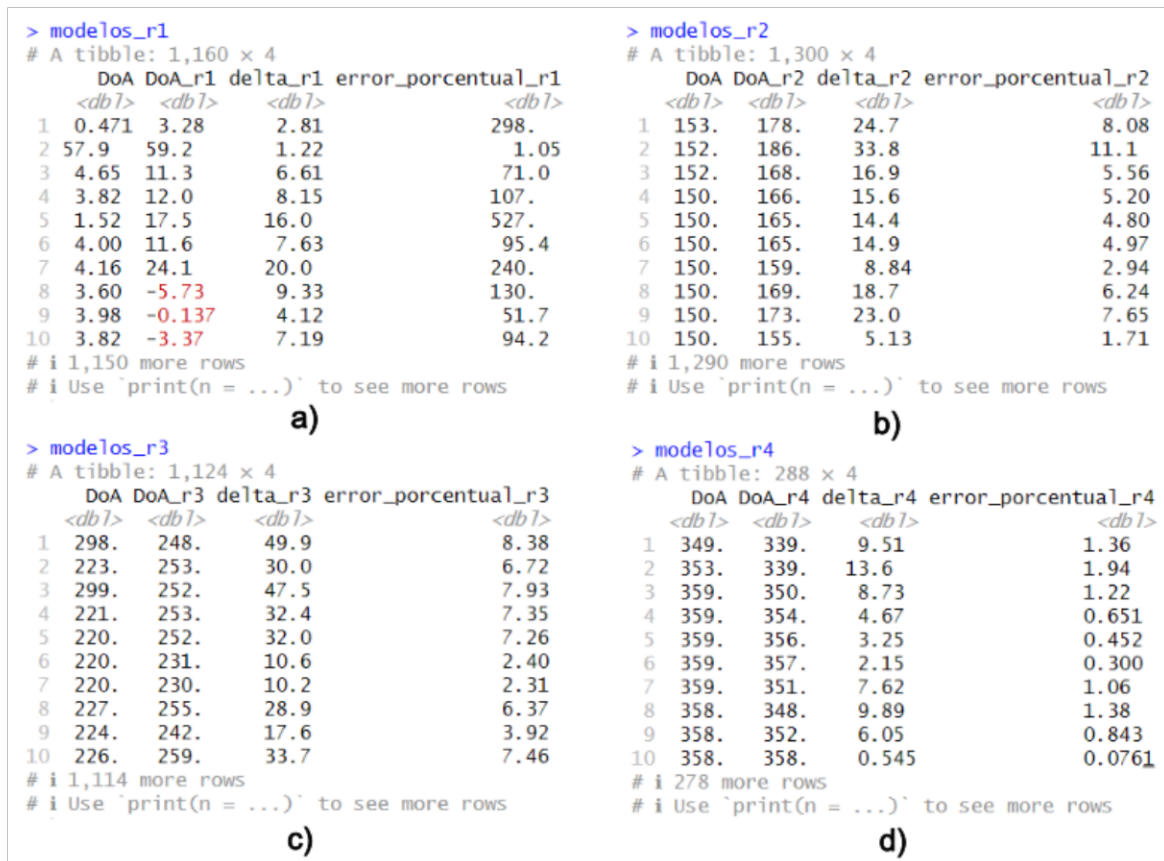


Figura 3.20. a) Variación y error porcentual rango de 0° a 150° b) Variación y error porcentual rango de 150° a 220° c) Variación y error porcentual rango de 220° a 300° d) Variación y error porcentual rango de 300° a 360°.

Como se puede observar mediante la **Figura 3.19**, mediante el RMSE de cada modelo implementado correspondiente a cada rango, se puede visualizar que el modelo ajustado para el rango de 150° a 220° destaca indicando que es el más preciso, por otro lado los demás modelos también permiten realizar predicciones del DoA para cada rango definido previamente, con una alta precisión y una gran variabilidad entre los atributos predictores y el DoA. Además, se observa que el modelo ajustado para el rango de 0° a 150° posee la variabilidad más alta indicando una alta relación entre los atributos predictores y el DoA a través del modelo, como resultado final se puede concluir que se obtuvieron modelos relativamente adecuados para la predicción del DoA mediante una técnica de regresión lineal permitiendo relacionar los parámetros de radiofrecuencia y observando su influencia para la estimación del DoA.

Mediante la **Figura 3.20**, se observan las variaciones y error porcentual de una cantidad determinada de datos entre la predicción y los valores reales del DoA, indicando que sus errores y variaciones son bajas y esto mejora conforme el modelo se estabiliza prediciendo

de mejor manera datos concentrados dentro del rango delimitado para cada tipo de modelo implementado.

3.1.3 RESUMEN DE RESULTADOS

Se procede a presentar un resumen de los resultados obtenidos en el estudio de la estimación del DoA, esto se observa mediante la **Tabla 3.6**.

Tabla 3.6. Resumen de modelos para estimación del DoA.

Modelo	Rango	Ecuación	Evaluación	
			RMSE	R2
General	0° a 360°	$-1.7757 * rssi_strongest + 2.4631 * rssi - 1.297 * rsrq$ $- 0.3945 rsnr - 0.1925 * Distancia_m$ $+ 303.7432$	82,593 06	0,055 5636 9
	0° a 150°	$(-4.969 * 10^{-7} * Distancia_m + 9.921 * 10^{-5} rssi) * rsnr * rsrq$ $* rssi_strongest * rssi - 3.487 * 10^3$	17,931 62	0,887 0832
Rangos	150° a 220°	$2.693 * 10^{-7} * rsnr * rssi * rssi_strongest * rsrq * Distancia_m$ $+ 3.737 * 10^3$	13,114 57	0,547 9492
	220° a 300°	$(-1.528 * 10^{-3} * Distancia_m - 8.819 * 10^{-4} * rsnr) * rssi$ $* rssi_strongest + 1.791 * 10^{-8} * rsnr * rsrq$ $* Distancia_m * rssi * rssi_strongest + 2.136$ $* 10^3$	17,266 73	0,343 2788
	300° a 360°	$(-8.540 * 10^{-5} rssi_strongest - 1.791 * 10^{-5} Distancia_m)$ $* rsnr * rsrq * rssi + 3.270 * 10^3$	14,414 02	0,471 0272

3.2 CONCLUSIONES

- Mediante el análisis de las mediciones de radiofrecuencia en una red real de comunicaciones móviles y la aplicación de una técnica de Machine Learning para la estimación del DoA, se identificaron patrones que influyen en la precisión de la estimación y en el comportamiento de dicho parámetro. Se observó que la estimación del DoA depende en gran medida de la distancia y dentro de qué direccionamiento del lóbulo de radiación se encuentra la estación móvil.
- Con la implementación de los modelos específicos mediante rangos para la estimación del DoA, se obtuvo un valor mínimo de error de 13.11457 en el rango de 150° a 220° y un valor máximo de error de 17.93162 en el rango de 0° a 150°

indicando su variación debido a que en el valor mínimo la región de medición es abierta, sin densidad de vegetación y con apuntamiento directo de antenas, mientras que en el valor máximo, la región de medición posee obstáculos, una gran densidad de vegetación afectando de mayor manera con errores en la toma de mediciones.

- Se concluye que la obtención de las características físicas de apuntamiento de las antenas de la estación y la obtención de modelos mediante rangos permiten mejorar su precisión.
- Tanto la parte física como el modelo general elaborado durante este trabajo corresponden al rango definido de 150° a 220° aproximadamente, esto varía debido a varios factores como la topografía del terreno, su robustez de vegetación, irregularidades, clima y entre otros factores, permitiendo delimitar los rangos para la obtención de modelos y estimación del DoA.
- Se concluye que para la elaboración de los modelos por rangos el lóbulo de radiación de las antenas influye de gran manera concentrando la mayor cantidad de datos en su máximo alcance y disminuyendo el mismo conforme nos alejamos de la región hacia donde apuntan sus antenas correspondientes, afectando de manera similar a los parámetros de radiofrecuencia.
- Los modelos ajustados mediante rangos muestran una mejor precisión de las predicciones del DoA en comparación a enfoques generalizados, esto se evidencia a través de la disminución del RMSE y el aumento del R², indicando la existencia de una mayor correlación entre los valores reales y las estimaciones del DoA.
- La elección de atributos predictores y el correcto proceso de preprocesamiento de la información o datos son parte fundamental e influyen en gran parte en la generación de modelos eficientes, además, las matrices de correlación y los modelos ajustados a cada rango determinado del DoA permiten identificar y hacer uso de atributos predictores que influyen en gran manera a la estimación del DoA.
- El enfoque en el proceso de evaluación de los modelos planteados usando mediciones estadísticas como el RMSE y el R² facilita en gran manera la comparación directa del rendimiento de un modelo, permitiendo identificar los modelos más precisos para cada rango del DoA, y además, facilitan la comprensión de la variabilidad entre atributos predictores y el DoA.

- El estudio correspondiente permitió demostrar el potencial existente en el uso del aprendizaje automático para la estimación del DoA, llegando al nivel de poder aplicar los modelos obtenidos en entornos prácticos para el mejoramiento de sistema de localización y posicionamiento, obteniendo una gran capacidad en la estimación del DoA con una alta precisión enfocado en el desarrollo de tecnologías avanzadas.

3.3 RECOMENDACIONES

- Se recomienda la adopción de técnicas de mayor complejidad para la estimación del DOA en la planificación y optimización de redes celulares para mejorar la cobertura y la calidad del servicio.
- En futuros trabajos se debería explorar la existencia de posibles modelos mediante el uso de regresiones logarítmicas o polinómicas incluyendo la aplicación de nuevas técnicas de Machine Learning para abordar las limitaciones encontradas en este estudio.
- Se sugiere realizar el desarrollo de aplicaciones de software que permitan incorporar los modelos obtenidos en el presente estudio para facilitar la implementación y uso en entornos de campo prácticos.
- Se recomienda la inserción de nuevos métodos estadísticos con mayor complejidad para la evaluación del rendimiento de los modelos para técnicas de Machine Learning, permitiendo optimizar de mejor manera cada modelo planteado asegurando que exista una gran precisión en la predicción del DoA.

4 REFERENCIAS BIBLIOGRÁFICAS

- [1] A. Dhabale, "Direction Of Arrival (DOA) Estimation Using Array Signal Processing", University of California Riverside, 2018. [En línea]. Disponible en: https://escholarship.org/content/qt1w13p71p/qt1w13p71p_noSplash_9f69e592ae9b76ba11f4349daa613333.pdf
- [2] T. S. Rappaport, *Wireless communications: principles and practice*, 2nd ed. en Prentice Hall communications engineering and emerging technologies series. Upper Saddle River, N.J: Prentice Hall PTR, 2002. [En línea]. Disponible en: <https://telkom2013.files.wordpress.com/2014/02/wireless-comm-princip-n-practice-theodoresrappaport.pdf>
- [3] D. Tse y P. Viswanath, *Fundamentals of Wireless Communication*, 1a ed. Cambridge University Press, 2005. doi: 10.1017/CBO9780511807213.

- [4] A. A. Abdelhak, "DIRECTION OF ARRIVAL ESTIMATION OF WIDEBAND RF SOURCES", Near East university, Nicosia, 2018. [En línea]. Disponible en: <http://docs.neu.edu.tr/library/6718703723.pdf>
- [5] F. Bravo y P. Badilla, "Introducción a Machine Learning", p. 62.
- [6] *Aprendizaje Machine-Learning*. [En línea]. Disponible en: <https://riptutorial.com/Download/machine-learning-es.pdf>
- [7] A. Cabrera, "EVALUACIÓN Y COMPARACIÓN DE LAS TÉCNICAS DE POSICIONAMIENTO EN 1D Y 2D BASADAS EN ALGORITMOS PARA LA RESOLUCIÓN DE LA DIRECCIÓN DE ARRIBO (DOA)", Escuela Superior Politécnica de Chimborazo, Riobamba, 2019. [En línea]. Disponible en: <http://dspace.esepoch.edu.ec/bitstream/123456789/10962/1/98T00234.pdf#page=23&zoo=100,129,228>
- [8] F. A. Sarría, "Sistemas de Información Geográfica", [En línea]. Disponible en: <https://www.um.es/geograf/sigmur/sigpdf/temario.pdf>
- [9] M. Montes de Oca, *Topografía*, Cuarta edición. México, DF: Alfa Omega, 1989. [En línea]. Disponible en: https://www.academia.edu/19198116/Topografia_Miguel_Montes_de_Oca
- [10] "Coordenadas geográficas", *Wikipedia, la enciclopedia libre*. el 17 de enero de 2024. Consultado: el 26 de enero de 2024. [En línea]. Disponible en: https://es.wikipedia.org/w/index.php?title=Coordenadas_geogr%C3%A1ficas&oldid=157298979
- [11] M. Skolnik, *Introduction to radar systems*, Tercera edición. Tata McGraw-Hill. [En línea]. Disponible en: <https://picture.iczhiku.com/resource/eetop/wHIGAzagLdPtpXcM.pdf>
- [12] M. Skolnik, *Introduction to radar systems*, Segunda edición. Tata McGraw-Hill, 1981. [En línea]. Disponible en: <https://soaneemrana.org/onewebmedia/INTRODUCTION%20TO%20RADAR%20SYSTEM%20BY%20MERRIL,%20I%20SKLOINK%20%284%29.pdf>
- [13] A. D. Pierce, *Acoustics: an introduction to its physical principles and applications*. en McGraw-Hill series in mechanical engineering. New York, NY: McGraw-Hill, 1981. [En línea]. Disponible en: <https://jontallen.ece.illinois.edu/Public/BOOKS/Pierce-Acoustics.81.pdf>
- [14] V. Rivera, "Estimación de la Dirección de Arribo y de los Coeficientes Óptimos en un Sistema de Antenas Inteligentes Usando Redes Neuronales", Instituto Tecnológico y de Estudios Superiores de Monterrey, Monterrey, 2000. [En línea]. Disponible en: https://repositorio.tec.mx/bitstream/handle/11285/572128/Rivera%20Arzola_TesisPDF_A.pdf?sequence=12&isAllowed=y
- [15] E. Dahlman, S. Parkvall, y J. Sköld, *4G LTE/LTE-Advanced for Mobile Broadband*. Elsevier, 2011. [En línea]. Disponible en: https://www.academia.edu/38394302/_ebook_4G_LTE_LTE_Advanced_for_Mobile_Broadband_pdf
- [16] J. Analuisa, "DISEÑO DE UNA RED 4G LONG TERM EVOLUTION (LTE) EN REDES MÓVILES", Universidad Técnica de Ambato, Ambato, 2014. [En línea]. Disponible en: https://repositorio.uta.edu.ec/bitstream/123456789/7796/1/Tesis_t905ec.pdf
- [17] Arcotel, "Boletín estadístico", Agencia de Regulación y Control de las Telecomunicaciones, Ecuador, 6, 2018. [En línea]. Disponible en: https://www.arcotel.gob.ec/wp-content/uploads/2015/01/BOLETIN-ESTADISTICO-Diciembre-2018-v4_4.pdf
- [18] R. Agüero y L. Muñoz, "Dimensionado de sistemas celulares", [En línea]. Disponible en: <https://www.tlmat.unican.es/siteadmin/submaterials/658.pdf>
- [19] "Ángulo acimut | PVEducation". Consultado: el 27 de enero de 2024. [En línea]. Disponible en: <https://www.pveducation.org/es/fotovoltaica/2-propiedades-de-la-luz-del-sol/%C3%A1ngulo-acimut>
- [20] "Machine Learning - Libro online de IAAR". Consultado: el 28 de enero de 2024. [En línea]. Disponible en: <https://iaarbook.github.io/ML/>

- [21] A. Páez, “Estudio de la calidad de la conexión en redes de telefonía celular en base a mediciones de campo y técnicas de Machine Learning”, Escuela Politécnica Nacional, Quito, 2023. [En línea]. Disponible en: <https://bibdigital.epn.edu.ec/bitstream/15000/25188/1/CD%2013807.pdf>
- [22] R Development Core Team, “Introducción a R”. [En línea]. Disponible en: <https://cran.r-project.org/doc/contrib/R-intro-1.1.0-espanol.1.pdf>
- [23] H. Mehta, P. Kanani, y P. Lande, “Google Maps”, *Int. J. Comput. Appl.*, vol. 178, pp. 41–46, may 2019, doi: 10.5120/ijca2019918791.
- [24] D. Hari, “Estimation of Rooftop Rainwater Harvesting Potential using Applications of Google Earth Pro and GIS”, *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, pp. 2278–3075, jul. 2019, doi: 10.35940/ijitee.I7853.078919.
- [25] J. López, “Wikiloc : Software libre y APIs de Google Maps para visualizar y compartir rutas GPS”, p. 6.
- [26] V. P. Villamar, P. Lupera, y F. Grijalva, “Análisis estadístico general de mediciones de parámetros de la red celular en Quito”, *Rev. Investig. En Tecnol. Inf.*, vol. 10, núm. 21, pp. 120–131, 2022.
- [27] S. Hemani y M. Oussalah, “Mobile Location System Using Netmonitor and MapPoint server”.
- [28] B. D. Nelms *et al.*, “CellMapper: rapid and accurate inference of gene expression in difficult-to-isolate cell types”, *Genome Biol.*, vol. 17, núm. 1, p. 201, sep. 2016, doi: 10.1186/s13059-016-1062-5.
- [29] A. Diaz, “Calculadora Geoespacial”, *Concurso Trab. Estud.*, p. 11, 2012.
- [30] “Dirección Distancia calcular entre dos espacios en el planeta directamente en el mapa”. Consultado: el 29 de enero de 2024. [En línea]. Disponible en: <https://www.sunearthtools.com/es/tools/distance.php#top>
- [31] “RPubs - Correlación y Regresión”. Consultado: el 11 de febrero de 2024. [En línea]. Disponible en: <https://rpubs.com/osoramirez/316691>
- [32] M. Andina, *Introducción a estadística con R*. Consultado: el 11 de febrero de 2024. [En línea]. Disponible en: <http://matiasandina.netlify.com/>
- [33] “RPubs - Regresión Simple”. Consultado: el 11 de febrero de 2024. [En línea]. Disponible en: <https://rpubs.com/paraneda/regresion>