

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA

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

**ANÁLISIS MEDIANTE TÉCNICAS DE MACHINE LEARNING DE
PARÁMETROS TÉCNICOS DE LA RED CELULAR LTE PARA
DETERMINAR PATRONES DE COMPORTAMIENTO EN FUNCIÓN
DE LAS HORAS DEL DÍA EN UN SECTOR DE LA CIUDAD DE
QUITO**

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

GYNA MARIBEL POZO ANDRADE

gyna.pozo@epn.edu.ec

DIRECTOR: PhD. PABLO ANIBAL LUPERA MORILLO

pablo.lupera@epn.edu.ec

DMQ, Febrero 2022

CERTIFICACIONES

Yo, GYNA MARIBEL POZO ANDRADE 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.

GYNA MARIBEL POZO ANDRADE

Certifico que el presente trabajo de integración curricular fue desarrollado por GYNA MARIBEL POZO ANDRADE, bajo mi supervisión.

PhD. Pablo Lupera Morillo
DIRECTOR

DECLARACIÓN DE AUTORÍA

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

GYNA MARIBEL POZO ANDRADE

PhD. PABLO LUPERA MORILLO

DEDICATORIA

El presente trabajo de integración curricular le dedico a la memoria de mis abuelitos que han sido el más grande motivo para llevar a cabo la realización de cada una de las etapas de mi desarrollo tanto personal como profesional.

AGRADECIMIENTO

Mi principal agradecimiento es para Dios que me ha bendecido con la salud y la vida para iniciar y ahora estar culminando esta etapa de mi vida.

Le agradezco a mi mamita linda Gyna Andrade, que sin importar limitaciones ni obstáculos, me ha dado la gran oportunidad de desarrollarme dentro de esta carrera universitaria, que como ella me dice siempre es la herramienta para iniciar el camino profesional, le agradezco por su compañía en este camino, por su apoyo, y su inmenso amor con el que me ha fortalecido día a día durante cada semestre y durante el desarrollo del presente trabajo.

A mis hermanos, Napito y Mile que me han acompañado con mucho cariño durante días, noches y madrugadas enteras, hasta terminar cada uno de los proyectos y trabajos que he realizado a lo largo de esta carrera universitaria.

A mi novio Ricardo que me ha acompañado, me ha apoyado incondicionalmente y que con mucha alegría y amor celebra cada uno de mis pequeños y grandes logros cada día.

A mis tías y al resto de mi familia que han estado al pendiente de cada avance que compone el desarrollo del presente trabajo, en especial a mi tía Maribel y a mis primas bellas Brendita, Andreita y Gaby que desde el inicio hasta el final de la carrera me han apoyado.

Expreso también mi más sincero agradecimiento al Dr. Pablo Lupera, que además de ser mi tutor, ha sido mi guía fundamental para llevar a cabo cada una de las etapas que han comprendido este trabajo.

Finalmente, agradezco a mis compañeros que formaron parte del desarrollo de proyecto general, ya que en esta etapa ha sido donde más he comprendido el valor del apoyo entre todos, en especial le agradezco profundamente a Michael que aparte de compañero le considero como un gran amigo y a Cristian que con su ayuda y consejos he avanzado más rápido de lo que esperaba para culminar con éxito este trabajo.

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.....	3
1.4.1 HERRAMIENTAS PARA LA RECOLECCIÓN DE LOS DATOS	3
1.4.1.1 NetMonitor.....	3
1.4.1.2. NetMonitor Cell Signal Logging	5
1.4.1.3 SetEdit	7
1.4.2 TERMINAL MÓVIL	10
1.4.3 HERRAMIENTAS PARA EL ANÁLISIS DE LOS DATOS	11
1.4.3.1 Definición del software R	11
1.4.3.2 Características de R.....	11
1.4.3.3 Ventajas del uso del Software R.....	12
1.4.3.4 Definición de Software RStudio	12
1.4.3.5 Secciones de la página principal de RStudio [8].....	12
1.4.3.6 Características de RStudio:.....	13
2 METODOLOGÍA.....	13
2.1 ESTABLECIMIENTO DE LAS CONDICIONES DE MEDICIÓN.....	13
2.1.1 OPERADORA.....	13

2.1.2	TIPO DE RED	14
2.1.3	ESTADO DE FUNCIONAMIENTO DEL TERMINAL	14
2.2	DEFINICIÓN DE LOS DATOS A RECOLECTAR Y PROCESAR	14
2.2.1	RECEIVED SIGNAL STRENGTH INDICATOR (RSSI)	14
2.2.2	REFERENCE SIGNAL RECEIVED QUALITY (RSRQ)	14
2.2.3	HORAS DEL DÍA	14
2.2.4	DÍA DE LA SEMANA	15
2.3	DEFINICIÓN DE LA ZONA A CONSIDERAR	15
2.4	PLANIFICACIÓN DE LA RECOLECCIÓN DE DATOS	17
2.5	PROCESO PARA EL ANÁLISIS DE LOS DATOS RECOLECTADOS ...	18
2.5.1	FASES PARA EL ANÁLISIS DE DATOS:	18
2.5.2	TÉCNICA DE CLASIFICACIÓN DE MACHINE LEARNING	18
2.5.3	PARÁMETROS SELECCIONADOS COMO VARIABLES DE ENTRADA.	18
2.5.4	PARÁMETRO SELECCIONADO COMO VARIABLE DE SALIDA	19
2.5.5	PROCESO DE DEPURACIÓN DE DATOS	19
2.5.6	RANGOS DE CATEGORIZACIÓN DE LA VARIABLE DE SALIDA	20
2.5.7	DESCRIPCIÓN DE COMANDOS UTILIZADOS EN EL PROCESO DEL ANÁLISIS DE DATOS	20
2.5.8	RESULTADOS OBTENIDOS	25
2.5.8.1	MATRIZ FINAL	25
2.5.8.2	Variables de la matriz "final"	25
2.5.8.3	Resumen estadístico de RSSI y RSRQ en semana 1	26
2.5.8.4	Resumen estadístico de RSSI y RSRQ en semana 2	26
2.5.8.5	Resumen estadístico de RSSI y RSRQ en semana 3	26
2.5.8.6	Resumen estadístico de RSSI y RSRQ en semana 4	26
3	ANÁLISIS DE RESULTADOS, CONCLUSIONES Y RECOMENDACIONES	27
3.1	RESULTADOS	27
3.1.1	ANÁLISIS GENERAL DE DATOS RECOLECTADOS:	27
3.1.1.1	Código general desarrollado en el proceso de análisis de datos	27
3.1.1.2	Árbol de decisión General	28
3.1.1.3	Precisión del modelo obtenido	29
3.1.1.4	Matriz de confusión	29
3.1.1.5	Resumen de reglas importantes	30
3.1.2	ANÁLISIS DE DATOS RECOLECTADOS EN LA SEMANA 1	30
3.1.2.1	Árbol de decisión de la semana 1	30

3.1.2.2	Precisión del modelo obtenido de la semana 1	31
3.1.2.3	Matriz de confusión de semana 1	31
3.1.2.4	Resumen estadístico de RSSI y RSRQ de la semana 1	32
3.1.2.5	Histograma de la semana 1	32
3.1.3	ANÁLISIS DE DATOS RECOLECTADOS EN LA SEMANA 2	33
3.1.3.1	Árbol de decisión de la semana 2.....	33
3.1.3.2	Precisión del modelo obtenido de la semana 2	34
3.1.3.3	Matriz de confusión de la semana 2	34
3.1.3.4	Resumen estadístico de RSSI y RSRQ de la semana 2	35
3.1.3.5	Histograma de la semana 2.....	35
3.1.4	ANÁLISIS DE DATOS RECOLECTADOS EN LA SEMANA 3	36
3.1.4.1	Árbol de decisión de la semana 3.....	36
3.1.4.2	Precisión del modelo obtenido de la semana 3	37
3.1.4.3	Matriz de confusión de la semana 3	38
3.1.4.4	Resumen estadístico de RSSI y RSRQ de la semana 3	38
3.1.4.5	Histograma de la semana 3.....	38
3.1.5	ANÁLISIS DE DATOS RECOLECTADOS EN LA SEMANA 4	39
3.1.5.1	Árbol de decisión de la semana 4.....	39
3.1.5.2	Precisión del modelo obtenido de la semana 4	40
3.1.5.3	Matriz de confusión de semana 4	40
3.1.5.4	Resumen estadístico de RSSI y RSRQ de la semana 4	41
3.1.5.5	Histograma de la semana 4:.....	41
3.2	CONCLUSIONES.....	42
3.3	RECOMENDACIONES	44
4	REFERENCIAS BIBLIOGRÁFICAS	46
5	ANEXOS.....	48

RESUMEN

En el presente estudio se propone un análisis de datos recolectados en un sector urbano de la ciudad de Quito para obtener tendencias de los parámetros técnicos variables de radiofrecuencia (RF) de las redes celulares en función de las horas del día. Para llevar a cabo el desarrollo de este estudio se han definido tres etapas principales, las cuales se describen a continuación; la primera es la toma de mediciones durante cuatro semanas en cuatro puntos ubicados a distancias determinadas de la estación base en el sector mencionado, la segunda fase se refiere al análisis de los datos mediante una técnica de clasificación de Machine Learning con el uso de las herramientas R y Rstudio y en la tercera fase se detectan y describen las relaciones obtenidas a través del análisis de los parámetros técnicos para la obtención de un modelo que permita establecer el comportamiento de la red en el transcurso del tiempo. Dichas fases se han desarrollado en tres capítulos.

En el primer capítulo se considera el marco teórico, donde se definen los conceptos, características y formas de uso de las aplicaciones para la toma de mediciones y para el análisis de datos. Además, se describen las características técnicas principales del terminal móvil utilizado para la toma de mediciones.

En el segundo capítulo se presenta la metodología que se ha empleado para llevar a cabo el desarrollo de cada una de las fases mencionadas anteriormente. Inicialmente se definen los parámetros y se establecen las condiciones en las que se va a realizar la toma de mediciones, después se define la zona de estudio y se presenta la planificación de la recolección de datos, también se describe la técnica de clasificación de Machine Learning y se definen las variables de entrada y de salida, y finalmente se describen los comandos utilizados en el proceso del análisis de datos y se muestran los resultados de la clasificación.

Por último, en el tercer capítulo se presentan los resultados a partir del modelo obtenido, las conclusiones y las recomendaciones.

PALABRAS CLAVE: Machine Learning, LTE, técnica de clasificación, variación temporal, parámetros técnicos.

ABSTRACT

In the present study, an analysis of data collected in an urban sector of the city of Quito is proposed to obtain trends of the variable technical parameters of radio frequency (RF) of cellular networks depending on the hours of the day. To carry out the development of this study, three main stages have been defined, which are described below; the first is the taking of measurements for four weeks at four points located at certain distances from the base station in the aforementioned sector, the second phase refers to the analysis of the data through a classification technique of Machine Learning with the use of the tools R and Rstudio and in the third phase the relationships obtained through the analysis of the technical parameters are detected and described to obtain a model that allows establishing the behavior of the network over time. These phases have been developed in three chapters.

In the first chapter, the theoretical framework is considered, where the concepts, characteristics and forms of use of the applications for taking measurements and for data analysis are defined. In addition, the main technical characteristics of the mobile terminal used for taking measurements are described.

The second chapter presents the methodology that has been used to carry out the development of each of the phases mentioned above. Initially, the parameters are defined and the conditions under which the measurements will be taken are established, then the study area is defined and the data collection planning is presented, the Machine Learning classification technique is also described and the input and output variables are defined, and finally the commands used in the data analysis process are described and the results of the classification are shown.

Finally, the third chapter presents the results from the model obtained, the conclusions and recommendations.

KEYWORDS: Machine Learning, LTE, classification technique, temporal variation, technical parameters.

1 INTRODUCCIÓN

Actualmente las redes celulares LTE se encuentran en constante desarrollo e implementación en la mayoría de países, ya que presentan características mejoradas con la finalidad de brindar un servicio de calidad al acceder a la red, donde las principales se refieren a la mayor cobertura y velocidades de uplink y downlink que ofrecen, sin embargo, existen diferencias en el funcionamiento correcto de estas redes para garantizar adecuados niveles de calidad de servicio, esto debido a varios factores, entre los cuales se puede mencionar el impacto que tiene la variación que presentan ciertos parámetros en la percepción de la calidad de la comunicación celular.

Por otra parte, las técnicas de Machine Learning están teniendo gran impacto en la actualidad, ya que gracias a sus características permiten determinar modelos predictivos a partir de reglas obtenidas en base a las relaciones encontradas entre las variables de entrada y de salida del proceso de aprendizaje.

Teniendo en cuenta las ventajas anteriormente mencionadas, se puede hacer uso del Machine Learning con el objetivo de obtener un modelo, que sea apto para funcionar con nuevos conjuntos de datos y de esta forma presentar un modelo útil que sirva para el monitoreo del funcionamiento de la red celular LTE. En este trabajo se presenta una propuesta de análisis de datos recolectados en un sector determinado de la ciudad de Quito, con la finalidad de encontrar patrones de comportamiento que relacionen parámetros técnicos tales como: RSSI y RSRQ en función de las horas del día usando aplicaciones móviles y software de análisis de datos.

1.1 OBJETIVO GENERAL

Analizar los datos recolectados de parámetros técnicos de la red celular LTE durante cuatro semanas en un sector de la ciudad de Quito y determinar patrones de comportamiento en función de las horas del día.

1.2 OBJETIVOS ESPECÍFICOS

1. Determinar aplicaciones móviles adecuadas a las capacidades del terminal móvil para la medición de los datos.

2. Seleccionar diferentes ubicaciones para el registro de datos, tomando en cuenta que entre cada punto exista una distancia considerable para tener un registro variado de datos.
3. Diseñar un cronograma detallando las actividades que se realizarán durante la etapa de recolección de datos, de forma que puedan ser organizadas, de acuerdo al tiempo establecido.
4. Definir la técnica de Machine Learning según el tipo de datos que se van a analizar.
5. Analizar los datos de los parámetros que varían en función de las horas del día y buscar posibles relaciones entre sí, para obtener tendencias que aporten con información útil y actualizada con respecto al funcionamiento de la red celular LTE en la ciudad de Quito.

1.3 ALCANCE

En el presente trabajo se propone realizar el análisis de datos de parámetros técnicos variables de la red celular LTE para determinar patrones de comportamiento en función de las horas del día [11]. Para llevar a cabo este análisis, inicialmente, se aborda una etapa de recolección de datos de parámetros técnicos de la red antes mencionada, definiendo como ambiente de medición el sector Monjas – urbanización Las Orquídeas de la ciudad de Quito, después se plantea una fase en la que se van a observar y definir los parámetros que presenten mayor variación y ciertas relaciones entre sí de acuerdo a los datos obtenidos. A continuación, se va a ejecutar una etapa de depuración que consiste en detectar datos atípicos y corregir errores. En la siguiente fase se va a realizar el análisis de los datos mediante el uso de una herramienta de simulación, y finalmente, se considera ejecutar una fase para determinar patrones de comportamiento o tendencias de acuerdo a los resultados obtenidos de las relaciones de los parámetros técnicos en función de las horas del día.

Se planifica que la fase de recolección de datos de parámetros técnicos de la red celular LTE, sea desarrollada durante cuatro semanas, en cuatro puntos diferentes, en cada uno de los cuales se realizarán mediciones durante una semana. Estas mediciones se realizarán en la zona perteneciente al sector Centro-Sur de la ciudad de Quito, ubicada en la parroquia Puengasí, urbanización Las Orquídeas, con el uso de la aplicación NetMonitor Cell Signal Logging [3], instalada en un teléfono móvil Huawei Y9 Prime (2019) [6], el mismo que se va a mantener estático, sin la conexión de ningún tipo de servicio y con la

configuración necesaria para funcionar en modo de red LTE por medio de la aplicación SetEdit [4]. Posteriormente, con las herramientas R y Rstudio se aplicará la técnica de clasificación de reglas de Machine Learning, para realizar el análisis de aquellos datos de los parámetros técnicos que presenten mayor variación de acuerdo con las 24 horas del día y de esta forma obtener un modelo predictivo ilustrado en un árbol de decisiones, para finalmente, determinar patrones de comportamiento o tendencias actuales de la red LTE, que sean útiles para futuros estudios relacionados directa o indirectamente con el volumen de tráfico de las redes celulares en esta zona de la ciudad de Quito. Este estudio también será útil para las etapas de monitoreo del funcionamiento de este tipo de redes [8], [9].

1.4 MARCO TEÓRICO

1.4.1 HERRAMIENTAS PARA LA RECOLECCIÓN DE LOS DATOS

1.4.1.1 NetMonitor

Se tienen varias aplicaciones útiles para el registro de parámetros técnicos de las radiobases, por ejemplo, una muy útil y completa es NetMonitor, que permite conocer varios detalles de la conexión como; intensidad de las señales de las antenas que se encuentran disponibles alrededor, así como también, permite identificar la ubicación concreta de las mismas y la muestra en el mapa. Además, NetMonitor es una herramienta gratuita y se la puede descargar, instalar y usar en cualquier sistema operativo Android [1].

El proceso de instalación y uso de la herramienta consiste en:

- Se realiza la búsqueda e instalación de la aplicación, ingresando a Play Store desde cualquier dispositivo con sistema operativo Android.
- Se abre la aplicación de acuerdo a lo que se muestra en la Figura 1.1, se aceptan las condiciones y permisos de seguridad y posteriormente se observa una lista de información de las señales de las antenas de las redes de comunicación móvil disponibles. La lista de antenas disponibles también permite conocer el operador de la red a la que se encuentra conectado el dispositivo, el tipo de tecnología que se está usando (3G, GPRS, LTE, etc), la señal que se recibe de cada una de ellas, la ubicación y un grupo de números de identificación. Como por ejemplo, en la página inicial de la herramienta se observa que el teléfono móvil usado está conectado a una antena Claro del operador CONECEL, se identifica que la antena

pertenece a dicha empresa, ya que al observar en la parte de “operador”, se puede visualizar el número 21401, correspondiente a la red de Claro.



Figura 1.1. Página principal de NetMonitor.

- Se selecciona el segundo ícono en la parte inferior de la ventana principal como se puede ver en la Figura 1.2 y es posible observar una lista en la que se puede editar el detalle de las diferentes antenas para identificarlas en el dispositivo con mayor facilidad.



Figura 1.2. Detalles de antenas cercanas identificadas.

- Al seleccionar el tercer ícono de la parte inferior de la pantalla principal, se mostrará la ubicación en el mapa geográfico de todas las antenas al alcance del dispositivo, de tal forma que permite observar gráficamente dichas antenas, ver Figura 1.3.

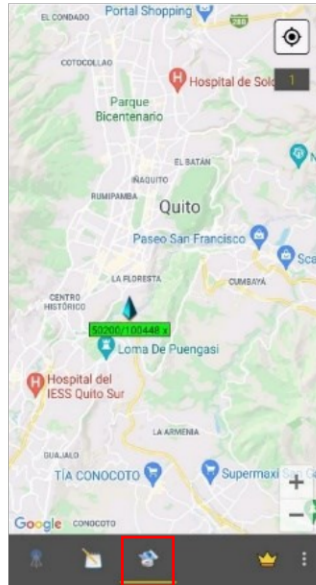


Figura 1.3. Ubicación geográfica de radiobases cercanas.

1.4.1.2. NetMonitor Cell Signal Logging

Es una aplicación gratuita disponible para teléfonos inteligentes, que es útil para monitorear parámetros técnicos presentes en las redes celulares. Entre sus funcionalidades permite obtener datos acerca del tipo de tecnología que está usando el terminal móvil (GSM, WCDMA, LTE), datos acerca de la intensidad de la señal, así como información detallada acerca de la operadora a la que se conecta el terminal móvil y datos de parámetros importantes para el análisis de redes celulares tales como RSSI, RSRQ y RSSNR [3].

Net Monitor Lite es una herramienta muy importante para el desarrollo del presente trabajo, ya que se adapta a la necesidad de estar en funcionamiento, tanto con datos móviles y con WiFi desactivado, que es lo que inicialmente se planteó como condición para iniciar con la etapa de recolección de datos.

Adicionalmente, como otra de las ventajas que NetMonitor Cell Signal Logging proporciona, es que permite la exportación de los datos registrados en formato .csv que es ideal para la posterior importación y análisis de los datos que se realizará con el uso de las herramientas R y Rstudio.

Los pasos para instalación y uso básico de NetMonitor Cell Signal Logging son:

- Para la instalación de Net Monitor Lite, se accede a Play Store y se realiza su descarga.
- Se abre la aplicación y en la parte superior derecha de la pantalla principal se encuentra los 3 íconos principales que se describe a continuación en la Figura 1.6.

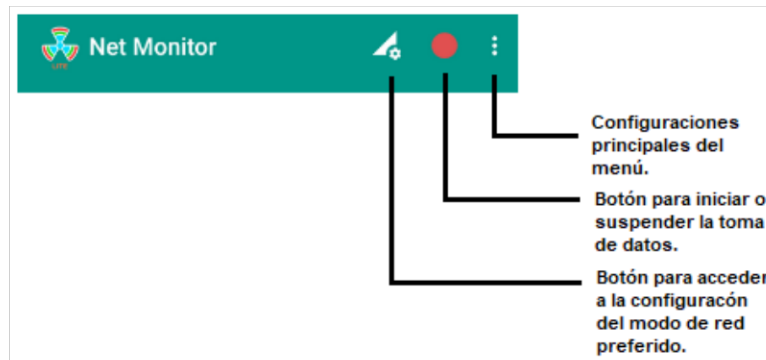


Figura 1.6. Funcionalidades principales de Net Monitor Lite.

- Bajo de las funcionalidades principales se encuentra la barra de pestañas con cada una de las secciones que brindan la información de cada parámetro técnico considerado como se puede ver en la Figura 1.7.

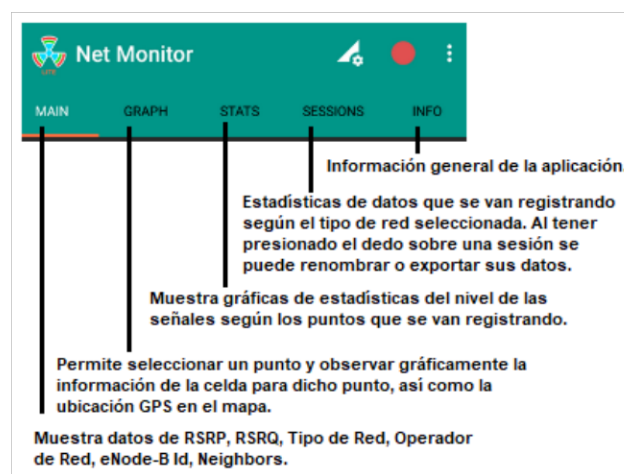


Figura 1.7. Pestañas con secciones que brinda la aplicación.

- Adicional, en la parte superior derecha de la página del menú principal también se encuentra cuatro íconos que se relacionan con las configuraciones previas del terminal móvil, como se puede ver en la Figura 1.8.

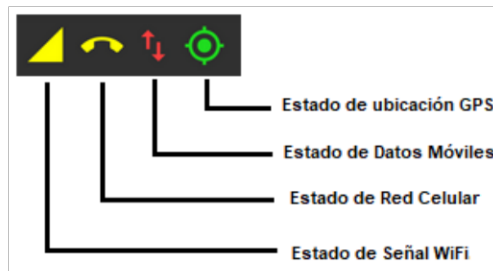


Figura 1.8. Íconos de Configuraciones en Menú Principal.

Para realizar las mediciones en el presente trabajo inicialmente, es importante desactivar los datos móviles y WiFi, activar el GPS y dar clic en el botón para iniciar la toma de datos. Al concluir con la recolección, se selecciona el botón que permite suspender el registro de dichos datos, y finalmente, para su exportación se selecciona la pestaña SESSIONS, donde se encuentran todos los registros de datos, manteniendo presionado el registro que se desea exportar para que se despliegue una lista, cuyas opciones permiten cambiar el nombre del registro de datos y seleccionar el formato en el que se desea exportar la sesión, como se puede ver en la Figura 1.9.

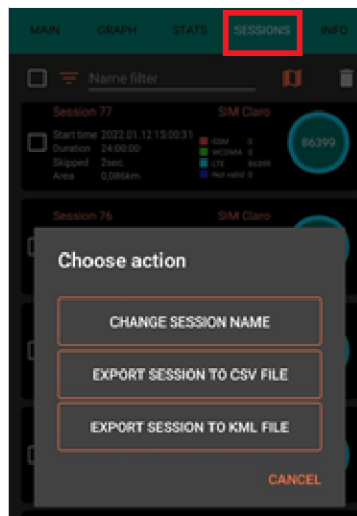


Figura 1.9. Íconos para cambio de nombre del registro y exportación de datos.

1.4.1.3 SetEdit

Es una aplicación que sirve para forzar al teléfono a una conexión única con la red LTE, funciona perfectamente en teléfonos móviles de la marca Huawei, lo que no sucede con aplicaciones como ForceLTE y 4 LTE Forzar solo 4G. Para su instalación se realiza la descarga de dicha aplicación en Play Store y se procede de forma fácil a abrir y realizar su correspondiente configuración, como se muestra a continuación [4]:

- En la página principal de la aplicación se muestran todas las configuraciones del sistema del dispositivo, por lo cual se debe tener mucho cuidado al momento de manipular alguna característica, ver Figura 1.10.



Figura 1.10. Página principal de SetEdit.

- Para forzar la red LTE en el dispositivo móvil, es importante modificar la opción “hw_networkmode_preference”, colocando el número 11, de esta forma se aumentan opciones en la pestaña de “Modo de red preferido” en los ajustes del celular, dónde ya se encuentra una opción de LTE, ver Figura 1.11.

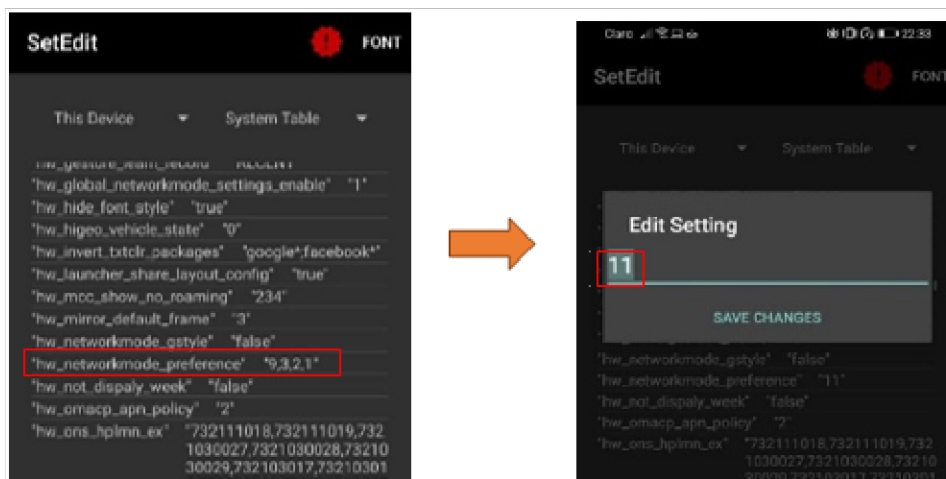


Figura 1.11. Modificación de valores en línea: “hw_networkmode_preference”.

- Posteriormente, se recomienda reiniciar el celular para que se actualicen los cambios realizados y a continuación se ingresa en Ajustes>Redes móviles>Datos móviles>Modo de red preferido y se observa que se encuentran disponibles más modos de red, es aquí donde se selecciona la opción LTE, ver Figura 1.12.

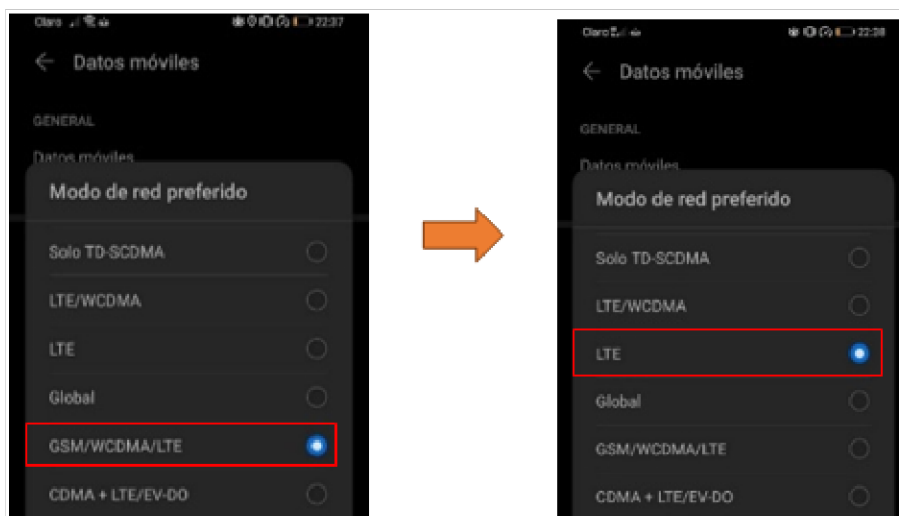


Figura 1.12. Selección de tipo de red LTE.

- Finalmente, se desactiva el WiFi en caso de estar activo y se activan los datos móviles, donde se observa que se ha forzado al teléfono móvil para conectarse a la red 4G, ver Figura 1.13.

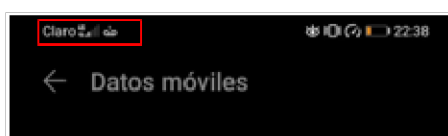


Figura 1.13. Funcionamiento de red 4G.

1.4.2 TERMINAL MÓVIL

Las características del terminal móvil que se utilizará para la recolección de los datos son las siguientes:

Tabla 1.1. Especificaciones Técnicas de Terminal Móvil [6].

HUAWEI Y9 Prime (2019)	CARACTERÍSTICAS GENERALES:
	<p>Released 2019 Modelo: STK-LX3 RAM: 4,0 GB Almacenamiento interno: 128 GB Resolución: 2340 x 1080 Versión Kernel: 4.14.116 Color: Verde esmeralda Dual SIM</p>
ESPECIFICACIONES:	
RED	<p>Tecnología: GSM/HSP/LTE Banda 2G: GSM 850 / 900 / 1800 / 1900 - SIM 1 & SIM 2 (Modelo doble SIM) Banda 3G: HSDPA 800 / 850 / 900 / 1700(AWS) / 1900 / 2100 Banda 4G: 1, 2, 3, 4, 5, 7, 8, 18, 19, 20, 26, 28, 38, 40, 41 Velocidad: HSPA 42.2/5.76 Mbps, LTE-A (2CA) Cat12 600/50 Mbps</p>
APARIENCIA	<p>Dimensiones: 163.5 x 77.3 x 8.9 mm (6.44 x 3.04 x 0.35 in) Peso: 196.8 g (6.95 oz) Diseño: Frente de vidrio, respaldo de plástico SIM: Simple SIM (Nano-SIM) o doble SIM (Nano-SIM, modo doble de espera)</p>
DISPLAY	<p>Tipo: IPS LCD Tamaño: 6.59 inches, 106.6 cm² (~84.3% relación pantalla - cuerpo) Resolución: 1080 x 2340 pixels, 19.5:9 proporción (~391 densidad ppi)</p>
HARDWARE	<p>Sistema Operativo: Android 10 Chipset: Kirin 710F (12 nm) CPU: Octa-core (4x2.2 GHz Cortex-A73 & 4x1.7 GHz Cortex-A53) GPU: Mali-G51 MP4</p>
MEMORIA	<p>Ranura para tarjetas: microSDXC (ranura dedicada) Interna: 128GB 4GB</p>
COMUNICACIONES	<p>WLAN: Wi-Fi 802.11 a/b/g/n/ac, doble-banda, Wi-Fi Directo, Punto de acceso Bluetooth: 5.0, A2DP, LE GPS: Sí, con A-GPS, GLONASS, GALILEO, BDS, QZSS NFC: No Radio: radio FM USB: USB Tipo-C 2.0, USB On-The-Go</p>
CARACTERÍSTICAS	<p>Sensores: Huella digital (montada en la parte trasera), acelerómetro, giroscopio, proximidad, brújula</p>
BATERÍA	<p>Tipo: Li-Po 4000 mAh, no removible</p>

1.4.3 HERRAMIENTAS PARA EL ANÁLISIS DE LOS DATOS

1.4.3.1 Definición del software R

R es un software de distribución libre que, por medio de su entorno de desarrollo integrado, brinda la oportunidad de realizar el análisis directo de grandes cantidades de datos y obtener resultados en base a gráficas, modelos estadísticos, de clasificación, de agrupamiento, entre otros [8].

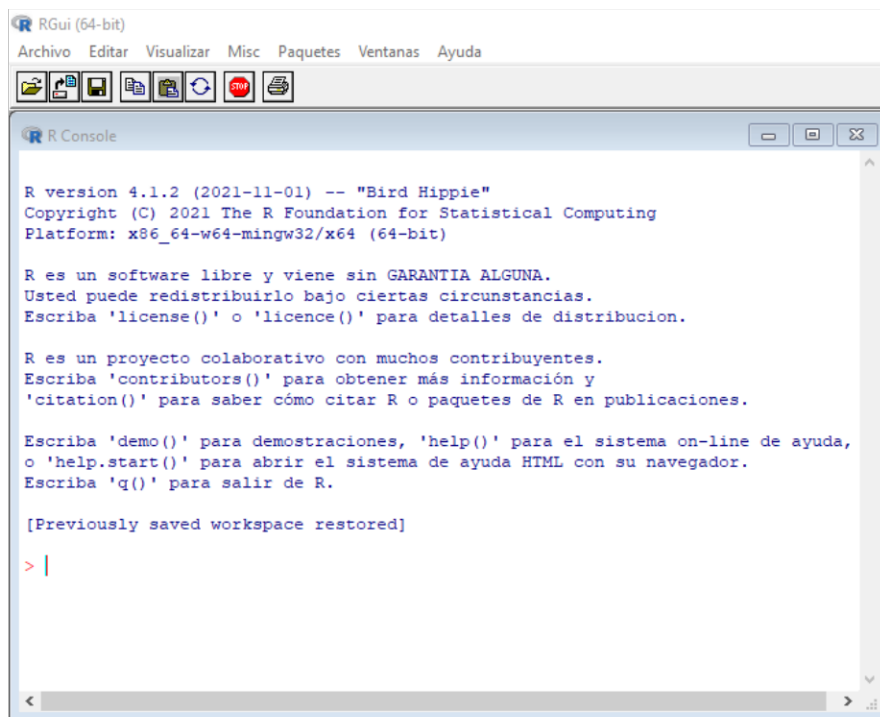


Figura 1.14. Pantalla inicial de software R.

1.4.3.2 Características de R

A continuación, se describen las principales características de R [9]:

- Software gratuito
- Permite la manipulación, configuración y almacenamiento de datos.
- Ejecuta cálculos con matrices empleando gran variedad de operadores.
- Contiene una lista de herramientas para realizar análisis de datos.
- Permite la visualización de resultados por medio de gráficas.
- Contiene comandos y funcionalidades que permiten un análisis avanzado de los datos.
- Permite la generación de documentación en formato físico y digital.

- Permite la importación de archivos con formato .csv.
- Es compatible con lenguajes de programación tales como; C, C++ o Fortran.

1.4.3.3 Ventajas del uso del Software R

A continuación, se describen las ventajas de R [8]:

- **Replicabilidad:** Permite la fácil comprensión y réplica exacta del análisis de los datos.
- **Eficiencia:** La sintaxis de R permite unir varias tareas en una misma línea de comandos, lo cual hace más eficiente al programa.
- **Control:** Permite hacer modificaciones de las bases de datos y detección de errores de forma rápida, ya que acepta la ejecución de las líneas del código de forma individual.

1.4.3.4 Definición de Software RStudio

Es una interfaz para el uso de R, dónde su estructura que se presenta dividida en 4 secciones principales permite interactuar y visualizar de forma ordenada cada uno de los procesos que se van llevando a cabo en el análisis de los datos de forma simultánea [8].

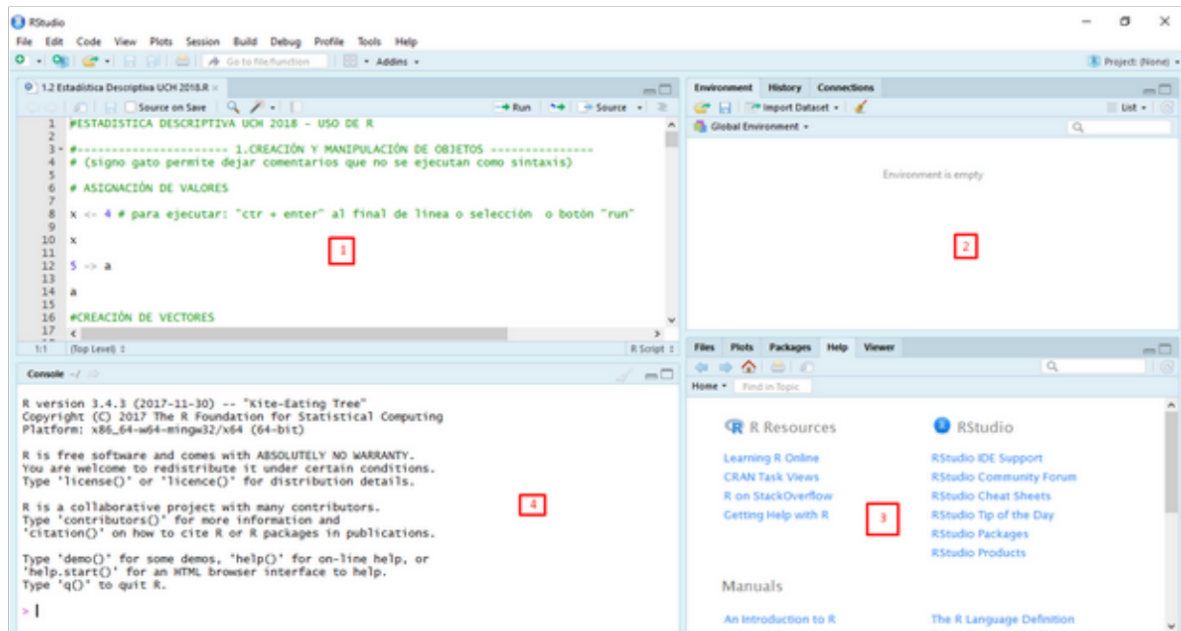


Figura 1.15. Pantalla principal de Software RStudio.

1.4.3.5 Secciones de la página principal de RStudio [8]

- **Ventana (1):** ventana para editar la sintaxis.

- **Ventana (2):** entorno de trabajo donde se muestra el conjunto de datos, resultados, variables, gráficos, etc.
- **Ventana (3):** tiene las siguientes opciones:
 - **files** permite ver el historial de archivos trabajados con el programa.
 - **plots** permite visualizar los gráficos que se generen.
 - **packages** permite ver los paquetes descargados y guardados en el disco duro, así como gestionar su instalación o actualización.
 - **help** permite acceder a la página oficial del software que ofrece manuales para el usuario, cursos on line, información general, descarga de paquetes, información de los paquetes instalados, etc.
 - **viewer** muestra los resultados al construir reportes mediante funcionalidades tipo *rmarkdown*.
- **Ventana (4):** es la consola o el software R en su versión básica, que permite ejecutar las operaciones realizadas desde el editor de sintaxis.

1.4.3.6 Características de RStudio:

- Permite abrir varios scripts simultáneamente
- Permite ejecutar cada línea del código de forma individual
- Muestra el workspace
- Muestra el historial
- Muestra los objetos del workspace
- Integra la ayuda
- Integra la gestión de librerías

2 METODOLOGÍA

2.1 ESTABLECIMIENTO DE LAS CONDICIONES DE MEDICIÓN

Las condiciones de medición son las siguientes:

2.1.1 OPERADORA

La red operadora a la que se conecta el terminal móvil para la recolección de datos es de una de las operadoras que brindan servicio en la ciudad. Se puede mencionar que en la zona de estudio las tecnologías vigentes son; GSM, 3G, HSPA +, 4G LTE y 4.5G [10].

2.1.2 TIPO DE RED

La red sobre la cual se realizarán las mediciones es una red de cuarta generación con tecnología LTE. Desde sus inicios en el año 2004, se desarrolló para mejorar el acceso de radio terrestre de UMTS y en la actualidad hace posible que los usuarios tengan acceso a un servicio de Internet móvil de alta velocidad. Entre las características mejoradas se encuentra el incremento de la velocidad, alcanzando velocidades teóricas de 150 Mbps en el enlace de downlink y 50 Mbps en uplink. Lo cual se reforzó con el desarrollo de la nueva versión de LTE denominada LTE-Advanced [11].

2.1.3 ESTADO DE FUNCIONAMIENTO DEL TERMINAL

Durante las mediciones el servicio de transmisión de datos del teléfono móvil se encuentra desactivado, ya que para el presente estudio no es necesario que el terminal se conecte a Internet. La recolección de datos mediante mediciones de los parámetros técnicos se realiza mientras no se está ejecutando la conexión de ningún tipo de servicio.

2.2 DEFINICIÓN DE LOS DATOS A RECOLECTAR Y PROCESAR

Los parámetros que se van a medir son los siguientes:

2.2.1 RECEIVED SIGNAL STRENGTH INDICATOR (RSSI)

Es un parámetro que se emplea en redes inalámbricas para medir la cantidad promedio de potencia de una señal recibida, se tiene como unidad de medida los dBm. Se define como un valor de mejor señal recibida cuando se tiene una medida de 0dBm, mientras que, al tener una medida de -100 dBm o menor, se considera un valor débil de señal recibida [12].

2.2.2 REFERENCE SIGNAL RECEIVED QUALITY (RSRQ)

Es un parámetro técnico que determina la calidad de una señal recibida, este valor es importante para definir las celdas disponibles en función de la calidad de la señal. En caso de ser necesario se usa la RSRQ para la selección de una nueva celda o en el caso que el nivel de Reference Signal Received Power (RSRP)¹ no sea suficiente para realizar procesos de handover [12].

2.2.3 HORAS DEL DÍA

Se considera realizar la recolección de datos durante las 24 horas del día, para lo cual se configura la aplicación NetMonitor Cell Signal Logging para que registre datos cada segundo.

¹ RSRP: Es el nivel de la potencia de la señal recibida desde la estación base, este parámetro se mide en dBm [11].

2.2.4 DÍA DE LA SEMANA

La recolección de los datos se realizará durante los 7 días de la semana en cada uno de los puntos que se describen más adelante. Este parámetro se tomará dependiendo de la fecha de registro de los datos y será incluido dentro de las variables que se van a analizar para observar su impacto en el comportamiento de la conexión.

2.3 DEFINICIÓN DE LA ZONA A CONSIDERAR

Se definió como el escenario de estudio para el presente trabajo una zona perteneciente al sector Centro-Sur de la ciudad de Quito, ubicada en la parroquia Puengasí, urbanización Orquídeas.

- Se determinó como punto de partida para iniciar la recopilación de datos el Punto 1, ubicado en la calle Thomas Roseau E15-63 y Gonzalo Escudero, cuyas coordenadas son; -0,22527523351481865, -78,48871821171500, 2763 m.
- Se continúa con la recolección de datos en el Punto 2, que se encuentra a 100 metros hacia el Oeste, a partir de la zona inicial (Punto 1), en el domicilio ubicado en la calle E15 Gonzalo Escudero S1-191 S1L y Thomas Roseau, cuyas coordenadas son; -0,225162, -78,4893476, 2766 m.
- El Punto 3 de recopilación de datos se encuentra a 200 metros hacia el Oeste a partir del Punto 1, en el domicilio ubicado en la calle Gonzalo Escudero y Thomas Roseau, cuyas coordenadas son; -0,22558081261775914, -78,49013372252930, 2785 m.
- El Punto 4 de recopilación de datos se encuentra a 400 metros hacia el Norte, a partir del Punto 1, en el domicilio ubicado en el Pasaje Rodrigo Jácome E14-481 y Gonzalo Escudero, cuyas coordenadas son; -0,22439379552111724, -78,48815743768200, 2721 m.

Es importante mencionar, que al ubicar el terminal móvil en cada uno de los cuatro puntos definidos, se observó por medio de la aplicación Netmonitor que el móvil se enlazó tanto a la radiobase eNodeB 100400 como a la radiobase eNodeB 100448, esto se puede observar en la Figura 2.1. La conexión a la estación base, depende de la ubicación de cada punto y de la saturación de la red.

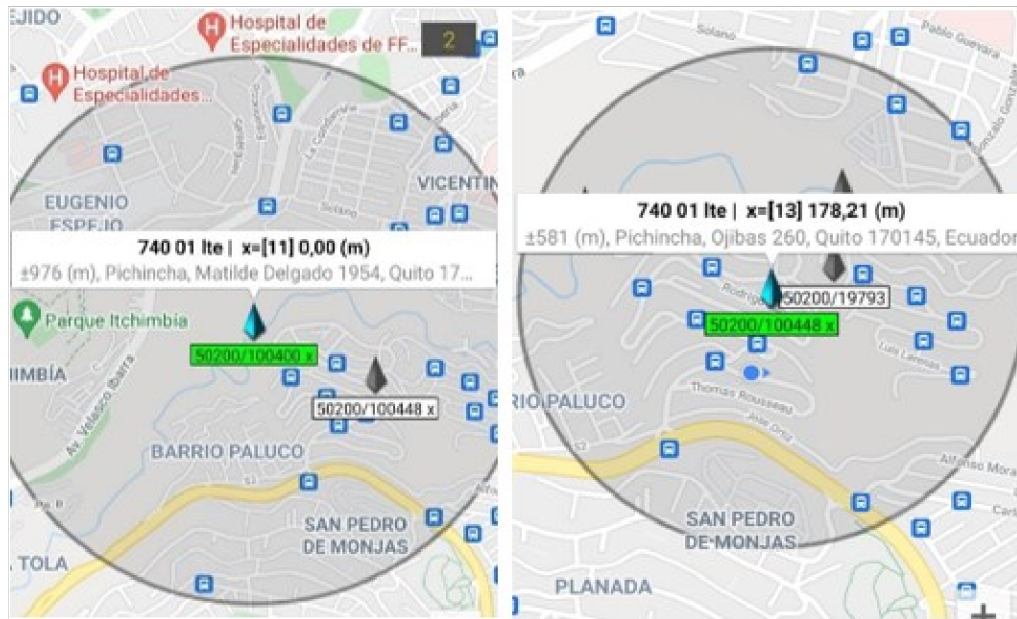


Figura 2.1. a) Radiobase eNodeB 100400. b) Radiobase eNodeB 100448.

Los cuatro puntos definidos para la recolección de datos se han seleccionado de forma en que se encuentren separados una distancia apreciable entre sí, con el objetivo de recolectar datos variables y útiles para el posterior análisis y evaluación. Por lo tanto, se decidió realizar la toma de datos durante los 7 días de la semana en cada uno de los diferentes puntos definidos que se pueden identificar en la Figura 2.2.

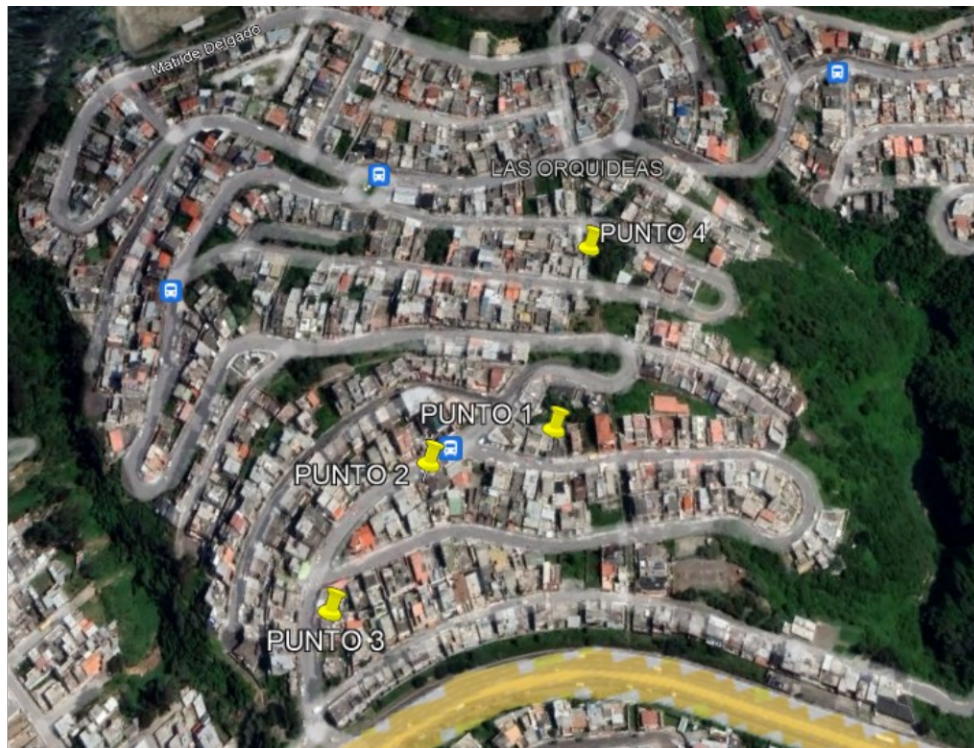


Figura 2.2. Escenario de estudio y ubicación de los puntos para recolección de datos.

2.4 PLANIFICACIÓN DE LA RECOLECCIÓN DE DATOS

Una vez que se definió la zona de interés para la recolección de datos, se identificaron los posibles inconvenientes que se pueden presentar para darles solución antes de proceder con la recopilación de información. En este caso después de definir las ubicaciones idóneas para llevar a cabo el proceso de recolección de información, se procedió a pedir los permisos necesarios a los diferentes propietarios de las viviendas en las que se planeó dejar el terminal móvil estático durante los 7 días en cada punto, debido a que el análisis de los datos debe relacionarse con las horas del día.

Se definió como fecha de inicio de la recolección de datos el 1 de diciembre del 2021 y fecha de finalización el 29 de diciembre del 2021, de forma que se cumplan cuatro semanas de recopilación de información en la zona seleccionada.

Posterior a la recolección de datos, se definió un periodo de tiempo para realizar depuración, detección de errores o datos atípicos y análisis de datos, de manera que se puedan obtener datos útiles que permitan determinar tendencias basadas en parámetros técnicos que presenten variaciones significativas en función de las diferentes horas del día. Para la fase de depuración y corrección de errores se empleó la herramienta Excel, mientras que, para la fase del procesamiento de datos se utilizó las herramientas R y Rstudio.

Cada una de las fases mencionadas para la recopilación de datos ha sido plasmada de forma gráfica en un diagrama de Grantt, como se puede ver en la Figura 2.3.

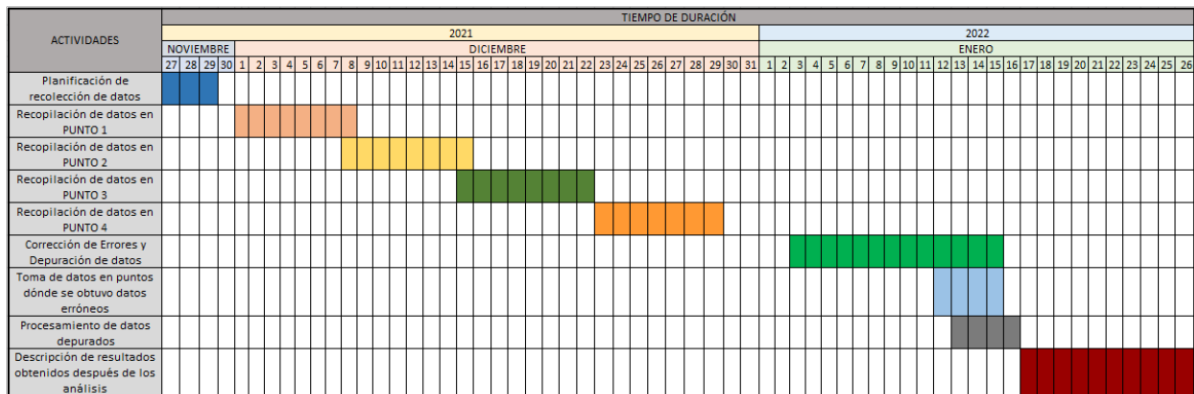


Figura 2.3. Cronograma del proceso de recolección de datos.

2.5 PROCESO PARA EL ANÁLISIS DE LOS DATOS RECOLECTADOS

Para realizar la fase de análisis de datos se usó el software RStudio debido a su entorno de desarrollo integrado que permite la manipulación directa de los datos, realización de operaciones y obtención de resultados que se muestran fácilmente en formato de texto y en formato gráfico [8].

2.5.1 FASES PARA EL ANÁLISIS DE DATOS:

A continuación, se enumeran las fases para el análisis de datos [9]:

- **Importar los datos:** se pueden importar bases de datos, archivos de texto u otros.
- **Depuración y corrección de errores de los datos:** se eliminan datos duplicados, erróneos, valores atípicos, etc.
- **Análisis de los datos:** desarrollo y aplicación de modelos predictivos de clasificación o de agrupamiento a los datos seleccionados.
- **Obtención de resultados:** visualización y exportación de gráficas y resultados estadísticos obtenidos, para posteriormente determinar las conclusiones.

2.5.2 TÉCNICA DE CLASIFICACIÓN DE MACHINE LEARNING

En el presente trabajo se ha tomado en cuenta el modelo de predicción basado en el árbol de clasificación para el análisis de los datos recolectados durante cuatro semanas en diferentes puntos del sector Monjas – Orquídeas. Se utiliza este modelo debido a que permite predecir la pertenencia de los objetos a una clase, sobre la base de variables explicativas cuantitativas y cualitativas, lo cual resultó ideal porque la variable que se definió como salida es de tipo carácter [14].

En este estudio la finalidad del uso del método de clasificación para el análisis de datos consiste en obtener el diagrama del árbol de decisión, que muestra una serie de resultados en base a relaciones obtenidas de todas las comparaciones posibles entre los datos seleccionados [15].

2.5.3 PARÁMETROS SELECCIONADOS COMO VARIABLES DE ENTRADA

Los parámetros definidos como variables de entrada son; día, hora y RSSI, de los cuales los dos primeros han sido seleccionados en función del objetivo general que se especifica en el componente del presente estudio y en el caso del RSSI, ha sido elegido en función de la variación de las medidas obtenidas con relación a la variable de salida.

	A	B	C	D	E	F	G
1	report	date	day	time	rssl_stronge:rssl		
2	0	12/1/2022	miércoles	12:32:00	-116	-103	-8
3	1	12/1/2022	miércoles	12:32:01	-116	-103	-8
4	2	12/1/2022	miércoles	12:32:02	-117	-103	-5
5	3	12/1/2022	miércoles	12:32:03	-117	-103	-5
6	4	12/1/2022	miércoles	12:32:04	-121	-114	-6
7	5	12/1/2022	miércoles	12:32:05	-121	-114	-6
8	6	12/1/2022	miércoles	12:32:06	-121	-112	-8
9	7	12/1/2022	miércoles	12:32:07	-121	-112	-8
10	8	12/1/2022	miércoles	12:32:08	-121	-112	-8
11	9	12/1/2022	miércoles	12:32:09	-118	-105	-7
12	10	12/1/2022	miércoles	12:32:10	-118	-105	-7
13	11	12/1/2022	miércoles	12:32:11	-118	-105	-7
14	12	12/1/2022	miércoles	12:32:12	-119	-112	-8
15	13	12/1/2022	miércoles	12:32:13	-119	-112	-8
16	14	12/1/2022	miércoles	12:32:14	-111	-112	-16
17	15	12/1/2022	miércoles	12:32:15	-111	-112	-16
18	16	12/1/2022	miércoles	12:32:16	-117	-113	-13
19	17	12/1/2022	miércoles	12:32:17	-117	-113	-13
20	18	12/1/2022	miércoles	12:32:18	-117	-113	-13
21	19	12/1/2022	miércoles	12:32:19	-123	-115	-10
22	20	12/1/2022	miércoles	12:32:20	-123	-115	-10

Figura 2.4. Parámetros considerados de entrada.

2.5.4 PARÁMETRO SELECCIONADO COMO VARIABLE DE SALIDA

El parámetro definido como variable de salida es el RSRQ, el cual se ha seleccionado principalmente en función de la relación que presenta con RSSI.

	A	B	C	D	E	F	G
1	report	date	day	time	rssl_stronge:rssl		rsrq
2	0	12/1/2022	miércoles	12:32:00	-116	-103	-8
3	1	12/1/2022	miércoles	12:32:01	-116	-103	-8
4	2	12/1/2022	miércoles	12:32:02	-117	-103	-5
5	3	12/1/2022	miércoles	12:32:03	-117	-103	-5
6	4	12/1/2022	miércoles	12:32:04	-121	-114	-6
7	5	12/1/2022	miércoles	12:32:05	-121	-114	-6
8	6	12/1/2022	miércoles	12:32:06	-121	-112	-8
9	7	12/1/2022	miércoles	12:32:07	-121	-112	-8
10	8	12/1/2022	miércoles	12:32:08	-121	-112	-8
11	9	12/1/2022	miércoles	12:32:09	-118	-105	-7
12	10	12/1/2022	miércoles	12:32:10	-118	-105	-7
13	11	12/1/2022	miércoles	12:32:11	-118	-105	-7
14	12	12/1/2022	miércoles	12:32:12	-119	-112	-8
15	13	12/1/2022	miércoles	12:32:13	-119	-112	-8
16	14	12/1/2022	miércoles	12:32:14	-111	-112	-16
17	15	12/1/2022	miércoles	12:32:15	-111	-112	-16
18	16	12/1/2022	miércoles	12:32:16	-117	-113	-13
19	17	12/1/2022	miércoles	12:32:17	-117	-113	-13
20	18	12/1/2022	miércoles	12:32:18	-117	-113	-13
21	19	12/1/2022	miércoles	12:32:19	-123	-115	-10
22	20	12/1/2022	miércoles	12:32:20	-123	-115	-10

Figura 2.5. Parámetro considerado de salida.

2.5.5 PROCESO DE DEPURACIÓN DE DATOS

Previo a la fase de análisis de datos se eliminaron las columnas de parámetros que no son tomados en cuenta en dicho proceso, con el fin de simplificar la matriz general y reducir los tiempos de procesamiento del software Rstudio, además, se han reemplazado y corregido datos erróneos, los mismos que se presentaron por problemas relacionados con la aplicación usada para la recolección de datos.

2.5.6 RANGOS DE CATEGORIZACIÓN DE LA VARIABLE DE SALIDA

Para definir los rangos de las tres categorías de RSRQ, que se especificaron en el código principal, se usaron los valores recomendados en [16] y que se muestran en la Figura 2.6.

		RSSI	SINR (dB)	RSRQ (dB)	RSRP (dBm)	EC/IO (dB)
Technology		LTE and 3G	LTE Only	LTE Only	LTE Only	HSPA+ and EVDO
Signal Quality 	Excellent	> -65	> 12.5	> -5	> -84	> -2
	Good	-65 to -75	10 to 12.5	-9 to -5	-85 to -102	-2 to -5
	Fair	-75 to -85	7 to 10	-12 to -9	-103 to -111	-5 to -10
	Poor	< -85	< 7	< -12	< -111	< -10

Figura 2.6. Rangos recomendados de RSRQ [16].

Tomando en cuenta la cantidad de datos recolectados, se especificaron para el análisis únicamente 3 categorías, tal y como se muestra en la tabla 2.1.

Tabla 2.1. Categorías definidas para análisis de datos.

Categorización	Rango
Excelente	RSRQ \geq 5
Media	-9 \leq RSRQ $<$ -5
Mala	RSRQ $<$ -9

2.5.7 DESCRIPCIÓN DE COMANDOS UTILIZADOS EN EL PROCESO DEL ANÁLISIS DE DATOS

a) Definición del directorio en el que se encuentran los datos recolectados durante las cuatro semanas

```
setwd("D:/EPN/TESIS/TRABAJO DE INTEGRACIÓN CURRICULAR/DATOS/DEPURADOS")
```

Figura 2.7. Definición del directorio dónde se encuentran los archivos con los datos recolectados.

b) Importación de datos recolectados

Los archivos de los datos correspondientes a las 24 horas, de los 7 días de la semana durante 4 semanas han sido inicialmente depurados y unidos en cuatro archivos, de tal forma que en cada uno de ellos se encuentren los datos correspondientes a los 7 días de la semana, dichos archivos se muestran en la Figura 2.8 con las instrucciones utilizadas. Por tanto, con el comando `read.csv` se puede importar cada uno de los archivos deseados y se emplea el comando `sep` para separar cada una de las columnas ya que inicialmente se muestran todos los datos en una sola columna.

```

datosemana1 <- read.csv("semana1.csv", sep=";")
datosemana2 <- read.csv("semana2.csv", sep=";")
datosemana3 <- read.csv("semana3.csv", sep=";")
datosemana4 <- read.csv("semana4.csv", sep=";")

```

Figura 2.8. Código para la importación de datos.

c) Unión de datos de cada semana en una sola matriz

El comando `rbind`, usado en la Figura 2.9 permite unir varios archivos de bases de datos y ha sido empleado para unir los cuatro archivos y obtener una sola matriz denominada “final”, que va a ser usada durante todo el proceso del análisis de datos [17]. De esta forma se tiene la ventaja de obtener una sola matriz con mayor cantidad de datos y por ende se espera que la precisión del árbol de decisiones obtenido sea alta.

```

final <- rbind(datosemana1, datosemana2, datosemana3, datosemana4)

```

Figura 2.9. Código para unir matrices.

d) Librerías que se van a usar

A continuación se presentan las librerías que han sido necesarias para el desarrollo del programa y la forma de programar se encuentra en la Figura 2.10 [17], [18]:

- **Dplyr.**- Permite la manipulación de data frames².
- **Rpart.**- Permite crear árboles de clasificación en base a datos y fórmulas de predicción.
- **Rpart.plot.**- Grafica los resultados de `rpart`.
- **Rattle.**- Permite acceder y modificar datos.
- **RColorBrewer.**- Permite dar color a los árboles de decisión.

```

library(dplyr) #Permite manipulación con data frames
library(rpart) #Permite crear con los datos y fórmulas de predicción árboles de clasificación
library(rpart.plot) #Grafica los resultados de rpart
library(rattle) #Permite acceder y modificar datos

library(RColorBrewer) #Permite dar color a los árboles de decisión

```

Figura 2.10. Uso de librerías.

e) Creación de la variable de salida

La variable que se ha definido como salida es “RSRQ”, por tanto al ser de tipo entero, se ha creado una variable de tipo carácter denominada “calidadrx”, como se puede ver en la Figura 2.11.

² Data Frames: Es el conjunto de datos que se toma para ser manipulado en la fase de análisis. Estos conjuntos de datos pueden estar compuestos por instancias y variables [22].

```
final$calidadrx <- "NA"
```

Figura 2.11. Creación de variable de salida.

f) Categorización de la variable de salida

Se han definido valores y rangos de acuerdo a la variación de los datos registrados del parámetro RSRQ, definiendo que se trata de excelente calidad de señal de referencia recibida cuando se tienen valores mayores o iguales a -5, media calidad de señal de referencia recibida entre valores de -5 y -9, y mala calidad de señal de referencia recibida cuando se tiene valores menores a -9, como se puede observar en la Figura 2.12.

```
final$calidadrx[final$rsrq >=(-5)] <- "Excelente"  
final$calidadrx[final$rsrq <(-5) & final$rsrq >=(-9)] <- "Media"  
final$calidadrx[final$rsrq <(-9)] <- "Mala"
```

Figura 2.12. Categorización y dimensionamiento de la variable de salida.

g) Selección de las variables de entrada y de salida

Las variables de entrada se han definido de acuerdo a la mayor variación de parámetros y a su relación con la variable de salida. Se han tomado como variables de entrada; day, time y RSSI, y como variable de salida RSRQ, como se puede ver en la Figura 2.13.

```
final <- select(final,c(day, time, rssi, calidadrx))
```

Figura 2.13. Selección de variables de entrada y de salida.

h) Cambio de formato de parámetro "time"

Para solucionar el problema de que al ingresar el parámetro time, la hora ingresa como variable de tipo caracter, se cambia el formato a tipo entero y se condiciona para que el programa solo trabaje con las horas y ya no con los minutos y segundos como se puede ver en la Figura 2.14 [17].

```
final$time <- strptime(final$time, format = "%H:%M:%S")  
final$time <- format(as.POSIXct(final$time), format="%H")  
final$time <- as.integer(final$time)
```

Figura 2.14. Cambio del formato de parámetro "time".

i) Creación del conjunto de datos de entrenamiento y de prueba

Se crea la función, dónde se toma el 80% de datos para entrenamiento y 20% para prueba, la forma de programación se muestra en la Figura 2.15 [19].


```

create_train_test <- function(data, size =0.8, train=TRUE) {
  n_row = nrow(data)
  final_row = size * n_row
  train_sample <-1: final_row
  if (train == TRUE) {
    return (data[train_sample, ])
  } else {
    return (data[-train_sample, ])
  }
}

```

Figura 2.15. Creación de la función de entrenamiento y de prueba.

j) Muestreo de datos

Es importante realizar la mezcla aleatoria de datos para un adecuado análisis, de esta forma se determina la información más relevante. Por tanto, con la función “sample” se puede realizar este paso, como se puede ver en la Figura 2.16. La función sample está tomando en cuenta que los datos se mezclen desde la primera fila hasta la última, es por esto que se pone el número 1 dentro del argumento de sample, a continuación, se utiliza el comando nrow que toma las filas y las columnas de todo el arreglo que corresponde a la matriz “final” [17].

```

final <- final[sample(1:nrow(final)), ]

```

Figura 2.16. Muestreo de datos.

k) Creación y designación de matrices de entrenamiento y de prueba

Como se mencionó anteriormente, se definió que el 80% de datos es para entrenamiento y 20% para prueba. A continuación se crean los conjuntos de datos para estas dos tareas [19].

```

train <- create_train_test(final, 0.8, train = TRUE)
test <- create_train_test(final, 0.8, train = FALSE)

```

Figura 2.17. Creación y designación de matrices de entrenamiento y de prueba.

l) Creación del árbol de clasificación

Se usa el comando “fit” para ajustar el modelo exponencial a datos de incidencia³, el comando rpart para crear el árbol por el método de clasificación, los comandos minsplit y minbucket se usan para que el modelo se vuelva más preciso, de tal forma que entre menores son los valores que se definen para estos comandos, el modelo es más preciso, por tanto, se definió un valor de 1 para cada uno de estos parámetros [19].

³ Es el proceso de construcción de una ecuación exponencial para que se ajuste a un conjunto de datos [22].

```
fit <- rpart(calidadrx ~., data = train, method = "class", minsplit = 1, minbucket = 1)
```

Figura 2.18. Creación del árbol de clasificación.

m) Predicción del modelo

Se realiza la predicción del modelo de acuerdo a las relaciones encontradas entre las variables de entrada y la de salida, para esto se emplea el comando “predict” como se observa en la Figura 2.19 [18], [19].

```
prueba <- predict(fit, test, type = 'class')
```

Figura 2.19. Predicción del modelo.

n) Gráfica del árbol de clasificación

Con el comando fancyRpartPlot se grafica el árbol de clasificación de acuerdo a las predicciones realizadas [17], [20].

```
fancyRpartPlot(fit, caption = NULL)
```

Figura 2.20. Gráfica del árbol de clasificación.

o) Cálculo de la precisión del modelo obtenido

Con el comando “accuracy”⁴, Figura 2.21 se obtiene el valor de la precisión del modelo de clasificación obtenido [19].

```
accuracy_Test <- sum(diag(matriz)) / sum(matriz)
```

Figura 2.21. Precisión del modelo obtenido.

p) Matriz de confusión

Con el comando “matriz”, Figura 2.22 se presenta la matriz de confusión del modelo obtenido, lo cual es muy importante observar para analizar los resultados obtenidos de acuerdo a las predicciones realizadas [19].

```
matriz
```

Figura 2.22. Comando para mostrar matriz de confusión.

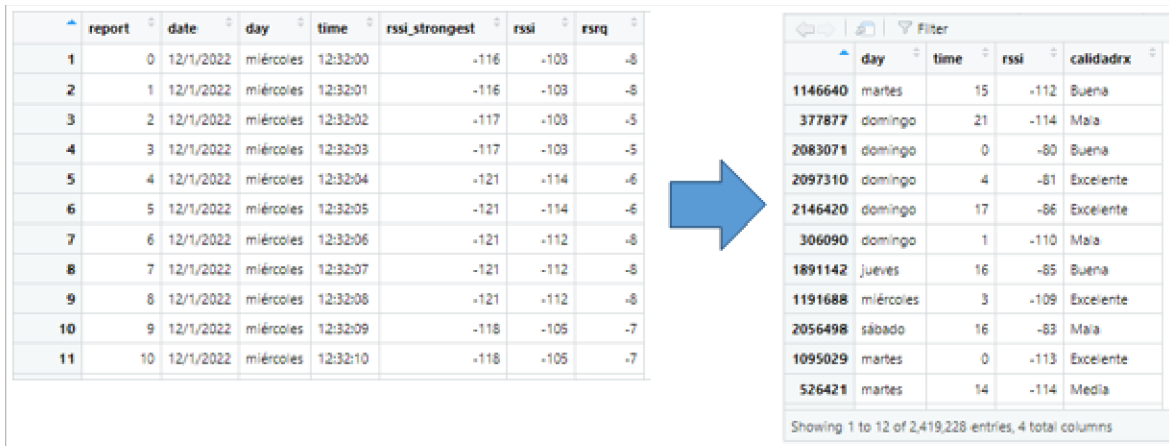
⁴ Accuracy: Se refiere al porcentaje de las instancias en el cual al modelo obtenido predijo el valor de forma correcta. Para determinar un pronóstico preciso del buen funcionamiento de un modelo es necesario probar el modelo con datos nuevos, es decir que no se usaron al ajustar el modelo [22].

2.5.8 RESULTADOS OBTENIDOS

A continuación se presentan los resultados obtenidos que se analizarán en el tercer capítulo.

2.5.8.1 MATRIZ FINAL

Después de realizar las modificaciones en el conjunto inicial de datos se obtiene la matriz que va a emplear el programa de análisis de datos, como se puede observar a continuación.



report	date	day	time	rssi_strongest	rssi	rsrq	
1	0	12/1/2022	miércoles	12:32:00	-116	-103	-8
2	1	12/1/2022	miércoles	12:32:01	-116	-103	-8
3	2	12/1/2022	miércoles	12:32:02	-117	-103	-5
4	3	12/1/2022	miércoles	12:32:03	-117	-103	-5
5	4	12/1/2022	miércoles	12:32:04	-121	-114	-6
6	5	12/1/2022	miércoles	12:32:05	-121	-114	-6
7	6	12/1/2022	miércoles	12:32:06	-121	-112	-8
8	7	12/1/2022	miércoles	12:32:07	-121	-112	-8
9	8	12/1/2022	miércoles	12:32:08	-121	-112	-8
10	9	12/1/2022	miércoles	12:32:09	-118	-105	-7
11	10	12/1/2022	miércoles	12:32:10	-118	-105	-7

day	time	rssi	calidadrx
1146640	martes	15	-112 Buena
377877	domingo	21	-114 Mala
2083071	domingo	0	-80 Buena
2097310	domingo	4	-81 Excelente
2146420	domingo	17	-86 Excelente
306090	domingo	1	-110 Mala
1891142	jueves	16	-85 Buena
1191688	miércoles	3	-109 Excelente
2056498	sábado	16	-83 Mala
1095029	martes	0	-113 Excelente
526421	martes	14	-114 Media

Figura 2.23. a) Matriz “final” antes de modificaciones de parámetros. b) Matriz “final” después de modificaciones en parámetros.

2.5.8.2 Variables de la matriz “final”

Después de las modificaciones realizadas en la matriz inicial, se obtuvo el cambio del tipo de variables de acuerdo con los parámetros a ser usados como variables de entrada y de salida en el posterior análisis de los datos como se puede ver en la Figura 2.24.



```

a) final | 2419228 obs. of 7 variables
$ report : fnt 0 1 2 3 4 5 6 7 8 9 ...
$ date   : chr "12/1/2022" "12/1/2022" "12/1/2022" "12/1/2022" ...
$ day    : chr "miércoles" "miércoles" "miércoles" "miércoles" ...
$ time   : chr "12:32:00" "12:32:01" "12:32:02" "12:32:03" ...
$ rssi_strongest: fnt -116 -116 -117 -117 -121 -121 -121 -121 -121 ...
$ rssi   : fnt -103 -103 -103 -103 -114 -114 -112 -112 -105 ...
$ rsrq  : fnt -8 -8 -5 -5 -6 -6 -8 -8 -7 ...

b) final | 2419228 obs. of 4 variables
$ day    : chr "miércoles" "miércoles" "miércoles" "miércoles" ...
$ time   : fnt 12 12 12 12 12 12 12 12 12 ...
$ rssi   : fnt -103 -103 -103 -103 -114 -114 -112 -112 -105 ...
$ calidadrx: chr "media" "media" "excelente" "excelente" ...
    
```

Figura 2.24. a) Variables de Matriz "final" antes de modificaciones de parámetros. b) Variables de Matriz "final" después de modificaciones de parámetros.

2.5.8.3 Resumen estadístico de RSSI y RSRQ en semana 1

Con el comando `summary` se puede mostrar un resumen estadístico. Por tanto, se ha usado este comando para mostrar las estadísticas de los datos recolectados en la semana 1, como se muestra en la Figura 2.25.

```
> #Resumen estadístico de RSSI y RSRQ
> summary(semana1$rssi)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-122.0 -113.0 -111.0 -111.6 -110.0 -103.0

> summary(semana1$rsrq)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -22     -12     -9      -10     -8      -5
```

Figura 2.25. Resumen estadístico de RSSI y RSRQ de semana 1.

2.5.8.4 Resumen estadístico de RSSI y RSRQ en semana 2

A continuación, se presenta las estadísticas de los datos recolectados en la semana 2:

```
> #Resumen estadístico de RSSI y RSRQ
> summary(semana2$rssi)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-124.0 -113.0 -110.0 -110.3 -108.0 -95.0

> summary(semana2$rsrq)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-22.000 -9.000 -7.000 -7.601 -6.000 -4.000
```

Figura 2.26. Resumen estadístico de RSSI y RSRQ de semana 2.

2.5.8.5 Resumen estadístico de RSSI y RSRQ en semana 3

A continuación, se presenta las estadísticas de los datos recolectados en la semana 3:

```
> #Resumen estadístico de RSSI y RSRQ
> summary(semana3$rssi)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-105.00 -98.00 -95.00 -94.92 -92.00 -87.00

> summary(semana3$rsrq)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-19.000 -11.000 -9.000 -9.493 -8.000 -5.000
```

Figura 2.27. Resumen estadístico de RSSI y RSRQ de semana 3.

2.5.8.6 Resumen estadístico de RSSI y RSRQ en semana 4

A continuación, se presenta las estadísticas de los datos recolectados en la semana 4:

```

> #Resumen estadístico de RSSI y RSRQ
> summary(semana4$rssi)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-120.00 -87.00  -84.00  -84.78  -83.00  -65.00

> summary(semana4$rsrq)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-21.000  -8.000  -7.000  -7.478  -6.000  -3.000

```

Figura 2.28. Resumen estadístico de RSSI y RSRQ de semana 4.

3 ANÁLISIS DE RESULTADOS, CONCLUSIONES Y RECOMENDACIONES

3.1 RESULTADOS

3.1.1 ANÁLISIS GENERAL DE DATOS RECOLECTADOS:

Inicialmente se ha desarrollado un código general, con el uso del modelo de predicción basado en el árbol de clasificación para el análisis de los datos, en el cual se unieron los 2419200 datos recolectados durante las cuatro semanas, con la finalidad de disponer de una base con mayor cantidad de datos; y así obtener, un patrón de comportamiento o un modelo con mayor precisión. Es importante mencionar que se ha configurado la aplicación NetMonitor Cell Signal Logging para que tome datos cada segundo.

3.1.1.1 Código general desarrollado en el proceso de análisis de datos

Se desarrolló un código general en Rstudio que se presenta en la siguiente figura, donde se tomaron en cuenta los datos recolectados durante las cuatro semanas, adicional se decidió emplear un código individual para obtener el análisis de datos de cada semana y de esta forma analizar detalladamente los datos registrados para obtener los resultados del trabajo de integración curricular.

```

1 #Escuela Politécnica Nacional
2 #Facultad de Ingeniería Eléctrica y Electrónica
3 #Trabajo de Integración Curricular
4 #Código General
5 #Modelo de Clasificación del Árbol de Decisiones
6 #Nombre: Gyna Pozo
7
8 #Se define el directorio donde se encuentran datos recolectados de 4 semanas
9 setwd("D:/EPN/TESIS/TRABAJO DE INTEGRACIÓN CURRICULAR/DATOS/DEPURADOS")
10
11 #Se cargan datos de las 4 semanas
12 datosemana1 <- read.csv("semana1.csv", sep=";")
13 datosemana2 <- read.csv("semana2.csv", sep=";")
14 datosemana3 <- read.csv("semana3.csv", sep=";")
15 datosemana4 <- read.csv("semana4.csv", sep=";")
16
17 #Unión de matrices de 4 semanas en una sola (final):
18 final <- rbind(datosemana1, datosemana2, datosemana3, datosemana4)
19
20 #Librerías que se van a usar:
21 library(dplyr) #Permite manipulación con data frames
22 library(rpart) #Permite crear con los datos y fórmulas de predicción árboles de clasificación
23 library(rpart.plot) #Grafica los resultados de rpart
24 library(rattle) #Permite acceder y modificar datos
25 library(RcolorBrewer) #Permite dar color a los árboles de decisión
26
27 #Categorización de parámetro de salida:
28 final$calidadrx <- "NA"
29 final$calidadrx[final$rsrq >=(-5)] <- "Excelente"
30 final$calidadrx[final$rsrq <(-5) & final$rsrq >=(-9)] <- "Media"
31 final$calidadrx[final$rsrq <(-9)] <- "Mala"
32
33 #Selección de variables de entrada y de salida:
34 final <- select(final,c(day, time, rss1, calidadrx))
35
36 #Cambio de tipo de variable "time":
37 final$time <- strptime(final$time, format = "%H:%M:%S")
38 final$time <- format(as.POSIXct(final$time), format="%H")
39 final$time <- as.integer(final$time)
40
41 #Creación de función para entrenamiento de datos:
42 create_train_test <- function(data, size =0.8, train=TRUE) {
43   n_row = nrow(data)
44   final_row = size * n_row
45   train_sample <-1: final_row
46   if (train == TRUE) {
47     return (data[train_sample, ])
48   } else {
49     return (data[-train_sample, ])
50   }
51 }
52
53 #Muestreo de matriz principal:
54 final <- final[sample(1:nrow(final)), ]
55
56 #Definición de 80% de datos para entrenamiento y 20% de datos para prueba:
57 train <- create_train_test(final, 0.8, train = TRUE)
58 test <- create_train_test(final, 0.8, train = FALSE)
59
60 #Predicción y construcción de árbol de clasificación:
61 fit <- rpart(calidadrx ~., data = train, method = "class", minsplit = 1, minbucket = 1)
62 prueba <- predict(fit, test, type = 'class')
63
64 #Creación de matriz de confusión:
65 matriz <- table(test$calidadrx, prueba)
66
67 #Graficación de árbol de clasificación:
68 fancyRpartPlot(fit, caption = NULL)
69
70 #Precisión de modelo de predicción:
71 accuracy_test <- sum(diag(matriz)) / sum(matriz)
72 length(fit$frame$var)
73 print(paste('Accuracy for test', accuracy_test))
74
75 #Matriz de confusión
76 matriz

```

Figura 3.1. Código general desarrollado en Rstudio.

3.1.1.2 Árbol de decisión General

El árbol de decisiones general obtenido después del análisis de la totalidad de los datos es el siguiente:

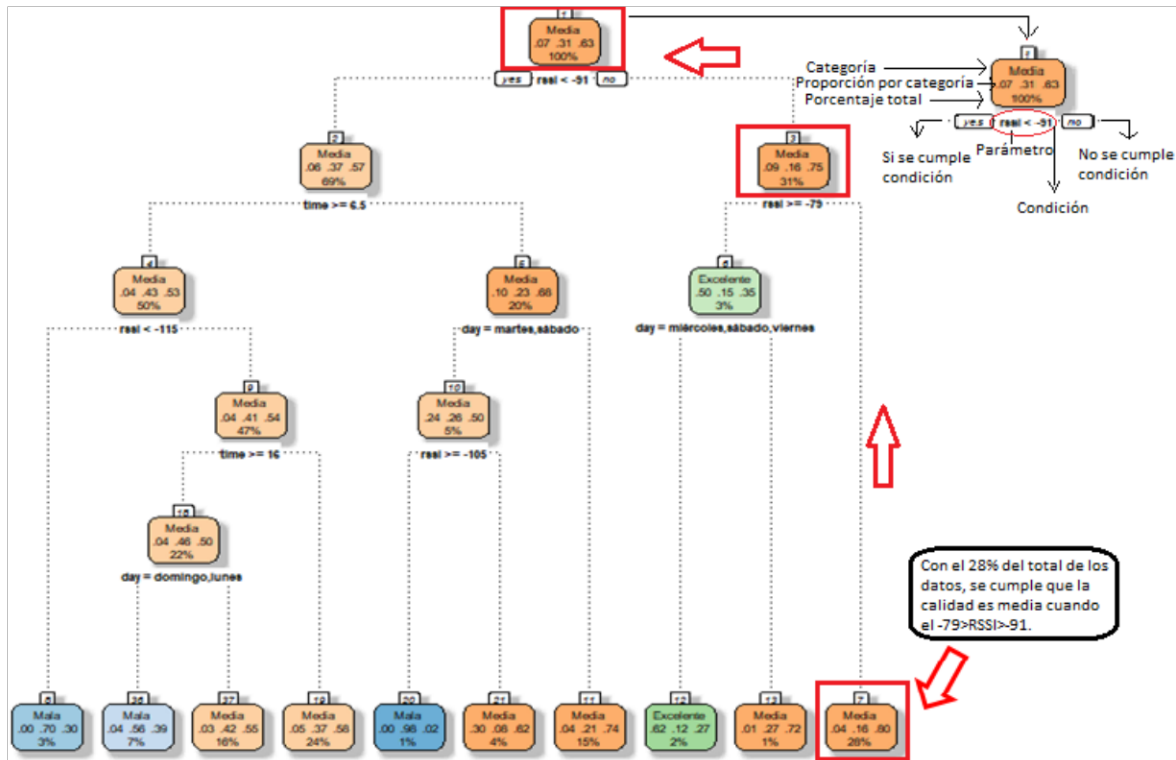


Figura 3.2. Árbol de decisión general.

3.1.1.3 Precisión del modelo obtenido

La precisión del modelo de predicción que se ha obtenido en la fase de análisis general de datos ha sido de 66.85%, como se puede ver en la Figura 3.3. Por tanto, tomando en cuenta que el valor de precisión obtenido es de aproximadamente 67%, se asume que se ha conseguido la predicción de un modelo confiable, ya que aproximadamente 7 de cada 10 predicciones se han acertado correctamente [24].

```
> print(paste('Accuracy for test', accuracy_Test))
[1] "Accuracy for test 0.66846062590163"
```

Figura 3.3. Precisión del modelo general obtenido.

3.1.1.4 Matriz de confusión

Es una herramienta disponible en R de evaluación del modelo que permite visualizar de forma clara las predicciones del modelo y también permite la evaluación de este. Es así como, del análisis de la matriz general, 6961 datos de 11160 usados en la categoría “Excelente” han sido acertados, 32337 datos de 50470 usados en la categoría “Mala” han sido acertados y 284134 datos de 422216 usados en la categoría “Media” han sido acertados, como se puede ver en la Figura 3.4. Por tanto, tomando en cuenta que más del

50% de datos son acertados en las tres categorías se confirma que el modelo predictivo es bueno.

```
> #Matriz de confusión
> matriz
      prueba
      Excelente  Mala  Media
Excelente    6961  1417  23125
Mala         1240 32337 114957
Media        2959 16716 284134
```

Figura 3.4. Matriz de confusión general.

3.1.1.5 Resumen de reglas importantes

A continuación, se presenta un resumen de las reglas del modelo general, el cual se obtuvo al ingresar el comando `rpart.rules(fit)` en la consola de Rstudio. Por ejemplo, la primera de las reglas que se muestra en la Figura 3.5 indica que la calidad de la señal recibida es excelente cuando el RSSI es mayor o igual a -79 y mientras el día sea; miércoles, viernes o sábado.

```
> rpart.rules(fit)
calidadrx  Exc  Mal  Med
Excelente [.62 .12 .27] when rssi >= -79 & day is miércoles or sábado or viernes
Mala [.04 .56 .39] when rssi is -115 to -91 & time >= 16 & day is domingo or lunes
Mala [.00 .70 .30] when rssi < -115 & time >= 7
Mala [.00 .98 .02] when rssi is -105 to -91 & time < 7 & day is martes or sábado
Media [.03 .42 .55] when rssi is -115 to -91 & time >= 16 & day is jueves or martes or miércoles or sábado or viernes
Media [.05 .37 .58] when rssi is -115 to -91 & time is 7 to 16
Media [.30 .08 .62] when rssi < -105 & time < 7 & day is martes or sábado
Media [.01 .27 .72] when rssi >= -79 & day is domingo or jueves or lunes or martes
Media [.04 .21 .74] when rssi < -91 & time < 7 & day is domingo or jueves or lunes or miércoles or viernes
Media [.04 .16 .80] when rssi is -91 to -79
```

Figura 3.5. Resumen de reglas obtenidas del modelo de predicción.

3.1.2 ANÁLISIS DE DATOS RECOLECTADOS EN LA SEMANA 1

Es necesario indicar que durante la semana 1 se recolectaron los datos solamente en el PUNTO 1 localizado en la ubicación indicada anteriormente.

3.1.2.1 Árbol de decisión de la semana 1

El árbol de decisiones obtenido con los datos recolectados durante la semana 1 es el siguiente

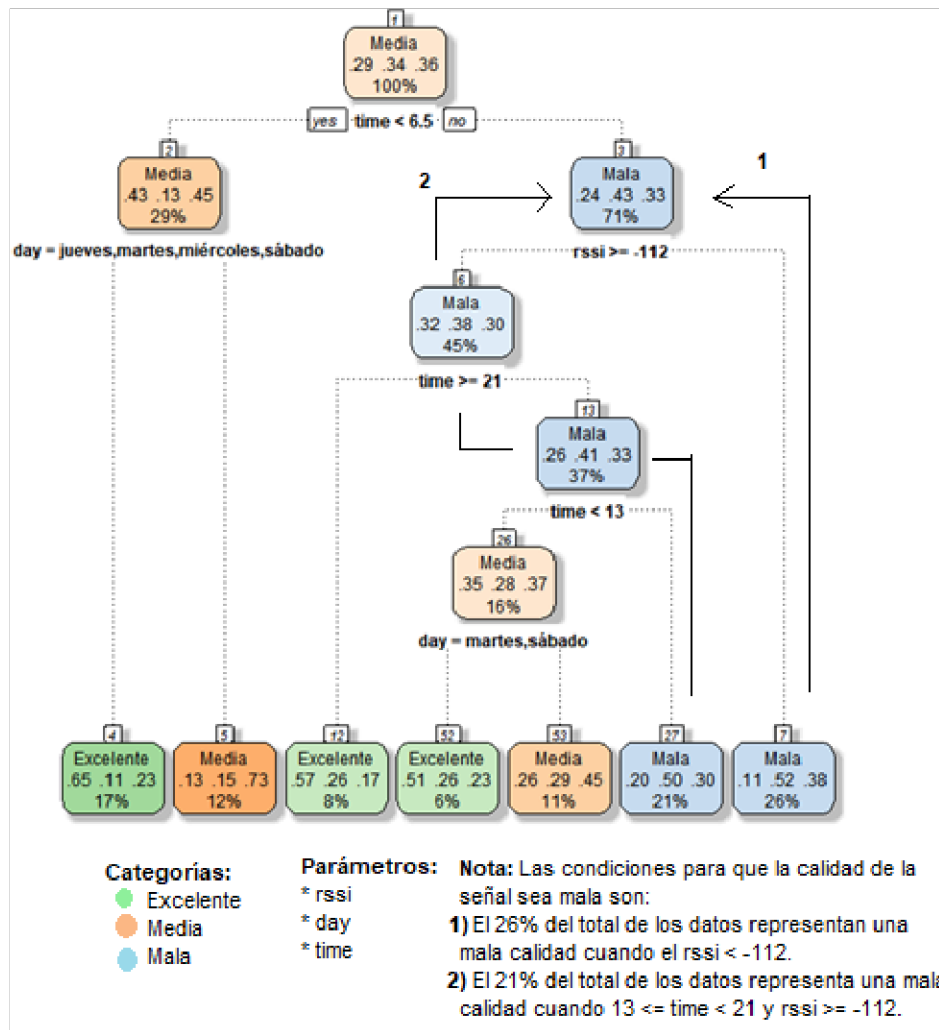


Figura 3.6. Árbol de decisión de semana 1.

3.1.2.2 Precisión del modelo obtenido de la semana 1

La precisión del modelo de predicción que se ha obtenido en la fase de análisis de datos de la semana 1 ha sido de 56.01%, como se puede ver en la Figura 3.7. Por tanto, se asume que se ha conseguido la predicción de un modelo aceptable, ya que aproximadamente 6 de cada 10 predicciones se han acertado correctamente [24].

```
> print(paste('Accuracy for test', accuracy_Test))
[1] "Accuracy for test 0.560151121839917"
```

Figura 3.7. Precisión del modelo de la semana 1.

3.1.2.3 Matriz de confusión de semana 1

Del análisis de datos de la matriz correspondiente a la semana 1, se ha obtenido que de 35871 datos usados para prueba en la categoría “Excelente”, 21708 han sido acertados, de 57232 datos usados en la categoría “Mala”, 29264 han sido acertados y de 27859 datos

usados en la categoría “Media”, 16785 han sido acertados, como se puede ver en la Figura 3.8. Por tanto, considerando que del total de datos que se han tomado para prueba, más del 50% han resultado acertados en las tres categorías, se confirma que el modelo predictivo es bueno.

```
> #Matriz de confusión
> matriz
      prueba
      Excelente  Mala Media
Excelente  21708  8345  5132
Mala       6365  29264  5942
Media      7798  19623  16785
```

Figura 3.8. Matriz de confusión de semana 1.

3.1.2.4 Resumen estadístico de RSSI y RSRQ de la semana 1

Se obtuvo un resumen estadístico de los valores máximos, mínimos, media y mediana de los parámetros RSSI y RSRQ obtenidos durante la semana 1, como se muestra en la Tabla 3.1.

Tabla 3.1. Estadísticas de los valores de RSSI y RSRQ durante semana 1.

Parámetro	Valor mínimo	Mediana	Media	Valor máximo
RSSI	-122	-111	-111.6	-103
RSRQ (dB)	-22	-9	-10	-5

Por tanto, comparando los valores sugeridos para estos parámetros que se muestran en la referencia [17] y que se presentan en la Tabla 3.2 con los valores obtenidos en el análisis de datos de este estudio, se comprueba que de acuerdo con el valor medio obtenido en RSRQ se comprueba que la calidad de la señal recibida en el PUNTO 1 de forma general es media.

Tabla 3.2. Valores de parámetros de calidad de señal sugeridos para red LTE [16].

Categoría	RSSI	RSRQ (dB)
Excelente	> -65	> -5
Buena	-65 a -75	-9 a -5
Media	-75 a -85	-12 a -9
Pobre	< -85	< -12

3.1.2.5 Histograma de la semana 1

En el histograma obtenido de los datos recolectados en el PUNTO 1, que fue donde se recolectaron los datos durante la semana 1, se ha graficado la distribución de las barras

con la frecuencia de repetición de los datos en función de la variable de salida (RSRQ) se obtienen siete gráficos de los siete días de la semana, como se puede ver en la Figura 3.9. De los histogramas se observa que el comportamiento de los datos de RSRQ obtenidos en este punto durante los siete días varía entre excelente, media y mala calidad de señal recibida, sin embargo, prevalece la calidad media, coincidiendo esto con el diagrama del árbol obtenido.

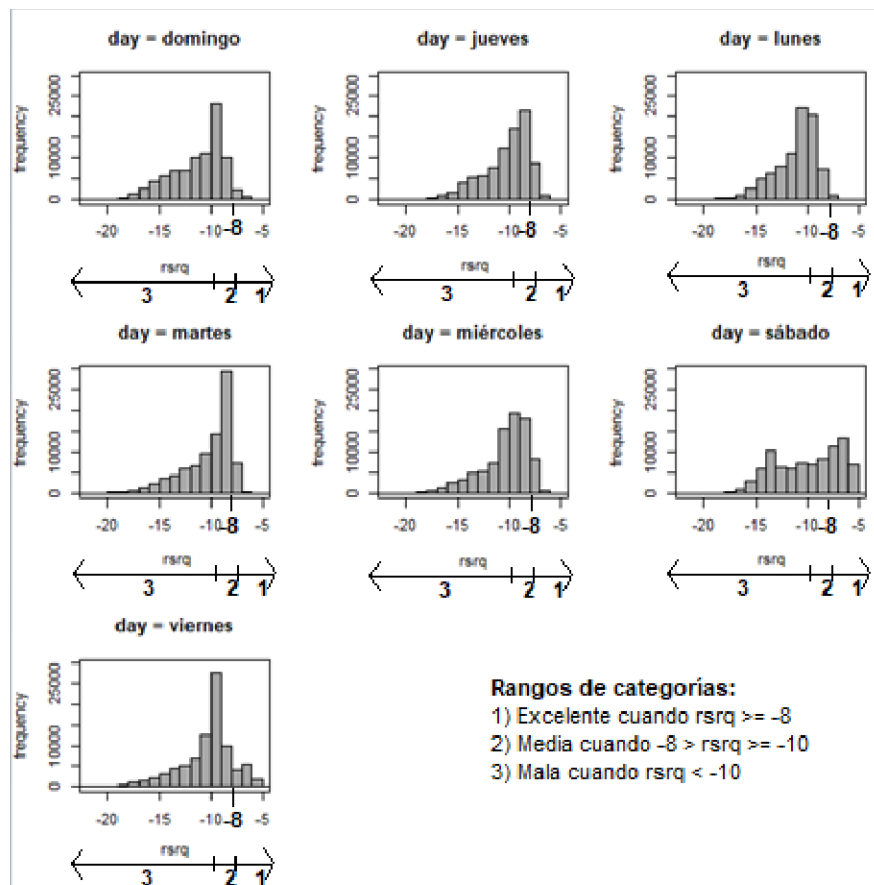


Figura 3.9. Histograma de semana 1.

3.1.3 ANÁLISIS DE DATOS RECOLECTADOS EN LA SEMANA 2

Es necesario indicar que durante la semana 2 se recolectaron los datos solamente en el PUNTO 2 localizado en la ubicación indicada anteriormente.

3.1.3.1 Árbol de decisión de la semana 2

El árbol de decisiones obtenido con los datos recolectados durante la semana 2, después del análisis de los datos es el siguiente:

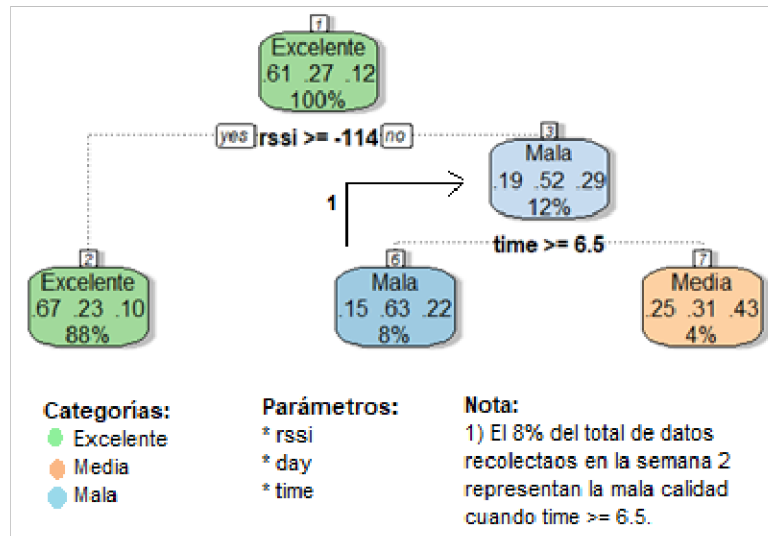


Figura 3.10. Árbol de decisión de semana 2.

3.1.3.2 Precisión del modelo obtenido de la semana 2

La precisión del modelo de predicción que se obtuvo en la fase de análisis de datos de la semana 2 es de 65.27%, como se puede ver en la Figura 3.11. Por tanto, considerando que se obtuvo un valor de precisión de aproximadamente 65%, se asume que se ha conseguido la predicción de un modelo aceptable, ya que aproximadamente 7 de cada 10 predicciones se han acertado correctamente [24].

```
> print(paste('Accuracy for test', accuracy_Test))
[1] "Accuracy for test 0.652709115259338"
```

Figura 3.11. Precisión del modelo de la semana 2.

3.1.3.3 Matriz de confusión de la semana 2

Del análisis de datos de la matriz correspondiente a la semana 2, se obtiene que de 106197 datos usados para prueba en la categoría “Excelente”, 70554 han sido acertados, de 9588 datos usados en la categoría “Mala”, 6105 han sido acertados y de 5177 datos usados en la categoría “Media”, 2294 han sido acertados, como se puede ver en la Figura 3.12. Por tanto, considerando que del total de datos que se han tomado para prueba, más del 50% han resultado acertados en las dos primeras categorías y aproximadamente el 50% en el caso de la tercera categoría, de esta forma se puede indicar que con el modelo predictivo obtenido se logra obtener esa precisión.

```

> #Matriz de confusión
> matriz
      prueba
      Excelente Mala Media
Excelente 70554 1477 1331
Mala      25217 6105 1552
Media     10426 2006 2294

```

Figura 3.12. Matriz de confusión de la semana 2.

3.1.3.4 Resumen estadístico de RSSI y RSRQ de la semana 2

Se determinó en la Tabla 3.3 resumen estadístico de los valores máximos, mínimos, media y mediana de los parámetros RSSI y RSRQ obtenidos durante la semana 2.

Tabla 3.3. Estadísticas de los valores de RSSI y RSRQ durante la semana 2

Parámetro	Valor mínimo	Mediana	Media	Valor máximo
RSSI	-124	-110	-110.3	-95
RSRQ (dB)	-22	-7	-7.601	-4

Comparando los valores sugeridos anteriormente en la Tabla 3.2 con los valores obtenidos en el análisis estadístico de datos, se comprueba que de acuerdo con el valor medio obtenido en RSRQ se comprueba que la calidad de la señal recibida en el PUNTO 2 de forma general es media.

3.1.3.5 Histograma de la semana 2

En el histograma obtenido de los datos recolectados en el PUNTO 2 se presenta la distribución de las barras de la frecuencia de repetición de los datos en función de la variable de salida (RSRQ), mostrando siete gráficos, cada uno con respecto a cada día de la semana como se puede ver en la Figura 3.13. Se observa que el comportamiento de los datos de RSRQ obtenidos en este punto durante los siete días varía entre excelente, media y mala calidad de señal recibida; de esta forma, se puede comprobar que el parámetro día no es un factor que tiene mayor influencia en las decisiones de las cuales se formó el diagrama del árbol obtenido.

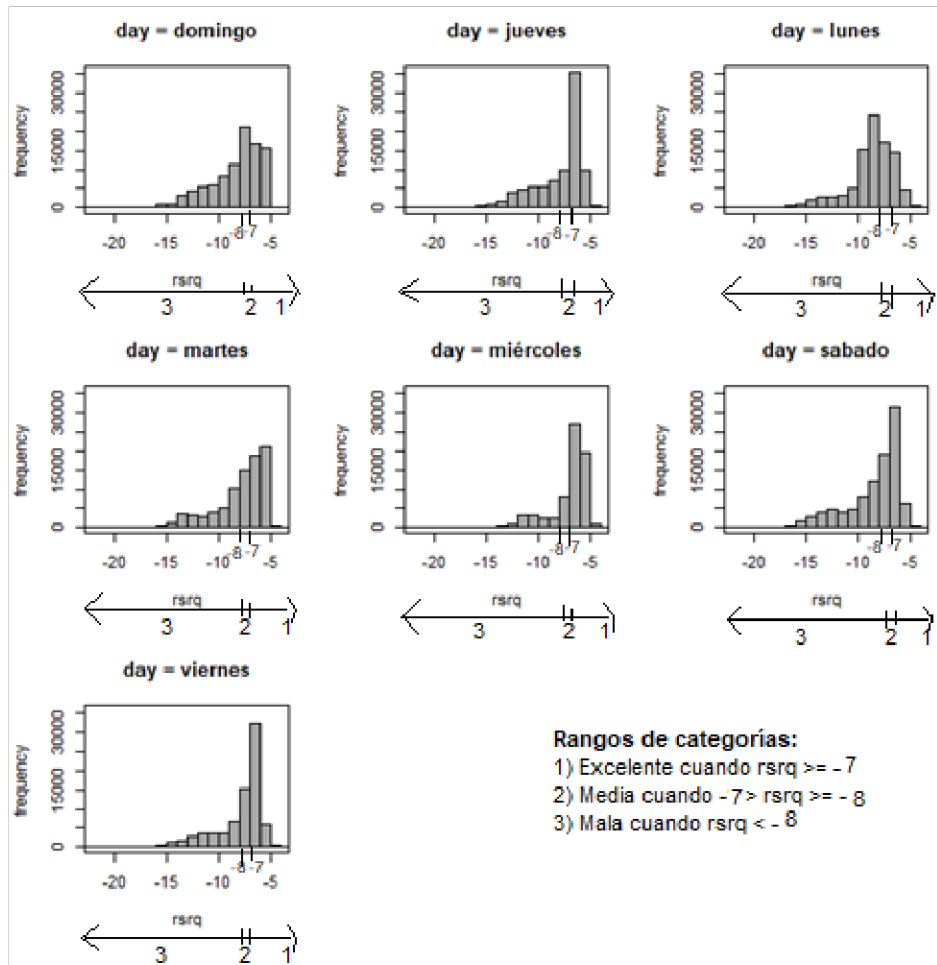


Figura 33 Histograma de la semana 2.

3.1.4 ANÁLISIS DE DATOS RECOLECTADOS EN LA SEMANA 3

Es necesario indicar que durante la semana 3 se recolectaron los datos solamente en el PUNTO 3 localizado en la ubicación indicada anteriormente.

3.1.4.1 Árbol de decisión de la semana 3

El árbol de decisiones obtenido con los datos recolectados durante la semana 3 es el siguiente:

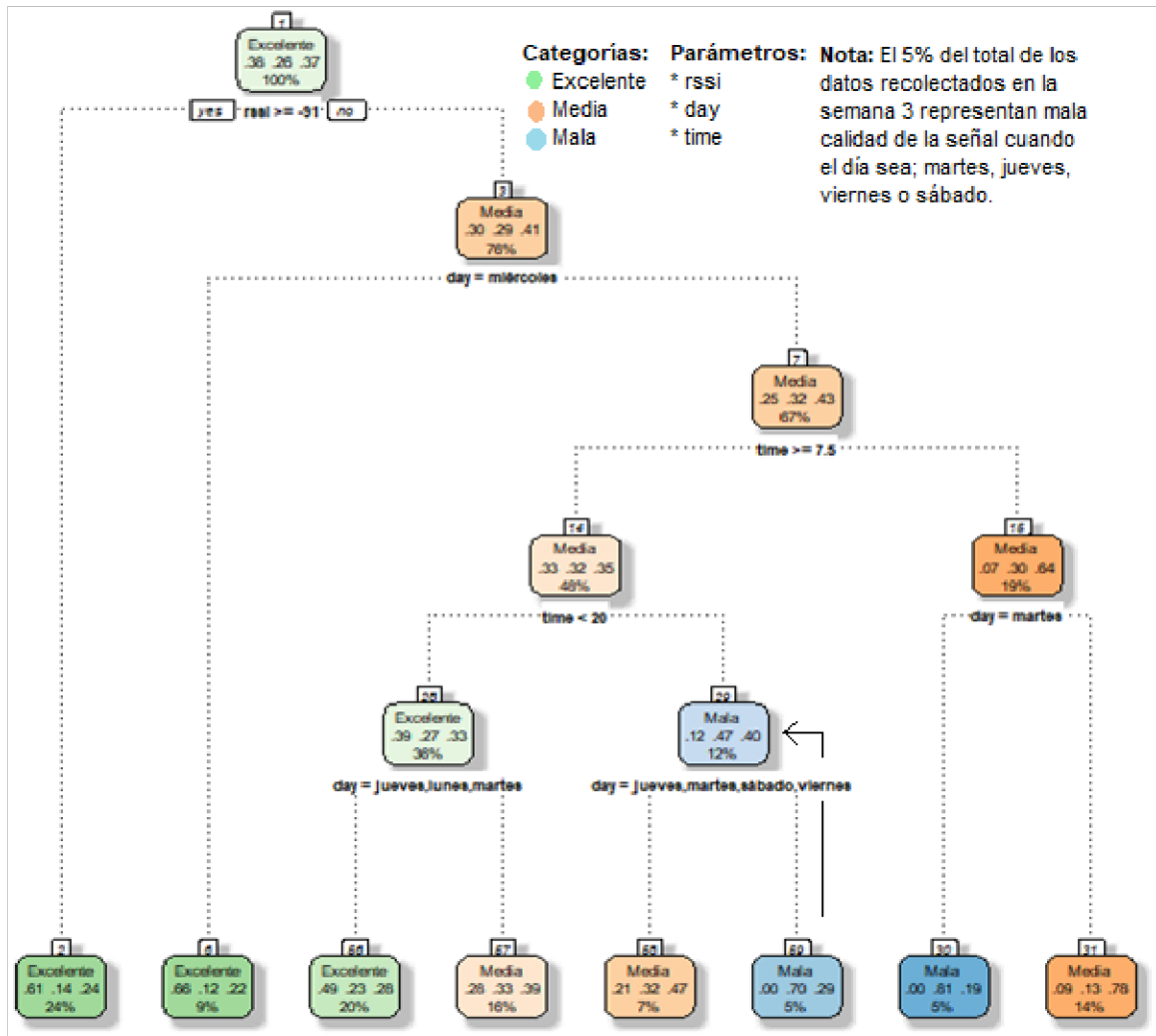


Figura 3.14. Árbol de decisión de la semana 3.

3.1.4.2 Precisión del modelo obtenido de la semana 3

La precisión del modelo de predicción que se ha obtenido en la fase de análisis de datos de la semana 3 es de 58.64%, como se puede ver en la Figura 3.15. Por tanto, considerando que se obtuvo un valor de precisión de aproximadamente 59%, se asume que se ha conseguido la predicción de un modelo aceptable, ya que aproximadamente 6 de cada 10 predicciones se han acertado correctamente [24].

```
> print(paste('Accuracy for test', accuracy_Test))
[1] "Accuracy for test 0.586407301466576"
```

Figura 3.15. Precisión del modelo de la semana 3.

3.1.4.3 Matriz de confusión de la semana 3

Del análisis de datos de la matriz correspondiente a la semana 3, se ha obtenido que de 63995 datos usados para prueba en la categoría “Excelente”, 36718 han sido acertados, de 11648 datos usados en la categoría “Mala”, 8814 han sido acertados y de 45319 datos usados en la categoría “Media”, 25401 han sido acertados, como se puede ver en la Figura 3.16. Por tanto, se tiene que del total de datos que se han tomado para prueba, más del 50% han resultado acertados en las dos primeras categorías y aproximadamente el 50% en el caso de la tercera categoría.

```
> #Matriz de confusión
> matriz
      prueba
      Excelente Mala Media
Excelente 36718 19 8664
Mala      10918 8814 11254
Media     16359 2815 25401
```

Figura 3.16. Matriz de confusión de la semana 3.

3.1.4.4 Resumen estadístico de RSSI y RSRQ de la semana 3

Se presenta en la Tabla 3.4 un resumen estadístico de los valores máximos, mínimos, media y mediana de los parámetros RSSI y RSRQ obtenidos durante la semana 3.

Tabla 3.4. Estadísticas de los valores de RSSI y RSRQ durante la semana 3.

Parámetro	Valor mínimo	Mediana	Media	Valor máximo
RSSI	-105	-95	-94.92	-87
RSRQ (dB)	-19	-9	-9.493	-5

Por tanto, comparando los valores sugeridos para estos parámetros que se muestran en la Tabla 3.2 con los valores obtenidos en el análisis de datos, se comprueba que de acuerdo con el valor medio obtenido en RSRQ se comprueba que la calidad de la señal recibida en el PUNTO 3 de forma general es media.

3.1.4.5 Histograma de la semana 3

En el histograma obtenido de los datos recolectados en el PUNTO 3 se presenta la distribución de las barras de la frecuencia de repetición de los datos en función de la variable de salida (RSRQ), se muestran siete gráficos, uno por cada día de la semana, como se puede ver en la Figura 3.17. Se observa que el comportamiento de los datos de RSRQ obtenidos en este punto el día martes, muestra mala calidad de señal recibida, ya

que la barra que sobresale se encuentra en valores menores a -10, mientras que en el gráfico correspondiente al comportamiento de datos el día miércoles muestra calidad de la señal recibida media, ya que la barra que sobresale se encuentra entre valores de -8 t -10.

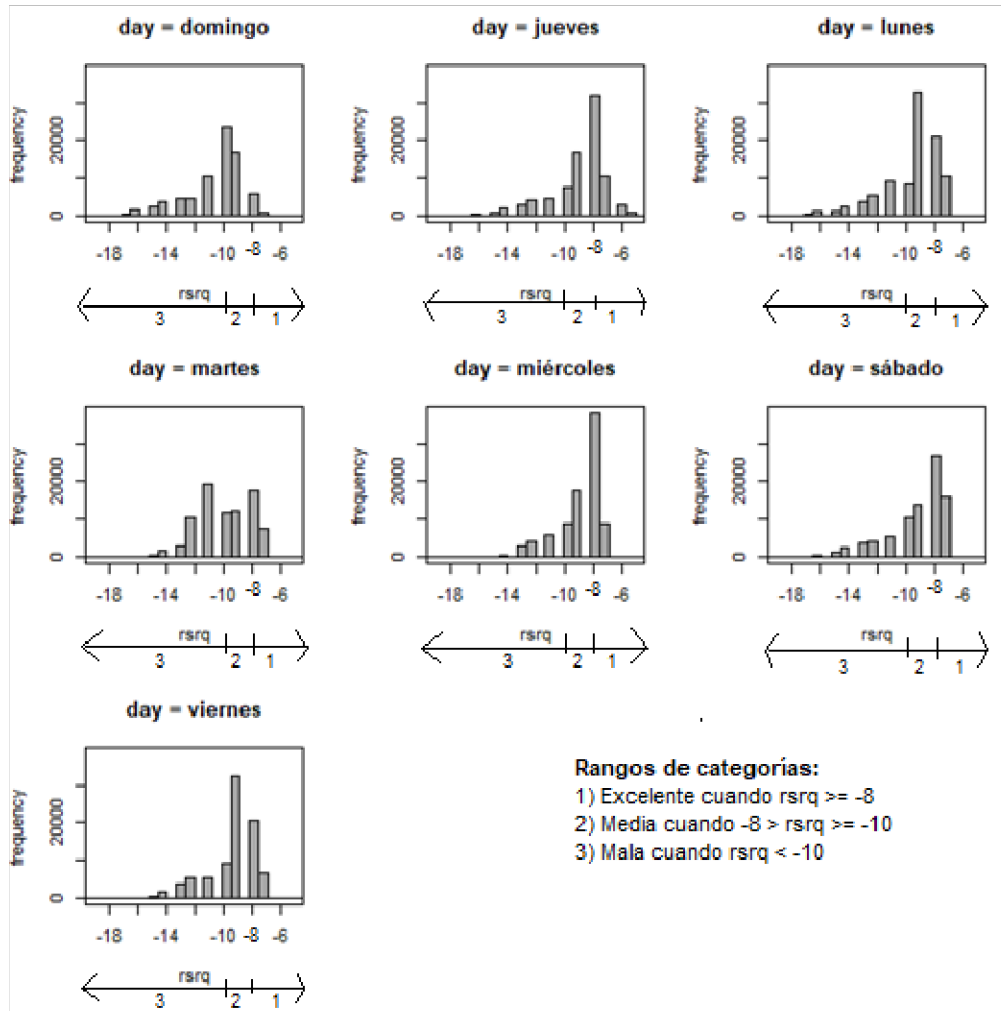


Figura 3.17. Histograma de la semana 3.

3.1.5 ANÁLISIS DE DATOS RECOLECTADOS EN LA SEMANA 4

Es necesario indicar que durante la semana 4 se recolectaron los datos solamente en el PUNTO 4 localizado en la ubicación indicada anteriormente

3.1.5.1 Árbol de decisión de la semana 4

El árbol de decisiones obtenido con los datos recolectados durante la semana 4 es el siguiente:

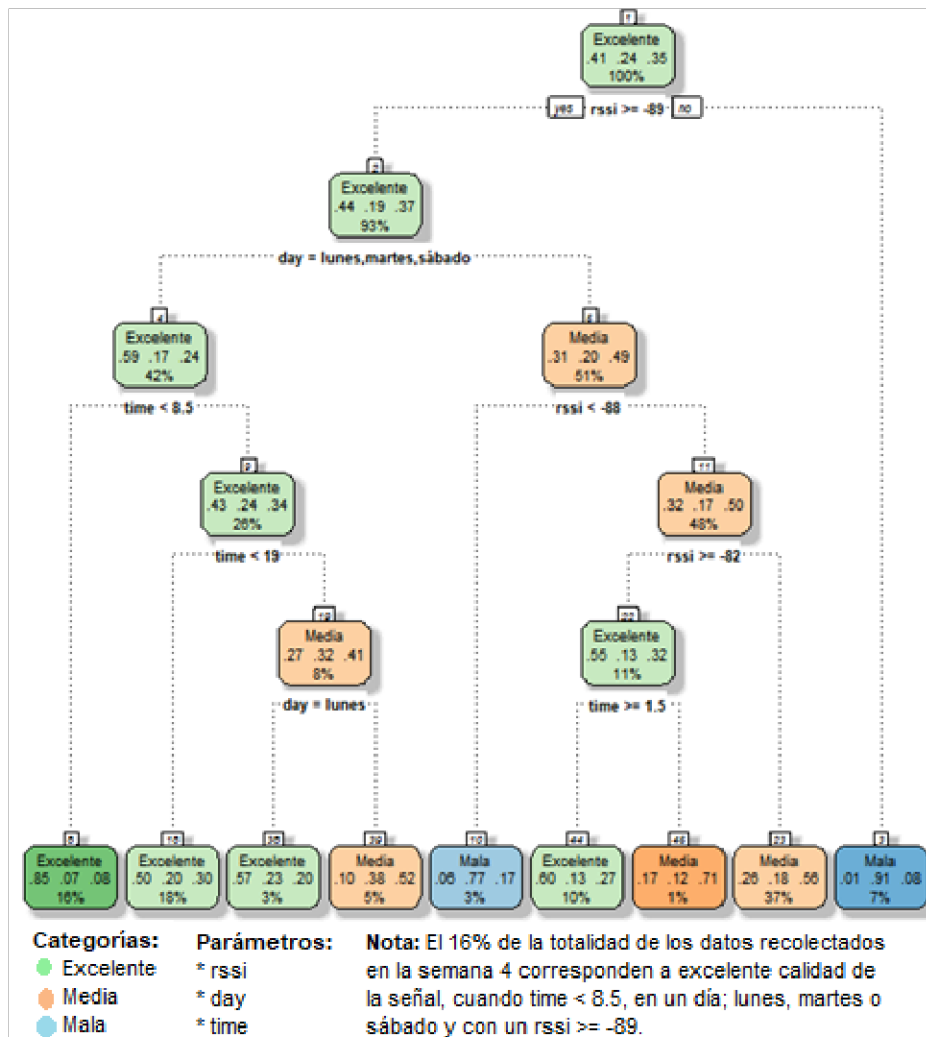


Figura 3.18. Árbol de decisión de la semana 4.

3.1.5.2 Precisión del modelo obtenido de la semana 4

La precisión del modelo de predicción que se ha obtenido en la fase de análisis de datos de la semana 4 es de 63.17%, como se puede ver en la Figura 3.19. Por tanto, considerando que se obtuvo un valor de precisión mayor al 60%, ya que aproximadamente 6 de cada 10 predicciones se han acertado correctamente [24].

```
> print(paste('Accuracy for test', accuracy_Test))
[1] "Accuracy for test 0.631669449909889"
```

Figura 3.19. Precisión del modelo de la semana 4.

3.1.5.3 Matriz de confusión de semana 4

Del análisis de datos de la matriz correspondiente a la semana 4, se ha obtenido que de 56206 datos usados para prueba en la categoría “Excelente”, 36438 han sido acertados,

de 11636 datos usados en la categoría “Mala”, 10207 han sido acertados y de 53120 datos usados en la categoría “Media”, 29763 han sido acertados, como se puede ver en la Figura 3.20. Por tanto, se observa que del total de datos que se han tomado para prueba, más del 50% han resultado acertados en las tres categorías.

```

> #Matriz de confusión
> matriz
      prueba
      Excelente Mala Media
Excelente 36438 276 12472
Mala      7871 10207 10885
Media     11897 1153 29763

```

Figura 3.20. Matriz de confusión de la semana 4.

3.1.5.4 Resumen estadístico de RSSI y RSRQ de la semana 4

Se ha determinado un resumen estadístico de los valores máximos, mínimos, media y mediana de los parámetros RSSI y RSRQ obtenidos durante la semana 4, como se muestra en la Tabla 3.5.

Tabla 3.5. Estadísticas de los valores de RSSI y RSRQ durante semana 4

Parámetro	Valor mínimo	Mediana	Media	Valor máximo
RSSI	-120	-84	-84.78	-65
RSRQ (dB)	-21	-7	-7.478	-3

Por tanto, comparando los valores sugeridos para estos parámetros que se muestran en la Tabla 3.2 con los valores obtenidos en el análisis de datos, se comprueba que de acuerdo con el valor medio obtenido en RSRQ se comprueba que la calidad de la señal recibida en el PUNTO 4 de forma general es media.

3.1.5.5 Histograma de la semana 4:

En el histograma de la Figura 3.21 obtenido de los datos recolectados en el PUNTO 4 se presenta la distribución de las barras de la frecuencia de repetición de los datos en función de la variable de salida (RSRQ), se muestran siete gráficos, cada uno corresponde a un día de la semana. Se observa que en los días; lunes, martes y sábado, las barras con respecto a los datos de RSRQ obtenidos que más sobresalen, se encuentran en RSRQ \geq -6, correspondiente a una categoría de excelente calidad de señal recibida, lo cual también se puede confirmar con el diagrama del árbol de decisiones obtenido.

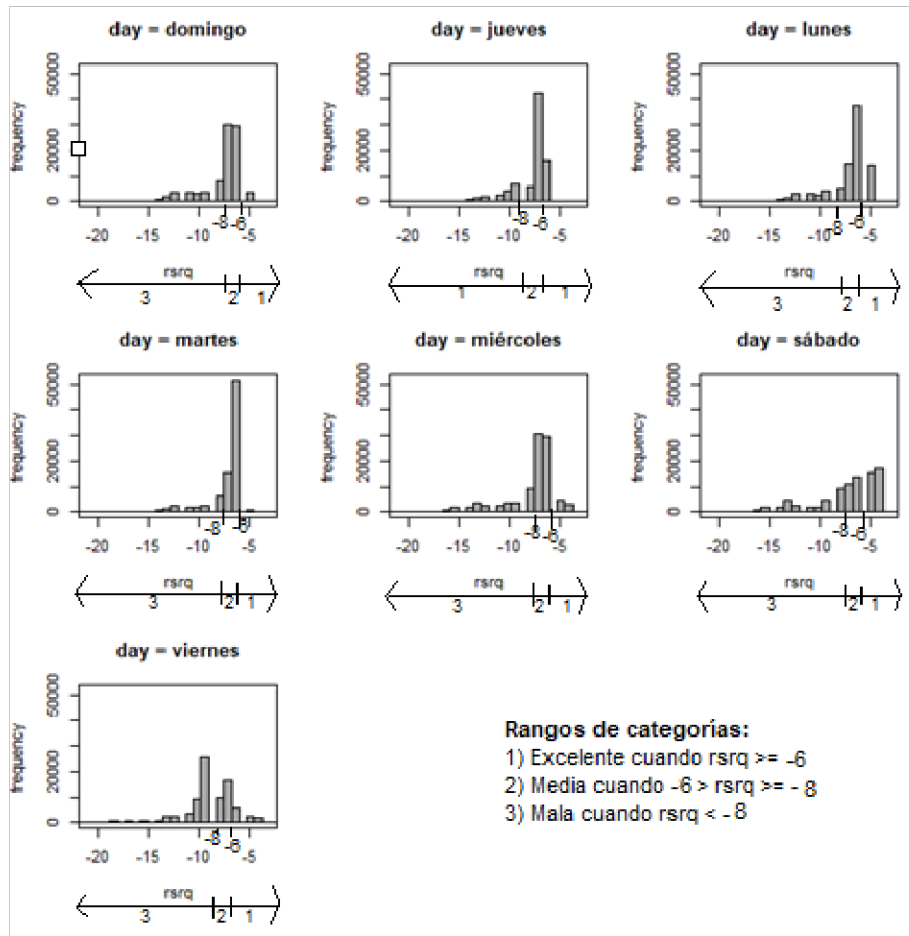


Figura 3.21. Histograma de la semana 4.

3.2 CONCLUSIONES

Tomando en cuenta que durante las mediciones no se ejecutó la conexión de ningún tipo de servicio, y que, en esta fase de recolección, el móvil se encontró en estado estático, se identificó que las variaciones más sobresalientes de las señales de referencia recibidas en el terminal móvil se presentaron en los siguientes parámetros: RSSI y RSRQ y se observa que las variaciones aparecen en el transcurso de las horas del día y en el transcurso de los días de la semana.

Con el análisis general de datos recolectados durante las cuatro semanas en la zona de estudio se obtuvo un modelo predictivo ilustrado en un árbol de clasificación, que categoriza como una excelente calidad de señal recibida cuando se cumplen las siguientes condiciones: el RSSI mayor o igual a -79 y el día puede ser miércoles, viernes o sábado.

Los resultados del análisis general de los datos muestran que en el sector Monjas – Orquídeas de la ciudad de Quito se tiene mala calidad de señal recibida de referencia cuando se tiene las siguientes condiciones: RSSI se encuentra entre valores de -91 y -115

y la hora del día es después de las 4pm de un día domingo o lunes. Esto puede deberse a que actualmente muchas personas aún nos encontramos en modalidad de teletrabajo o telestudio, por tanto, resulta más fácil y cómodo comunicarnos por medio de llamadas. Adicionalmente, se debe mencionar que en la mayoría de los casos, se deja la organización de la agenda de la semana para el domingo, incluso hasta el día siguiente, provocando de esta manera una saturación de la capacidad de las radio bases que brindan el servicio celular en esta zona. También de los resultados se observa que hay mala calidad de señal recibida cuando el RSSI se encuentra entre -91 y -105 y la hora es antes de las 7am de un día martes o sábado; esto posiblemente se debe a que, en los días mencionados, en horas de la madrugada se pueden reducir los recursos de la radio base debido a que no se utiliza con mayor frecuencia el servicio en esos momentos o también se puede deber a un posible mantenimiento establecido por la operadora en esos horarios y días.

Por otra parte, del análisis general se determina que los días; martes, miércoles, jueves, viernes y sábado a partir de las 4pm y al tener un RSSI entre -91 y -115, se tiene calidad de señal recibida media, esto probablemente se debe a que la mayoría de usuarios de las redes móviles empieza a llegar de los trabajos a partir de esa hora y surge la necesidad de comunicarse progresivamente con familiares y amigos; sin embargo, como no todos los usuarios acceden al servicio simultáneamente y por tanto no existe una degradación de la calidad de la señal.

Con la herramienta para análisis de datos Rstudio fue posible mostrar el porcentaje de precisión del modelo de predicción, donde se obtuvo un 66.85% de precisión del modelo del análisis general de datos, 56.01% de precisión del análisis de datos de la semana 1, 65.27% de precisión del análisis de datos de la semana 2, 58,64% de precisión del análisis de datos de la semana 3 y 63.17% de precisión del análisis de datos de la semana 4, por lo que se obtiene que en los 5 casos la precisión del modelo predictivo es mayor al 50%.

El análisis realizado en la matriz de confusión para cada uno de los casos de análisis de datos es otra forma de evaluación del modelo predictivo que se ha obtenido, por tanto, en los cinco casos analizados se observa que en casi todas las categorías de cada análisis se acierta en más del 50% de datos definidos para prueba.

Durante las 24 horas del día, en los puntos de estudio en el sector Monjas – Orquídeas de la ciudad de Quito, se tiene generalmente una calidad media de señal recibida, lo cual se determina de acuerdo a las relaciones obtenidas en el árbol de clasificación y se lo confirma desde la experiencia de los habitantes del barrio, además, esto coincide con la ubicación geográfica de las antenas de las radio bases cercanas con respecto al barrio mencionado,

ya que se tiene la presencia de ciertos objetos intermedios que obstruyen y debilitan la señal en algunos de los puntos dónde se recolectaron los datos.

El análisis de datos de cada semana de forma individual ha sido muy importante, ya que permite observar que dependiendo del punto en el que se han recolectado los datos, los parámetros que se obtienen varían en rangos diferentes, es así como, para el análisis de datos de cada semana se han definido rangos de categorización de la señal recibida de acuerdo a los valores recomendados para redes LTE, pero ajustándolos a los datos obtenidos en cada punto; esto también se realizó debido a los datos con que se cuenta para el análisis y para evitar errores que se presentaron inicialmente en el software Rstudio. En cambio, al realizar el análisis general de datos se definieron los rangos de categorización de RSRQ, únicamente de acuerdo a los valores sugeridos para redes LTE y no fue necesario realizar ajustes de los valores indicados.

3.3 RECOMENDACIONES

Se recomienda a las operadoras de servicios móviles que brindan acceso a la red en el sector de Monjas – Orquídeas de la ciudad de Quito, aumentar la capacidad de los recursos los días domingos y lunes, especialmente pasado las 4pm.

Para futuros trabajos, se recomienda realizar estudios de redes LTE que impliquen la recolección y análisis de una mayor cantidad de datos con la ayuda de aplicaciones móviles, pero ejecutando la conexión de algunos servicios y posteriormente realizar la comparación con los resultados obtenidos en el presente trabajo.

En caso de realizar trabajos donde sea necesario el uso de herramientas móviles de recolección continua de datos para su posterior análisis, se recomienda el uso de la aplicación NetMonitor Cell Signal Logging que funciona muy bien durante periodos de tiempo de hasta 24 horas, sin embargo si se presenta el caso en el que sea necesario tomar datos continuos durante varios días, la solución es configurar el móvil para que se reinicie automáticamente cada 24 horas y posterior realizar la configuración de la aplicación para que cada vez que se encienda el móvil, automáticamente se inicie una nueva sesión de recolección de datos.

Para obtener un mejor modelo predictivo con la técnica de clasificación de Machine Learning es muy importante contar con una base de gran cantidad de datos y también asignar las condiciones adecuadas de acuerdo a los valores obtenidos en la etapa de recolección.

Se recomienda el uso de las herramientas Rstudio y R, ya que son muy útiles para realizar el análisis de datos y porque tienen varias funcionalidades que hacen que la interacción con el usuario sea amigable, debido a que muestran los resultados de forma gráfica o de texto, por otra parte, permiten el análisis de grandes cantidades de datos y hacen posible la ejecución individual de cada línea del código, lo cual ayuda a detectar errores desde el inicio.

4 REFERENCIAS BIBLIOGRÁFICAS

- [1] «Obtén más detalles sobre las antenas disponibles con NetMonitor», Adslzone. [En línea]. Disponible en: <https://www.adslzone.net/moviles/obten-mas-detalles-sobre-las-antenas-disponibles-con-netmonitor/>. [Accedido: 09-oct-2021].
- [2] Parizene, «NetMonitor», Google Play. [En línea]. Disponible en: https://play.google.com/store/apps/details?id=com.parizene.netmonitor&hl=es_EC&gl=US. [Accedido: 09-oct-2021].
- [3] «NetMonitor Cell Signal Logging», Google Play, 10-feb-2021. [En línea]. Disponible en: https://play.google.com/store/apps/details?id=ru.v_a_v.netmonitor&hl=es_EC&gl=US. [Accedido: 13-oct.2021]
- [4] «SetEdit», Google Play, 11-abr-2021. [En línea]. Disponible en: https://play.google.com/store/apps/details?id=by4a.setedit22&hl=es_EC&gl=US. [Accedido: 22-oct-2021].
- [5] «FORZAR RED 4G LTE ONLY EN CUALQUIER TELEFONO HUAWEI [SIN ROOT]», Youtube, 20-abr-2021. [En línea]. Disponible en: <https://www.youtube.com/watch?v=4Vdeht0Dfx8&list=LL&index=1&t=59s>. [Accedido: 21-oct-2021].
- [6] «Huawei Y9 Prime (2019)», GSMARENA, [En línea]. Disponible en: [https://www.gsmarena.com/huawei_y9_prime_\(2019\)-9742.php](https://www.gsmarena.com/huawei_y9_prime_(2019)-9742.php). [Accedido: 25-oct-2021].
- [7] «Huawei Y9 Prime (2019)», NDTV – Gadgets 360, 25-oct-2021. [En línea]. Disponible en: <https://gadgets.ndtv.com/huawei-y9-prime-2019-price-in-india-13058>. [Accedido: 26-oct-2021].
- [8] «RStudio para Estadística Descriptiva en Ciencias Sociales», Bookdown, 16-ene-2022. [En línea]. Disponible en: <https://bookdown.org/gboccardo/manual-ED-UCH/como-definir-que-y-cuantos-software-de-analisis-estadistico-manejar.html#r-el-temido>. [Accedido: 16-ene-2022].
- [9] «Lenguaje R», UNIR, 16-ene-2022. [En línea]. Disponible en: <https://www.unir.net/ingenieria/revista/lenguaje-r-big-data/>. [Accedido: 16-ene-2022].
- [10] «Nuestra identidad corporativa», Claro, 2020. [En línea]. Disponible en: <https://www.claro.com.ec/personas/institucional/quienes-somos/>. [Accedido: 26-nov-2021].
- [11] F. Lema, “Estudio para la recomendación de parámetros técnicos a utilizarse en el control de redes LTE 4G en Ecuador”. Tesis de Ingeniería, Departamento de Electrónica, Telecomunicaciones y Redes de Información, Escuela Politécnica Nacional, Quito, ECUADOR, 2017.

- [12] «Descripción de RSSI y SNR para equipos WLAN de Huawei», HUAWEI, 15-mar-2019. [En línea]. Disponible en: <https://forum.huawei.com/enterprise/es/descripci%C3%B3n-de-rssi-y-snr-para-equipos-wlan-de-huawei/thread/510787-100239>. [Accedido: 26-nov-2021].
- [13] S. Grau, "Medición y Análisis de las Redes de Comunicaciones Móviles 4G LTE en Cullera". Tesis de Ingeniería Sistemas de Telecomunicaciones, Universidad Politécnica de Valencia, GANDIA, 2019.
- [14] «Árboles de clasificación y de regresión», XLSTAT, 16-ene-2022. [En línea]. Disponible en: <https://www.xlstat.com/es/soluciones/funciones/arboles-de-clasificacion-y-de-regresion>. [Accedido: 16-ene-2022].
- [15] «Qué es un diagrama de árbol de decisión», LUCIDCHART, 16-ene-2022. [En línea]. Disponible en: <https://www.lucidchart.com/pages/es/que-es-un-diagrama-de-arbol-de-decision>. [Accedido: 16-ene-2022].
- [16] «Understanding LTE Signal Strength», USAT, 2020. [En línea]. Disponible en: <https://usatcorp.com/faqs/understanding-lte-signal-strength-values/>. [Accedido: 28-ene-2022].
- [17] «Search all R packages on CRAN and Bioconductor», RDocumentation. [En línea]. Disponible en: <https://www.rdocumentation.org/>. [Accedido: 6-ene-2022].
- [18] J. Bosco, «Arboles de decisión con R - Clasificación», RPubS, 23-abr-2018. [En línea]. Disponible en: https://rpubs.com/jboscomendoza/arboles_decision_clasificacion. [Accedido: 7-ene-2022].
- [19] «Árbol de decisión en R | Árbol de clasificación y código en R con ejemplo», EBOOKS ONLINE. [En línea]. Disponible en: <https://ebooksonline.es/arbol-de-decision-en-r-arbol-de-clasificacion-y-codigo-en-r-con-ejemplo/>. [Accedido: 7-feb-2022].
- [20] R. González, «Modelo de clasificación con arboles de decisión en R con ejemplo en RStudio», 3-jul-2020. [En línea]. Disponible en: <https://gonzalezgouveia.com/modelo-de-clasificacion-con-arboles-de-decision-en-r-con-ejemplo-en-rstudio/>. [Accedido: 8-feb-2022].
- [21] «Manipulación básica de dataframes», Gauss del Serv de Apoyo Esadístico (SAE) del SAI. [En línea]. Disponible en: <http://gauss.inf.um.es:8080/tabular-dataframe/>. [Accedido: 7-feb-2022].
- [22] «Ajuste de modelos exponenciales a datos», las-matematicas.com, 2020. [En línea]. Disponible en: <https://las-matematicas.com/6-9-ajuste-de-modelos-exponenciales-a-datos/>. [Accedido: 7-feb-2022].
- [23] R.J. Hyndman, and A.B. Koehler, "Another look at measures of forecast accuracy". International Journal of Forecasting, vol. 22, no. 4, pp. 115-117, Dic. 2006.
- [24] «¿Cómo sé si mi modelo de predicción es realmente bueno?», datos.gob.ec, 26-ene-2021. [En línea]. Disponible en: <https://datos.gob.es/es/blog/como-se-si-mi-modelo-de-prediccion-es-realmente-bueno>. [Accedido: 28-ene-2022].

5 ANEXOS

ANEXO I. Cronograma de actividades

ANEXO II. Archivos de datos en formato digital

ANEXO I

ACTIVIDADES	TIEMPO DE DURACIÓN (SEMANAS)																TOTAL DE HORAS DE ACTIVIDADES
	2021								2022								
	NOVIEMBRE				DICIEMBRE				ENERO				FEBRERO				
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	
Recopilación de información acerca de herramientas para recolección de datos	6																6 horas
Descripción y envío de documento escrito de las herramientas para recolección de datos	5	2															7 horas
Descarga, instalación y prueba de herramientas para recolección de datos		6															6 horas
Presentación acerca de las herramientas de recolección de datos		2															2 horas
Búsqueda de fuentes bibliográficas relacionadas con el componente designado		6	1														7 horas
Recopilación de información de parámetros técnicos de redes celulares con tecnología LTE				6													6 horas
Definición y gestión de puntos para recolección de datos				9													9 horas
Descripción de etapa de recolección de datos				4	2												6 horas
Recolección de datos					13	15	15	15									58 horas
Selección de datos que se van a analizar									4								4 horas
Recopilación de información de herramientas para el análisis de datos								5	1								6 horas

