

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 DE LA RELACIÓN EXISTENTE ENTRE LA VELOCIDAD
DE CONEXIÓN Y LA VELOCIDAD DEL TERMINAL MÓVIL EN UNA
CONEXIÓN PERMANENTE MEDIANTE MEDICIONES DE CAMPO Y
TÉCNICAS DE MACHINE LEARNING.**

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

MICHAEL ALEXANDER CURIPALLO MARTÍNEZ

DIRECTOR: Ph.D. PABLO ANÍBAL LUPERA MORILLO
pablo.lupera@epn.edu.ec

FEBRERO 2022

CERTIFICACIONES

Yo, MICHAEL ALEXANDER CURIPALLO MARTINEZ 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.

MICHAEL ALEXANDER CURIPALLO MARTINEZ

Certifico que el presente trabajo de integración curricular fue desarrollado por MICHAEL ALEXANDER CURIPALLO MARTINEZ, bajo mi supervisión.

PABLO ANÍBAL LUPERA MORILLO
DIRECTOR

Certificamos que revisamos el presente trabajo de integración curricular.

SORAYA LUCIA SINCHE MAITA
REVISOR1 DEL TRABAJO DE
INTEGRACIÓN CURRICULAR

LUIS FELIPE URQUIZA AGUIAR
REVISOR2 DEL TRABAJO DE
INTEGRACIÓN CURRICULAR

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.

MICHAEL ALEXANDER CURIPALLO MARTINEZ

PABLO ANÍBAL LUPERA MORILLO

DEDICATORIA

Quiero dedicar este proyecto y todo logro profesional, a mi familia, quienes siempre me apoyaron en cada decisión que he tomado y sin ellos no estaría donde estoy ahora. A mi madre, Carmita Martínez, la mujer que me ha demostrado que no existen imposibles. A mi hermana, Johanna Curipallo, el mejor regalo que la vida me pudo dar. Y especialmente quiero dedicar este proyecto a quien más le emocionaba mi triunfo y que lamentablemente ya no está con nosotros, a mi padre, Marco Curipallo.

Michael Curipallo

AGRADECIMIENTO

En primer lugar, quiero agradecer a Dios, por darme salud y la fuerza necesaria para superar todo obstáculo. A mi maravillosa madre, quien siempre me ha sostenido y ha limpiado mis lágrimas en todo el proceso, sin ella no estaría donde estoy ahora y cada éxito en mi vida es el resultado de su esfuerzo. También quiero agradecer a mi hermana por estar siempre conmigo, por ser mi mejor amiga, por motivarme y contagiarme de su alegría en momentos difíciles. Nada me hace más feliz que tenerlas conmigo, las amo infinitamente. Quiero agradecer a mi padre, al mejor hombre que he conocido, él dejó en mi un legado y recuerdos maravillosos, le quiero agradecer inmensamente porque él siempre creyó en mi desde que inicie esta carrera universitaria, y aunque ya no este conmigo sé que está muy orgulloso de mi, porque fue testigo de cada esfuerzo que hice. Este y todos los logros que alcance en mi vida serán por y para él. Gracias padre, sé que me amas tanto como yo te amo a ti, te extraño demasiado.

Mis agradecimientos también van dirigido a mis abuelos, tíos, primos y cada integrante de mi familia que en algún momento me brindo ánimos para seguir adelante y gracias a ellos pude sentirme acompañado y respaldado en todo momento.

Quiero también agradecer a los amigos que encontré en este hermoso trayecto, Daysi Aimacaña, Evelyn Curicho, Erick Moreira, Gyna Pozo, Alexander Lara y Jhon Jirón, quienes no solo fueron compañeros de clase, sino que con su amistad hicieron que el trayecto sea más hermoso. Al mejor amigo que pude encontrar en mi vida politécnica, Franklin Zambrano, que a pesar de que somos muy diferentes hemos sido muy buena dupla en trabajos académicos y situaciones personales y sin su ayuda, el camino hubiese sido más complicado. A todos mis compañeros de clases, profesores, personal y autoridades de la prestigiosa Escuela Politécnica Nacional, todos aportaron positivamente en mi rendimiento académico y estoy muy agradecido con cada uno de ellos. Pero especialmente, quiero agradecer al Ph.D. Lupera Pablo, quien fue un excelente tutor y profesor y supo guiarme y ayudarme durante la realización de este proyecto.

Finalmente, quiero agradecer a mi pequeño amuleto de la buena suerte, mi compañera durante todas las noches de desvelo, quien se dormía en mi escritorio acompañándome hasta que termine mis deberes, solo me bastaba con verla para motivarme y seguir trabajando, gracias Lucky.

Michael Curipallo

ÍNDICE DE CONTENIDO

CERTIFICACIONES	I
DECLARACIÓN DE AUTORÍA	II
DEDICATORIA	III
AGRADECIMIENTO	IV
ÍNDICE DE CONTENIDO	V
RESUMEN.....	VII
ABSTRACT	VIII
1. INTRODUCCIÓN.....	1
1.1 OBJETIVO GENERAL.....	2
1.2 OBJETIVOS ESPECÍFICOS.....	2
1.3 ALCANCE	2
1.4 MARCO TEÓRICO	3
1.4.1 HERRAMIENTAS UTILIZADAS PARA LA RECOLECCION DE DATOS	3
1.4.1.1 Netmonitor	4
1.4.1.2 ForeLTE	5
1.4.1.3 Cellmapper	6
1.4.1.4 Network Cell info Lite.....	8
1.4.1.5 G Net-Track Pro	9
1.4.2 TERMINAL MÓVIL	11
1.4.3 HERRAMIENTA DE ANÁLISIS DE DATOS.....	13
1.4.3.1 R Y RSTUDIO.....	13
1.4.3.2 Google Eart Pro.....	15
2. METODOLOGÍA	16
2.1 DEFINICIÓN DE LOS DATOS A RECOLECTAR Y PROCESAR.....	16
2.2 DEFINICIÓN DE LA ZONA A CONSIDERAR Y/O LAS RUTA	17
2.3 PLANIFICACIÓN DE LA RECOLECCIÓN DE DATOS.....	19
2.4 TÉCNICAS DE MACHINE LEARNING	20

2.5 CÓDIGO EN R PARA EL ANÁLISIS DE LOS DATOS RECOLECTADOS	22
3. PRUEBAS, RESULTADOS, CONCLUSIONES Y RECOMENDACIONES.....	21
3.1 PRUEBAS	27
3.1.1 DATOS PARA EL ENTRENAMIENTO Y PRUEBA.....	27
3.1.2 PARÁMETROS DE ENTRADA Y SALIDA.....	27
3.1.2.1 R y RSTUDIO.....	28
3.1.2.2 Google Eart Pro.....	30
3.1.3 MODELOS DE ENTRENAMIENTO CON DIFERENTES PARÁMETROS DE ENTRADAS.	29
3.1.3.1 Parametros de entrada.....	30
3.1.3.2 Parametros de salida.....	31
3.1.3.2.1 Velocidad del terminal móvil y RSRP	31
3.1.3.2.2 Velocidad del terminal móvil y SNR	33
3.1.3.2.3 Velocidad del terminal móvil y RSSI	34
3.1.3.3 Tres parámetros de entrada y un parámetro de salida	35
3.2 RESULTADOS	39
3.2.1 ANÁLISIS DE LOS MODELOS CON MEJORES RESULTADOS.....	39
3.2.1.1 Entradas: velocidad del terminal móvil y RSRP – salida: velocidad de conexión	40
3.2.1.2 Entradas: velocidad del terminal móvil, RSRQ, RSSI y SNR Salida: velocidad de conexión	42
3.3 PRUEBA DEL MODELO CON NUEVOS DATOS ALEATORIOS	45
3.4 CONCLUSIONES.....	48
3.5 RECOMENDACIONES.....	50
4. REFERENCIAS BIBLIOGRÁFICAS.....	50
5. ANEXOS.....	53

RESUMEN

En el presente trabajo se realiza el análisis de la relación existente entre la velocidad de conexión con la velocidad del terminal móvil, mediante la técnica de Machine Learning conocida como aprendizaje supervisado, con clasificación de datos mediante las técnicas de reglas basadas en partición, mejor conocidas como árboles de decisión.

En el primer capítulo se describen las funcionalidades y la información que otorgan las herramientas, también se presenta un resumen con las características más importantes del dispositivo móvil con el que se realizarán las mediciones. Finalmente, se describen las herramientas que se utilizarán para analizar los datos recolectados.

En el segundo capítulo se presenta un resumen de los datos otorgados por las herramientas para definir los parámetros que se utilizarán como entradas de los modelos, en esta sección también se definen las zonas exactas donde se realizarán las mediciones. Después se describe la técnica de Machine Learning y las técnicas de reglas basadas en partición. Por último, se presenta el código realizado en RStudio, donde se define que el modelo tendrá una sola salida que corresponde a la velocidad de la conexión con tres posibles valores: excelente, regular y mala.

En el tercer capítulo se plantean tres escenarios donde se modifica el número de entradas para seleccionar un modelo que permita relacionar las velocidades. Después, se comprueba la eficiencia del modelo utilizando datos aleatorios a las rutas definidas. Por último, se presentan las conclusiones y recomendaciones a las cuales se llegó después del desarrollo del proyecto.

PALABRAS CLAVE: velocidad de conexión, velocidad del terminal móvil, aprendizaje supervisado, árbol de decisión, parámetros RF.

ABSTRACT

In this project, the analysis of the relationship between the connection speed and the speed of the mobile terminal is carried out, using the Machine Learning technique known as supervised learning, with data classification using partition-based rule techniques, better known as decision trees.

In the first chapter the functionalities and the information provided by the tools are described, a summary is also presented with the most important characteristics of the mobile device with which the measurements will be made. Finally, the tools that will be used to analyze the collected data are described.

The second chapter presents a summary of the data provided by the tools to define the parameters that will be used as inputs to the models, in this section the exact areas where the measurements will be made are also defined. Later, the Machine Learning technique and partition-based rule techniques are described. Finally, the code made in RStudio is presented, where it is defined that the model will have a single output with three possible values: excellent, fair, and poor.

In the third chapter, three scenarios are proposed where the number of inputs is modified to select a model that allows the speeds to be related. Then, its efficiency will be checked using random data to the defined routes. Finally, the conclusions and recommendations that were reached after the development of the project are presented.

KEYWORDS: connection speed, mobile terminal speed, supervised learning, decision tree (DTs).

1. INTRODUCCIÓN

Las redes de datos móviles han evolucionado con el tiempo a pasos agigantados, mejorado las capacidades y tecnologías para satisfacer los requerimientos de los usuarios. Actualmente, en el Ecuador se utiliza la red celular 4G LTE con mayor predominancia, permitiendo a millones de personas conectarse a internet para subir o descargar archivos sin problemas, además, los usuarios cada vez requieren de más servicios, ya que estos se vuelven dependientes de la red, actualmente el ser humano utiliza un dispositivo móvil para conectarse a la red en todo momento, ya sea en situaciones estáticas o en movimiento. Es por esta razón que se realiza el presente proyecto, ya que en este se busca analizar la relación entre la velocidad de conexión con la velocidad del terminal móvil para definir la calidad de conexión, estableciendo categorías en función de las velocidades de descarga. Además, el presente proyecto servirá de datos básicos de campo referenciales para los procesos de monitoreo y planificación de las redes.

Existen diferentes estudios en donde se tiene un propósito similar al de este proyecto, por ejemplo, el artículo titulado “Estudio de cobertura de señal exterior y convergencia de red de TD-LTE y TD-SCDMA en alta velocidad ferroviaria” desarrollado en China por Qun Hou, Yifan Zhou, Xin Peng y Long Zhang. En este artículo se establece que, cuando la velocidad del móvil supera los 200 Km/h, la calidad de conexión empieza a bajar rápidamente y a presentar problemas, debido al cambio de frecuencia Doppler, la selección y reelección de celda, la conmutación y la pérdida de penetración [1]. Por lo anterior y teniendo en cuenta el tráfico vehicular, las señales de tránsito, las características de las calles y edificaciones dentro del centro Histórico de Quito se puede dar una idea de que la velocidad del terminal móvil no influiría en la calidad de conexión a la red móvil en este sector de la ciudad, ya que la velocidad máxima permitida es de 50 Km/h. Además, dentro de la zona definida existen una gran cantidad de radio bases, estas están distribuidas de tal manera que se pensaría que son suficientes para garantizar una conexión permanente de alta capacidad sin importar la velocidad vehicular. Para comprobar la hipótesis mencionada, se recolectarán datos de mediciones con un dispositivo móvil, mediante las herramientas que servirán para la recolección de datos y las herramientas para el análisis de las mediciones. Estas mediciones se realizarán en zonas específicas dentro del sector seleccionado, considerando diferentes rangos de velocidades del terminal móvil y manteniendo un tipo de comunicación que garantiza la conexión permanente a la red (por ejemplo: una descarga de archivos). Para el análisis de los datos se definen las entradas y salidas del modelo y se observan las variaciones de los diferentes parámetros. Durante las mediciones se observa que las velocidades de conexión, RSRP (Reference Signal Received Power),

RSRQ (Reference Signal Received Quality), SNR (Signal-to-noise ratio) y RSSI (Received Signal Strength Indicator) son parámetros que varían en toda la medición, por lo que estos parámetros se considerarán como posibles entradas de los modelos que se van a crear. Por otro lado, las velocidades de conexión serán la única salida del modelo, esta salida tendrá tres posibles valores: calidad mala para velocidades menores a 1000 Kbps, calidad regular para velocidades mayores a 1000 Kbps, pero menores a 5000 Kbps y calidad excelente para velocidades de conexión mayores a 5000 Kbps. Para la creación del modelo se usará Rstudio, donde se realizará el entrenamiento mediante la técnica de Machine Learning conocida como aprendizaje supervisado, con clasificación de datos mediante las técnicas de reglas basadas en partición, mejor conocidas como árboles de decisión, donde se establecerán reglas y condiciones basadas en los parámetros de entrada. El análisis permitirá identificar los parámetros que se manifiestan cuando la velocidad de transmisión se encuentra en las categorías mala, regular y excelente, de tal manera que será posible establecer las probables razones de la disminución de la calidad.

1.1 OBJETIVO GENERAL

Deducir la dependencia que existe entre la velocidad de conexión y la velocidad del terminal móvil en una conexión permanente, mediante técnicas de aprendizaje automático utilizando datos recolectados en el centro histórico de Quito.

1.2 OBJETIVOS ESPECÍFICOS

1. Obtener mediciones de parámetros técnicos de la señal del terminal móvil en el centro histórico de Quito, considerando diferentes rutas y velocidades del terminal.
2. Depurar los datos por rutas y rangos de velocidades,
3. Clasificar los datos mediante las técnicas de reglas basadas en partición estableciendo reglas y condiciones.
4. Analizar los datos mediante la técnica de aprendizaje supervisado.

1.3 ALCANCE

En el proyecto se definen las posibles relaciones existentes entre la velocidad de cobertura y la velocidad del terminal móvil en base a datos recolectados en varios puntos y a diferentes velocidades. Para realizar la recolección de datos, se utilizará la aplicación G-NetTrack Pro que otorga diferentes datos de la señal móvil incluyendo los valores de descarga de downlink y uplink, estas mediciones se toman cada 3 segundos sin necesidad de reiniciar la medición. Además, es importante mencionar que esta aplicación no otorga información detallada sobre las estaciones bases a la que el terminal se conecta, por esta razón solo analizaran modelos con los siguientes parámetros de entrada: Velocidad del

terminal móvil, RSRP, RSRQ, SNR y RSSI. En la primera etapa conocida como la recolección de datos, se establecen las tres rutas donde se realizarán las mediciones dentro del centro histórico de Quito. Para realizar las mediciones se establecen dos rangos principales; velocidades por debajo de los 20 Km/h y velocidades desde los 20 Km/h hasta los 40 Km/h. Por otro lado, para la tercera ruta, que comprende el viaducto 24 de mayo se tiene un rango adicional, velocidades de 40 Km/h hasta 80 Km/h. Se debe tener en cuenta diferentes factores que pueden afectar en las mediciones, como, por ejemplo; el tráfico vehicular, peatones, y señalización de tránsito, ya que estos pueden impedir que se tengan la misma cantidad de datos en los rangos establecidos, por lo tanto, se realizaron las mediciones a altas horas de la noche. Para la segunda etapa que comprende el procesamiento y análisis de los datos recolectados, primero se ordena y depuran los datos y se define las entradas y salidas del modelo. Para la creación del modelo se usará un lenguaje de programación con enfoque estadístico, específicamente se usará la interfaz de R llamada Rstudio. Mediante este lenguaje se realizará la programación de la técnica de aprendizaje automático (machine learning), mediante el uso de diferentes bibliotecas y paquetes con funcionalidades de cálculos y gráficas. El entrenamiento se realizará mediante la técnica de Machine Learning conocida como aprendizaje supervisado, con clasificación de datos mediante las técnicas de reglas basadas en partición, mejor conocidas como árboles de decisión, donde se establecerán reglas y condiciones basadas en los parámetros de entrada, y así obtener el modelo de entrenamiento. En la tercera etapa se realizará el análisis de los resultados. Primero se analizará la relación entre la velocidad de conexión y la velocidad del terminal móvil para poder asociar con el modelo obtenido. También, es importante analizar la relación que existe entre la o las entradas del modelo con las salidas, por esto, se hará un análisis independiente considerando: modelos con dos entradas y una salida y modelos de tres entradas y una salida. Estas combinaciones de entradas se harán de tal manera que el modelo resultante considere la velocidad del terminal móvil como condición principal para predecir el comportamiento del fenómeno.

1.4 MARCO TEÓRICO

1.4.1 HERRAMIENTAS UTILIZADAS PARA LA RECOLECCION DE DATOS

En este capítulo se presentan las herramientas y su funcionamiento que se utilizarán en el presente proyecto. A continuación, se presenta en la Figura 1.1 las herramientas instaladas en el dispositivo móvil.



Figura 1.1: Herramientas que se van a utilizar.

1.4.1.1 NetMonitor

Esta es una aplicación gratuita para el sistema operativo Android que sirve para obtener diferentes datos sobre la red móvil y el dispositivo que se está utilizando, principalmente sirve para monitorear redes CDMA, GSM, WCDMA, LTE, TD-SCDMA, 5G NR [2].

En una de las ventanas disponibles se presenta la información acerca de la operadora que ofrece servicio al dispositivo, además presenta los datos acerca del tipo de tecnología y el ancho de banda que se está utilizando. Además, se presenta una gráfica que muestra al usuario “la cobertura”. Finalmente, se obtiene información acerca de las antenas a las que está conectado el dispositivo para acceder a la red; en esta información se encuentran las características TAC¹, CI², PCI³, entre otras, y la distancia en metros a la que se encuentra.

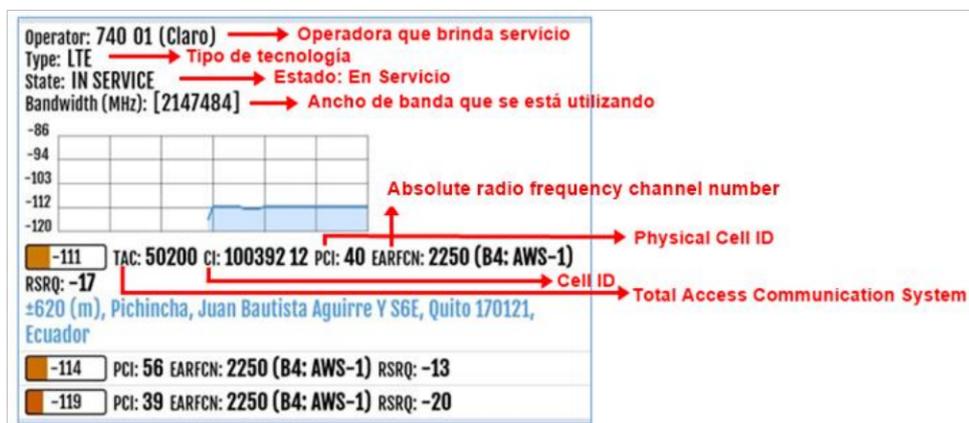


Figura 1.2: Información de cobertura y antenas otorgada por NetMonitor.

En una siguiente ventana se despliega la información de las antenas que ofrecen servicio al dispositivo, en esta información también se encuentra la distancia, ubicación y características de cada antena.

¹ TAC (Total Access Communication System): estándar analógico de telefonía móvil.

² CI (Cell ID): Identificador de la celda de servicio actual.

³ PCI (Physical Cell ID): Identificador físico de celda.

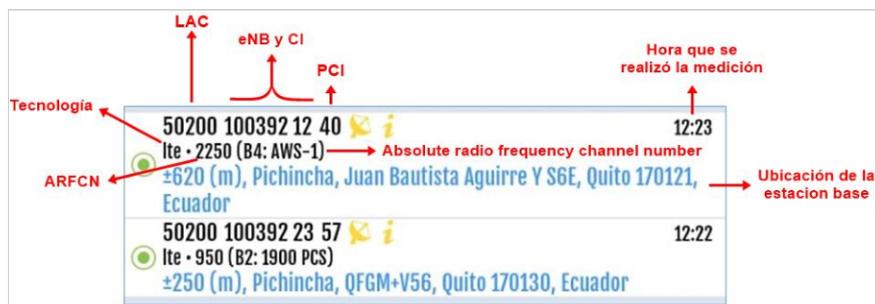


Figura 1.3: Información de las antenas servidoras otorgada por NetMonitor.

En la última ventana disponible del menú se encuentra la ubicación de las antenas dentro del mapa.



Figura 1.4: Ubicación de las antenas en Google Maps.

1.4.1.2 ForceLTE

Al igual que la aplicación anterior, se trata de una aplicación gratuita para Android. Sirve principalmente para gestionar la red con el objetivo de mantener el dispositivo en la red LTE o en la que el usuario desee (Figura 1.5). En la ventana principal llamada "4G ForceLTE" se presentan diferentes métodos para realizar la configuración, esto se debe a que los métodos son específicos para diferentes sistemas operativos y se debe seleccionar el que sea compatible con el dispositivo móvil que se está utilizando. Esta aplicación también ofrece la opción de configurar una segunda tarjeta SIM y se debe realizar la configuración individualmente [3].

Una vez seleccionado el método adecuado, se presentará la información del dispositivo, y dentro de esta información se encuentra la opción para "Establecer el tipo de red preferida", esta permite establecer la red que se desea brinde servicio al dispositivo.

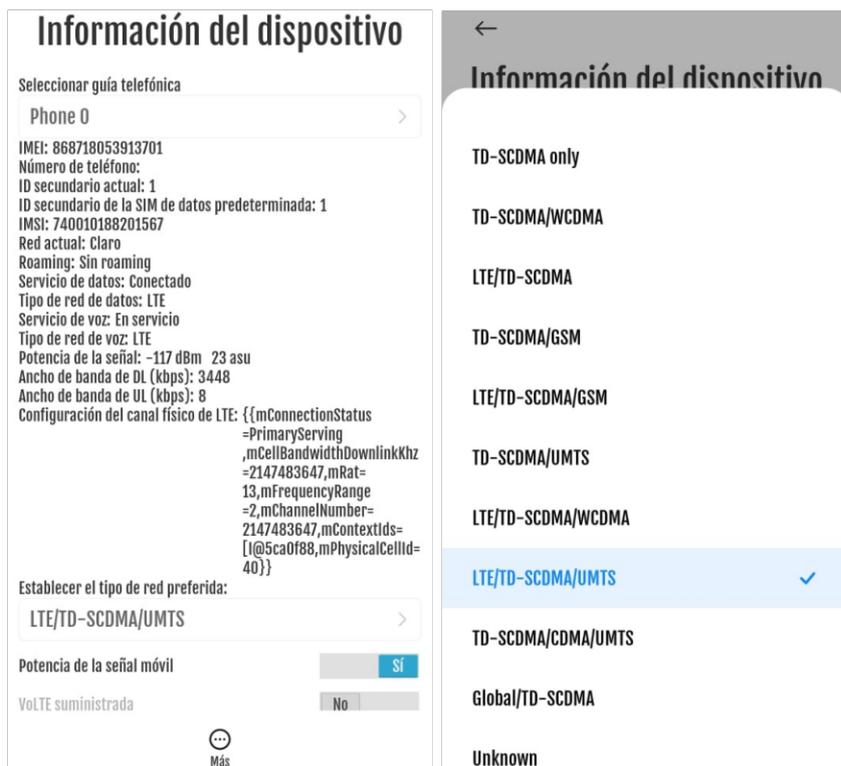


Figura 1.5: Información y configuración del dispositivo en ForceLTE.

La siguiente ventana presenta un medidor de señal que permite conocer la intensidad de la conexión de la red una vez realizada la configuración anterior. Finalmente, la tercera ventana es de configuraciones, donde además se tiene un manual de uso de la aplicación, se tiene la opción de cambiar la interfaz gráfica y se pueden leer las políticas de la aplicación.

1.4.1.3 CellMapper

Esta aplicación tiene como objetivo principal mostrar información avanzada de las redes celulares GSM/CDMA/UMTS/LTE/LTE-A, mapas de cobertura celular y ubicación de sus torres, además permite cargar este tipo de datos para contribuir a los mapas de cobertura celular [4]. Esta aplicación cuenta con varias opciones dentro de su menú principal, como por ejemplo la visualización de las celdas y torres dentro del mapa, entre otras. En la primera ventana llamada "celda" se obtiene información de las antenas más cercanas al punto donde se está utilizando el dispositivo, tal como se muestra a continuación:

eNB 100392:12 - LTE	-109 dBm
MCC: 740 MNC: 1 TAC: 50200	
LTE Reference Signal Received Power	-109 dBm
LTE Reference Signal Received Quality	-16 dB
LTE Receive Signal Strength	-73 dBm
Band Name	AWS-1
RX Frequency	2140.0 MHz
Band Number	4
TX Frequency	1740.0 MHz
LTE Cell Identity	25700364
(EA/UA/A)RFCN	2250
LTE Physical Cell Identity	40

Figura 1.6: Información presentada en la ventana “Celda” de CellMapper.

Entre la información que se presenta en la figura anterior se encuentra la potencia y frecuencia de recepción, números de identificación, entre otras características. Además, se presentan las opciones para subir o grabar datos.

Por otro lado, en la ventana “Mapa” se encuentran todas las antenas ubicadas en diferentes puntos dentro de la ciudad. Es importante mencionar que además de visualizar las antenas en el mapa, se puede obtener información detallada de dichos elementos, como, por ejemplo, en la figura que se muestra a continuación, se presenta tanto la ubicación, patrón de radiación e información de la antena. Para obtener esta información se debe seleccionar en el mapa la ubicación de la antena y desplazar hacia arriba la ventana que aparece en la parte inferior.

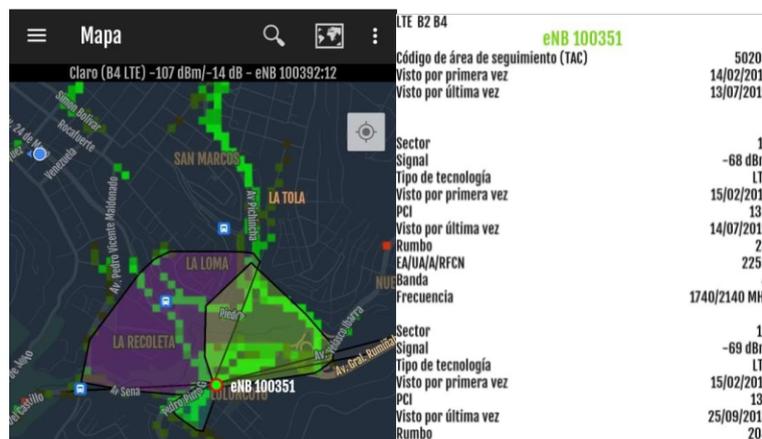


Figura 1.7: Antenas y sus datos en el mapa de CellMapper.

Además, esta aplicación permite realizar búsquedas usando diferentes filtros como por ejemplo por bandas, o también permite realizar búsqueda si es que se tiene información de las antenas, con el objetivo de facilitar el estudio que se está realizando.

1.4.1.4 Network Cell info Lite

La aplicación Network Cell Info Lite permite conocer la ubicación de la señal en el mapa obtener los datos de la potencia de la señal y señales vecinas. La ventana principal de esta aplicación es la ventana llamada "Medir", en esta ventana se presenta la intensidad de la señal junto con la información de la celda y celdas vecinas, como también el tipo de red inalámbrica y nombre del proveedor de red. En la siguiente figura se presenta un ejemplo de esta ventana:



Figura 1.8: Medir en Network Cell Info Lite.

Además, presenta información del estado del servicio, datos móviles e indicaciones de encendido/apagado de itinerancia⁴, consta de un indicador de intensidad de la señal y "LCD" de intensidad y calidad de la señal. Además, entrega detalles de identificación celular, información de señales Wifi, entre otra información [5].

La siguiente ventana es la llamada "Velocidad", en la que se puede visualizar información sobre la velocidad de la conexión, tal como se muestra a continuación.



Figura 1.9: Informe de velocidad de conexión en Network Cell Info Lite.

⁴ La itinerancia o Roaming hace referencia a la posibilidad de utilizar redes móviles fuera del área de servicio local de la propia compañía, es decir, se puede realizar o recibir llamadas y mensajes, o navegar por internet, dentro de la zona de servicio de otra empresa del mismo (Roaming Nacional) o diferente (Roaming Internacional) país [17].

Entre la información que se obtiene se puede mencionar [5]:

- **Velocidad de carga y descarga:** muestra qué tan rápido el dispositivo envía o extrae datos del servidor⁵ (Mbps).
- **Ping:** mide el tiempo de reacción de la conexión, es decir que muestra la rapidez con la que obtiene una respuesta después de enviar una solicitud. También es conocido como latencia.
- **Jitter:** mide la variación del ping a lo largo del tiempo.

Un resumen de la información obtenida se presenta en la ventana, es decir que, se puede visualizar la información de la red de una manera más amigable para el usuario.

En su última ventana “Mapa” se presentan las mediciones y además se proporciona la ubicación visual de la celda, de tal manera que, en el mapa se observa el punto donde se encuentra el dispositivo y la ubicación de la celda de servicio, es importante mencionar que esta no es la ubicación de la torre.



Figura 1.10: Mapa en Network Cell Info Lite

1.4.1.5 G-NetTrack Pro

Esta aplicación es un monitor de red inalámbrica para dispositivos con sistema operativo Android, a diferencia de las aplicaciones anteriores necesita de una licencia para trabajar con todas sus opciones. Principalmente permite monitorear y registrar los parámetros de la red móvil sin utilizar equipos especializados. Al igual que las aplicaciones anteriores, permite obtener una mejor visión de la red, mostrando información detallada de la misma. La diferencia entre esta aplicación con las demás, es que presenta datos adicionales en

⁵ Servidor más cercano al punto de medición, el dispositivo descargara y enviara un archivo a ese servidor para calcular la velocidad respectiva [18].

sus archivos kml⁶, como por ejemplo la velocidad de descarga y de subida, cuyas mediciones se realizan de manera continua.

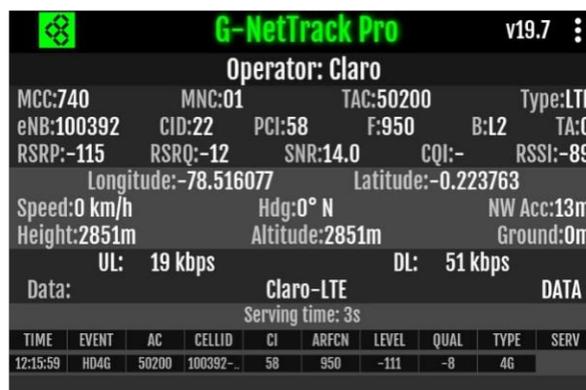


Figura 1.11: Ventana principal de G-NetTrack Pro.

En la figura 1.11 se pueden observar los datos que esta aplicación entrega en el dispositivo; al comparar con las aplicaciones descritas anteriormente se puede observar que presenta los mismos datos de la red, señal, antenas y distancias, por ejemplo : MCC (Mobile Country Code), MNC (Mobile Network Code), CID (Identificador de Celda), Tipo de red, RSRP (Reference Signal Received Power), RSRQ (Reference Signal Received Quality), SINR (Signal to Interference plus Noise Ratio), entre otros que se enumeran a acontinuacion [6] [7]:

- **Timestamp:** tiempo en que se realizó la muestra.
- **Longitude y Latitude:** Coordenadas GPS del dispositivo móvil.
- **Speed:** velocidad del dispositivo móvil expresados en km/h.
- **Operatorname:** nombre de la operadora celular.
- **CellId:** Serving cell for mobile device.
- **NetworkMode:** estándar de comunicación móvil (2G/3G/4G).
- **RSSI:** Representa la potencia recibida que incluye una celda de servicio, la interferencia y el ruido de otras fuentes.
- **RSRP:** Representa una potencia promedio sobre los símbolos de referencia específicos de celda transportados dentro de distintos RE⁷.
- **RSRQ:** Representa una relación entre RSRP y el indicador de intensidad de la señal recibida (RSSI).
- **SNR:** valor de la relación señal/ruido (dB).

⁶ Archivos KML: Formato de archivos para presentar datos en un navegador terrestre como Google Earth [19].

⁷ RE (Resource Element): unidad más pequeña de asignación de recursos, representa el uso de una portadora y un símbolo.

- **DL_bitrate:** tasa de descarga medida en el dispositivo expresada en kbps.
- **UL_bitrate:** tasa de enlace ascendente medida en el dispositivo expresada en kbps.

Una de las ventanas más importantes de esta aplicación es la llamada “MAP”, ya que permite tomar datos en pruebas OUTDOOR, en esta ventana (Figura 1.12) también se especifican por colores los niveles de la señal, además presenta la opción para especificar que señal se desea observar durante la medición. [6].

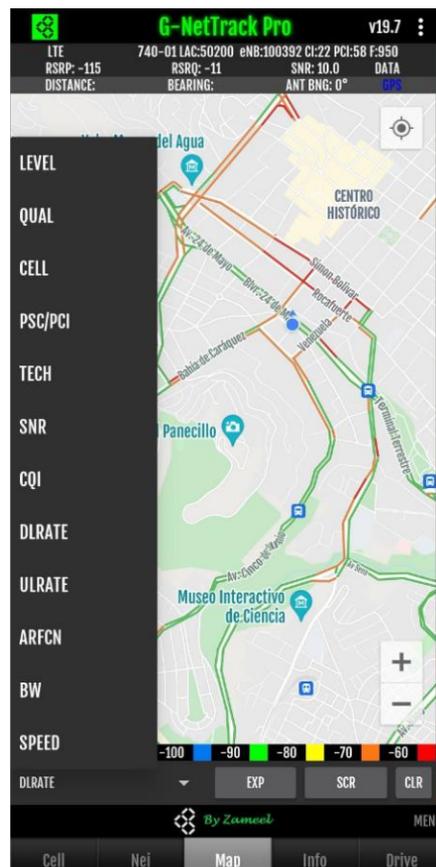


Figura 1.12: Opción MAP en G-NetTrack Pro.

1.1.2 TERMINAL MÓVIL

Las características del terminal móvil que se utilizarán para la recolección de datos son las siguientes [8]:

SISTEMA:

Modelo del dispositivo: Redmi Note 9 Pro

Memoria RAM: 6 GB

CPU: Ocho núcleos máximo 2.32 GHz

Sistema Operativo: MIUI V11 (Android 10)

Procesador: Qualcomm 720G.

TARJETA SIM:

Cantidad de tarjetas SIM soportadas: 2

Tipo de tarjeta SIM: Nano-SIM (4FF - fourth form factor, since 2012, 12.30 x 8.80 x 0.67 mm)

Otras características:

- Doble SIM en reposo (cuando una esté ocupada la otra no estará activa)

REDES:

LTE, GSM, CDMA, W-CDMA

Otras características:

- UMTS (384 kbit/s)
- EDGE
- GPRS
- HSPA+
- LTE Cat 12 (102.0 Mbit/s, 603.0 Mbit/s)
- EV-DO Rev. A (1.8 Mbit/s, 3.1 Mbit/s)
- TD-SCDMA TD-HSDPA

CONECTIVIDAD:

Estándares wifi-soportados:

802.11a - 802.11b - 802.11g - 802.11n – 802.11ac

Otras características:

- Sincronización con la computadora
- OTA sync
- Infrarrojo
- Anclaje a red (tethering)
- NFC
- VoLTE
- ViLTE y VoWiF

SEGUIMIENTO UBICACIÓN:

GPS, A-GPS, GLONASS, Galileo, BeiDou, NavIC

1.1.3 HERRAMIENTA DE ANÁLISIS DE DATOS

A continuación, se presentan las herramientas instaladas en el computador (Figura 1.13), mismas que se utilizarán para el análisis de datos en este estudio.



Figura 1.13: Herramientas para el análisis de datos.

1.4.3.1 R y RStudio

R es un software libre que permite realizar análisis estadísticos y gráficos. Es considerado como uno de los lenguajes de programación más completos y sencillos en el mercado, además, permite al usuario obtener prototipos de modelos con buenos resultados y en corto tiempo. Solo cuenta con una consola que se presenta en la Figura 1.14, donde se realizan las líneas de programación, en esta consola se puede utilizar diferentes comandos, algoritmos, operaciones entre otras funciones.

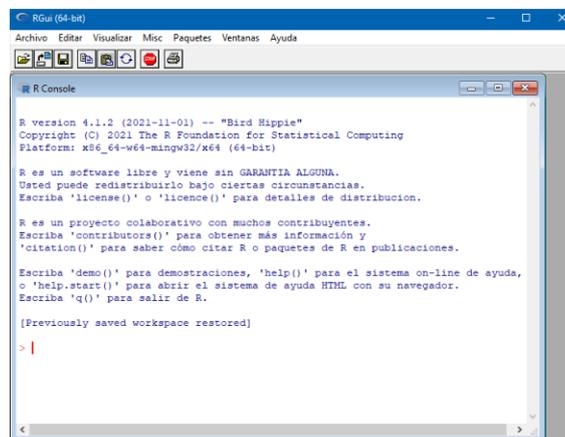


Figura 1.14: Consola de R

Por otro lado, se tiene RStudio, este es un IDE, es decir que es un software que integra R en una interfaz gráfica más amigable para los usuarios. RStudio presenta varios paneles que permiten mantener una mejor visibilidad al momento de programar, además presentan más opciones para sintetizar el trabajo y realizarlo en menos tiempo [9]. Sus paneles principales son:

- **Script:** es el cuaderno de trabajo donde se realiza la programación. Se sitúa en la parte superior-izquierda de RStudio tal como se muestra en la Figura 1.15. Las líneas de comandos escritas en este panel se pueden ejecutar; una a una utilizando el botón llamado Run o en bloques utilizando el comando Ctrl+r [9].
- **Consola:** se sitúa en la parte inferior-izquierda. En este panel se ejecutan las instrucciones creadas en el script y muestra los resultados o avisos de error. Si se escribe la instrucción en la consola, se debe presionar Enter para ejecutarla y obtener el resultado [9].
- **Entorno:** se sitúa en la parte superior-derecha del escritorio de RStudio y está compuesto por [9]:
 - **Environment:** se registran los objetos que se crean en el script. También presenta la opción para cargar y guardar una sesión de trabajo, importar datos y limpiar los objetos de la sesión.
 - **History:** se registran las instrucciones que se ejecutaron.
 - **Connections:** permite conectarse a diferentes fuentes de datos.
 - **Tutorial:** presenta un manual donde se presenta información del software para poder usarlo.
- **Archivos:** se sitúa en la parte inferior-derecha del escritorio de RStudio y está compuesto por [9]:
 - **Files:** es un explorador de archivos y se visualizan los archivos que se encuentran en la dirección establecida en Set Working Directory.
 - **Plots:** Ventana que presenta los gráficos resultantes. Consta de varias opciones como Zoom, Export y opciones para limpiar la ventana.
 - **Packages:** presenta un listado de los paquetes instalados y cargados en la sesión de trabajo. También presenta opciones para cargar e instalar nuevos paquetes.
 - **Help:** información de comandos, funciones o paquetes.

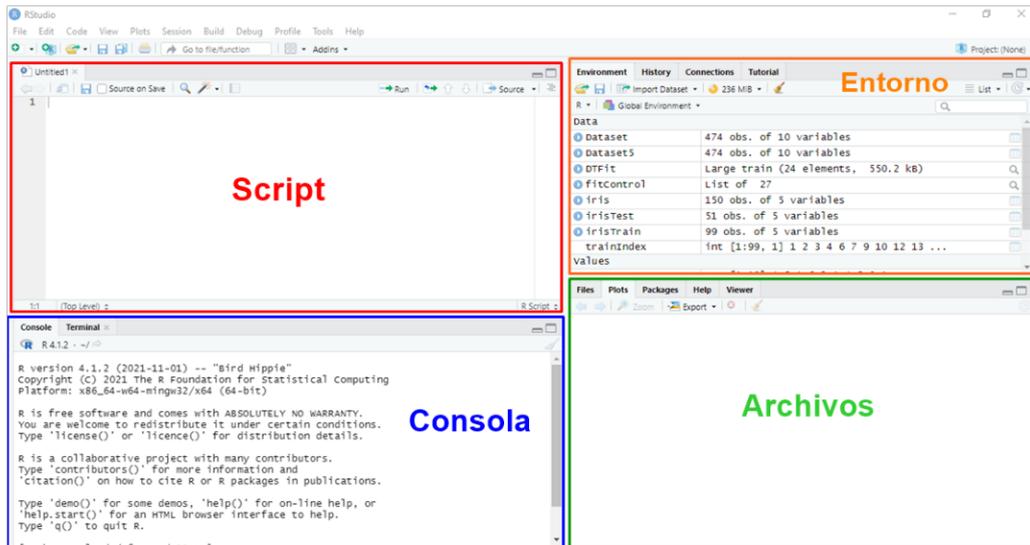


Figura 1.15: Paneles principales de RStudio.

1.4.3.2 Google Earth Pro

Este software libre se utilizará para visualizar los datos otorgados por G NetTrack Pro. Como se mencionó en la sección de herramientas para la recolección de datos, esta herramienta exporta archivos en formatos txt y kml. Los archivos kml se abren en Google Earth Pro, mostrando el recorrido realizado, además, se observan puntos con diferentes colores dependiendo de la señal y en cada punto se muestra una información detallada, tal como se muestra en la figura 1.16. Esta herramienta se utilizará principalmente para tener una noción acerca de la señal en las diferentes rutas y así poder realizar la clasificación del parámetro que se va a predecir.

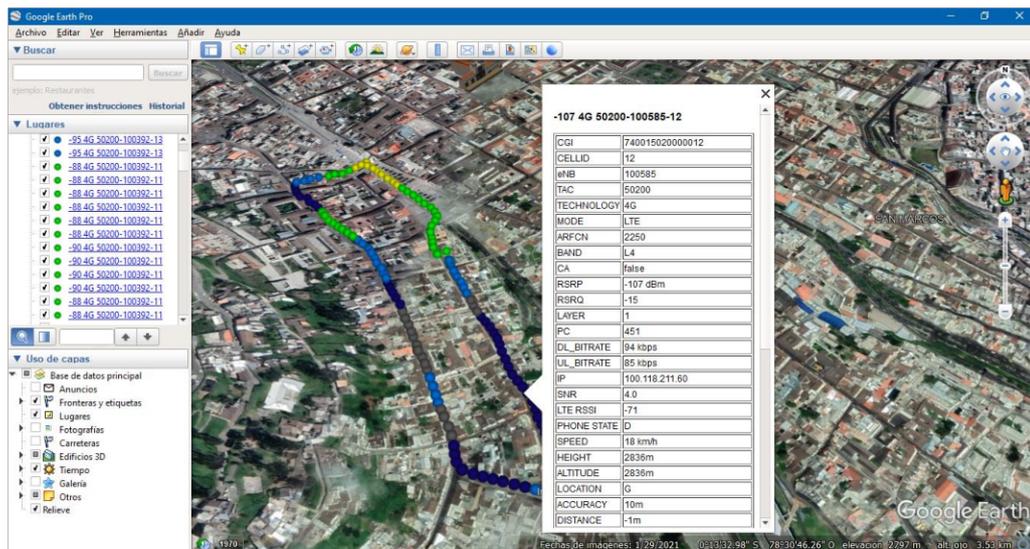


Figura 1.16: Análisis de datos en Google Earth Pro.

2. METODOLOGÍA

2.1 DEFINICIÓN DE LOS DATOS A RECOLECTAR Y PROCESAR

Con los datos recolectados se pretende encontrar el o los tipos de relaciones entre la velocidad de conexión y la velocidad de movimiento del terminal móvil y observar el comportamiento de los parámetros técnicos que pueden ser recolectados, mediante el uso de herramientas de monitoreo disponibles y usadas en este proyecto. La recolección de datos se realiza mientras se mantiene permanentemente un tipo de comunicación en la red celular, es decir, mediante la ejecución de una conexión de datos constante con la red.

Para lograr el objetivo mencionado, se deben identificar y definir todos los parámetros que se pueden obtener gracias al monitoreo de la red con las herramientas disponibles. Entre las herramientas que se utilizarán se tiene principalmente: **Network Cell Info Lite** y **CellMapper**, ambas ofrecen diferentes parámetros como, por ejemplo: mcc, mnc, cellid, lat, lon, entre otros. A continuación, se presenta un resumen de todos los parámetros otorgados por la herramienta Network Cell Info Lite:

Tabla 2.1 Parámetros obtenidos con la herramienta Network Cell Info Lite [10].

PARÁMETROS	DESCRIPCIÓN	UNIDADES / POSIBILIDADES
SIM	Operadora del terminal móvil.	0, 1, 2, o 10
RADIOTYPE	Estándares de telecomunicaciones de radio.	GSM, UMTS, LTE, CDMA.
RADIO	Tipo de radio real.	EDGE, 1xRTT, HSPA, LTE, etc.
MCC	Código de país móvil.	740 (Ecuador).
MNC	Código de red móvil.	0, 1, 2, o 10
AREA	Código de área de seguimiento/ubicación (LAC/TAC).	Valores hexadecimales con una longitud de dos octetos.
CELLID	Identidad celular (ECI).	Valor Único.
UNIT	Identidad celular física LTE (PCI).	Valor entre 0 y 503.
LAT	Ubicación de la medición (Latitud y longitud).	Grados.
LON		
SIGNAL	Potencia de la señal (RSRP).	dBm.
EXTRA	(Reservado)	-
ACC	Precisión de la ubicación del dispositivo, en el momento de la medición.	Metros.
TIME	Marca de tiempo de la medición.	Hora UTC en milisegundos.
SPEED	Velocidad del terminal móvil en el momento de la medición.	Km/h.

BEARING	Dirección de desplazamiento del dispositivo en el momento de la medición.	Grados.
ALT	Altitud del terminal móvil en el momento de la medición.	Metros.
API	Interfaz de programación de aplicaciones.	Versión de Android (29 para Android 10).
DEVICE	Nombre del Dispositivo.	Xiaomi Redmi Note 9 Pro

Por otro lado, G-NetTrack Pro otorga datos adicionales sobre la red LTE, estos datos serán los que permitan realizar el análisis del proyecto, ya que se toman secuencialmente sin interrupciones. A continuación, se presentan los datos de más importancia otorgados por la aplicación:

Tabla 2.2 Parámetros obtenidos con la herramienta G-NetTrack Pro. [6]

PARÁMETROS	DESCRIPCIÓN	UNIDADES
LEVEL	Potencia recibida de la señal de referencia. En una red celular LTE tiene el nombre de RSRP.	dBm.
QUAL	Calidad de la señal de referencia. En una red celular LTE tiene el nombre de RSRQ.	dB.
SNR	Relación señal / ruido. Se mide solo en LTE.	dB.
LTERSSI	Potencia recibida en toda la banda	dB.
UL	La velocidad de transferencia de datos de enlace ascendente actual.	Kbps.
DL	La velocidad de transferencia de datos de enlace descendente actual	Kbps.

En la tabla 2 se presentan los datos que se utilizarán en el proyecto, es decir que estos no son todos los datos otorgados por G-NetTrack Pro. También hay información acerca de la fecha, longitud y latitud del punto de medición, información de la estación base a la que se conecta, información de celdas vecinas, entre otras, pero, estos datos no se mencionan en la tabla para evitar redundancia con los datos presentados en la tabla 1.

2.2 DEFINICIÓN DE LA ZONA A CONSIDERAR Y/O LAS RUTA

Las mediciones se realizarán en el centro de Quito, considerando diferentes velocidades del terminal móvil en una misma ruta. Por lo tanto, se trabajará en días, horas y rutas de poco tráfico vehicular, ya que el objetivo principal es analizar las relaciones existentes entre las velocidades del móvil con las velocidades de conexión independientemente de factores externos como las características geográficas, de las edificaciones, de las calles, entre otras. A continuación, se indican las rutas seleccionadas:

- **Primera Ruta:** Ruta de la Figura 2.1 de 1.60 kilómetros.

La primera ruta se ubica entre las calles; Bahía de Caráquez y Antonio Borrero, esta ruta, al igual que las dos siguientes, se encuentra dentro del Centro Histórico de Quito. Además, es importante mencionar que en estas calles se encuentran dos semáforos y dos rompe velocidades, esta información permitirá realizar consideraciones en las mediciones.

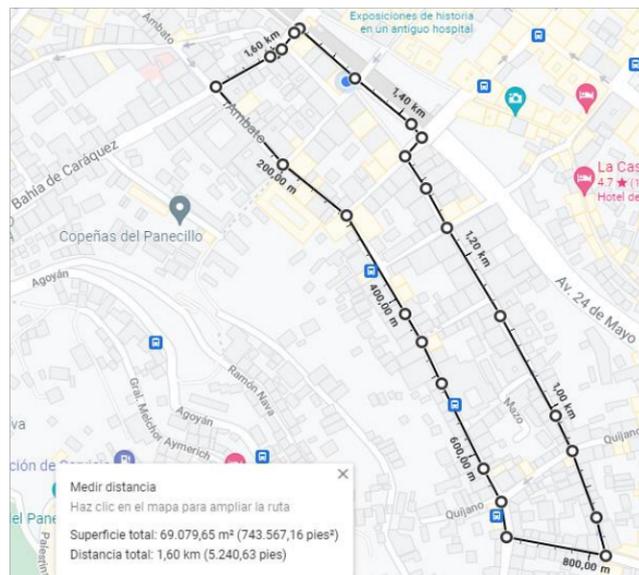


Figura 2.1 Ruta A de 1.60 Km

- **Segunda Ruta:** Ruta de la Figura 2.2 de 1.20 kilómetros.

La segunda ruta se ubica entre las calles; Imbabura y Cumandá. En esta ruta se encuentra un solo semáforo y varias intersecciones de vías con alto tráfico vehicular.

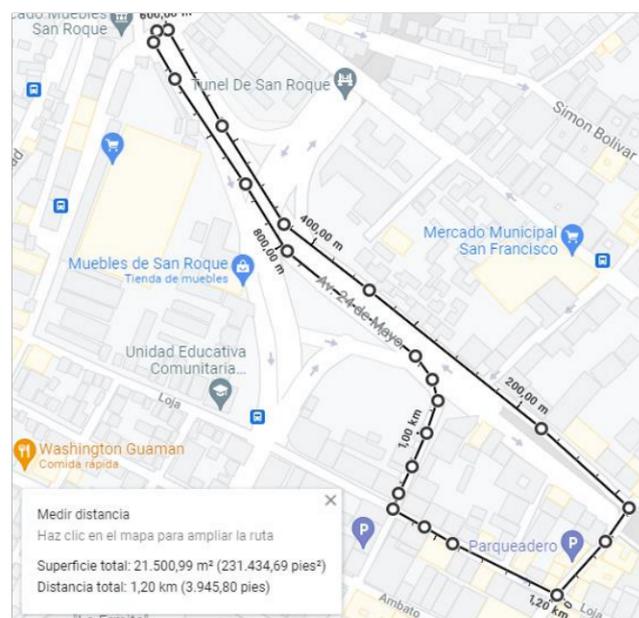


Figura 2.2 Ruta B de 1.20 Km

- **Tercera Ruta:** Ruta de la Figura 2.3 de 3.95 kilómetros.

Se establece esta tercera ruta, ya que permite realizar mediciones con velocidades por encima del máximo permitido (40Km/h) dentro de la ciudad. Comprende desde la calle Cumandá hasta la Avenida Santa Cruz. En esta ruta se encuentran 2 semáforos, 3 rompe velocidades y un redondel.

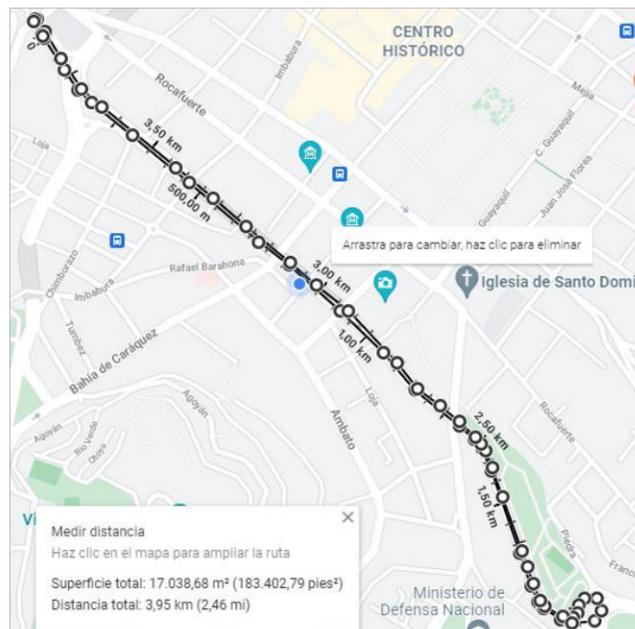


Figura 2.3 Ruta C de 3.95 Km

Como ya se mencionó anteriormente, estas rutas fueron escogidas debido a que presentan facilidades para el objetivo de la recolección de datos. Se buscaron rutas que permitan otorgar seguridad a los ciudadanos y al estudiante encargado de las mediciones.

2.3 PLANIFICACIÓN DE LA RECOLECCIÓN DE DATOS

Considerando que el tiempo que se ha programado para la etapa de recolección de datos, comprende desde el 15 de diciembre hasta el 7 de enero se establece que se utilizará un día para realizar mediciones en una misma ruta a dos diferentes velocidades. Se utilizará un día para organizar los archivos, donde se establece una base de datos organizada por fechas, rutas y velocidades. También se tiene previsto contar con un día para realizar correcciones en las mediciones, es decir, por ejemplo, si al organizar los archivos se observa que algún dato está erróneo, no se grabó correctamente o presenta cualquier inconveniente, se utilizará este día para volver a realizar las mediciones en dichas condiciones. Las velocidades previstas son; Para la ruta A y B, se utilizará una velocidad máxima de 40 Km/h (ya que la velocidad máxima permitida dentro de la ciudad es 50Km/h). Para la ruta C, se empleará hasta una velocidad de 80 Km/h, ya que esta ruta comprende

un túnel (Viaducto 24 de Mayo) en donde no existe tráfico vehicular ni controles de velocidad.

El cronograma propuesto para la etapa de recolección de datos es el presentado en el ANEXO 1.

2.4 TÉCNICAS DE MACHINE LEARNING

El análisis de los datos se realizará mediante la técnica de Machine Learning conocida como de aprendizaje supervisado. Mediante esta técnica, se realizará la clasificación de datos y específicamente se utilizará el método de reglas basadas en partición o mejor conocidas como árboles de decisión, donde se establecerán reglas y condiciones que permitan clasificar los datos para obtener un resultado.

- **Aprendizaje supervisado**

Es un tipo de algoritmo de aprendizaje en donde se necesita de un entrenamiento mediante el uso de datos. Con esta técnica de Machine Learning se encuentra una función matemática que relacione a los datos de entrada y salida, con el objetivo de predecir el comportamiento al ingresar nuevos datos. En estos datos además de contener una gran cantidad de ejemplos de entrenamiento, también deben constar las respuestas o predicciones que se buscan obtener con el algoritmo [11].



The diagram shows a table with 7 columns: Speed, RSRP, RSRQ, SNR, LTERSSI, and Calidad. A red bracket above the first five columns is labeled "Posibles entradas" with an upward arrow. A red arrow above the "Calidad" column is labeled "Salida".

	Speed	RSRP	RSRQ	SNR	LTERSSI	Calidad
1	0	-95	-19	0	-55	Optima
2	0	-95	-19	0	-55	Optima
3	0	-88	-14	9	-53	Optima
4	18	-88	-14	9	-53	Optima
5	14	-88	-14	10	-53	Optima
6	17	-88	-14	10	-53	Optima
7	14	-88	-14	11	-53	Optima
8	14	-88	-14	11	-53	Optima
9	16	-90	-10	11	-59	Optima
10	13	-90	-10	6	-59	Optima

Figura 2.4 Ejemplo de datos para utilizar en el algoritmo de Machine Learning.

En el ejemplo de la figura 2.4 se presentan algunos datos tomados en una de las rutas establecidas para este proyecto. Se puede observar que en las 5 primeras columnas se presentan diferentes datos como Speed (Velocidad del terminal móvil), RSRP, RSRQ, SNR y RSSI, mientras que en la última columna se presenta la predicción que se desea obtener.

El aprendizaje supervisado se subdivide en dos tipos:

- **Clasificación:** permite predecir las etiquetas de clase categóricas de nuevos registros con base al entrenamiento realizado. Se puede tener una clasificación binaria o multiclase. Cuando la etiqueta es discreta (solo dos clases) se trata de clasificación binaria, y si existen más de dos clases se tendrá una clasificación multiclase [11][12].
- **Regresión:** permite predecir un valor continuo mediante la relación entre variables dependientes e independientes, es decir que es un proceso estadístico predictivo que busca una ecuación en la que se sustituyen los valores de las variables y como resultado se obtiene el valor que se desea predecir [12][13].

Para realizar el entrenamiento mediante esta técnica, se pueden utilizar varios algoritmos, como, por ejemplo:

Para aprendizaje supervisado por clasificación se tienen las siguientes técnicas [11]:

- ✓ **Árboles de decisión:** Algoritmo donde se evalúan valores en cada nodo para llegar a una clasificación final, es decir que se asemeja a un diagrama de flujo que consta de varias raíces y ramas, tal como se muestra en la figura 2.5, además, su lectura se realiza de arriba hacia abajo. Para realizar el aprendizaje se necesita de una variable dependiente en el conjunto de entrenamiento.

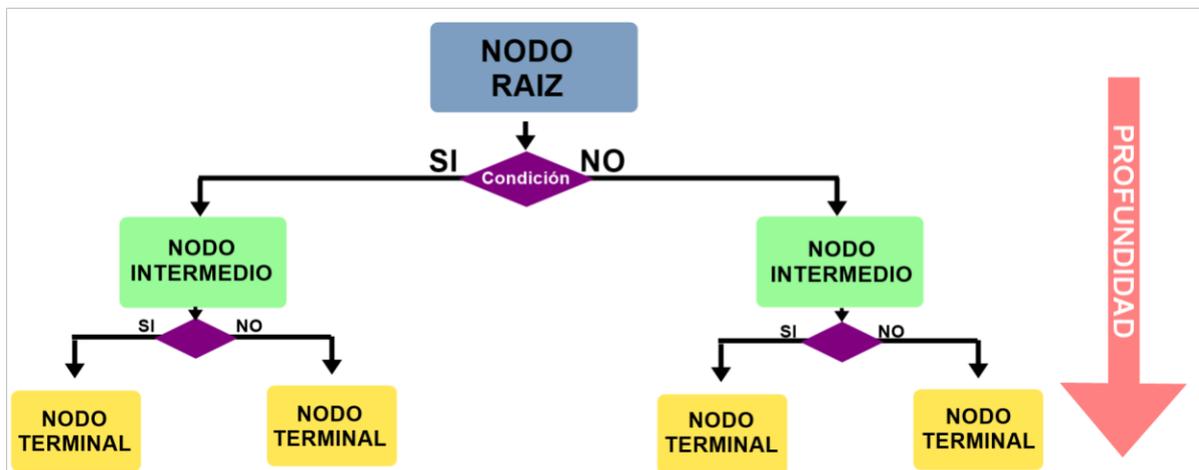


Figura 2.5. Partes de un árbol de decisiones

- ✓ **Support Vector Machine (SVM):** puede ser usando tanto para clasificación como para regresión. En este algoritmo se construye un hiperplano que separa las clases de datos.
- ✓ **Random Forest:** combina una gran cantidad de árboles de decisión independientes entre sí, por eso el nombre de "bosque". Su objetivo es reducir la varianza.

Para aprendizaje supervisado por regresión se pueden mencionar las siguientes técnicas:

- ✓ **Regresión lineal:** es considerado como uno de los algoritmos más sencillo y consiste en utilizar el método de los cuadrados mínimos para realizar el entrenamiento, además, otorga una línea recta al graficar la ecuación [14].
- ✓ **Regresión logística:** algoritmo usado en problemas de clasificación binaria, teniendo en cuenta que no deja de ser una regresión, debido a que el resultado de la ecuación es la probabilidad de que pertenezca a una clase, que se clasifica como positivo o negativo, dependiendo del umbral que se utilice [12][14].

En este proyecto se realizará el análisis mediante el aprendizaje supervisado por clasificación utilizando los árboles de decisión, donde se establecerán reglas y condiciones que permitan clasificar los datos para obtener un resultado determinado. Se utilizará este algoritmo debido a que presenta varias ventajas al programar en R, por ejemplo, son fáciles de construir, interpretar y visualizar, además, permite seleccionar las variables más importantes, es decir, que en dicho algoritmo no siempre se utilizan todas las entradas, esto permitirá que la predicción solo relacione variables que tengan relevancia [12].

2.5 CÓDIGO EN R PARA EL ANÁLISIS DE LOS DATOS RECOLECTADOS

Primero se establece el directorio donde se encuentran todas las mediciones, además se instalan y se cargan las librerías que se utilizarán:

```
1 #Michael Alexander Curipallo Martinez
2 #Se establece el directorio donde se encuentran los datos
3 setwd("C:/Users/Usuario/Desktop/EPN/Tesis/DATOS_Michael_Curipallo")
4 #Librerías
5 library(dplyr)
6 library(rpart)
7 #install.packages("rpart.plot")
8 library(rpart.plot)
9 #install.packages("rattle")
10 library(rattle)
11 library(RColorBrewer)
12 library(scatterplot3d)
13 library(rgl)
```

Figura 2.6 Líneas de código para establecer el directorio e instalar las librerías.

En las siguientes líneas de comando se importan y depuran los datos que se utilizarán para el entrenamiento, estos datos son recolectados en las tres rutas ya especificadas. Primero se abren todos los archivos csv que se encuentran en la carpeta, estos archivos cuentan con nombres fáciles de entender, su estructura es la siguiente:



Figura 2.7 Ejemplo de estructura del nombre de los archivos csv para entrenamiento.

Posteriormente, desde la línea 19 a la 24 de la Figura 2.8, se establecen las columnas que se van a analizar y se crea la columna que se va a predecir. Estos datos se unirán en una sola matriz llamada “total”. Por otro lado, para la condición que se va a predecir, se establecen tres rangos de acuerdo con la Tabla 3, evaluando la velocidad de descarga en la conexión, es por esa razón que este parámetro se elimina en la línea 39 de los parámetros de entrada.

Tabla 2.3. Condición para crear el parámetro de salida

Rangos de velocidades de descarga [kbps]	Calidad
Mayores a 5000	Excelente
De 1000 a 5000	Regular
Menores a 1000	Mala

```

15 #ENTRENAMIENTO
16 #Se cargan todas las mediciones en un solo objeto.
17 datos <- lapply(sys.glob("Ruta?/R??0_?.csv"), read.csv, sep="")
18 #Depuracion de datos
19 j<-length(datos)
20 for (i in 1:j) {
21   datos[[i]] <- datos[[i]] %>%
22     select(c(Speed, NetworkMode, Level,Qual, CQI, ARFCN, DL_bitrate))
23   datos[[i]]$calidad <- "Pesima"
24 }
25
26 #Mezclar los dataframes en la lista
27 total<- do.call("rbind",datos)
28 total<-rename(total,
29               RSRP=NetworkMode,
30               RSRQ=Level,
31               SNR = Qual,
32               LTERSSI=CQI,
33               DL_bitrate=ARFCN,
34               UL_bitrate=DL_bitrate)
35
36 #Condiciones para establecer las 3 categorias
37 total$calidad[total$DL_bitrate >= 5000]<- "Excelente"
38 total$calidad[total$DL_bitrate >= 1000 & total$DL_bitrate < 5000]<- "Regular"
39 total$calidad[total$DL_bitrate < 1000 ]<- "Mala"
40
41 #Se elimina las columnas que ya se usaron para las condiciones
42 total<-select(total, -c(DL_bitrate, UL_bitrate))

```

Figura 2.8 Líneas de código para crear la matriz total que sirve para el entrenamiento.

Después de crear la matriz total que contiene todos los datos necesarios para el entrenamiento, se procede a programar el árbol de decisión. Estableciendo primero la cantidad de datos destinados para entrenamiento y para pruebas. Para esto, se crea una función que permita considerar la división 80/20, esto significa que; el 80% del total de los datos se utilizará para entrenar el modelo, mientras que el 20% del total para hacer las predicciones. La función llamada “create_train_test” define tres argumentos [15]:

- data: conjunto de datos utilizados para entrenar el modelo.
- size: Tamaño del split. Por defecto, 0.8.
- train: si se establece en “TRUE”, la función crea el conjunto de entrenamiento, de lo contrario, el conjunto será para las predicciones. El valor predeterminado se establece en “TRUE”.

Es importante mencionar que; en la línea 45 del código, se utiliza la función “*sample*” para mezclar los datos almacenados en la matriz total, con el objetivo de tener una variación de datos cuando se vaya a realizar el modelo. Sin esta línea, siempre se utilizarán las dos primeras rutas para entrenamiento y la tercera ruta para las predicciones, por lo que, se producirá un modelo con baja exactitud, ya que se tiene falta de información para velocidades mayores a los 40 Km/h.

```

44 #Datos de entrenamiento y prueba
45 total <- total[sample(1:nrow(total)), ] #Se mezclan aleatoriamente los datos
46 create_train_test <- function(data, size =0.8, train=TRUE) {
47   n_row = nrow(data)#numero de filas
48   total_row = size * n_row #Devuelve la enésima fila para construir el conjunto de entrenamiento
49   train_sample <-1: total_row #Se selecciona desde la primera fila hasta la enésima fila
50   if (train == TRUE) {
51     return (data[train_sample, ])#Si es VERDADERO se crea un conjunto para entrenar el modelo
52   } else {
53     return (data[-train_sample, ])#Si es FALSO se crea el conjunto para realizar las pruebas
54   }
55 }
56 #80% de datos para el entrenamiento
57 train <- create_train_test(total, 0.8, train = TRUE)
58 #20% de datos para predicciones
59 test <- create_train_test(total, 0.8, train = FALSE)

```

Figura 2.9 Líneas de código para crear el conjunto de datos destinados al entrenamiento y pruebas.

Después de crear ambos conjuntos de datos, se crea el árbol de decisión, para esto se utiliza la función “*rpart*”, en esta función se deben establecer los siguientes parámetros [15]:

- *formula*: se especifica la variable que se va a predecir y con qué datos se va a trabajar.
- *data*: datos que se usarán para realizar el entrenamiento
- *method*: se define el algoritmo con el que se va a trabajar, puede ser *anova*, *poisson*, *class* o *exp* dependiendo de la aplicación, en este caso se usará *class* debido a que la salida es categórica.
- *minsplit*: el número mínimo de observaciones que deben existir en un nodo para que se intente una división.
- *minbucket*: el número mínimo de observaciones en cualquier nodo terminal.

Como ya se mencionó anteriormente, el objetivo de este proyecto es relacionar la velocidad de cobertura con la velocidad del terminal móvil, y, al ya utilizar la velocidad de cobertura para crear la condición que define el parámetro de salida, solo se tendría a la velocidad del

terminal móvil como única variable de entrada, pero, al realizar el entrenamiento con estas condiciones se obtiene el siguiente error:

```
Error in apply(model$frame$yval2[, yval2per], 1, function(x) x[1 + x[1]]) :
dim(x) must have a positive length
```

Figura 2.10 Error al usar solo la velocidad del terminal móvil como parámetro de entrada.

Este error se debe a que al utilizar la columna Speed (velocidad del terminal móvil) como única entrada, no es posible obtener un árbol con más de un nodo, es decir que se necesita de algún otro parámetro adicional para obtener un árbol con más ramas (mayor profundidad de análisis). Por esta razón, se especifica como parámetros de entrada algún otro parámetro adicional junto con la velocidad del terminal móvil, tal como se muestra en la figura 2.11:

```
61 #Creacion del arbol
62 arbol <- rpart(Calidad ~Speed +RSRQ+SNR+LTERSSI,
63 data = train,
64 method = "class",
65 minsplit = 1,
66 minbucket = 1)
67 fancyRpartPlot(arbol) #Gráfica del árbol
68 rpart.rules(arbol) #Condiciones del árbol
69 arbol #Se muestra en Consola los datos de cada rama
```

Figura 2.11 Líneas de código para definir los atributos de entrada.

Finalmente, se utilizan comandos que permite visualizar diferente información acerca del árbol de decisión, como, por ejemplo, el gráfico, sus condiciones y también se imprimirá en consola el modelo creado.

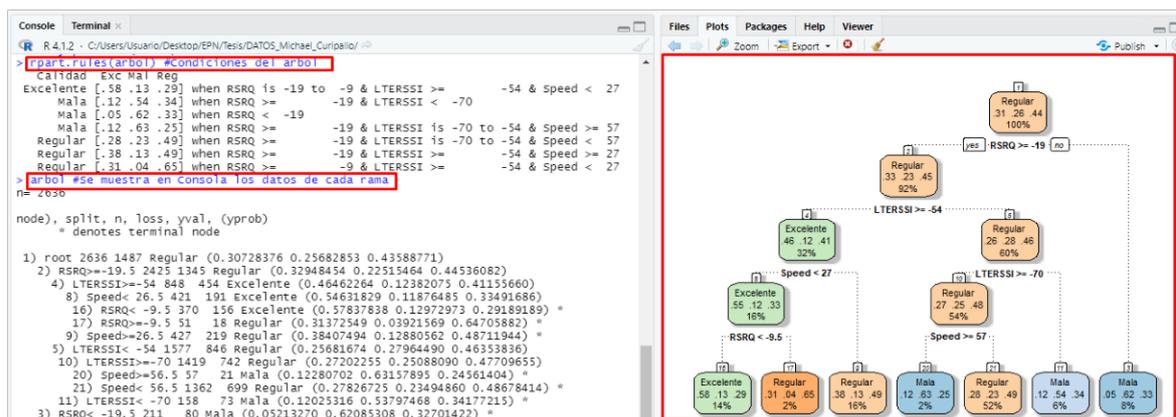


Figura 2.12 Información del árbol de decisión obtenida con 4 entradas y 3 salidas.

En la figura 2.12 se presenta en consola la información acerca de las condiciones y datos del modelo, mientras que en la sección de archivos se presenta el gráfico del árbol de decisión creado mediante el entrenamiento. En este árbol se pueden observar las tres posibilidades de salidas obtenidas que son: Calidad Excelente, Calidad Regular y Calidad Mala, también se muestra que la clasificación se realiza utilizando las variables de entrada,

y de forma general; se tiene el primer nodo, nodos intermedios y los nodos terminales, es decir que cumple con las características del algoritmo. Un análisis más detallado se realizará en el capítulo de resultados.

Al continuar con el código de programación se tiene la etapa de predicción. Para este propósito se utilizará la función “predict”, donde se especifica el algoritmo creado, los datos que se van a utilizar y el tipo de algoritmo con el que se creó el árbol de decisiones. Además, se crea la matriz de confusión para evaluar el modelo creado. Este análisis también se puede realizar mediante la valoración de la exactitud calculada, este valor representa el porcentaje de predicciones correctas frente al total, es decir, es el valor que indica la cantidad de los datos que están bien clasificados por el modelo. Es importante mencionar que se deben tomar en cuenta los valores presentes en la diagonal de la matriz de confusión.

```

71 #Prediccion
72 prueba <- predict(arbol, test, type = 'class')
73
74 #Matriz de confusion
75 matriz <- table(test$calidad, prueba)
76 accuracy_Test <- sum(diag(matriz)) / sum(matriz)
77 print(paste('Exactitud del modelo', accuracy_Test))
78 matriz

```

Figura 2.13 Líneas de código para la predicción y evaluación de la exactitud del modelo.

Al ejecutar estas últimas líneas de código, se obtiene en la consola el siguiente resultado, que se presenta como un ejemplo en esta parte:

```

> print(paste('Exactitud del modelo', accuracy_Test))
[1] "Exactitud del modelo 0.471212121212121"
> matriz

```

		prueba			Predicciones
		Excelente	Mala	Regular	
Instancias en la clase real	Excelente	37	8	146	Verdaderos positivos
	Mala	15	53	104	
	Regular	37	39	221	

Figura 2.14 Resultado en la consola de R después de la aplicación del modelo.

El análisis de los resultados obtenidos se explicará en el capítulo correspondiente.

3. PRUEBAS, RESULTADOS, CONCLUSIONES Y RECOMENDACIONES

En este capítulo se presenta el proceso realizado en la etapa de prueba y descripción de los resultados obtenidos, teniendo en cuenta todas las consideraciones mencionadas en capítulos anteriores.

3.1 PRUEBAS

3.1.1 DATOS PARA EL ENTRENAMIENTO Y PRUEBA

Como ya se mencionó, los datos obtenidos en las tres rutas establecidas se utilizarán en el entrenamiento y en la prueba del modelo, estos datos se encuentran almacenados por rutas y se presentan en los anexos: ANEXO II, ANEXO III y ANEXO IV respectivamente. A continuación, se presenta un ejemplo de cada medición, estas capturas fueron tomadas en Google Earth al abrir los archivos kml otorgados por G-NetTrack Pro.

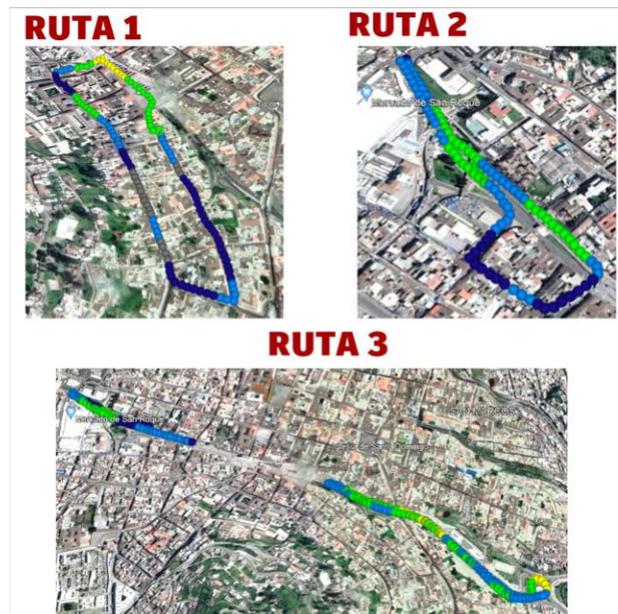


Figura 3.1 Ejemplos de mediciones en cada ruta.

En la figura 3.1 se observa cada punto donde la aplicación tomó una muestra de datos, además, se observa que estos puntos cambian de color dependiendo de la intensidad de la señal. Por otro lado, en la ruta 3 se puede observar un pequeño trayecto sin datos, esto se debe a que la aplicación no puede realizar mediciones dentro del túnel.

3.1.2 PARÁMETROS DE ENTRADA Y SALIDA

3.1.2.1 Parámetros de entrada

Como se mencionó en la metodología, a pesar de que el objetivo es relacionar las velocidades, se consideran otros parámetros de entrada, estos datos serán los que presenten variaciones durante los recorridos, específicamente los datos seleccionados son: velocidad del terminal móvil, RSRP, RSRQ, SNR y/o RSSI. Esta consideración se realizó ya que, al utilizar la velocidad del terminal móvil como único parámetro de entrada, se tiene una carencia de relaciones para formar un árbol de varias ramas. Esto se debe a que la función *rpart* crea su modelo independientemente al objetivo que se tenga, es decir

que realiza el entrenamiento con los datos más importantes que permitan crear un árbol de decisiones con una profundidad aceptable. Para comprobar que no existe una relación predominante entre las velocidades, se presentan los diagramas de dispersión:

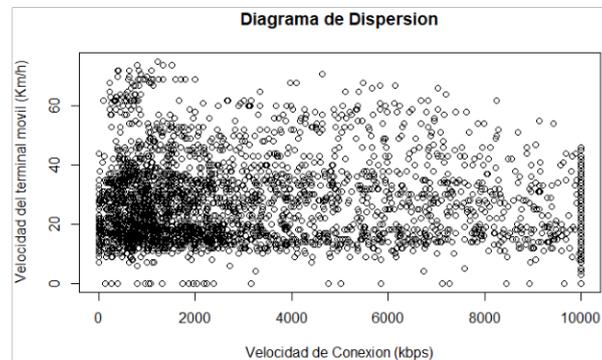


Figura 3.2 Diagrama de dispersión entre la velocidad de cobertura (velocidad de transmisión) y la velocidad del terminal móvil (movimiento).

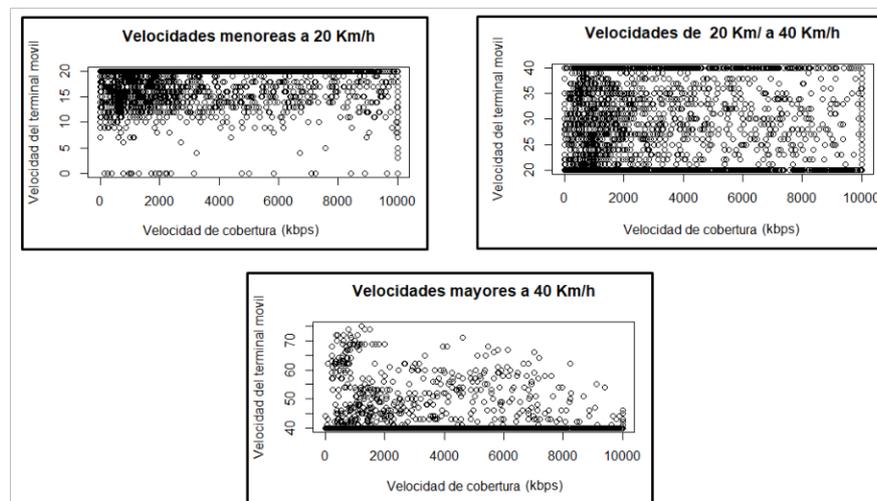


Figura 3.3 Diagramas de dispersión en los rangos de velocidades establecidos en las mediciones.

En las dos figuras mostradas anteriormente se puede observar el diagrama de dispersión entre ambas velocidades, donde; en la figura 3.2 se observan todos los datos obtenidos en el proceso de medición, y en la figura 3.3 se clasifican los datos en los tres rangos de velocidades móviles; es decir, de 0 a 20 Km/h, 20 a 40 Km/h y mayor a 40 Km/h. Al analizar estas figuras se observa que no existe una relación evidente entre ambos parámetros, esto también se puede comprobar calculando la correlación entre estos dos datos:

```
> correlacion<-cor(x=total$speed, y=total$DL_bitrate) #Correlacion
> print(paste('Correlacion entre velocidad de conexion y velocidad del terminal movil', correlacion))
[1] "Correlacion entre velocidad de conexion y velocidad del terminal movil -0.13969446891693"
```

Figura 3.4 Correlación lineal entre la velocidad de conexión y la velocidad del terminal móvil.

Al obtener la correlación entre ambas velocidades se comprueba una vez más que no existe una relación, ya que al obtener una correlación de -0.13 (próximo a 0) indica que no hay relación lineal entre las dos variables. Por otro lado, se puede analizar la cantidad de datos, en los diagramas de dispersión se observa que existe una gran cantidad de datos para velocidades menores a 2 Mbps cuando la velocidad del terminal móvil se encuentra entre los 10 Km/h hasta los 40 Km/h aproximadamente.

3.1.2.2 Parámetros de salida

Como salida se tiene la categoría denominada "Calidad" que cuenta con 3 posibles valores: Excelente, Regular y Mala, la cual se evaluó con rangos de velocidad de cobertura tal como se explicó en la metodología.

Los datos que serán utilizados para crear y probar el modelo deberán estar en una matriz total, la cual tiene todas las entradas y el valor a predecir.

3.1.3 MODELOS DE ENTRENAMIENTO CON DIFERENTES PARÁMETROS DE ENTRADAS.

Se obtendrán diferentes modelos de entrenamiento utilizando varias combinaciones de parámetros de entrada, con el objetivo de tener un modelo utilizando la velocidad del terminal móvil como condición principal. Después de analizar cada uno de estos modelos se seleccionará el que presente mejores condiciones de análisis y se lo detallará en la sección de Resultados.

3.1.3.1 Cinco parámetros de entrada y 1 salida

Como primer modelo se tiene el modelo que involucra todos los datos de la matriz final, es decir que en este entrenamiento se utilizan como parámetros de entrada todos los datos mencionados, esto permite tener una idea general sin limitar nuestros objetivos.

Al utilizar los cinco datos como parámetros de entrada se tiene el siguiente modelo:

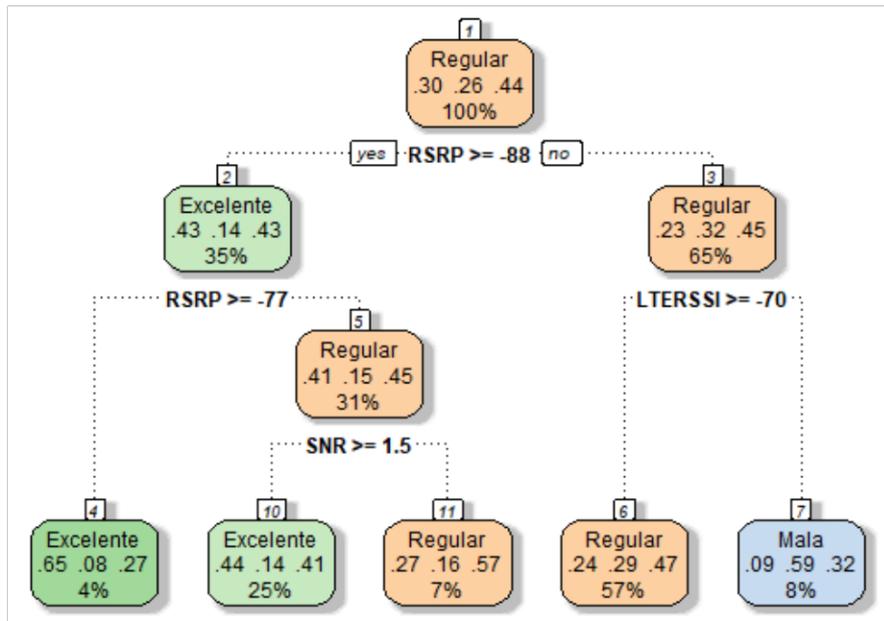


Figura 3.5 Árbol de decisión al utilizar los 5 datos como parámetros de entrada.

En este primer modelo se tienen 6 nodos y 3 condiciones. Se observa que el 31% del total de datos es de calidad excelente, el 44% calidad regular y el 26% de calidad mala. Las condiciones de este árbol son las siguientes:

Para calidad excelente:

- RSRP mayor o igual a -77 dBm.
- RSRP mayor o igual a -88 dBm, pero menor a -77 dBm y SNR mayor o igual a 1.5 dB.

Para calidad regular:

- RSRP mayor o igual a -88 dBm, pero menor a -77 dBm y SNR menor a 1.5 dB.
- RSRP menor a -88 dBm, RSSI mayor o igual a -70 dB.

Para calidad mala:

- RSRP menor a -88 dBm, RSSI menores a -70 dB.

Como se puede observar de acuerdo con las condiciones de la figura 3.5; a pesar de que se están utilizando 5 parámetros de entrada, solo 3 de estos forman parte del árbol de decisión: es decir, que en sus condiciones no se toma en cuenta la velocidad del terminal móvil (Speed) ni el parámetro RSRQ. Como ya se explicó anteriormente, esto significa que los tres parámetros considerados son suficientes para que el algoritmo cree su modelo. Además, estas condiciones se presentan en orden de importancia, por lo que, se puede

comprobar la falta de relación entre la velocidad del terminal móvil y la velocidad de conexión, haciendo que este sea un parámetro sin tanta importancia al utilizar los demás datos. Por esta razón, se decide realizar el análisis utilizando dos y tres parámetros de entrada, siempre y cuando la velocidad del terminal móvil sea parte de las condiciones.

3.1.3.2 Dos parámetros de entrada y una salida

El objetivo de estos análisis es escoger el modelo que permita cumplir con el componente establecido, por lo que, a continuación, se presentan solo los modelos que en sus condiciones cuentan con la velocidad del terminal móvil. Además, se realizará un análisis rápido y de forma general de cada caso.

3.1.3.2.1 Velocidad del terminal móvil y RSRP

Si los parámetros de entrada son la velocidad del terminal móvil y RSRP se obtiene el siguiente árbol de decisión:

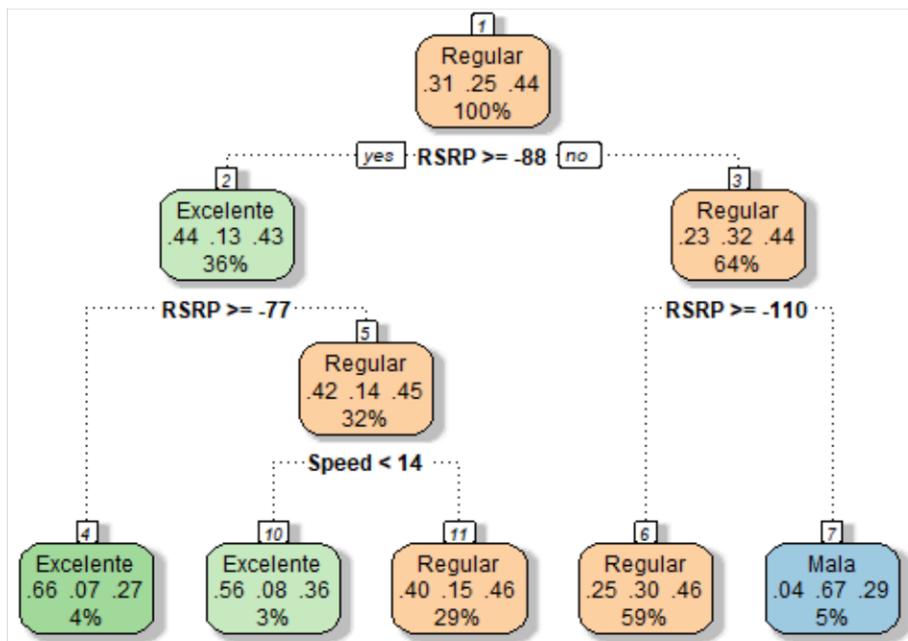


Figura 3.6 Árbol de decisión al utilizar la velocidad del terminal móvil y RSRP como parámetros de entrada.

En este modelo se tienen 8 nodos y 4 condiciones. Al igual que el árbol anterior, se observa que el 31% del total de datos es de calidad excelente, el 44% calidad regular y el 26% calidad mala, esto se debe a que en la raíz aún no se evalúa ninguna condición, solo muestra información acerca de las categorías de los datos recolectados, por lo que esta información se observará en todos los modelos.

A continuación, se presenta un análisis de forma general sobre las condiciones de este árbol; lo más importante de este modelo es que ambas entradas intervienen para definir la calidad. Analizando sus condiciones se observa que se tendrá una calidad de descarga “Excelente” cuando:

- La potencia de referencia RSRP es mayor o igual a -77 dBm.
- RSRP está entre los -88 dBm y -77 dBm y la velocidad del terminal móvil es menor que 14 Km/h.

Por otro lado, se tendrá una calidad de conexión “Regular” si:

- RSRP está entre los -88 dBm y -77 dBm y la velocidad del terminal móvil es mayor a los 14 Km/h.
- RSRP es menor a -88 dBm, pero mayor o igual que -110 dBm.

Y finalmente, cuando se tiene potencias de RSRP menores a -110 dBm se conseguirá una calidad “Mala”.

Si se compara este modelo con el modelo de 5 entradas, se puede ver que ambos cuentan con dos condiciones que involucran a la potencia de referencia RSRP. Además, estas condiciones cumplen con la teoría de calidad de señal LTE, donde en [16] se establece que:

Tabla 3.1 Calidad de señal LTE dependiendo de RSRP [16].

RSRP	Intensidad de señal	Descripción
>= -80dBm	Excelente	Señal fuerte con velocidades máximas de datos
-80dBm a -90dBm	Buena	Señal fuerte con buenas velocidades de datos
-90 dBm a -100 dBm	Regular a pobre	Señal con velocidades confiables, pero es posible tener fallas.
<= -100dBm	Sin señal	Desconexión

3.1.3.2.2 Velocidad del terminal móvil y SNR

Si los parámetros de entrada son la velocidad del terminal móvil y SNR se obtiene el siguiente árbol de decisión:

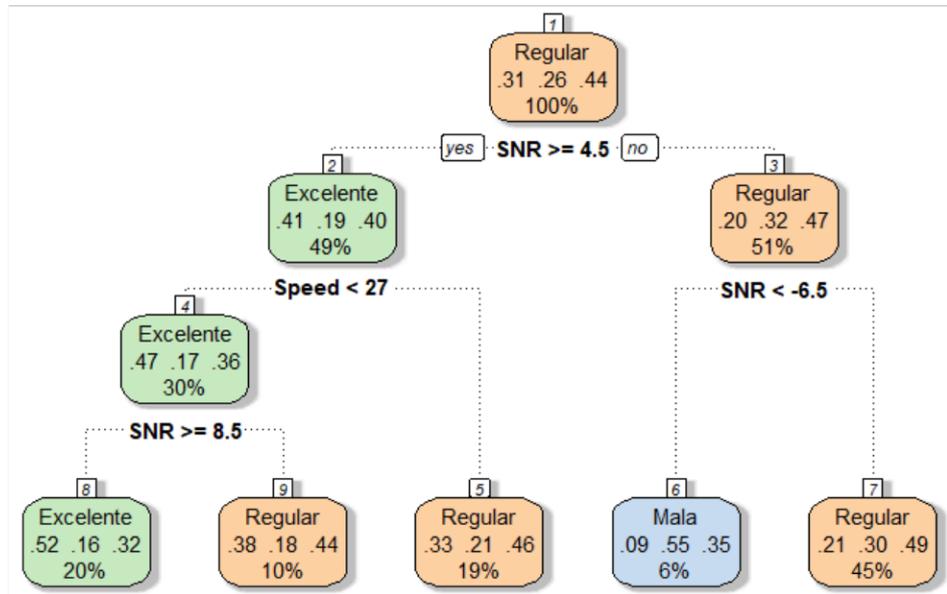


Figura 3.7 Árbol de decisión al utilizar la velocidad del terminal móvil y SNR como parámetros de entrada.

En este árbol se observa que se tendrá una calidad de conexión “Excelente” cuando la relación SNR sea mayor o igual a 8.5 dB y la velocidad del terminal móvil sea menor a los 27 Km/h. Mientras que, se define una calidad regular cuando:

- SNR sea mayor o igual a 4.5 dB y la velocidad del terminal móvil sea mayor a los 27 Km/h.
- SNR sea mayor o igual a 4.5 dB, pero menor a 8.5 dB y la velocidad del terminal móvil sea menor a los 27 Km/h.
- SNR sea menor a 4.5 dB, pero mayor que -6.5 dB.

Finalmente, la condición para que la calidad sea mala, es que el SNR menor que -6.5 dB. A igual que en el caso anterior, estas condiciones cumplen con la teoría de calidad de señal LTE; donde se establece que; a mayor SNR mejor calidad de señal y por tanto mayor velocidad de transmisión [16].

3.1.3.2.3 Velocidad del terminal móvil y RSSI

Si los parámetros de entrada son la velocidad del terminal móvil y RSSI se obtiene el siguiente árbol de decisión:

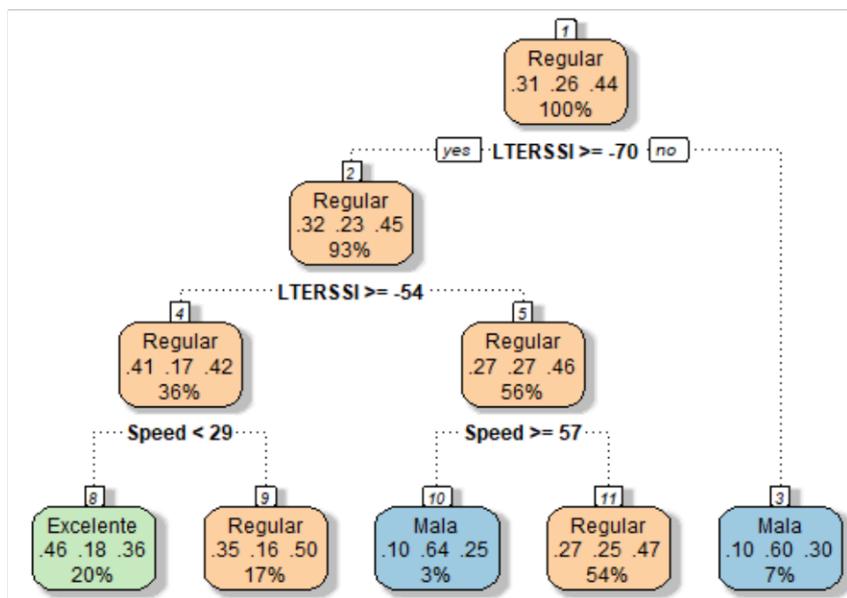


Figura 3.8 Árbol de decisión al utilizar la velocidad del terminal móvil y RSSI como parámetros de entrada.

En el modelo de la figura 3.8 se observa que se tendrá una calidad de conexión excelente cuando:

- El indicador de fuerza de la señal RSSI sea mayor o igual a -54 dBm y la velocidad del terminal móvil sea menor a los 29 Km/h.

Mientras que para definir que una señal es de calidad regular se tienen las siguientes condiciones:

- RSSI mayor o igual a -54 dBm y la velocidad del terminal móvil es mayor a 29 Km/h y menor 54 Km/h.
- RSSI menor a -54 dBm y mayor a -70 dBm y la velocidad del terminal móvil es menor a 57 Km/h.

Y finalmente las condiciones para una calidad mala son:

- Para RSSI menores a -70 dBm.
- RSSI entre los -54 dBm a -70 dBm y velocidades del terminal móvil mayores a los 57 Km/h.

A igual que en el caso anterior, estas condiciones cumplen con la teoría de calidad de señal LTE presentada en [16] en función de RSSI, donde se establece que:

Tabla 3.2 Calidad de señal LTE dependiendo de RSSI [16].

RSSI	Intensidad de señal	Descripción
> -65dBm	Excelente	Señal fuerte con velocidades máximas de datos
-65 dBm a -75 dBm	Buena	Señal fuerte con buenas velocidades de datos
-75dBm a -85dBm	Media	Se pueden lograr velocidades intermedias pero útiles, rápidas y confiables, pero es posible obtener fallos.
-85dBm a -95dBm	Pobre	El rendimiento caerá drásticamente

3.1.3.3 Tres parámetros de entrada y un parámetro de salida

Con el análisis de los modelos de dos entradas se pudo tener una idea clara acerca del comportamiento de los datos recolectados en la zona de estudio y se pudo comprobar que la calidad depende de estos parámetros junto con la velocidad del terminal móvil. A continuación, se realiza un nuevo análisis aumentando el número de entradas para poder observar la relación entre las velocidades. Por lo que en esta parte se procede a utilizar tres parámetros de entrada y una salida. A igual que el análisis anterior, se presentan solo los árboles de decisiones que en sus condiciones cuentan con la velocidad del terminal móvil. Además, se realizará un análisis rápido y de forma general.

3.1.3.3.1 Velocidad del terminal móvil, SNR y RSRQ

Si los parámetros de entrada son la velocidad del terminal móvil, el SNR y el RSSI se obtiene el siguiente árbol de decisión:

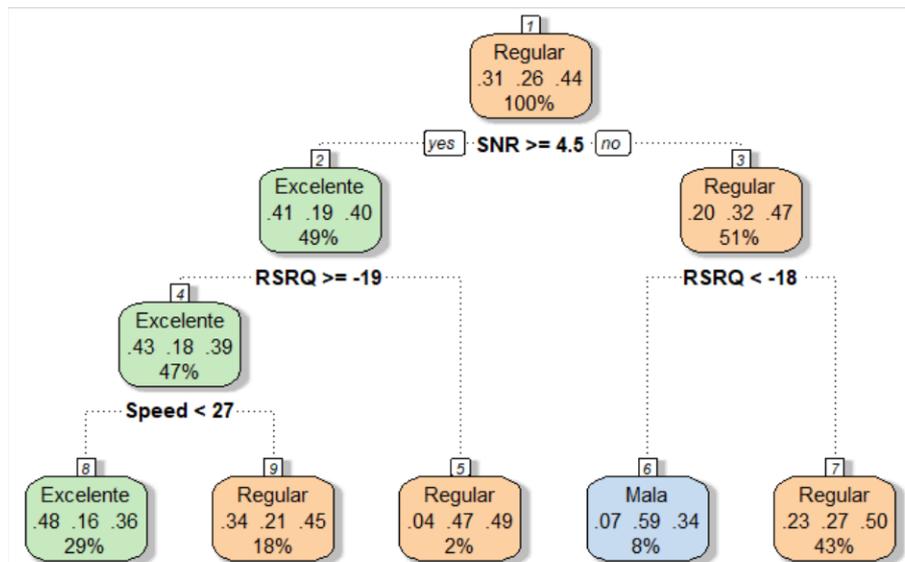


Figura 3.9 Árbol de decisión al utilizar la velocidad del terminal móvil, SNR y RSRQ como parámetros de entrada.

En este primer árbol con tres entradas y una salida se observa que se tendrá una calidad de conexión excelente cuando se cumpla que el SNR sea mayor o igual a 4.5 dB, el RSRQ es mayor o igual a -19 dB y La velocidad del terminal móvil sea menor a 27 Km/h. Mientras que, para una calidad regular se tienen las siguientes condiciones:

- SNR mayor o igual a 4.5 dB y RSRQ menor a -19 dB.
- SNR mayor o igual a 4.5 dB, RSRQ mayor o igual a -19 dB y velocidades del terminal móvil mayores a 27 Km/h.
- SNR menor a 4.5 dB y RSRQ mayor a -18 dB.

Finalmente, cuando el SNR es menor a 4.5 dB y RSRQ menor a -18 dB, se obtendrá una mala calidad de conexión. Analizando este modelo se puede concluir que el valor de RSRQ es un dato de importancia para predecir la calidad de la señal, pero como se pudo observar en el análisis de dos entradas no se presentó un modelo que incluya a este parámetro, esto se puede explicar teniendo en cuenta el funcionamiento de rpart, ya que al utilizar RSRQ y la velocidad del terminal móvil como únicas entradas, se obtenía un modelo que no presentaba a la velocidad del terminal en sus condiciones. Por otro lado, este árbol también cumple con la teoría de calidad de señal LTE presente en [16], donde se establece que:

Tabla 3.3 Calidad de señal LTE dependiendo de RSRQ [16].

RSRQ	Calidad de la señal	Descripción
>= -10dB	Excelente	Señal fuerte con velocidades máximas de datos
-10dB a -15dB	Buena	Señal fuerte con buenas velocidades de datos
-15dB a -20dB	Regular a pobre	Se pueden lograr velocidades de datos confiables, pero es posible tener fallos.
<= -20dB	Sin señal	Desconexión

3.1.3.3.2 Velocidad del terminal móvil, SNR y RSSI

Si los parámetros de entrada son la velocidad del terminal móvil, el SNR y el RSSI se obtiene el siguiente árbol de decisión:

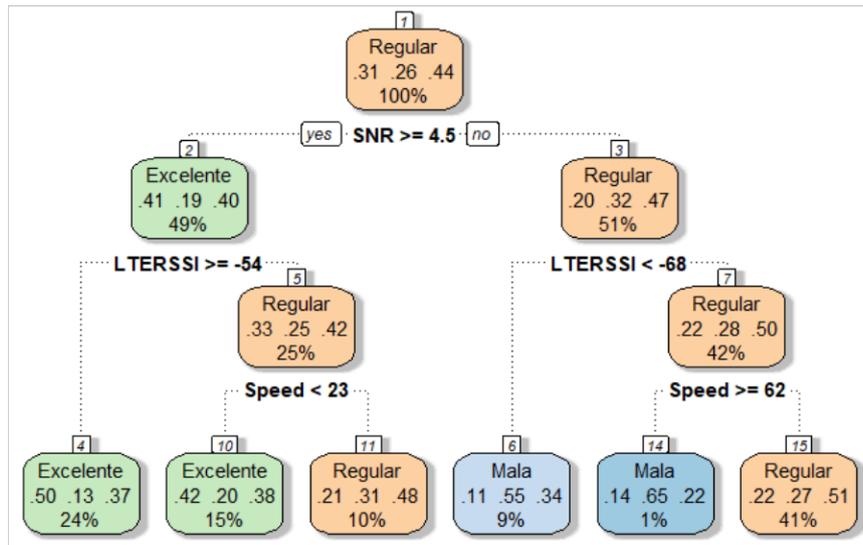


Figura 3.10 Árbol de decisión al utilizar la velocidad del terminal móvil, SNR y RSSI como parámetros de entrada.

En el modelo de la figura 3.10 se observa que se tendrá una calidad de conexión excelente cuando se cumpla que:

- El RSSI sea mayor o igual a -54 dBm y el SNR sea mayor o igual a 4.5 dB.
- SNR sea mayor o igual a 4.5 dB, RSSI sea menor a -54 dBm y la velocidad del terminal móvil sea menor a 23 Km/h.

Por otro lado, se tendrá una calidad de conexión regular cuando se cumplan las siguientes condiciones:

- SNR sea mayor o igual a 4.5 dB, RSSI sea menor a -54 dBm y la velocidad del terminal móvil sea mayor a 23 Km/h.
- SNR sea menor a 4.5 dB, RSSI sea mayor a -68 dBm y la velocidad del terminal móvil es menor a 62 Km/h.

Finalmente, se va a tener una mala calidad de conexión en cualquiera de las siguientes condiciones:

- SNR sea menor a 4.5 dB, RSSI sea menor a -68 dB.
- SNR sea menor a 4.5 dB, RSSI sea mayor a -68 dB y la velocidad del terminal móvil es mayor o igual a 62 Km/h.

Con este modelo se pueden concluir varias relaciones: se presenta una combinación de condiciones de los modelos de dos entradas, también, se sigue cumpliendo la teoría de

calidad para LTE, y, por último, la velocidad del terminal móvil en todos los modelos presentados hasta el momento influye para la categoría denominada “excelente”, cuando se tienen velocidades bajas.

3.1.3.3 Velocidad del terminal móvil, RSRQ y RSSI

Si los parámetros de entrada son la velocidad del terminal móvil, el RSRQ y el RSSI se obtiene el siguiente árbol de decisión:

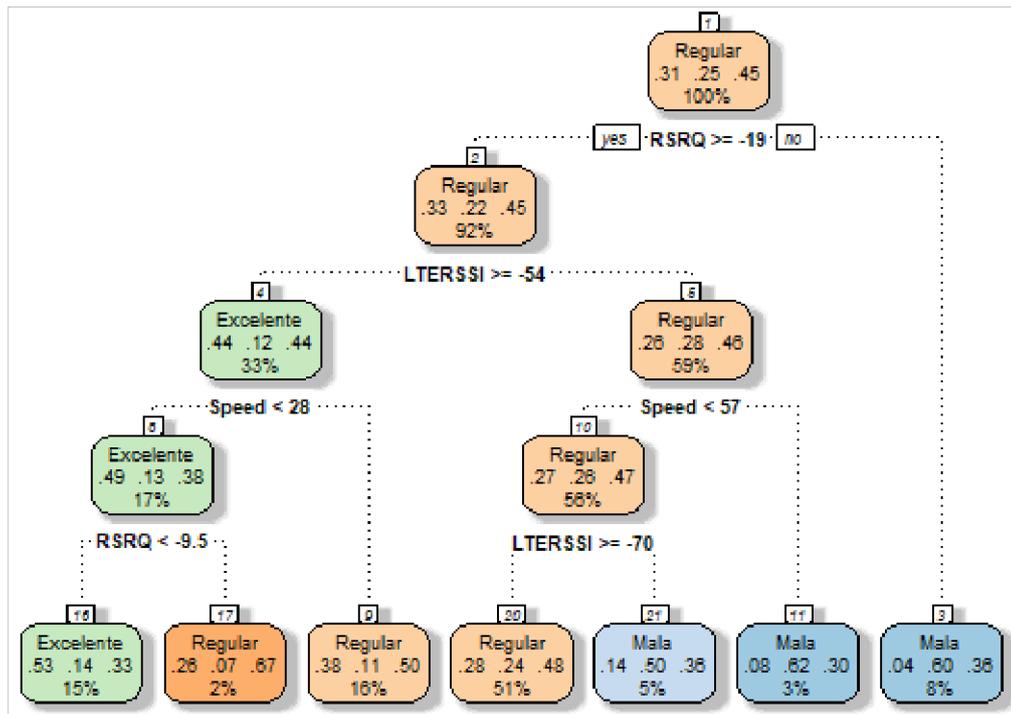


Figura 3.11 Árbol de decisión al utilizar la velocidad del terminal móvil, RSSI y RSRQ como parámetros de entrada.

En el modelo de esta última combinación de entradas se observa que, a diferencia de los árboles anteriores, este presenta 12 nodos, 6 condiciones y profundidad 5. Analizando este árbol se tiene; una calidad de conexión “excelente” cuando se cumplan que el RSRQ es mayor o igual a -19 dB, pero menor a -9.5 dB, RSSI mayor o igual a -54 dBm y la velocidad del terminal móvil menor a los 28 Km/h. Mientras que para una calidad regular se debe cumplir:

- RSRQ mayor o igual a -19 dB, RSSI mayores o iguales a -54 dBm y velocidades del terminal móvil mayores a 28 Km/h
- RSRQ mayores a -9.5 dB, RSSI mayores o iguales a -54 dBm y velocidades del terminal móvil menores a 28 Km/h.

- RSRQ mayor o igual a -19 dB, RSSI entre los -70 dBm a -54 dBm y velocidades del terminal móvil menores a 57 Km/h.

Mientras que para obtener la categoría “Mala” se tienen tres posibles condiciones:

- RSRQ menor a -19 dB.
- RSRQ mayor o igual a -19 dB, velocidad del terminal móvil mayores a 57 Km/h y RSSI menor a -54 dBm.
- RSRQ mayor o igual a -19 dB, RSSI menores a -70 dBm y velocidades del terminal móvil menores a 57 Km/h.

De igual manera, con este modelo se puede observar las mismas conclusiones del árbol anterior; las condiciones de los modelos anteriores se repiten en este último modelo, manteniendo la teoría de calidad para LTE, también, se sigue observando que a velocidades del terminal móvil menores a los 28 Km/h (velocidades bajas) se obtiene una calidad “excelente”.

3.2 RESULTADOS

3.2.1 ANÁLISIS DE LOS MODELOS CON MEJORES RESULTADOS

Tomando en cuenta todos los modelos que se obtuvieron anteriormente se pudo observar que los modelos con dos y tres entradas presentan las mismas condiciones principales que el modelo con 5 entradas. Es por esta razón que en esta parte se analizarán detalladamente los modelos que se consideran más importantes para cumplir con el objetivo del componente, estos son; un modelo considerando el dato más importante (RSRP) con la velocidad del terminal móvil y un modelo utilizando los cuatro datos restantes (RSRQ, SNR, RSSI) y la velocidad del terminal móvil.

3.2.1.1 Entradas: velocidad del terminal móvil y RSRP – salida: velocidad de conexión

En el modelo presentado en la Figura 3.5 que pertenece al modelo con 5 entradas y 3 salidas, se observa que el parámetro más importante es RSRP, también, se observó que este parámetro es de igual importancia en el modelo de la figura 3.6 que pertenece al modelo de 2 entradas (RSRP y Velocidad del terminal móvil), ya que la mayoría de las ramas dependen de esta entrada, haciendo que el algoritmo lo considere primero antes que algún otro parámetro, esto debido a su gran relación con la calidad, para comprobar se calcula la correlación entre estos dos parámetros:

```
> correlacion<-cor(x=total$RSRP, y=total$DL_bitrate) #Correlacion
> print(paste('Correlacion entre Velocidad de conexion y RSRP', correlacion))
[1] "Correlacion entre velocidad de conexion y RSRP 0.25570350724488"
```

Figura 3.12 Correlación lineal entre la velocidad de conexión RSRP.

Al tener una correlación de 0.26 se puede asegurar que existe una relación lineal positiva entre la velocidad de conexión y RSRP. Por otro lado, si se analiza la relación existente entre RSRP con la velocidad del terminal móvil se tiene el siguiente diagrama de dispersión:

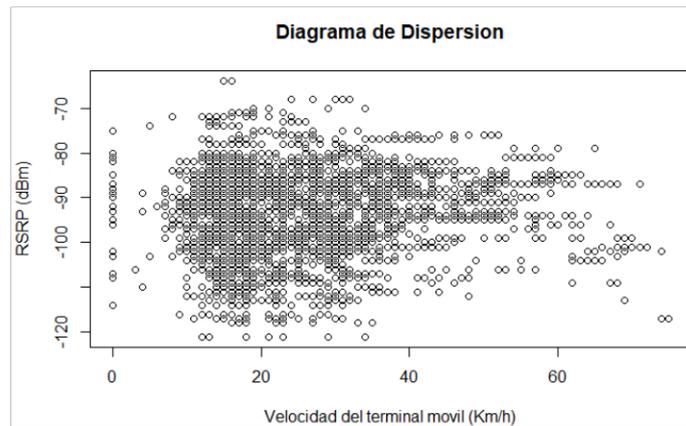


Figura 3.13 Diagrama de dispersión entre la velocidad del terminal móvil y RSRP.

Según la teoría de calidad de LTE y las condiciones obtenidas de los modelos presentados anteriormente, se tendrá una calidad de conexión excelente cuando el RSRP es mayor a -80 dBm y las velocidades del terminal móvil menores a 14 Km/h. Si se observa la figura 3.13, se puede observar que para estos rangos no existen muchos datos. Mientras que la mayor cantidad de datos se presentan en el rango de -80 dBm < RSRP < -110 dBm y velocidades del terminal móvil mayores a 10 Km/h y menores a 40 Km/h, es decir categoría regular, esto también se puede comprobar en la consola de R:

```
node), split, n, loss, yval, (yprob)
* denotes terminal node
1) root 2636 1482 Regular (0.30500759 0.25720789 0.43778452)
2) RSRP>=-88.5 926 510 Excelente (0.44924406 0.13498920 0.41576674)
4) RSRP>=-77.5 97 31 Excelente (0.68041237 0.08247423 0.23711340) *
5) RSRP< -77.5 829 467 Regular (0.42219542 0.14113390 0.43667069)
10) Speed< 13.5 90 37 Excelente (0.58888889 0.06666667 0.34444444) *
11) Speed>=13.5 739 408 Regular (0.40189445 0.15020298 0.44790257) *
3) RSRP< -88.5 1710 941 Regular (0.22690058 0.32339181 0.44970760)
6) RSRP< -100.5 590 329 Mala (0.21525424 0.44237288 0.34237288) *
7) RSRP>=-100.5 1120 553 Regular (0.23303571 0.26071429 0.50625000) *
```

Figura 3.14 Información del primer modelo en la consola de RStudio.

En la figura 3.14 se presenta la información otorgada en consola sobre el modelo obtenido. En esta información se describen el número de datos que se evalúa por nodo y se especifican las condiciones. El número 1 se trata del primer nodo, donde indica que se está

trabajando con 2636 datos de entrenamiento, de este total, el 43.778% de datos (1154 datos) corresponden a la categoría regular. Las siguientes dos condiciones importantes se evalúan con el resto de los datos, que son 1482 datos. Si se analiza las condiciones que llevan a una calidad excelente, podemos comprobar que se tienen pocos datos. Por ejemplo, en el nodo 10, se establece la condición Speed <13.5, de los 90 datos que se evalúan en esa condición, 53 datos son de calidad excelente y representan el 3% de los 2636 datos. Con este informe se puede corroborar el análisis sobre el diagrama de dispersión, a continuación, se presenta el diagrama de dispersión en 3D para poder visualizar la relación entre los tres parámetros:

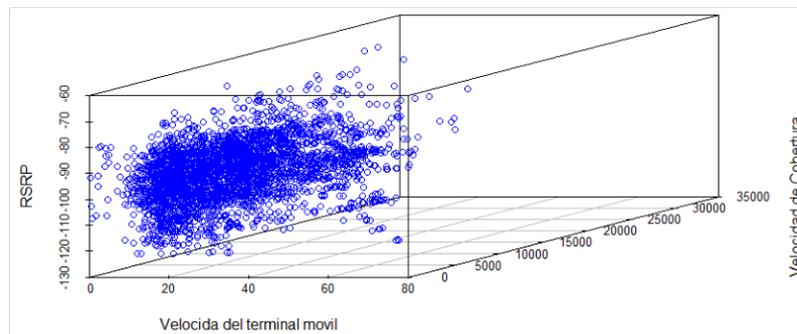


Figura 3.15 Diagrama de dispersión entre la velocidad del terminal móvil, RSRP y la velocidad de cobertura.

Los resultados con este modelo son:

- Exactitud del modelo

```
> print(paste('Exactitud del modelo', accuracy_Test))
[1] "Exactitud del modelo 0.465151515151515"
```

Figura 3.16 Exactitud del primer modelo obtenido en la consola de RStudio.

- Matriz de confusión

Tabla 3.4 Matriz de confusión para el primer modelo con 2 entradas y 1 salida

	Excelente	Mala	Regular
Excelente	23	2	175
Mala	3	20	153
Regular	12	8	264

Al analizar la precisión del modelo podemos concluir que es un modelo con una precisión del 47%. Por otro lado, analizando la diagonal se pueden observar que existen más aciertos que errores, ya que se obtienen varios datos correspondientes con los valores estimados de forma correcta por el modelo.

3.2.1.2 Entradas: velocidad del terminal móvil, RSRQ, RSSI y SNR Salida: velocidad de conexión

Analizando los modelos de dos y tres entradas se pudo observar que las condiciones se mantienen, por lo que se decidió escoger un modelo con 4 entradas y 1 salida, ya que este modelo reunirá información de cada parámetro y se llegará al mismo objetivo. Primero se analizará la relación que existe entre estos datos con la velocidad de conexión.

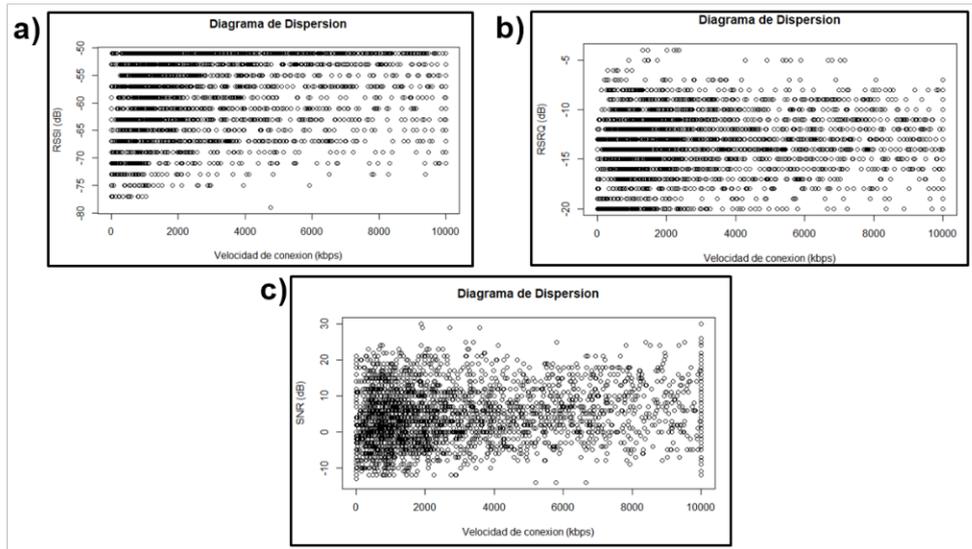


Figura 3.17 Diagramas de dispersión entre los parámetros de entrada y la velocidad de conexión.

En la figura 3.17 se presentan los diagramas de dispersión para poder analizar los datos obtenidos y como estos se relacionan con la velocidad de conexión. La figura 3.17 a) representa el diagrama de dispersión entre la velocidad de conexión y el RSSI, en esta se observa que para velocidades mayores a 5 Mbps se tienen una mayor cantidad de datos en -50 dB, también se observa que de forma general se tiene una concentración de datos en velocidades menores a 1Mbps. La figura b) representa el diagrama de dispersión entre la velocidad de conexión y RSRQ, se observa que se tiene una concentración alta de datos en velocidades menores a los 5 Mbps distribuida en todo el rango de RSRQ. Mientras que, para velocidades mayores a 5 Mbps (categoría excelente) se tiene una concentración en $-10 \text{ dB} < \text{RSRQ} < -15 \text{ dB}$. Por último, en la figura c), se presenta el diagrama de dispersión de la velocidad de conexión y SNR. Al igual que en los diagramas anteriores se observa que la mayoría de los datos se encuentra en velocidades menores a los 5 Mbps distribuida en todo el rango de RSRQ. En este diagrama, a diferencia de los anteriores, se puede observar una mejor relación entre los datos, es decir que es visible una relación positiva, para comprobar esto, se calcula la correlación lineal para los tres datos:

```

> correlacion<-cor(x=total$LTERSSI, y=total$DL_bitrate) #Correlacion
> print(paste('Correlacion entre velocidad de conexion y RSSI', correlacion))
[1] "Correlacion entre velocidad de conexion y RSSI 0.192320470353285"
> correlacion<-cor(x=total$RSRQ, y=total$DL_bitrate) #Correlacion
> print(paste('Correlacion entre velocidad de conexion y RSRQ', correlacion))
[1] "Correlacion entre velocidad de conexion y RSRQ 0.0904095403465492"
> correlacion<-cor(x=total$SNR, y=total$DL_bitrate) #Correlacion
> print(paste('Correlacion entre velocidad de conexion y SNR', correlacion))
[1] "Correlacion entre velocidad de conexion y SNR 0.240873858827799"

```

Figura 3.18 Correlación entre los diferentes parámetros con la velocidad de conexión.

Con los valores de correlación se puede confirmar lo dicho anteriormente, donde el SNR tiene una mejor relación lineal positiva comparado con los otros dos datos, este presenta un valor de 0.24, mientras que para RSSI y RSRQ se tienen valores cercanos a cero, indicando su poca relación con la velocidad de conexión. Por lo tanto, el modelo final es el siguiente:

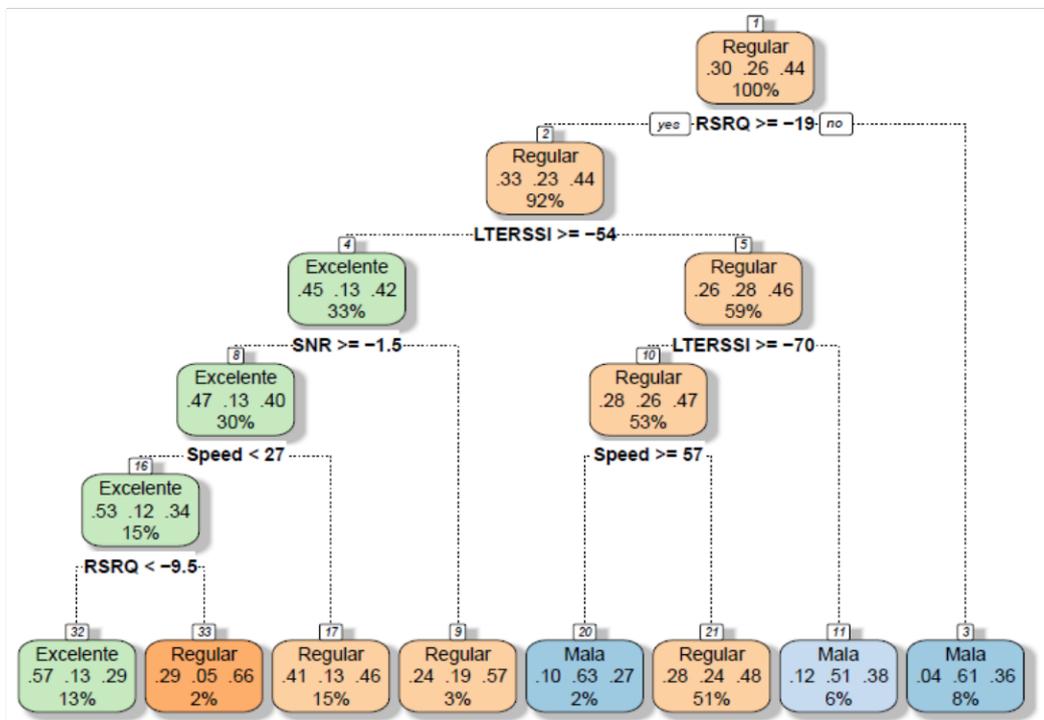


Figura 3.19 Modelo de entrenamiento final.

En el modelo de entrenamiento con 4 entradas y 3 salidas, se tiene un árbol de decisión con 14 nodos, 7 condiciones y una profundidad de 6. En la rama se observa que el 30% del total de datos es de calidad excelente, el 44% calidad regular y el 26% de calidad mala. Las condiciones de este árbol son las siguientes:

Para calidad excelente:

- RSRQ mayor o igual -19 dB, pero menor a -9.5 dB, RSSI mayor o igual a -54 dB, SNR mayor o igual a 1.5 dB y velocidad del terminal móvil menor a 27 Km/h.

Para calidad regular:

- RSRQ mayor o igual a -19 dB, RSSI mayor o igual a -54 dBm y SNR menor a -1.5 dB.
- RSRQ mayor o igual a -19 dB, RSSI mayor o igual a -54 dBm y SNR mayor o igual a -1.5 dB y velocidades mayores a 27 Km/h.
- RSRQ mayor a -9.5 dB, RSSI mayor o igual a -54 dBm y SNR mayor o igual a -1.5 dB y velocidades mayores a 27 Km/h.
- RSRQ mayor a -9.5 dB, RSSI menor a -54 dBm y mayor o igual a -70 dBm y velocidad del terminal móvil menores a 57 Km/h.

Para calidad mala:

- RSRQ menores a -19 dB.
- RSRQ mayor a -9.5 dB y RSSI menor a -70 dBm.
- RSRQ mayor a -9.5 dB, RSSI menor a -54 dBm, pero mayor o igual a -70 dBm y velocidad del terminal móvil mayores o iguales a 57 Km/h.

Como se puede visualizar, estas condiciones son las mismas que se presentaron en los modelos individuales, cumpliéndose así nuestro objetivo.

Los resultados de este modelo de acuerdo a la matriz de confusión son los siguientes:

- Matriz de confusión

Tabla 3.5 Matriz de confusión para el modelo con 4 entradas y 1 salida

	Excelente	Mala	Regular
Excelente	47	14	138
Mala	9	64	95
Regular	22	29	242

- Exactitud del modelo

```
> print(paste('Precisión del modelo', accuracy_Test))  
[1] "Precisión del modelo 0.534848484848485"
```

Figura 3.20 Exactitud del modelo con 4 entradas y 1 salida, obtenida en la consola de RStudio.

El modelo presenta una precisión del 53.48%, superando el valor mínimo esperado. Por otro lado, analizando la diagonal se pueden observar que, al igual que el modelo anterior, existen más aciertos que errores. Al comparar entre los dos modelos seleccionados, se

obtiene una mejor precisión cuando se utilizan 4 entradas y una salida, ya que presenta una mayor precisión y más aciertos al momento de predecir, por lo tanto, el algoritmo de este segundo modelo predice de mejor manera.

3.3 PRUEBA DEL MODELO CON NUEVOS DATOS ALEATORIOS

Para comprobar si las condiciones del modelo obtenido son correctas, se analizará con datos aleatorios dentro del sector establecido, estos datos no serán recolectados precisamente en las tres rutas que se usaron para el análisis anterior, esto se decidió ejecutar, ya que el análisis es independiente de las estaciones bases a las que se conecta el terminal móvil y de la ubicación del terminal, infraestructura de edificios, hora del día, entre otros parámetros. Por lo tanto, se utilizarán recorridos aleatorios dentro del centro histórico de Quito, tal como se muestra en el siguiente ejemplo:



Figura 3.21 Datos aleatorios recolectados dentro del Centro Histórico de Quito para pruebas.

Los archivos de estas mediciones se encuentran en el ANEXO V. Para poder abrir los archivos en RStudio, se los nombra con la siguiente estructura:

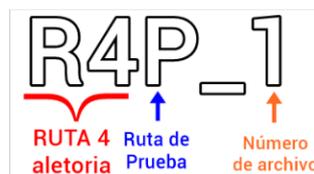


Figura 3.22. Ejemplo de estructura del nombre de los archivos csv con los datos para pruebas.

En resumen: para el entrenamiento de este modelo se utilizarán los datos obtenidos en las 3 rutas mencionadas en el capítulo 2, estos datos se almacenarán en la matriz llamada “total”, mientras que para la parte de prueba se usarán los datos aleatorios guardados en la matriz llamada “total2”, teniendo un total de:

total	3296 obs. of 6 variables
total2	933 obs. of 6 variables

Figura 3.23 Matriz de entrenamiento con 3296 datos y matriz de prueba con 933 datos de prueba.

Como se puede observar en la figura 3.23 se realizará el entrenamiento con 3296 datos, estos comprenden a todos los datos recolectados en las tres rutas, por lo que, para la parte de programación se debe omitir desde la línea 44 del código presentado en la figura 2.8 y ser reemplazada con las líneas de la figura 3.24, esto se realiza ya que, para este caso de comprobación, no se dividirán los datos en porcentajes. Por otro lado, se utiliza el mismo procedimiento realizado en el entrenamiento, es decir, que se seleccionarán las columnas de entradas, se creará una columna llamada “Calidad”, se evaluará la calidad dependiendo de la velocidad de descarga y finalmente se eliminarán las columnas que se utilizaron al crear la condición, tal como se muestra a continuación.

```

48 #PRUEBA
49
50 #Importación de Datos
51 #Se cargan todas las mediciones para pruebas en un solo objeto.
52 datos2 <- lapply(Sys.glob("Ruta4/R4P_?.csv"), read.csv, sep="")
53
54 #Depuración de datos
55 j<-length(datos2)
56 for (i in 1:j) {
57   datos2[[i]] <- datos2[[i]] %>% #Funcion para poder manipular los datos dentro de una lista de matrices
58   select(c(Speed, NetworkMode, Level,Qual, CQI, ARFCN, DL_bitrate)) #Columnas que se van a utilizar
59   datos2[[i]]$calidad <- "Mala" #se crea una columna llamada calidad, definida con la categoria mas baja
60 }
61
62 #Mezclar los dataframes en la lista
63 total2<- do.call("rbind",datos2) #Se unen todos los datos en una sola matriz
64 total2<-rename(total2, RSRP=NetworkMode,
65               RSRQ=Level,
66               SNR = Qual,
67               LTERSII=CQI,
68               DL_bitrate=ARFCN,
69               UL_bitrate=DL_bitrate)#se renombran las columnas
70
71 total2$Calidad[total2$DL_bitrate >= 5000]<- "Excelente"
72 total2$Calidad[total2$DL_bitrate >= 1000 & total2$DL_bitrate < 5000]<- "Regular"
73 total2$Calidad[total2$DL_bitrate < 1000 ]<- "Mala"
74 #Se eliminan los datos que se usaron para la clasificación
75 total2<-select(total2, -c(DL_bitrate, UL_bitrate))

```

Figura 3.24. Líneas de código para la creación de la matriz de prueba con datos aleatorios.

Las siguientes tres líneas de código se utilizan para asegurar que las variables de entrada de la matriz de prueba sean del mismo tipo que las variables de entrada de la matriz de entrenamiento con el objetivo de no presentar errores al momento de utilizar los datos en el algoritmo creado.

```

77 #Se cambian los tipos de datos para que las matrices de entrenamiento y prueba sean iguales
78 total2$RSRQ<-as.integer(total2$RSRQ) #int
79 total2$LTERSSI<-as.integer(total2$LTERSSI) #int
80 total2$SNR<-as.numeric(total2$SNR) #num

```

Figura 3.25. Líneas de código para cambiar el tipo de datos.

Por último, se presentan las líneas de código que se utilizan para la predicción y evaluación del modelo. Para la predicción se utiliza la función “predict”, donde se especifica el algoritmo creado, los datos que se van a utilizar y el tipo de algoritmo con el que se creó el árbol de decisiones. Finalmente, se crea la matriz de confusión y se calcula la exactitud del para evaluar el modelo creado.

```

82 prueba <- predict(arbol, total2, type = 'class') #Predicción usando los datos de prueba
83 matriz <- table(total2$calidad, prueba) #Matriz de confusión
84 accuracy_Test <- sum(diag(matriz)) / sum(matriz) #Precisión de la Matriz de confusión
85 print(paste('Accuracy for test', accuracy_Test)) #Imprimir la precisión
86 matriz

```

Figura 3.26. Líneas de código para la predicción y evaluación de la exactitud del modelo.

Para este modelo de prueba se consideran las mismas 4 entradas y 1 salida del último modelo presentado, ya que se observó que tiene una alta exactitud y presenta un conjunto de condiciones principales de los modelos pequeños inicialmente considerados. El modelo obtenido es el siguiente:

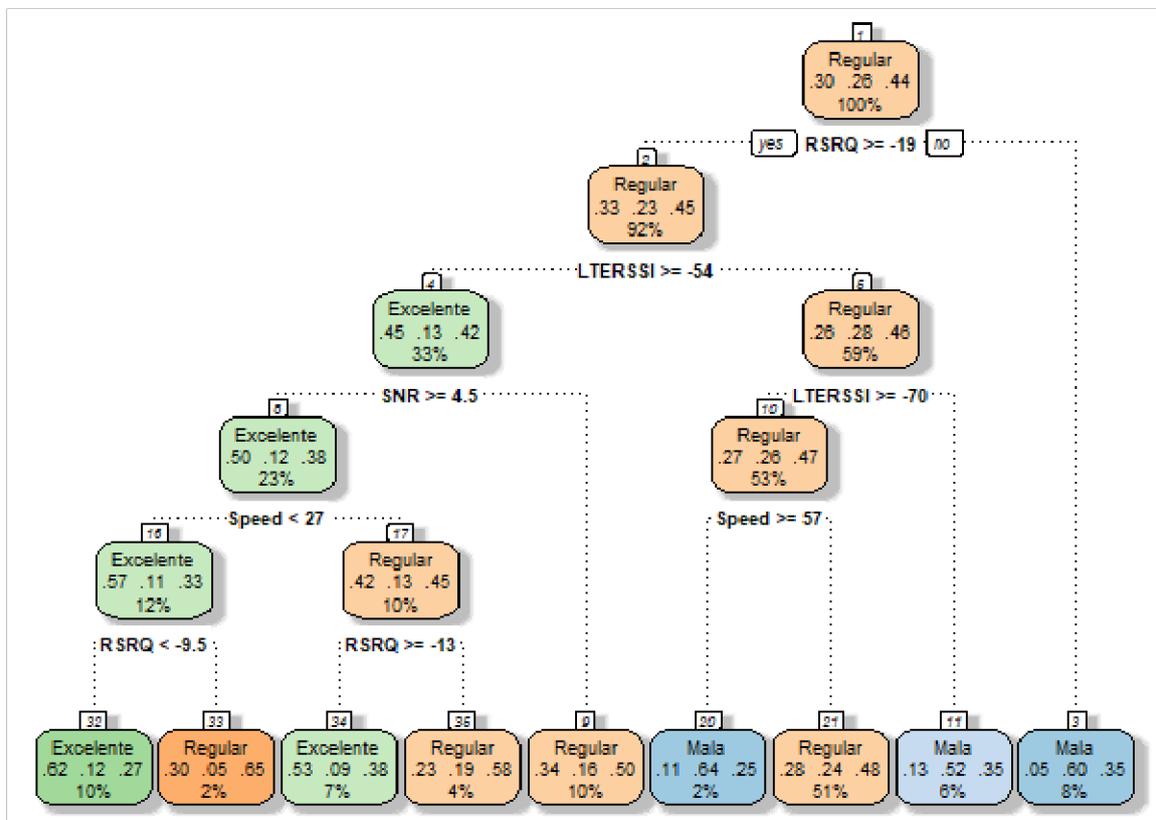


Figura 3.27 Modelo creado con datos aleatorios.

Analizando este modelo se puede observar que presenta las mismas condiciones que el modelo presentado en la figura 3.19, a pesar de que los datos no se obtuvieron en las mismas rutas. Concluyéndose que el modelo seleccionado anteriormente es el más adecuado, ya que con este modelo se comprueba que independientemente de factores externos se llegarán a las mismas condiciones cuando se trabajen con los mismos datos. Por otro lado, si se analiza la exactitud y la matriz de confusión, se tiene:

Tabla 3.6. Matriz de confusión del nuevo modelo.

	Excelente	Mala	Regular
Excelente	48	34	135
Mala	20	118	93
Regular	49	166	270

[1] "Exactitud del nuevo modelo 0.467309753483387"

Figura 3.28. Exactitud y Matriz de confusión del modelo con datos aleatorios.

Si se compara esta información con el modelo anterior, se puede observar que existe una reducción de exactitud, además, en la diagonal de esta nueva matriz se obtienen más errores que en el modelo anterior. Por lo que se puede concluir que a pesar de que las condiciones se mantienen, las predicciones pueden fallar por falta de información, como, por ejemplo, la distancia a la estación base con respecto al punto donde se realizaron las mediciones, esta información dará una visión mucho más amplia al relacionar las velocidades.

3.4 CONCLUSIONES

- En este estudio se propone una metodología de recolección de datos y su respectivo análisis para detectar degradaciones en la velocidad de descarga de datos desde la red móvil durante conexiones con el terminal en movimiento. Con esto, es posible detectar los puntos con problemas y sus posibles razones, información que podría servir de referencia para plantear soluciones y mejorar los niveles de calidad de servicio de la red.
- Tras el análisis del modelo con 5 entradas y 1 salida, se puede deducir que la calidad de conexión no depende directamente de la velocidad del terminal móvil, ya que estos dos parámetros cuentan con una correlación de -0.13, significando que no existe una relación lineal. Por esta razón este modelo deja de tener importancia para el análisis de este proyecto, ya que, al utilizar datos con correlaciones más altas, el algoritmo ignora la velocidad del terminal y no lo considera para sus modelos.

- La velocidad del terminal móvil será parte de las condiciones del modelo cuando se utilicen datos con correlación no tan alta, cuando esto sucede, se puede observar que se tendrán velocidades de conexión en la categoría excelente para velocidades menores a los 27 Km/h, mientras que, para velocidades mayores a este valor, se empieza a tener una velocidad de conexión en la categoría regular o mala. Concluyendo que; a menores velocidades del terminal móvil se tienen, mayores velocidades de conexión y viceversa.
- Tras los modelos obtenidos, se puede asegurar que se obtendrá una calidad excelente cuando el SNR sea mayor o igual a 4.5 dB, el parámetro RSSI sea mayor o igual a -54 dBm y el RSRQ sea menor a -9.5 dB. Además, cuando el parámetro de RSRP sea mayor o igual a -77 dBm se asegurará una alta velocidad de conexión independientemente de los otros parámetros mencionados. Mientras que, cuando el RSRQ sea menor a -19 dB y el RSSI sea menor a -70 dBm se asegura una velocidad de conexión menor a 1 Mbps. Y, por último, se obtendrá una calidad de descarga regular cuando los parámetros se encuentren en rangos intermedios a los mencionados.
- La falta de relación entre las velocidades de conexión con las del terminal móvil se puede justificar por varios factores: falta de información (recolección de mayor cantidad de datos), como por ejemplo la distancia a la estación base a la que se conecta el terminal en el punto de la medición, ya que este factor podría tener una mayor relación con la salida; también esta falta de relación se puede deber a la falta de mediciones en velocidades mayores a 40 Km/h en las dos primeras rutas.
- El modelo con 4 entradas y 1 salida es el conjunto de condiciones de los modelos que tienen menos entradas, es decir que, los modelos con 2 o 3 entradas presentan siempre las mismas condiciones cuando se analizan los mismos datos, y esto se puede ver en el modelo completo, ya que este presenta las condiciones principales de los modelos pequeños. Además, se observó que con este modelo se obtiene una exactitud de 53.84 %, ya que el modelo utiliza todos los datos de entrada a excepción del RSRP, considerando varias condiciones que se relacionan con el objetivo propuesto en este trabajo.
- Se comprobó que al utilizar datos independientes a las rutas establecidas (datos aleatorios en el mismo sector de estudio), se obtuvo el mismo modelo, pero la exactitud disminuyó a 46.73%. En base a esto, se puede mencionar que para que el modelo pueda ser utilizado con datos aleatorios se necesitan otros parámetros de entrada.

- Creemos que el modelo podría aumentar su exactitud si se utilizan otros datos de entrada, como por ejemplo la distancia a la estación base.

3.5 RECOMENDACIONES

- Se recomienda que se recolecte mayor cantidad de datos en una misma ruta, además de que en todas las rutas se pueda tener los mismos rangos de velocidades del terminal móvil, ya que esta falta de datos puede afectar al momento de obtener los resultados. También es recomendable utilizar otra herramienta que permita obtener la información de las estaciones base a las que se conecta, ya que esta información puede ser de gran importancia para relacionar con las velocidades de conexión.
- Se recomienda limpiar los datos almacenados antes de evaluar un nuevo modelo.
- Otra recomendación para evitar errores al realizar el código es que se debe analizar primero la correlación entre los datos para establecer si existe la posibilidad de crear un árbol de decisión.
- Por último, para evitar errores en la recolección de datos se recomienda utilizar las versiones Pro de las aplicaciones, ya que las versiones lite presentan limitaciones.

4. REFERENCIAS BIBLIOGRÁFICAS

- [1] Hou, Q., Zhou, Yifan. F., Peng, X. y Zhang, L, "Study of TD-LTE Coverage and Uplink/Downlink Power Budget in High-Speed Train", in Applied Mechanics and Materials, 2014, pp. 1-5.
- [2] Google Play, "Netmonitor", 2021. [En línea]. Disponible en: <https://play.google.com/store/apps/details?id=com.parizene.netmonitor&hl=en&gl=US>.
- [3] Google Play, "ForceLTE", 2021. [En línea]. Disponible en: https://play.google.com/store/apps/details?id=com.xsquarestudio.forcelte&hl=es_EC&gl=US.
- [4] Google Play, "CellMapper", 2021. [En línea]. Disponible en: https://play.google.com/store/apps/details?id=cellmapper.net.cellmapper&hl=es_EC&gl=US.
- [5] m2catalyst, "Network Cell Info Manual", 2020. [En línea]. Disponible en: <https://m2catalyst.com/apps/network-cell-info/manual>

- [6] Gyokov Solutions, "G-NetTrack Pro manual", 2021. [En línea]. Disponible en: <https://gyokovsolutions.com/manual-g-nettrack/>
- [7] R. Tupiza, "Desarrollo de una herramienta para medición, envío y presentación de parámetros de calidad y cobertura de una red 4G para ambientes indoor", Ingeniería en Electrónica y Telecomunicaciones, Tesis, EPN, Quito, ECU, 2018.
- [8] De características, "Xiaomi Redmi Note 9", 2020. [En línea]. Disponible en: <https://decaracteristicas.com/xiaomi-redmi-note-9-pro/>
- [9] RGG, "Qué es R y RStudio?", 2019. [En línea]. Disponible en: https://gonzalezgouveia.com/que-es-r-y-rstudio/#Que_es_R_y_RStudio
- [10] wilysis, " Network Cell Info (Common) Export Database Formats", 2018. [En línea]. Disponible en: <http://wilysis.com/networkcellinfo/25-nci-export-db>.
- [11] Zhi-Hua Zhou, A brief introduction to weakly supervised learning, National Science Review, vol. 5, no. 1, pp. 44 53, 2017
- [12] Ciencia & Datos, " Aprendizaje Supervisado: Introducción a la Clasificación y Principales Algoritmos", 2019. [En línea]. Disponible en: <https://medium.com/datos-y-ciencia/aprendizaje-supervisado-introducci%C3%B3n-a-la-clasificaci%C3%B3n-y-principales-algoritmos-dadee99c9407>
- [13] Towards Data Science, " A Brief Introduction to Supervised Learning", 2019. [En línea]. Disponible en: <https://towardsdatascience.com/a-brief-introduction-to-supervised-learning-54a3e3932590>
- [14] IBM, "Supervised Learning", 2020. [En línea]. Disponible en: <https://www.ibm.com/cloud/learn/supervised-learning>
- [15] Guru99, "Decision Tree in R | Classification Tree & Code in R with Example", 2021. [En línea]. Disponible en: <https://www.guru99.com/r-decision-trees.html#5>
- [16] TELTONIKA, "Mobile Signal Strength Recommendations", 2022. [En línea]. Disponible en: https://wiki.teltonika-networks.com/view/Mobile_Signal_Strength_Recommendations
- [17] SENATEL, "Roaming Nacional Automático", Boletín No. 3, pp 1, enero 2015.

- [18] New York City Fiber-Optic Internet | Pilot Fiber, "Guide To Speed Tests: How To Run, Read, And Use Them", 2021. [En línea]. Disponible en: <https://www.pilotfiber.com/blog/speed-tests>
- [19] Google Developers, "Tutorial de KML", 2014. [En línea]. Disponible en: https://developers.google.com/kml/documentation/kml_tut?hl=es-419#:~:text=KML%20es%20un%20formato%20de,y%20Google%20Maps%20para%20m%C3%B3viles.

5. ANEXOS

ANEXO I. Cronograma propuesto para la etapa de recolección de datos.

ANEXO II. Datos recolectados en la Ruta A.

ANEXO III. Datos recolectados en la Ruta B.

ANEXO IV. Datos recolectados en la Ruta C.

ANEXO V. Datos recolectados en rutas aleatorias dentro del centro histórico de Quito.

ANEXO I

ACTIVIDADES	DICIEMBRE 2021														TOTAL DE HORAS POR ACTIVIDADES
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	
Realizar 3 mediciones considerando velocidades inferiores a los 20 km/h en la ruta A	2														2
Clasificar los archivos csv recolectados en una base de datos. Organizando por ruta y velocidad	1														1
Realizar 3 mediciones considerando velocidades mayores a 20km/h e inferiores a los 40 km/h en la ruta A		2													2
Clasificar los archivos csv recolectados en una base de datos. Organizando por ruta y velocidad		1													1
Realizar 3 mediciones considerando velocidades inferiores a los 20 km/h en la ruta B					2										2
Clasificar los archivos csv recolectados en una base de datos. Organizando por ruta y velocidad					1										1
Realizar 3 mediciones considerando velocidades mayores a 20km/h e inferiores a los 40 km/h en la ruta B						2									2
Clasificar los archivos csv recolectados en una base de datos. Organizando por ruta y velocidad						1									1
Realizar 3 mediciones considerando velocidades inferiores a los 20 km/h en la ruta C								2							2
Clasificar los archivos csv recolectados en una base de datos. Organizando por ruta y velocidad								1							1
Realizar 3 mediciones considerando velocidades mayores a 20km/h e inferiores a los 40 km/h en la ruta C								2							2
Clasificar los archivos csv recolectados en una base de datos. Organizando por ruta y velocidad								1							1
Realizar 3 mediciones considerando velocidades mayores a 40km/h e inferiores a los 70 km/h en la ruta C												2			2
Clasificar los archivos csv recolectados en una base de datos. Organizando por ruta y velocidad												1			1
Realizar varias mediciones a diferentes velocidades dentro del centro historico para pruebas												2	3		5
TOTAL DE HORAS POR DÍA	3	3	0	0	3	3	0	6	0	0	0	5	3	0	