

## Predicción experimental de cobertura en una red LTE mediante el uso de algoritmos de machine learning

### Experimental coverage prediction in an LTE network using machine learning algorithms

Jorge Enrique Herrera Rubio , German Arley Portilla González ,  
Ana María Osorio Velazco   
Universidad de Pamplona, Colombia

OPEN  ACCESS

**Recibido:**  
15/07/2024  
**Aceptado:**  
18/09/2024  
**Publicado:**  
25/11/2024

**Correspondencia:**  
jherrera@unipamplona.edu.co

**DOI:**  
<https://doi.org/10.17081/invinno.12.2.7215>



Copyright 2024 by  
Investigación e Innovación en  
Ingenierías

#### Resumen

**Objetivo:** Predecir de forma experimental la cobertura de una red Long Term Evolution (LTE) mediante el uso de algoritmos de machine learning. **Metodología:** La investigación es de tipo descriptiva por el análisis de las causas y factores que influyen en las variables, según el método cuasi-experimental por las mediciones de potencia en condiciones controladas de campo. Se realiza el estudio de las variables físicas que determinan la cobertura en la prestación del servicio de una red móvil mediante el trazado de cuatro rutas; los datos son evaluados, pre procesados y filtrados para probarlos en mediante modelos matemáticos clásicos, luego se aplican los métodos estadísticos de predicción, como también los algoritmos de predicción con la ayuda del aprendizaje automático (ML) para la generación de modelos que categoricen el uso de cada celda dentro de la agrupación, entre los métodos a usar son: estadística descriptiva, Random Forest (RD), cuantificación de densidad, Isolation Forest (IF) y Multivariable Outliers (MO). Posteriormente se realiza la validación cruzada y la predicción del uso de la celda. **Resultados:** Con los valores obtenidos se establece la correlación entre el comportamiento de los parámetros de potencia y distancia, y con las variables óptimas de cobertura se compara con los modelos tradicionales para una celda estudiada. **Conclusiones:** El proceso de la integración de técnicas matemáticas y computacionales facilita la toma de decisiones de los ingenieros de radio en el diagnóstico, gestión, planificación y redimensionamiento de redes LTE bajo condiciones en donde las redes se hacen mucho más heterogéneas y los servicios móviles requieren de más recursos de la infraestructura tecnológica.

**Palabras claves:** Algoritmos, cobertura, uso, predicción, experimental, red.

#### Abstract

**Objective:** Experimentally predict the coverage of a Long Term Evolution (LTE) network using machine learning algorithms. **Methodology:** The research is descriptive through the analysis of the causes and factors that influence the variables, according to the quasi-experimental method through power measurements under controlled field conditions. The study of the physical variables that determine the coverage in the provision of a mobile network service is carried out by tracing four routes; the data is evaluated, pre-processed and filtered to test it using classical mathematical models, then statistical prediction methods are applied, as well as prediction algorithms with the help of machine learning (ML) to generate models that categorize the use of each cell within the grouping, among the methods to use are: descriptive statistics, Random Forest (RD), density quantification, Isolation Forest (IF) and Multivariable Outliers (MO). Subsequently, cross-validation and prediction of cell usage are performed. **Results:** With the values obtained, the correlation between the behavior of the power and distance parameters is established, and with the optimal coverage variables it is compared with the traditional models for a studied cell. **Conclusions:** The process of integrating mathematical and computational techniques facilitates the decision making of radio engineers in the diagnosis, management, planning and resizing of LTE networks under conditions where networks become much more heterogeneous and mobile services require of more technological infrastructure resources.

**Keywords:** Algorithms, coverage, use, prediction, experimental, network.

**Como citar (IEEE):** J.E. Herrera Rubio, G.A. Portilla González, A.M. Osorio Velazco "Predicción experimental de cobertura en una red LTE mediante el uso de algoritmos de machine learning", Investigación e Innovación en Ingenierías, vol. 12, no. 2, pp. 178-198, 2024, doi: <https://doi.org/10.17081/invinno.12.2.7215>

## Introducción

La ocupación y sobrecarga de tráfico de las redes celular por el aumento de los servicios de datos y multimedia generan congestión y fallos por falta de cobertura, al requerirse más ancho de banda por parte de los usuarios al prestador de la red, lo que lleva a la sección de ingeniería de red plantear soluciones para mejorar el servicio e ir atendiendo los problemas de escalabilidad y redimensionamiento de infraestructura, en su defecto la calidad del canal de servicio [1] se deteriora para ello se requiere de la asignación de recursos de hardware y software en cada estación base móvil [2].

Las mediciones de campo eléctrico y su correspondencia a los niveles de potencia en sitio pueden facilitar el proceso del redimensionamiento de la red para determinar los límites de cobertura que proporciona un celda en servicio; conociendo el rendimiento del enlace de subida y de bajada se puede realizar la asignación de recursos de forma dinámica en el nodo, para optimizar el ancho de banda [3] y predecir el rendimiento de la red de forma anticipada, sin embargo, se deben de considerar factores implícitos como: el ambiente urbano, la densidad telefónica, altas horas pico de tráfico, el tipo de tecnología, la clase de modulación y el tipo de antenas entre otros [4].

En entornos reales la propagación se vuelve compleja, lo que hace difícil predecir la intensidad de señal debido a que las ondas electromagnéticas se propagan en condiciones adversas y las pérdidas aumentan de forma significativa afectando el rendimiento de los enlaces de radio [5], por tanto, con las campañas de medición en el área de la ingeniería de radio puede apoyarse para facilitar la construcción, ajuste y evaluación de dichos enlaces a través de modelos matemáticos y con la simulación verificar el funcionamiento de la cobertura en diferentes tipos de ambientes [6].

Generalmente existen modelos experimentales y estadísticos que utilizan variables y parámetros conocidos con ecuaciones de tipo determinísticas para establecer un grado de precisión dependiendo del tipo de escenario de prueba y requieren de mucha información para convertirse en modelos de referencia holísticos [7], que han resultado ser buenos predictores, sin embargo, con métodos de aprendizaje automático [8] por medio de un conjunto de datos, es más ágil el proceso de las predicciones para solucionar los problemas de clasificación y regresión [9] de las muestras de los niveles de potencia que se analizan en un enlace de radio.

Con la evaluación de los datos obtenidos en la estación de radio móvil en el casco urbano en la ciudad de Pamplona, se someten a pruebas con técnicas analíticas [10] en función de potencia, campo eléctrico, distancia y las celdas activas que hacen parte del clúster de la red a evaluar con el fin de comparar y predecir el comportamiento del nivel de intensidad de potencia e indicar de forma sintética cual se ajusta mejor al ambiente caracterizado.

## Estado del arte

Para el análisis de predicción de cobertura [10] proporcionan un estudio exhaustivo y actualizado de las diversas técnicas de predicción de cobertura de red de acuerdo a los diferentes terrenos para cada tipo de generación móvil más adecuada de acuerdo a los fenómenos de propagación y modelos estándar más conocidos [11]. Por su parte, [12] realiza la predicción de pérdidas en una ruta utilizando métodos de aprendizaje automático mediante dos mecanismos de entrenamiento y los datos medidos se utilizan para verificar el diseño.

Un referente de redes inalámbricas que requieren de modelos avanzados para encontrar la mayor precisión de la pérdida de trayectoria en entornos complejos

es analizado por [13] utilizando aprendizaje automático para modelar combinando tres técnicas: el análisis de componentes principales (PCA) para reducir el número de características del conjunto de datos y simplificar el modelo de aprendizaje, luego el regresión multidimensional basada en redes neuronales artificiales (RNA) para definir la estructura de pérdida de trayectoria y finalmente el análisis de varianza basado en procesos gaussianos para aprender el efecto de sombra [8].

Con nuevas técnicas se puede lograr una gran precisión en la predicción de la intensidad de la señal considerando la cantidad de obstáculos y las trayectorias de las ondas que se propagan, como lo estudian [14] usando un algoritmo de regresión de aprendizaje automático mediante una aproximación basada en Clustering y algoritmos de regresión automática Random Forest, regresión AdaBoost, y regresión K-Nearest Neighbors. Para determinar la calidad y el rendimiento de la red en condiciones reales [15] también examinan dos enfoques: aprendizaje automático clásico con modelo de bosque aleatorio para predecir el rendimiento del caudal de datos dentro de la red móvil y el modelo de previsión de series temporales, finalmente con aprendizaje profundo obtienen un rendimiento de predicción mejor que las técnicas de previsión de series temporales.

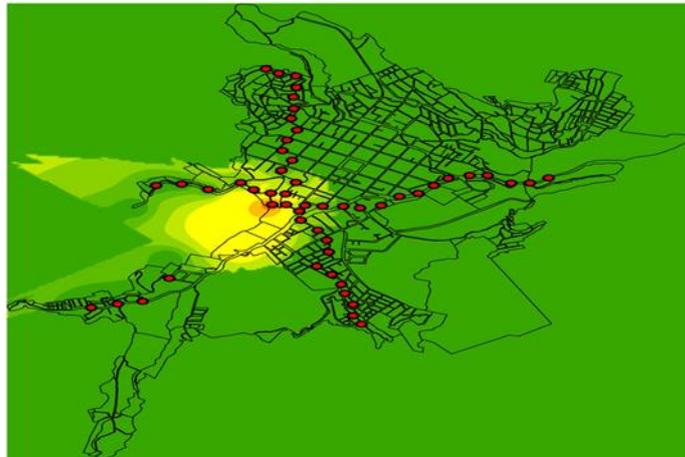
En su investigación [16] detallan la pérdida de atenuación de la señal de radio de una micro celda móvil mediante modelos de predicción de pérdidas del aprendizaje automático con una red neuronal de percepción multicapa (MLP) que es capaz de predecir la pérdida de trayecto de forma precisa hasta en un 50% con el ajuste de los hiper parámetros. De igual manera, [17] comparan el aprendizaje supervisado y de series temporales para predecir el tráfico de datos mensual en horas punta por celda en una red LTE activa, los métodos utilizados son: Random Forest, diferentes redes neuronales, regresión de vectores de soporte, media móvil integrada auto regresiva estacional y Holt-Winters aditivo.

Los autores [18] analizan la aplicación de técnicas de aprendizaje automático a problemas de predicción del rendimiento en redes inalámbrica, utilizando los métodos clásicos de aprendizaje automático, como la regresión de procesos gaussianos, el suavizado exponencial de series temporales y los bosques aleatorios, que pueden dar excelentes resultados de predicción; finalmente proponen tareas de predicción espacial, temporal y multidimensional. A su vez, [19] realizan una comparación entre el modelo de propagación Cost-231 y los métodos Support Vector (SVR), Random Forest (RF) y K-Nearest Neighbor (KNN) en un entorno urbano para comprobar los datos de pérdidas de trayecto resultados

## **Metodología**

La predicción de cobertura del uso de la celda se basa en las mediciones de la variable de potencia con respecto a la distancia, a través de un recorrido de cuatro rutas de una celda en particular confinado en un escenario de pruebas como se muestra en figura 1, se aplican las normas técnicas de medición del espectro para determinar la forma como se extraen los datos mediante mediciones en campo [20].

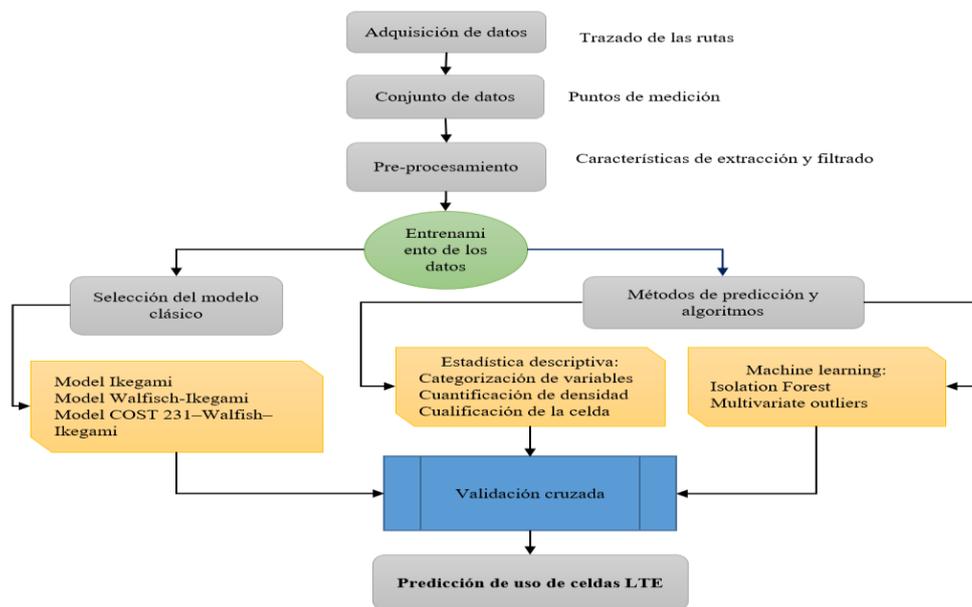
Figura 1. Escenario de estudio celda municipio de Pamplona.



Fuente: Elaboración propia.

Se define la metodología de trabajo de acuerdo al esquema de la figura 2 con sus respectivas etapas que surgen en el proceso.

Figura 2. Esquema del proceso metodológico.



Fuente: Elaboración propia.

### Adquisición de datos

Se establece la toma de muestras durante tres minutos realizando un recorrido por cada una de las rutas cada 100 metros de separación por punto de medición; considerando la línea de vista (LOS) en lo posible desde la radio base, se considera una heterogeneidad del 50%, un margen de error del 10% y un nivel de confianza del 90% [21].

### Conjunto de datos

El proceso de construcción del paquete de datos o dataset, parte desde la selección de cada una de las rutas propias del análisis para la cobertura de celdas

LTE como se presenta en la figura 1, de esta manera se ajusta el dispositivo móvil con soporte Android para capturar muestra por un intervalo de 180 segundos cada punto de medida dicho proceso de captura se hace en caminata con el dispositivo a una altura de 1,5 m. La información de las muestras se capturan en un archivo en texto plano con la información de la celda, longitud, latitud, potencia de la señal, tipo de red y parámetros propios de cada una de las celdas asociadas a la estación base y posteriormente llevada a una hoja de cálculo.

### **Pre-procesamiento**

Con el conjunto de variables de estudio se agregan casillas que permiten un análisis más práctico para el caso del campo eléctrico medido en dBuV para facilitar los procesos de graficación y análisis. Debido a que la información arrojada en la elaboración del Dataset [22] puede contener errores de captura y algunas variables tengan información perdida, faltante o errónea se aplica una técnica de Missing que garantice que las muestras tengan la calidad esperada; el método aplicado es el de media de valor faltante y complemento por valor de vecino superior, lo que corresponde a el reemplazo de datos que no coinciden mediante estas dos estrategias del pre procesamiento de la información y de esta forma eliminar anomalías en los datos que pueden afectar el análisis gráfico y técnica del experimento [23].

Algunos de los datos obtenidos en la casilla generada adicional como campo eléctrico presentó cambios relacionados al manejo del valor extremo en punto decimal, un ejemplo de la situación es la siguiente: al tener una conversión óptima de una de las capturas, en el último dígito luego de una secuencia de ceros después del punto decimal se encuentra un valor de 1, 2 o 3 por lo que el agrupamiento de las variables en valor de campo es afectado, por esta razón se hace necesario realizar un redondeo por debajo a la muestra que presente esta variabilidad.

### **Train Data**

Del volumen de los datos para cada una de las coordenadas se determina mediante Pareto una distribución de 80% para el proceso de pruebas, análisis estadístico y demás estrategias prácticas que permitan analizar el comportamiento de las redes LTE en el uso de celdas, utilizando algoritmos y métodos de predicción [24] y se hace una referencia corta con modelos de propagación estándar para la predicción de potencia que más sea justa al estudio en campo, con el propósito de contrastar el comportamiento con los algoritmos de aprendizaje supervisado. Los datos restantes son usados para procesos de validación, comparación y demás metodologías de comparación que garantice la validación de lecturas y resultados arrojados.

### ***Selección del modelo de clasificación***

Desde el punto de vista del diseño de estaciones radio base para coberturas en ambientes urbanos con la tecnología LTE existen modelos de tipo semi empíricos que facilitan la predicción de los niveles de potencia en ambientes urbanos [25], para el caso de estudio se seleccionan los siguientes modelos con el fin de comprar su funcionalidad y practicidad en el desarrollo de las mediciones en campo.

### Modelo Ikegami

Considera las pérdidas en recepción entre un enlace de radiocomunicaciones por medio de una aproximación matemática de la forma como se recibe una señal de radio considerando los factores de distancia, frecuencia, ángulo de propagación con línea de vista, la altura del móvil, la estación radio base y las pérdidas por reflexión por efecto de las calles, por medio de la ecuación 1 [19].

$$L(R) = 10 \log f + 10 \log(\sin\phi) + 20 \log(h - hr) - 10 \log w + 20 \log \left\{ 1 + \left[ \frac{3}{Lr^2} \right] \right\} - 5,8 \quad (1)$$

Donde:  $f$ : frecuencia en Mega Hertz MHz.  $\phi$ : ángulo entre la calle y la línea directa de la base al móvil.  $Lr$ : pérdida de reflexión y típicamente se asume como 0,25.  $h$ : altura media de los edificios circundantes, en metros.  $hr$ : altura del móvil.  $w$ : ancho de la calle, en metros.

El modelo es útil para estimar la atenuación de la señal y predecir la intensidad de la señal recibida en el diseño y la planificación de redes móviles, ya que permite optimizar la cobertura y la capacidad del sistema.

### Modelo Walfisch-Ikegami

Es el resultado de la fusión de los modelos de J. Walfisch y F. Ikegami, el estudio es de tipo empírico fue desarrollado y mejorado años siguientes. La precisión del modelo es bastante alta en los entornos urbanos especialmente en la propagación sobre los tejados [26], examina el rango de frecuencias de 800 MHz a 2000MHz, la altura del transmisor ( $htx$ ) desde los 4m a 50m, la altura del receptor ( $htr$ ) entre 1m a 3m y la distancia ( $d$ ) entre el transmisor y el receptor. Para el caso de estudio se aplicó la condición del modelo cuando no hay línea de vista, utilizando la ecuación 2.

$$Lb = L_o + L_{rts} + L_{msd} \quad (2)$$

Dónde:  $L_o$  son pérdidas por el espacio libre,  $L_{msd}$  las pérdidas por difracción de múltiples esquinas de los techos de los edificios (Walfisch) y  $L_{rts}$  Pérdidas debido a una única difracción final cuando la onda se propaga hacia la calle (Ikegami) y la orientación de estas con respecto a las antenas transmisoras [27]. El valor de  $Lb$  puede llegar a ser mínimo de  $L_o$  cuando  $L_{rts} + L_{msd} \leq 0$ . El desarrollo de la ecuación de este modelo se integra en el modelo compacto que se detalla a continuación

### Modelo COST 231-Walfish-Ikegami

Se desarrolló para sistemas micro celular para enlaces con línea de vista (LoS) y enlaces sin línea de vista (NLoS) [4], para este caso las pérdidas de trayecto expresan en función de los siguientes parámetros: la estación móvil ( $ms$ ), estación base ( $bs$ ), distancia en metros ( $d$ ), la altura de la estación base ( $bs$ ) sobre el nivel del mar ( $hb$ ) para valores entre  $4 \leq hb \leq 50$  (m), la altura de la antena de la estación móvil ( $hm$ ) sobre el nivel del mar para valores entre  $1 \leq hm \leq 3$  (m), altura nominal del techo de los edificios en metros ( $hRoof$ ), la altura de  $bs$  con respecto a los tejados ( $\Delta hb = hb - hRoof$ ), altura de  $MS$  con respecto a los tejados ( $\Delta hm = hRoof - hm$ ), ancho de las calles en metros ( $w$ ), separación de edificios en metros ( $b$ ) y ángulo de la onda incidente que se forma con la calle en grados( $\phi$ ) [28].

En casos donde no se tienen los datos reales de la infraestructura de los edificios y las carreteras, los valores por defecto son:  $b = 20\text{m}$  hasta  $50\text{m}$ ,  $w = b/2$  (m),  $\phi = 90^\circ$ , y  $h_{Roof} = 3 \times \text{número de pisos} + \text{techo}$  (m), donde el techo =  $3\text{m}$  cuando es inclinado y  $0\text{ m}$  si es plano.

Las pérdidas de trayectoria por NLoS son:

$$L_p(dB) = \begin{cases} L_o + L_{rts} + L_{msd}, & \text{para } L_{rts} + L_{msd} \geq 0 \\ L_o, & \text{para } L_{rts} + L_{msd} < 0 \end{cases} \quad (3)$$

Se define la pérdida de espacio libre como:

$$L_o = 32.4 + 20 \log_{10}\{d\} + 20 \log_{10}\{f_c\} \quad (4)$$

La difracción y la pérdida de dispersión del techo a calle representan el acoplamiento de la propagación de la onda a lo largo de la trayectoria de la multipantalla en la calle donde se encuentra la ms, y viene dada por [29]:

$$L_{rts} = -16.9 - 10 \log_{10}\{w\} + 10 \log_{10}\{f_c\} + 20 \log_{10}\{\Delta h_m\} + L_{ori} \quad (5)$$

Para [3] la pérdida de orientación en la calle:

$$L_{ori} = \begin{cases} -10 + 0.354(\phi) & , \quad 0 \leq \phi \leq 35^\circ \\ 2.5 + 0.075(\phi - 35^\circ) & , \quad 35^\circ \leq \phi \leq 55^\circ \\ 4.0 - 0.114(\phi - 55^\circ) & , \quad 55^\circ \leq \phi \leq 90^\circ \end{cases} \quad (6)$$

La pérdida por difracción en múltiples pantallas es:

$$L_{msd} = L_{bsh} + k_a + k_d \log_{10}\{d\} + k_f \log_{10}\{df_c\} - 9 \log_{10}\{b\} \quad (7)$$

Para:

$$L_{bsh} = \begin{cases} -18 \log_{10}\{1 + \Delta h_b\}, & h_b > h_{Roof} \\ 0 & , h_b \leq h_{Roof} \end{cases} \quad (8)$$

$L_{bsh}$ , es la ganancia de sombra (pérdida negativa) para los casos en que la antena de la bs está por encima de los tejados. Los parámetros  $k_a$  y  $k_d$  dependen de la longitud del trayecto  $d$  y de la elevación de la estación de base con respecto a los tejados  $\Delta h_b$ .

El término  $k_a$  son las pérdidas del trayecto cuando las antenas de la bs están situadas por debajo de los tejados de los edificios adyacentes y viene dado por [30]:

$$k_a = \begin{cases} 54 & , h_b > h_{Roof} \\ 5 - 0.8\Delta h_b, & d \geq 0.5 \text{ km and } h_b \leq h_{Roof} \\ 5 - \frac{0.8\Delta h_b d}{0.5}, & d < 0.5 \text{ km and } h_b \leq h_{Roof} \end{cases} \quad (9)$$

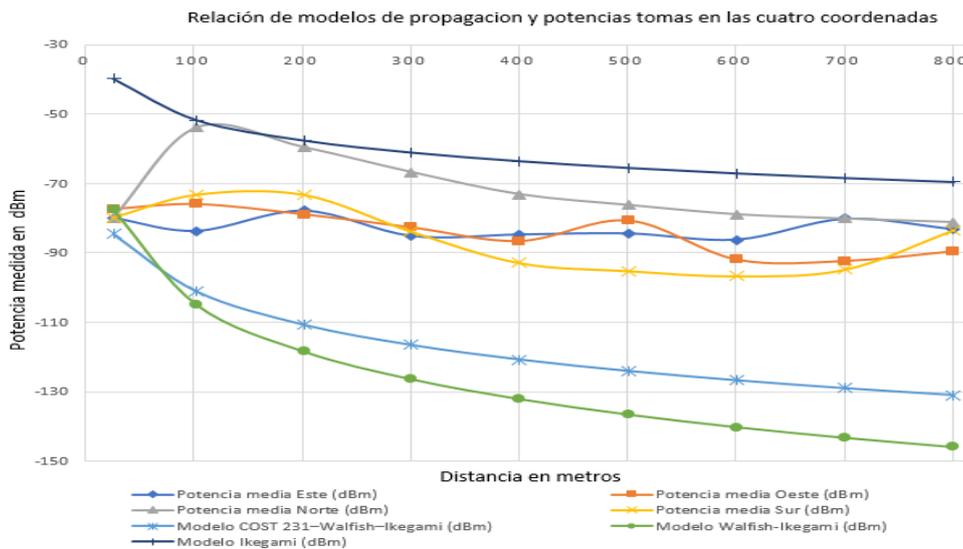
Los términos  $k_d$  y  $k_f$  representan la dependencia de la pérdida de difracción de la multi pantalla de la distancia y la frecuencia, definidos como:

$$k_d = \left\{ \begin{array}{l} 18, \quad h_b > h_{Roof} \\ 18 - \frac{15\Delta h_b}{h_{Roof}}, \quad h_b \leq h_{Roof} \end{array} \right\} \quad (10)$$

$$k_f = -4 + \left\{ \begin{array}{l} 0.7 \left( \frac{f_c}{925 - 1} \right) \quad \text{Ciudad media y suburbana (*)} \\ 1.5 \left( \frac{f_c}{925 - 1} \right) \quad \text{area metropolitana} \end{array} \right\} \quad (11)$$

Considerar que (\*) es utilizada para densidad moderada de arboles. Este modelo funciona mejor para el  $h_b \gg h_{Roof}$ . Se pueden esperar grandes errores de predicción para  $h_b \approx h_{Roof}$  [31], y se considera un modelo pobre cuando  $h_b \ll h_{Roof}$ , porque no consideran la guía de ondas en los cañones de las calles y la difracción en las esquinas de las calles [32], los modelos utilizados se aprecian en la figura 3.

Figura 3. Modelos de propagación para predicción del estudio de campo



Fuente: Elaboración propia.

En la comprobación de las mediciones realizadas en campo, se observa que el modelo de Ikegami es útil en micro celdas de corto alcance y considera las contribuciones del primer rayo difractado y el segundo rayo reflejado con el fin de calcular la potencia que proviene del rayo directo [33], esto indica que el ángulo entre las calles y la línea de vista directa como también al altura del móvil y el ancho de las calles son indicadores que facilitan que la señal de radiofrecuencia se reciba en el móvil para predecir la intensidad de potencia. Aunque las curvas de los otros dos modelos tiene una tendencia similar se detalla un decaimiento exponencial a medida que el móvil se aleja de la radio con pérdidas de potencia entre 70dB a 80 dB por debajo de los valores medidos, lo cual no son modelos candidatos para hacer este tipo de predicción.

### Métodos de predicción y algoritmos

Para este apartado resulta necesario aplicar una serie de algoritmos que permitan una visión clara y contundente del comportamiento de los datos que garanticen como resultado la medición de uso de las celdas LTE, teniendo como base las variables de señal, campo eléctrico e identificador de la misma, para tal fin detallan las técnicas y procedimientos que permiten el análisis de las datas capturadas como se relaciona en el parágrafo anterior.

### Estadística descriptiva

#### Categorización de variables

Las variables que componen el Dataset, están dispuestas sobre el archivo plano que es analizado mediante el entorno Jupyter [34] que usa lenguaje de programación Python, este permite el uso de herramientas, librerías y demás paquetes que minimizan los tiempos de ejecución y garantizan resultados de visualización óptimos para cada proceso [35]. En la figura 4, se observa el comportamiento estadístico de los datos capturados para cada una de las coordenadas, estos se ubican de izquierda a derecha para la primera fila al igual que para la segunda fila, con las coordenadas siguientes: arriba Este - Oeste y abajo Norte - Sur, tomando como referencia de análisis el Este el campo eléctrico tienen un conteo de menos 2170 lecturas que corresponden al 100% de los datos.

El promedio del campo eléctrico es de 84.54 dBuV mientras que para el caso de Oeste, imagen superior derecha el promedio está en 83.72 dBuV, lo que corresponde a una variabilidad de un 1% respectivamente, en estas dos trayectorias se mantiene en un rango con diferencia de  $\pm 1$ , e n el caso Norte con un valor promedio 90.4 dBuV y la coordenada Sur de 82.56 dBuV lo que muestra que se encuentra en un rango similar, pero distante en al menos 6 puntos respecto a Norte, por tanto existe un relación de mejora en el modo de propagación de la onda desde el punto de emisión o en su defecto una potencia más elevada en la celda respecto a las demás coordenadas.

Figura 4. Estadística descriptiva de los datos

Ruta Este	Distance	Cellid	Signal	dBuV	Voltios	Field
count	2170.000000	2.170000e+03	2170.000000	2170.000000	2170.000000	2170.000000
mean	331.751152	6.406648e+05	-83.026728	23.973272	0.049844	84.545622
std	239.382139	3.561993e+05	14.284704	14.284704	0.075806	14.669967
min	0.000000	4.783080e+05	-113.000000	-6.000000	0.000501	54.000000
25%	100.000000	4.783090e+05	-90.000000	17.000000	0.007079	77.000000
50%	300.000000	4.783590e+05	-84.000000	23.000000	0.014125	83.000000
75%	500.000000	4.783590e+05	-75.000000	32.000000	0.038611	97.000000
max	800.000000	1.640855e+06	-57.000000	50.000000	0.316228	110.000000

Ruta Oeste	Distance	Cellid	Signal	dBuV	Voltios	Field
count	939.000000	9.390000e+02	939.000000	939.000000	939.000000	939.000000
mean	246.645367	6.121085e+05	-83.225772	23.722045	0.048263	83.722045
std	278.851079	2.943165e+05	13.675969	13.730674	0.073176	13.730674
min	0.000000	4.783080e+05	-113.000000	-6.000000	0.000501	54.000000
25%	100.000000	4.783100e+05	-93.000000	14.000000	0.005012	74.000000
50%	100.000000	4.783590e+05	-84.000000	23.000000	0.014125	83.000000
75%	500.000000	5.999100e+05	-73.000000	34.000000	0.050119	94.000000
max	800.000000	1.640855e+06	-57.000000	50.000000	0.316228	110.000000

Ruta Norte	Distance	Cellid	Signal	dBuV	Voltios	Field
count	1592.000000	1.592000e+03	1592.000000	1592.000000	1592.000000	1592.000000
mean	414.572864	5.602000e+05	-76.596734	30.403266	0.120013	90.403266
std	297.297602	2.546282e+05	15.525310	15.525310	0.183788	15.525310
min	0.000000	4.780540e+05	-113.000000	-6.000000	0.000501	54.000000
25%	0.000000	4.783080e+05	-89.000000	18.000000	0.007943	78.000000
50%	500.000000	4.783080e+05	-77.000000	30.000000	0.031623	90.000000
75%	700.000000	4.783590e+05	-65.000000	42.000000	0.125893	102.000000
max	800.000000	1.640855e+06	-47.000000	60.000000	1.000000	120.000000

Ruta Sur	Distance	Cellid	Signal	dBuV	Voltios	Field
count	1044.000000	1.044000e+03	1044.000000	1044.000000	1044.000000	1044.000000
mean	257.662835	6.168017e+05	-84.436782	22.563218	0.044008	82.563218
std	265.079625	2.877928e+05	13.637227	13.637227	0.070594	13.637227
min	0.000000	4.783080e+05	-113.000000	-6.000000	0.000501	54.000000
25%	0.000000	4.783590e+05	-94.000000	13.000000	0.004467	73.000000
50%	200.000000	4.783590e+05	-85.000000	22.000000	0.012589	82.000000
75%	500.000000	5.999100e+05	-76.000000	31.000000	0.035481	91.000000
max	800.000000	1.640855e+06	-57.000000	50.000000	0.316228	110.000000

Fuente: Elaboración propia.

Para el caso de medición del 75% de los datos capturados la percepción es similar a la tomada sobre el total, puesto que se presentan las siguientes lecturas, se parte desde la coordenada Este con un valor de campo de 97 dBuV, coordenada Oeste 94 dBuV, coordenada Norte 102 dBuV y coordenada Sur 91 dBuV, el comportamiento difiere en un par de puntos porcentuales, para el caso de Este a Oeste la diferencia es de 3 unidades respecto al total de los datos, o promedio del apartado anterior y para el caso de Norte y Sur la diferencia es

de 11 puntos, similar al Norte y Sur del total del datos presentados anteriormente; este comportamiento del conjunto de las celdas LTE para los procesos de captura comparte similitud tanto en el método como en el resultado, los cambios de aumento observados en promedio y 75% solo obedecen a la distancia estudiada para cada caso entre 0 y 800 metros.

La presentación de datos para las cuatro coordenadas en sus valores máximos, soporta el comportamiento de la coordenada norte y el uso de las celdas alcanza mayor campo eléctrico respecto a las tres coordenadas restantes, resultados que serán comparados con modelos de propagación.

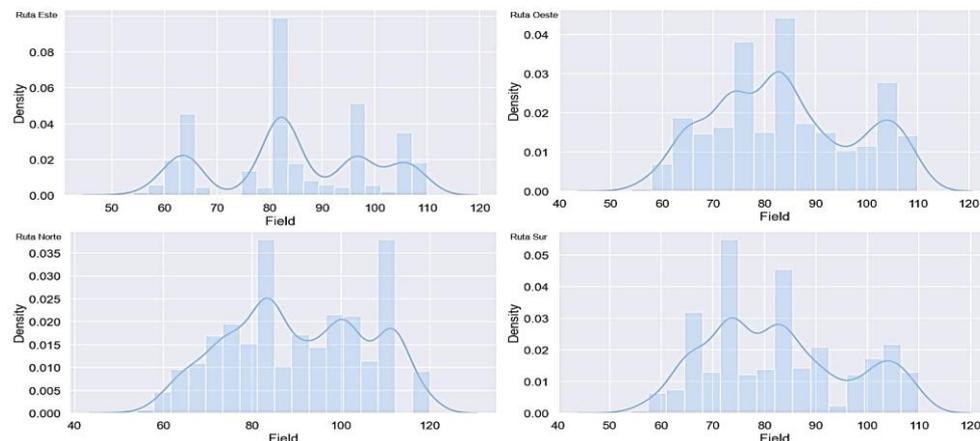
### Cuantificación de densidad

El comportamiento de los datos respecto a las variables tomadas para cada una de las muestras separadas en tiempo y distancia presentan cambios observables en una variable de control, como se muestra en la figura 5 se analiza la distribución de densidad [36] de acuerdo al valor del campo eléctrico; en este caso se cuantifica la cantidad de campo eléctrico que utiliza el móvil receptor a lo largo del recorrido [37].

Para el caso de la coordenada Este existen valores entre 50 dBuV y 60 dBuV que son valores mínimos de uso a lo largo del tramo, que corresponden al límite inferior de estudio, mientras que para el caso del límite superior su cuantificación de usabilidad corresponde a valores con baja frecuencia o repetibilidad, entre 100 dBuV y 110 dBuV. Para el caso de las demás coordenadas el comportamiento tiende a ser similar, con rangos en extremos dados por mínimos y máximos valores con una frecuencia de repetición más baja que los valores intermedios dados por valores de campo eléctrico para este tipo de tecnología, el análisis de densidad permite observar el comportamiento en términos de campo eléctrico.

Puntualmente en las cuatro trayectorias se presentan valores de 80 dBuV y 90 dBuV con las mayores frecuencias en cada una de las cuatro coordenadas, analizando el caso de Este, existe una concentración de red en límite inferior entre 60 dBuV y 70 dBuV, una concentración más alta para el caso de 80 dBuV a 90 dBuV y valores de 100 dBuV a 110 dBuV con una menor proporción del aporte de potencia, es estos casos es cuando las celdas cercanas aporten potencia para realizar un handover imperceptible al usuario móvil, sin embargo, por la distribución heterogénea del ambiente urbano existen niveles donde la señal se hace casi imperceptible razón por la cual al curva de distribución de densidad no es tan homogénea como en los otros trayectos.

Figura 5. Densidad de campo eléctrico dBuV



Fuente: Elaboración propia.

Para la situación planteada de coordenada Norte y Sur la frecuencia de uso de cada uno de los valores de campo eléctrico son más simétricos, esto indica que el uso de las celdas cercanas otorgan niveles de potencia aceptables en un rango estándar sin una marcada variabilidad, situación que es un poco desproporcionada en las coordenadas Este y Oeste, esto evidencia que existe sectores o puntos dentro del área de traslado de captura de la potencia y campo eléctrico que aumentan debido a los diversos parámetros y mecanismos de propagación, existiendo una aporte de potencia adicional cuando se realiza el traspaso de celda durante el recorrido del móvil, de ahí el aumento de densidad respecto al campo eléctrico [38].

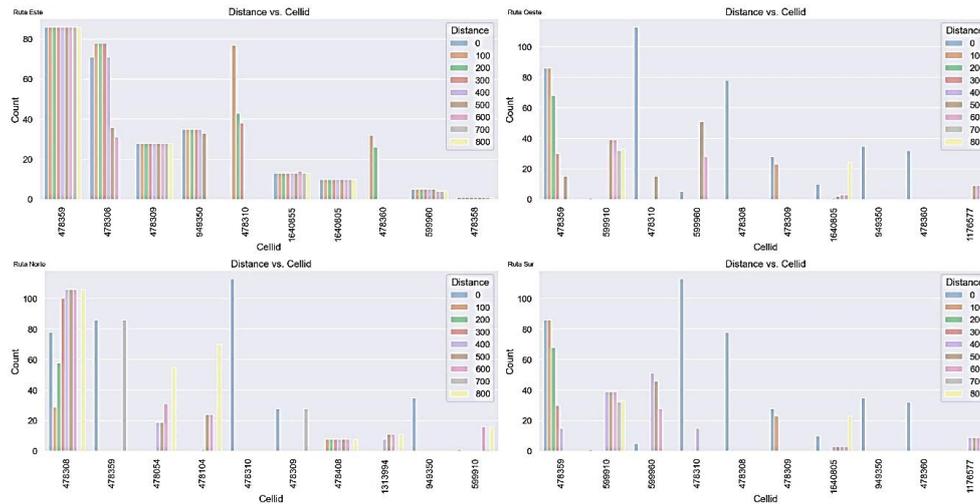
### *Cualificación de la celda*

A partir del conteo de grupos de uso [9] de celdas durante el recorrido del experimento se aprecia en la figura 6 que existen categorizadas 10 identificadores de celdas, que suministran niveles de potencias variables a lo largo de cada trayectoria, del mismo modo el uso de cada celda depende de parámetros de transmisión, ubicación y pérdidas presentes a lo largo del proceso, se dispone de 9 grupos de control como distancias para percibir de manera directa la cantidad de veces que es usada dicha celda en la trayectoria cerrada de 0 a 800 metros, de la siguiente manera.

Para el caso de la Cellid identificada con el número 478359 es la más usada en toda la ruta con un conteo de uso de 80; mientras que las otras celdas tienen características de usabilidad inferior, seguida de la Cellid 478308 la incidencia es de menos de 80 veces de uso, pero no en todas las distancias lo que permite denotar que la usabilidad decrece respecto a la celda de mayor uso, esto se debe a que al usar una celda esta será liberada para tomar otra que no afecte el proceso cobertura de la red, por consiguiente todas las celdas tendrán menor calidad de uso a lo largo de la trayectoria cerrada como se observa en la figura 5, ubicada en la parte superior izquierda que relaciona la coordenada Este.

Para el caso de la celda 478358 que es anterior a la celda de alta cualificación el aporte de conteo está por debajo de los 10 puntos lo que indica que la capacidad de la misma está sesgada por el constante uso de las celdas de mayor capacidad de potencia que dependen de los diferentes parámetros de propagación; para el caso de la coordenada Este el comportamiento prima sobre 5 celdas respecto a las demás y estas a su vez soportan la transmisión LTE con usos de más de 20 veces en la trayectoria marcada de 0 a 800 metros.

Figura 6. Conteo de uso de celda en trayectoria.



Fuente: Elaboración propia.

De la misma forma para las coordenadas Oeste y Sur se tiene presencia de la celda 478359 con una usabilidad inferior, lo que quiere decir que su configuración determina una buena calidad de señal, pero con el cambio notable de que no se usa en toda la trayectoria, sino que está enmarcada en 5 puntos máximos de toda la trayectoria.

Analizando el caso de Oeste y Sur el uso de esta celda de calidad está presente máximo hasta los 500 metros, la cobertura para las coordenadas diferentes a la del caso Este, dependen directamente de las demás celdas que dan aporte de cantidades inferiores a 40 puntos a excepción de Cellid 478310 que suministra aporte de más de 100 usos para las coordenadas Norte, Sur y Oeste lo que indica que la calidad de la misma es practica pero no garantiza niveles de cobertura óptimos para distancias superiores a los 100 metros, puede que luego de cruzar el umbral de distancia permitido la celda 478359 garantice la cobertura LTE en diferentes puntos de trayectoria.

### Machine learning:

#### Isolation Forest

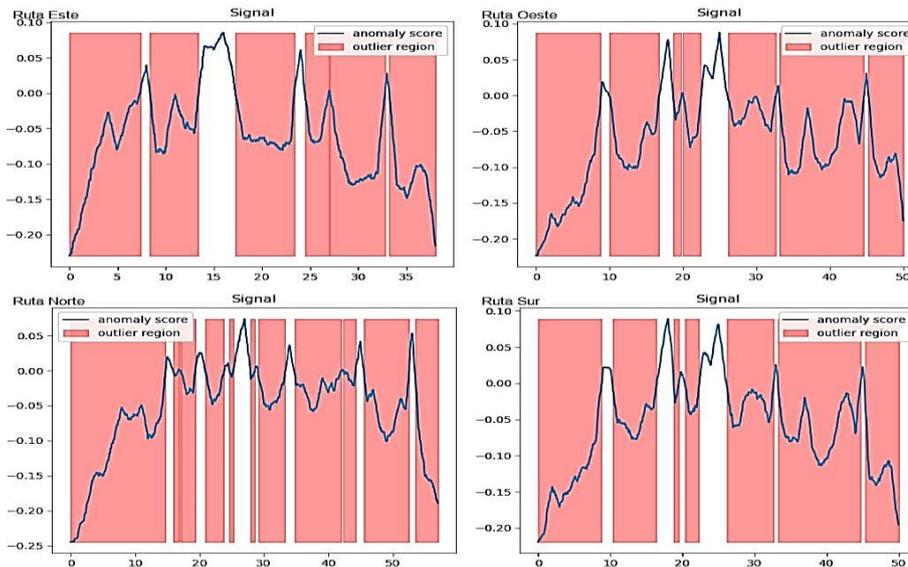
Mediante el algoritmo Isolation Forest [39] se busca obtener una tendencia de comportamiento de la señal, este algoritmo genera una segmentación aleatoria del conjunto total de los datos, los cuales corresponden al 80% de los capturados en la etapa preliminar, el número de divisiones seleccionada por el algoritmo depende directamente de la calidad de los mismos, el número de nodos generados en el árbol está dado por la siguiente ecuación.

$$c(n) = 2H_{n-1} - \frac{2(n-1)}{n} \quad (12)$$

Para el caso de estudio en la ruta Este se toma al menos 6 grupos, que relacionan directamente con la variable señal y su comportamiento a partir de las muestras particulares hacia adelante y atrás de cada una de las muestras para seleccionar anomalías con saltos relativamente cortos [40]. Esta partición entrega una región aislada puntuada del total de los datos y se enmarca por el color sólido, del mismo modo a partir de la región aislada del árbol sobre el conjunto de dato se puntúa la anomalía de señal a lo largo de las puntuaciones de región aislada del valor de señal a partir de las observaciones y luego se selecciona de manera aleatoria un

valor que es dividido entre los valores mínimos y máximos de la característica señal según este caso, como se detalla en la figura 7.

Figura 7. Patrón de comportamiento basado en el conjunto de anomalías Outliers



Fuente: Elaboración propia.

El algoritmo muestra una similitud entre las cuatro coordenadas y las regiones aisladas en promedio son las mismas, lo que difiere es el comportamiento de las anomalías presentes en la señal, pero se puede soportar debido a que la cobertura se relaciona directamente con el identificador de celda que suministra una potencia diferentes respecto a la distancia y ubicación del observador en la trayectoria cerrada; se resalta que en la coordenada Norte en la cuantificación de la trayectoria las anomalías son superiores al comprar con la otras coordenadas como lo evidencia la figura anterior.

### Multivariate Outliers

Se establece la medición del conjunto de vecinos K cercanos para las medidas capturadas respecto a la señal entregada por cada celda y se compara con los valores atípicos locales para cada una de las mediciones y de igual forma se comparan con cada conjunto de datos con un dato promedio de los vecinos K más cercanos respecto a la variable medida [41].

En la coordenada Este existe una diferencia, debido a que la relación de los vecinos es similar por la representación aproximada de la densidad de alcanzabilidad local (LRD) de cada punto y para el caso de los vecinos en los datos superiores es casi igual a 1 como factor unitario pero con alcance de 20 para el primer segmento, mientras que para el caso de la coordenada Oeste es de 30 como se observa en la figura 8, pero en este caso para valores atípicos superiores a los marcados por la coordenada inicial [42].

La información en cada uno de los grupos suministrados por el Valor Atípico Local (LOF) [43] se evidencia que los niveles de señal se mantienen en valores similares a lo largo de la ruta y las fluctuaciones respecto a las variabilidad de los valores atípicos se enmarca en las coordenadas Norte, Sur y Oeste añadiendo que su curva característica presenta una fluctuación de distancia media y un decremento en la distancia media hacia límite superior, comportamiento que no está presente en la

coordenada Este de manera marcada, sino que el fenómeno es casi nulo en distancias sobre los 400 metros hasta los 700 metros, ver figura 8.

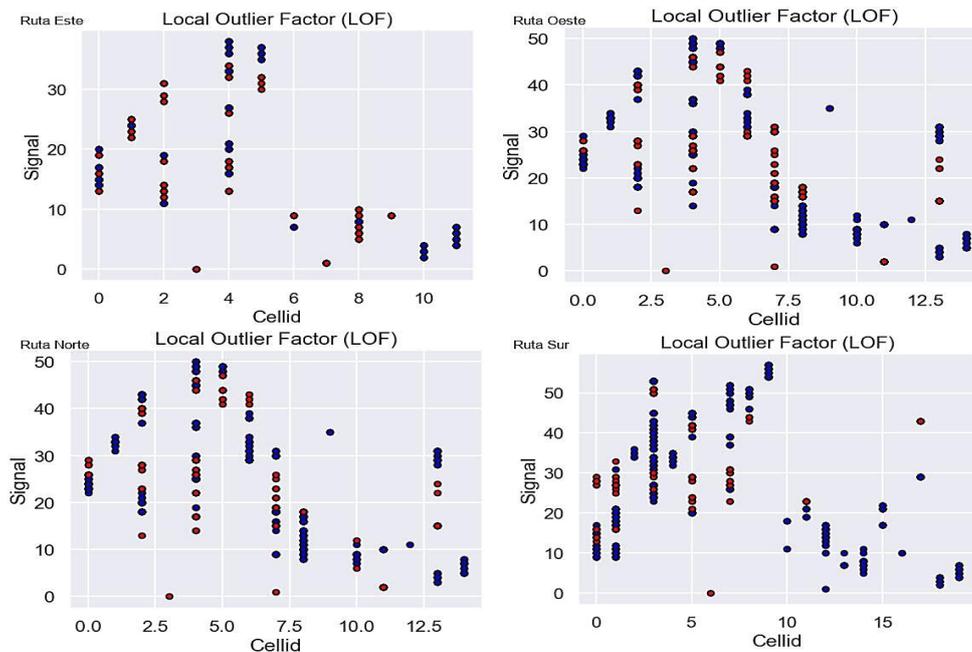
El parámetro dado por LDR está dado por:

$$LRD_k(A) = \frac{1}{\sum_{Xj \in N_k(A)} \frac{RD(A,Xj)}{\|N_k(A)\|}} \quad (13)$$

La ecuación anterior presenta de manera implícita la medida inversa promedio en que se tiene acceso a las celdas vecinas, quiere decir que existirá menor densidad de puntos de valores de potencia enmarcados en cada grupo de análisis, indicando que tan lejos está un celda con respecto a sus vecinas en el caso de la ruta Este; la no saturación de datos indica que existe un distanciamiento de los valores atípicos que son comunes en las demás coordenadas; con la ecuación 13 se calculan los LOF a lo largo de una trayectoria cerrada respecto al conjunto de valores atípicos en la ramas inferiores o superiores de cada coordenada [44].

$$LOF_k(A) = \frac{\sum_{Xj \in N_k} LRD_k(Xj)}{\|N_k(A)\|} X \frac{1}{LRD_k(A)} \quad (14)$$

Figura 8. Local Outliers Factor Cellid vs Signal



Fuente: Elaboración propia.

La mejor coordenada que representa el aporte de potencia de las celdas está en la ruta Este, debido que la separabilidad entre los valores atípicos corresponde a una poca fluctuación entre los valores de señal, mientras que para las otras coordenadas se presentan saturaciones incrementales y decrementales para cada uno de los grupos representados por el algoritmo LOF, pero que en forma de curva para el caso de Oeste, Norte y Sur es muy similar para no dar un criterio de igualdad, y se debe al comportamiento de las celdas a lo largo de trayectoria, mientras que la concentración de uso se limitaba al Cellid 478359 en el caso de Este, para los demás estaba distribuida en otro número de celdas pero con capacidades de radiación de potencia muy inferiores en la trayectoria, por lo que el uso de los grupos de celdas fue distribuido a lo largo de las 4 trayectorias para

analizar la manera como se mantiene la cobertura de la estación base que está radiando en esa ubicación.

## Resultados

### Validación cruzada

En el proceso de análisis de cada una de las muestras dentro de las rutas se presentan resultados que varían según el desplazamiento de los grupos de entrenamiento y predicción de cada uno de los métodos usados, en las primeras iteraciones se desplazan los bloques de datos sectorizados en proporciones iguales al 20%, en cada una de las etapas siguientes con el fin de que la precisión se ajuste al comportamiento real de la celda se realiza por lo menos 4 veces hasta que cada una de las muestras presentes por cada traza sea analizada; la medida de los grupos en el proceso de predicción arroja porcentajes poco distantes entre 80 %, 85% y 89% que relacionan la predicción de celda que cumple en gran medida, del mismo modo pueden emplearse procedimientos similares que permitan que la cantidad total de los datos sean analizados y comparados con el fin de minimizar el error que presente el modelo.

Si se toman los resultados de los algoritmos de aprendizaje de las predicciones de las celdas y se comparan con los entregados por el modelado matemático Ikegami el nivel de potencia suministrado por la celda disminuye respecto a la distancia alrededor de 50 dBm al igual que para el caso de uso de predicción; se observa que la celda 478359 es la más representativa en el recorrido de 0 a 800 metros, del mismo modo varias celdas son usadas a lo largo de toda la trayectoria de acuerdo al análisis realizado por su cuantificación de uso, algunas como la 478358 presentan valores mínimos de uso que se relaciona con el nivel de potencia que pueden suministrar hasta el punto de trayectoria final.

El modelo COST 231-Walfish-Ikegami presenta una atenuación similar a la presentada al modelo Walfish pero con una pendiente mayor y esto se debe a las consideraciones de parámetros que intervienen en el proceso, el comportamiento está relacionado directamente con la densidad de campo eléctrico en cada una de las rutas que indica el lugar donde se tiene la mayor potencia en el dispositivo móvil.

### Predicción de uso de la celda

Los métodos de predicción con la estadística descriptiva son buenos indicadores para resolver el problema de planificación y redimensionamiento de forma holística como sucede cuando se realiza el diseño para mejorar la precisión por agrupación de las estaciones radio bases acercándolas entre sí, hasta el punto donde la señal es débil; en este caso se utiliza un gráfico de decisión para establecer un umbral de peso que diferencia las celdas vecinas y los valores atípicos, como también el conteo de grupos de uso de celdas como se analizó en la figura 5 y 6. Pero se encuentran defectos como la incapacidad de generar gráficos eficaces para un conjunto de potencias medidas a pequeñas escalas y las curvas de decisión en la densidad de probabilidad del campo eléctrico son muy sensibles a variaciones momentáneas.

En el caso de modelos de predicción deterministas y empíricos como Ikegami, Walfisch-Ikegami, COST-231- Walfisch-Ikegami y otros; aunque son la fuente más real por las medidas que se realizan en campo, permiten realizar el ajuste de los parámetros y con ellos se estiman las pérdidas, porque consideran los factores externos, eléctricos y físicos que afecta la propagación, además ayudan de forma acertada en los posibles cambios que tiene la potencia espectral que se recibe en

punto específicos del sitio de medición, sin embargo, son más complejos por la cantidad de variables considerar.

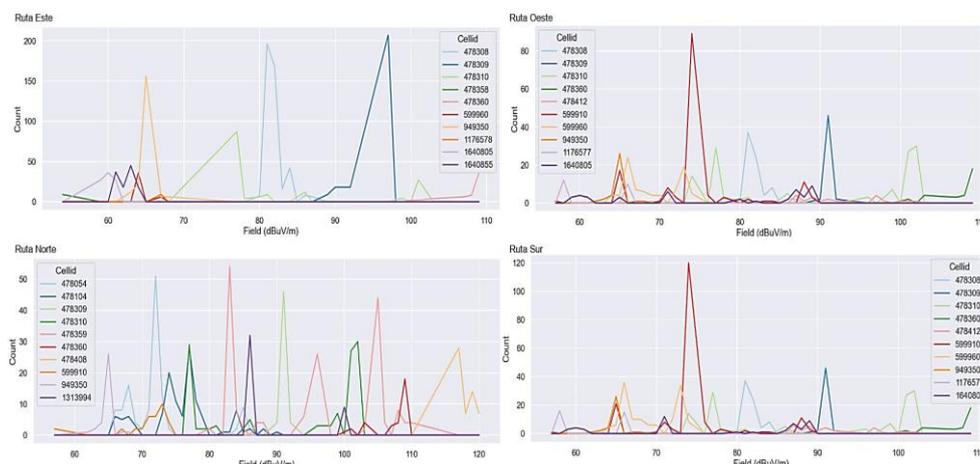
Ahora bien, los algoritmos de aprendizaje supervisado seleccionados para el caso de estudio permiten la integración con los modelos y método expuesto anteriormente, se tiene que el Isolation Forest detecta las anomalías de valores de potencia que facilita el aislamiento de datos que no son útiles; al construir los arboles de decisión de forma aleatoria con un entrenamiento selectivo se pueden descartar la celdas que mejor cobertura presentan, esto se logró con una división en función de las observaciones de cuantas divisiones son necesarias para separarlas del resto del grupo, lo que avala una buena selección, que al combinarla con los valores atípicos multivariantes se puede descartar la variable de potencia para generar regiones extremas en un espacio multidimensional de acuerdo con la potencia que emite la celda durante la ruta examinada.

Es aquí donde es posible discriminar los errores de medición, errores en la entrada o captura de los datos y fenómenos extraordinarios (físicos) en la realización de pruebas piloto que distorsionan los parámetros y las variables que se están estudiando, lo que lleva un sesgo del resultado de análisis como se discutió en el comentario de las figuras 7 y 8, que facilita la interpretación de los resultados.

El uso de las celdas que intervienen en el proceso son cuantificadas respecto a la cantidad de veces que se presentan en la trayectoria de 0 a 800 metros, en la figura 9 se pueden observar las 10 celdas relevantes, para el caso de la ruta Este se tienen al menos 5 celdas que con valores de campo eléctrico de 70 a 100 dBuV/m y las demás generan aportes pero la media obtenida es inferior al esperado para ser representativa

Al momento de generar un desplazamiento sobre los datos mediante la sectorización de información de los datos la predicción de uso de celda para el caso de este se encuentra entre la 478309, 478308 y 949350 con usos superiores a las 100 veces, para el caso de las rutas Oeste y Norte la cantidad de usos de celdas no supera las 90 veces, este resultado no quiere decir que el uso es mínimo a lo largo de la trayectoria sino que, muchas de las celdas que no generaban aporte significativo en la ruta Este, para el caso de Oeste y Norte superan la media de cruce y son observables en la figura de uso.

**Figura 9.** Comportamiento del campo eléctrico respecto a celdas de más uso.



Fuente: Elaboración propia.

Para el caso de Norte las 10 celdas tienen actividad representativa con aportes equivalentes, lo que quiere decir que están en continuo uso a lo largo del trayecto. Para la ruta Sur el comportamiento es similar a la ruta Este las celdas de aporte significativo son 10 pero solo el 50% supera la media, para el caso de la celda 599910 el aporte de campo eléctrico está en al menos 75dBuV/m con una incidencia de 120 veces, pero en el caso de la celda 478309 que presenta un nivel de campo eléctrico mayor, pero con uso a lo largo de la trayectoria menor, esto indica un mayor campo eléctrico en toda la celda pero no garantiza que se presente traspaso del celda que puede deberse a los factores intrínsecos y extrínsecos que afectan de manera positiva a unas celdas y a otras de manera negativa.

## Conclusión

Para este análisis diagnóstico se aprecia que las celdas de mayor potencia donde está concentrado en la configuración de las Cellid 468359, 478308 y 478309 aporta un mayor nivel de potencia de acuerdo al conteo de uso de celdas (cualificación de la celda) por trayectoria, evitando el traspaso de celda de forma continua y manteniendo un mejor relación de potencia cobertura por celda, que es verificable con el uso de los modelos de propagación en los primero 300 metros donde se tomaron los datos.

Esto permite la combinación de la técnica de selección de modelos de clasificación y los algoritmos de predicción para que el ingeniero de radio determine los niveles de potencia necesarios garantizando el servicio para no saturar los recurso de hardware del sistema, evitando la interferencia cocanal y aumentar el factor de reuso en micro celdas que no superan el radio de cobertura en un kilómetro, y que influye en una mejor redimensionamiento de la red.

Con el uso de los Multivariate Outliers se facilita la predicción de las celdas vecinas que mejor relación nivel de potencia aportan en un redimensionamiento de una red LTE, útil para situaciones que requieren cambio y configuraciones de hardware y software, basado en los valores atípicos de potencia y representando la mejor densidad de uso de la celda en condiciones de alta demanda de tráfico, situación que no es tan visible con el uso de los modelos de propagación estándar.

Para el caso del algoritmo Isolation Forest éste un buen indicador de las anomalías que se pueden presentar cuando sucede el traspaso de celdas en ambientes urbanos densos cuando se excede el límite del nivel de potencia permitido y sobre pasa el valor de -110dBm, es decir, la señal se considera mala de acuerdo al estándar internacional de LTE en condiciones exigentes; pero se debe considerar el tipo de servicio del dispositivo móvil al momento de hacer las pruebas piloto de radio frecuencia.

Finalmente, para desarrollar un trabajo en campo es recomendable realizar una combinación de las tres técnicas con una validación cruzada para integrar de forma practica el proceso de cobertura de una red LTE, es decir, debe haber una complementariedad con los algoritmo de aprendizaje para mejora la toma de decisiones en área de la ingeniería de radio; en su orden de jerarquía es aconsejable tener la data de entrenamiento, luego aplicar los métodos de estimación estadística, después probar los modelos de predicción y finalmente desarrollar y adaptar los algoritmos con Machine learning.

## Referencias bibliográficas

- [1] A. Kulkarni, A. Seetharam, A. Ramesh, and J. D. Herath, "DeepChannel: Wireless channel quality prediction using deep learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 443–456, 2020, doi: 10.1109/TVT.2019.2949954.

- [2] I.-R. Recommendation P-1411-9, Datos de propagación y métodos de predicción para la planificación de sistemas de radiocomunicaciones en interiores y redes radioeléctricas de área local en la gama de frecuencias de 900 MHz a 100 GHz Serie P, vol. 7. 2017.
- [3] I. M. M Mohamed, "Accurate Path-Loss Estimation for Wireless Cellular Networks," *J. Kejuruter.*, vol. 33, no. 2, pp. 317–328, 2021, doi: 10.17576/jkukm-2021-33(2)-16.
- [4] Q. He, A. Moayyedi, G. Dan, G. P. Koudouridis, and P. Tengkvist, "A Meta-Learning Scheme for Adaptive Short-Term Network Traffic Prediction," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 10, pp. 2271–2283, 2020, doi: 10.1109/JSAC.2020.3000408.
- [5] S. Bi, J. Lyu, Z. Ding, and R. Zhang, "Engineering radio maps for wireless resource management," *IEEE Wirel. Commun.*, vol. 26, no. 2, pp. 133–141, 2019, doi: 10.1109/MWC.2019.1800146.
- [6] T. S. Rappaport and S. Sandhu, "Radio wave propagation for emerging wireless personal communication systems," in *Wireless Personal Communications*, Springer, 2002, pp. 1–27.
- [7] Z. Chen and Y. Zhang, "Distributed spectrum monitoring system based on RSSI optimization algorithm," *Prog. Electromagn. Res. Symp.*, vol. 2017-Novem, pp. 2350–2355, 2017, doi: 10.1109/PIERS-FALL.2017.8293529.
- [8] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 4, pp. 3039–3071, 2019, doi: 10.1109/COMST.2019.2926625.
- [9] P. Campos, Á. Hernández-Solana, and A. Valdovinos-Bardají, "Machine Learning for Hidden Nodes Detection in Unlicensed LTE Networks," *Comput. Networks*, vol. 208, no. February, 2022, doi: 10.1016/j.comnet.2022.108862.
- [10] O. O. Erunkulu, A. M. Zungeru, C. K. Lebekwe, and J. M. Chuma, "Cellular Communications Coverage Prediction Techniques: A Survey and Comparison," *IEEE Access*, vol. 8, pp. 113052–113077, 2020, doi: 10.1109/ACCESS.2020.3003247.
- [11] M. D. Yacoub, *Foundations of Mobile Radio Engineering*. 2019.
- [12] Y. Zhang, J. Wen, G. Yang, Z. He, and J. Wang, "Path loss prediction based on machine learning: Principle, method, and data expansion," *Appl. Sci.*, vol. 9, no. 9, 2019, doi: 10.3390/app9091908.
- [13] H. S. Jo, C. Park, E. Lee, H. K. Choi, and J. Park, "Path loss prediction based on machine learning techniques: Principal component analysis, artificial neural network and gaussian process," *Sensors (Switzerland)*, vol. 20, no. 7, 2020, doi: 10.3390/s20071927.
- [14] C. E. G. Moreta, M. R. C. Acosta, and I. Koo, "Prediction of digital terrestrial television coverage using machine learning regression," *IEEE Trans. Broadcast.*, vol. 65, no. 4, pp. 702–712, 2019, doi: 10.1109/TBC.2019.2901409.
- [15] H. Elsherbiny, H. M. Abbas, H. Abou-Zeid, H. S. Hassanein, and A. Nouredin, "4G LTE Network Throughput Modelling and Prediction,"

2020 IEEE Glob. Commun. Conf. GLOBECOM 2020 - Proc., pp. 3–8, 2020, doi: 10.1109/GLOBECOM42002.2020.9322410.

[16]J. Isabona et al., “Development of a Multilayer Perception Neural Network for Optimal Predictive Modeling in Urban Microcellular Radio Environments,” *Appl. Sci.*, vol. 12, no. 11, 2022, doi: 10.3390/app12115713.

[17]C. Gijón, M. Toril, S. Luna-Ramírez, M. L. Marí-Altozano, and J. M. Ruiz-Avilés, “Long-term data traffic forecasting for network dimensioning in lte with short time series,” *Electron.*, vol. 10, no. 10, 2021, doi: 10.3390/electronics10101151.

[18]J. Riihijarvi and P. Mahonen, “Machine Learning for Performance Prediction in Mobile Cellular Networks,” *IEEE Comput. Intell. Mag.*, vol. 13, no. 1, pp. 51–60, 2018, doi: 10.1109/MCI.2017.2773824.

[19]N. Moraitis, L. Tsiipi, and D. Vouyioukas, “Machine learning-based methods for path loss prediction in urban environment for LTE networks,” *Int. Conf. Wirel. Mob. Comput. Netw. Commun.*, vol. 2020-  
Octob, 2020, doi: 10.1109/WiMob50308.2020.9253369.

[20]R. Bureau, “HANDBOOK Spectrum Monitoring,” *Int. Telecommun. Union*, p. 168, 2011.

[21]Francois Rancy, “Handbook on national spectrum management,” *Itu-R*, pp. 1–326, 2015.

[22]B. Charbuty and A. Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning,” *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.

[23]L. Alho, A. Burian, J. Helenius, and J. Pajarinen, “Machine learning based mobile network throughput classification,” *IEEE Wirel. Commun. Netw. Conf. WCNC*, vol. 2021-March, 2021, doi: 10.1109/WCNC49053.2021.9417365.

[24]L. Huang, L. Lu, and W. Hua, “A survey on next-cell prediction in cellular networks: Schemes and applications,” *IEEE Access*, vol. 8, pp. 201468–201485, 2020, doi: 10.1109/ACCESS.2020.3036070.

[25]Unión Internacional de Telecomunicaciones, “Recomendación UIT-R P.1411-7,” *Recom. UIT-R P.1411-7*, vol. 1, pp. 1–39, 2013, [Online]. Available: <https://www.itu.int/>.

[26]J. Herrera, “Metodología de medición del espectro con técnicas experimentales para obtener modelos de propagación,” vol. 1, p. 165, 2021, [Online]. Available: [https://www.unipamplona.edu.co/unipamplona/portallG/home\\_10/recursos/2015\\_s1/pag\\_contenido/02032015/libros\\_2015.jsp](https://www.unipamplona.edu.co/unipamplona/portallG/home_10/recursos/2015_s1/pag_contenido/02032015/libros_2015.jsp).

[27]A. Bhuvaneshwari, R. Hemalatha, and T. Satyasavithri, “Semi Deterministic Hybrid Model for Path Loss Prediction Improvement,” *Procedia Comput. Sci.*, vol. 92, pp. 336–344, 2016, doi: 10.1016/j.procs.2016.07.388.

[28]N. Heydarishahreza and N. Ansari, “Mobile Node Localization in Wireless Networks: Path-Loss Model, Trilateration, and Error Mitigation in a 5G Sub-6 GHz Scenario,” *J. Netw. Netw. Appl.*, vol. 3, no. 3, pp. 129–136, 2023, doi: 10.33969/j-nana.2023.030304.

- [29]M. Bouzidi, M. Mohamed, Y. Dalveren, A. Moldsvor, F. A. Cheikh, and M. Derawi, "Propagation Measurements for IQR Network in an Urban Environment," *Sensors*, vol. 22, no. 18, pp. 1–21, 2022, doi: 10.3390/s22187012.
- [30]S. Kumar, P., Patil, B., & Ram, "Selection of Radio Propagation Model for Long Term Evolution ( LTE ) Network," *Int. J. Eng. Res. Gen. Sci.*, vol. 3, no. 1, pp. 373–379, 2015, Accessed: Jul. 15, 2018. [Online]. Available: [www.ijergs.org](http://www.ijergs.org).
- [31]G. L. Stüber, *Principles of mobile communication: Fourth edition*. 2017.
- [32]A. L. Imoize and T. E. Ogunfuwa, "Propagation Measurements of a 4G LTE Network in Lagoon Environment," *Niger. J. Technol. Dev.*, vol. 16, no. 1, pp. 1–9, 2019, doi: 10.4314/njtd.v16i1.1.
- [33]S. Saunders and A. Aragon, *Antennas and propagation for wireless communications systems*. England, 2018.
- [34]B. Golman, "A set of Jupyter notebooks for the analysis of transport phenomena and reaction in porous catalyst pellet," *Comput. Appl. Eng. Educ.*, vol. 27, no. 3, pp. 531–542, 2019, doi: 10.1002/cae.22095.
- [35]I. Pintye, E. Kail, P. Kacsuk, and R. Lovas, "Big data and machine learning framework for clouds and its usage for text classification," *Concurr. Comput. Pract. Exp.*, vol. 33, no. 19, pp. 1–14, 2021, doi: 10.1002/cpe.6164.
- [36]Z. Wang, Z. Su, Y. Deng, J. Kurths, and J. Wu, "Spatial network disintegration based on kernel density estimation," *Reliab. Eng. Syst. Saf.*, vol. 245, no. August 2023, p. 110005, 2024, doi: 10.1016/j.ress.2024.110005.
- [37]G. K. Vishwakarma, C. Paul, and A. M. Elsayah, "An algorithm for outlier detection in a time series model using backpropagation neural network," *J. King Saud Univ. - Sci.*, vol. 32, no. 8, pp. 3328–3336, 2020, doi: 10.1016/j.jksus.2020.09.018.
- [38]Y. Pu, W. Yao, X. Li, and A. Alhudhaif, "An adaptive highly improving the accuracy of clustering algorithm based on kernel density estimation," *Inf. Sci. (Ny)*, vol. 663, no. August 2023, p. 120187, 2024, doi: 10.1016/j.ins.2024.120187.
- [39]O. Lifandali, N. Abghour, and Z. Chiba, "Feature Selection Using a Combination of Ant Colony Optimization and Random Forest Algorithms Applied to Isolation Forest Based Intrusion Detection System," *Procedia Comput. Sci.*, vol. 220, pp. 796–805, 2023, doi: 10.1016/j.procs.2023.03.106.
- [40]Ł. Gałka and P. Karczmarek, "Deterministic attribute selection for isolation forest," *Pattern Recognit.*, vol. 151, no. March, 2024, doi: 10.1016/j.patcog.2024.110395.
- [41]A. Al-Thaedan et al., "Downlink throughput prediction using machine learning models on 4G-LTE networks," *Int. J. Inf. Technol.*, vol. 15, no. 6, pp. 2987–2993, 2023, doi: 10.1007/s41870-023-01358-9.
- [42]M. Mayrhofer and P. Filzmoser, "Multivariate outlier explanations using Shapley values and Mahalanobis distances," *Econom. Stat.*, no. xxx, 2023, doi: 10.1016/j.ecosta.2023.04.003.

[43] A. Nowak-Brzezinska and C. Horyn, "Outliers in rules - The comparison of LOF, COF and KMEANS algorithms.," *Procedia Comput. Sci.*, vol. 176, pp. 1420–1429, 2020, doi: 10.1016/j.procs.2020.09.152.

[44] D. Zou et al., "Outlier detection and data filling based on KNN and LOF for power transformer operation data classification," *Energy Reports*, vol. 9, pp. 698–711, 2023, doi: 10.1016/j.egy.2023.04.094.