



UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA

FACSISTEL

INGENIERÍA EN TELECOMUNICACIONES

TRABAJO DE INTEGRACIÓN CURRICULAR

**Predicción de cobertura en redes móviles LTE
utilizando técnicas de aprendizaje automático
en Python**

Efraín Teodoro Barzola Guaranda

Dirigido por:
Ing. Daniel Jaramillo, M. Sc.

La Libertad – 2026



UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES

TRIBUNAL DE SUSTENTACIÓN

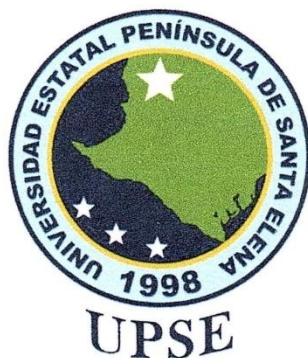
Ing. Ronald Rovira Jurado, Ph. D.
DIRECTOR DE LA CARRERA

Ing. Daniel Jaramillo, Mgtr.
TUTOR

Ing. Fernando Chamba, Mgtr.
DOCENTE ESPECIALISTA

Ing. Luis Amaya Farino, Mgtr.
DOCENTE GUÍA IUC

Ing. Corina Gonzabay De La A, Mgtr.
SECRETARIA



UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA FACULTAD DE SISTEMAS Y TELECOMUNICACIONES

CERTIFICACIÓN

Certifico que luego de haber dirigido científica y técnicamente el desarrollo y estructura final del trabajo, este cumple y se ajusta a los estándares académicos, razón por el cual apruebo en todas sus partes el presente trabajo de titulación que fue realizado en su totalidad por **Barzola Guaranda Efraín Teodoro**, como requerimiento para la obtención del título de Ingeniero en telecomunicaciones.

La Libertad, a los 04 días del mes de agosto del año 2026

TUTOR

Ing. Daniel Armando Jaramillo Chamba, Mgtr.



**UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

DECLARACIÓN DE RESPONSABILIDAD

Yo, **Barzola Guaranda Efraín Teodoro**

DECLARO QUE:

El trabajo de Titulación, **Predicción de cobertura en redes móviles LTE utilizando técnicas de aprendizaje automático en Python** de un previo a la obtención del título en Ingeniero en Telecomunicaciones, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría. En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

La Libertad, a los 04 días del mes de agosto del año 2026

EL AUTOR

Barzola Guaranda Efraín Teodoro



UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA FACULTAD DE SISTEMAS Y TELECOMUNICACIONES

CERTIFICACIÓN DE ANTIPLAGIO

Certifico que después de revisar el documento final del trabajo de titulación denominado **Predicción de cobertura en redes móviles LTE utilizando técnicas de aprendizaje automático en Python**, presentado por el estudiante, **Barzola Guaranda Efraín Teodoro** fue enviado al sistema antiplagio, presentando un porcentaje de similitud correspondiente al 06%, por lo que se aprueba el trabajo para que continúe con el proceso de titulación.



Certificado de análisis
Compilatio Magister+ | UPSE-ECU

Trabajo_de_Integracion_Curricular_Efraín_Barzola

ID: 66981b93b649afdee544ed2341d01a0601d76c8c



6%

Textos sospechosos

Nombre del fichero: Trabajo_de_Integracion_Curricular_Efraín_Barzola.txt
Tamaño del archivo original: 3,85 MB
Número de palabras: 28.999

Depositante: AMAYA FARIÑO LUIS MIGUEL
Fecha de depósito: 24 de agosto de 2020
Tipo de carga: interface
Fecha de fin de análisis: 24 de agosto de 2020

TUTOR

Ing. Daniel Armando Jaramillo Chamba, Mgtr.



UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA FACULTAD DE SISTEMAS Y TELECOMUNICACIONES

AUTORIZACIÓN

Yo, Barzola Guaranda Efraín Teodoro

Autorizo a la Universidad Estatal Península de Santa Elena, para que haga de este trabajo de titulación o parte de él, un documento disponible para su lectura consulta y procesos de investigación, según las normas de la Institución.

Cedo los derechos en línea patrimoniales de artículo profesional de alto nivel con fines de difusión pública, además apruebo la reproducción de este artículo académico dentro de las regulaciones de la Universidad, siempre y cuando esta reproducción no suponga una ganancia económica y se realice respetando mis derechos de autor

Santa Elena, a los 04 días del mes de agosto del año 2026

EL AUTOR

Barzola Guaranda Efraín Teodoro

AGRADECIMIENTO

Agradezco en primer lugar a mi familia, por su paciencia, su motivación constante y por sostenerme incluso en los momentos más difíciles. Sin su ejemplo y respaldo, este logro no habría sido posible.

Al Ing. Daniel Jaramillo, mi tutor, y al Ing. Fernando Chamba, mi especialista, por su guía, disposición y valiosos conocimientos brindados durante el desarrollo de este proyecto, así como por sus observaciones, recomendaciones y aportes que contribuyeron a mejorar y enriquecer el presente trabajo.

A mis compañeros de la carrera de Ingeniería en Telecomunicaciones, por las ideas, el trabajo en equipo y las anécdotas que hicieron más placentero este recorrido universitario.

A todos, mis más sinceros agradecimientos.

Efraín Teodoro Barzola Guaranda

DEDICATORIA

Dedico este trabajo, en primer lugar, a Dios, por brindarme la fortaleza y sabiduría para culminar esta etapa de mi vida.

A mis padres, Efraín Teódulo Barzola Merejildo y María Celinda Guaranda Perero, pilares fundamentales de mi vida, por su amor, sacrificio y apoyo incondicional; este logro también les pertenece.

A mis hermanos, por su cariño y compañía en cada momento de este camino.

A mis abuelos, cuya memoria permanece siempre en mi corazón y cuyas enseñanzas forman parte de quien soy hoy.

A los amigos y personas que conocí en este recorrido, por cada conversación y enseñanza compartida.

Con todo mi cariño y gratitud.

Efraín Teodoro Barzola Guaranda

RESUMEN

El presente trabajo tiene como finalidad predecir la cobertura en redes móviles LTE de CNT en el cantón La Libertad mediante técnicas de aprendizaje automático. El objetivo fue desarrollar, en Python, un modelo capaz de estimar parámetros relacionados con la calidad y el rendimiento de la señal LTE, como el RSRP, el RSRQ, el SINR, el CQI, el DL Throughput y el UL Throughput, a partir de mediciones reales de campo.

La metodología consistió en recolectar señales con un analizador de espectro y la aplicación G-NetTrack Pro, depurar y estructurar los datos, y entrenar modelos de Random Forest y XGBoost, complementados con un sistema interactivo de mapas de calor y perfiles topográficos. El conjunto final quedó conformado por 4 962 registros: 2 822 mediciones reales de campo y 2 140 escenarios aumentados, generados mediante un balance de enlace calibrado con las propias mediciones para incorporar variaciones de tilt, altura y potencia de la estación base.

Los resultados mostraron que Random Forest ofreció el mejor desempeño, con un error absoluto medio de 1,82 dB y un coeficiente de determinación de 0,964 en la predicción del RSRP, superando a XGBoost. Se concluye que el modelo constituye una herramienta confiable y replicable para apoyar la planificación y optimización de la red LTE en el cantón La Libertad.

Palabras clave: cobertura LTE, aprendizaje automático, Random Forest.

ABSTRACT

The purpose of this work is to predict the coverage in CNT LTE mobile networks in La Libertad canton using Machine Learning techniques. The objective was to develop, in Python, a model capable of estimating parameters related to LTE signal quality and performance, such as RSRP, RSRQ, SINR, CQI, DL Throughput, and UL Throughput, based on real field measurements.

The methodology consisted of collecting signals with a spectrum analyzer and the G-NetTrack Pro application, cleaning and structuring the data, and training Random Forest and XGBoost models, complemented with an interactive system of heat maps and topographic profiles. The final dataset comprised 4,962 records: 2,822 real field measurements and 2,140 augmented scenarios, generated through a link budget calibrated with the measurements themselves in order to account for variations in base station tilt, height and transmit power.

The results showed that Random Forest offered the best performance, with a mean absolute error (MAE) of 1.82 dB and a coefficient of determination of 0.964 in the RSRP prediction, surpassing XGBoost. It is concluded that the model constitutes a reliable and replicable tool to support the planning and optimization of the LTE network in La Libertad canton.

Keywords: LTE coverage, Machine Learning, Random Forest.

ÍNDICE GENERAL

TRIBUNAL DE SUSTENTACIÓN CERTIFICACIÓN	2
DECLARACIÓN DE RESPONSABILIDAD	4
CERTIFICACIÓN DE ANTIPLAGIO	5
AUTORIZACIÓN	6
AGRADECIMIENTO	7
DEDICATORIA	8
RESUMEN	9
ABSTRACT	10
ÍNDICE GENERAL	11
ÍNDICE DE TABLAS	16
ÍNDICE DE FIGURAS	18
ÍNDICE DE ECUACIONES	21
ABREVIATURAS	22
INTRODUCCIÓN	24
CAPÍTULO I: EL PROBLEMA DE LA INVESTIGACIÓN	26
1.1. Antecedentes	26
1.2. Descripción del proyecto	28
1.3. Objetivos	30
1.3.1. Objetivo general	30
1.3.2. Objetivos específicos	30
1.4. Justificación	30
1.5. Alcance del proyecto	32
1.6. Metodología	33
1.6.1. Metodología cuantitativa	33
1.6.2. Metodología descriptiva	34

1.6.3.	Metodología aplicada	34
1.6.4.	Metodología experimental	34
CAPÍTULO II: MARCO TEÓRICO		36
2.1.	Fundamentos de la tecnología LTE	36
2.1.1.	Arquitectura de la red LTE.....	36
2.1.2.	Principios de la transmisión LTE	37
2.1.3.	Parámetros clave de la señal LTE.....	38
2.1.4.	Cobertura y factores que la afectan en LTE.....	39
2.2.	Aplicación del aprendizaje automático	40
2.3.	Predicción del rendimiento de las redes LTE mediante el uso de aprendizaje automático	41
2.4.	Proceso de modelado predictivo	42
2.5.	Python	43
2.6.	Instrumentación para la recolección de datos LTE	44
2.6.1.	Narda SRM-3006	44
2.6.2.	G-NetTrack Pro	44
2.7.	Técnicas de aprendizaje automático para la predicción de niveles de cobertura y propagación en redes LTE.....	45
2.7.1.	Regresión lineal (Linear Regression).....	45
2.7.2.	Máquinas de vectores de soporte (Support Vector Machines).....	46
2.7.3.	Random Forest y XGBoost	46
2.7.4.	Redes neuronales artificiales (Artificial Neural Networks, ANN)	48
2.7.5.	Procesos gaussianos (Gaussian Process Regression, GPR).....	48
2.8.	Análisis comparativo de métodos de predicción utilizados en redes LTE.....	49
2.8.1.	Random Forest.....	49
2.8.2.	XGBoost.....	49

2.8.3.	Redes neuronales (ANN, RNN y CNN).....	50
2.8.4.	Interpolación espacial mediante Kriging	50
2.9.	Ubicación geográfica y herramientas para la representación espacial de datos.....	51
2.10.	Modelos de propagación empíricos	52
2.10.1.	Okumura-Hata	53
2.10.2.	Modelo COST-231 Hata.....	53
2.10.3.	Modelo de sombreado log-normal (Log-Normal Shadowing Model).54	
2.11.	Limitaciones de los modelos empíricos en zonas costeras	55
2.11.1.	Influencia de la humedad relativa y la salinidad	55
2.11.2.	Impacto de multitrayecto y superficie oceánica.....	55
2.11.3.	Restricciones de los modelos tradicionales y razonamiento del aprendizaje automático	56
2.12.	Comparativa de software para predicción de cobertura LTE	56
2.12.1.	ICS Telecom	56
2.12.2.	Atoll.....	57
2.12.3.	QualiPoc Android	57
2.12.4.	ZedBoard + Python	57
2.13.	Comparación de lenguajes y entornos de programación para el aprendizaje automático	59
2.13.1.	Python	59
2.13.2.	MATLAB.....	59
2.13.3.	C++.....	59
2.13.4.	Java	60
2.14.	Métricas de evaluación para modelos de predicción	61
2.14.1.	Error absoluto medio (MAE)	61
2.14.2.	Raíz del error cuadrático medio (RMSE)	62

2.14.3. Coeficiente de determinación (R^2).....	62
CAPÍTULO III: DESARROLLO DE LA PROPUESTA TÉCNICA	64
3.1. Diseño de la solución	64
3.2. Arquitectura del diseño	64
3.3. Infraestructura del hardware	66
3.3.1. Configuración de la estación base	66
3.3.2. Parámetros de la estación LTE analizada.....	67
3.4. Área de estudio	68
3.5. Recolección de las muestras y datos	71
3.5.1. Preparación de los datos	76
3.6. Definición de variables del sistema.....	78
3.6.1. Variables predictoras (entradas del modelo).....	78
3.6.2. Variables objetivo y variables de contexto	79
3.6.3. Entornos de desarrollo y librerías.....	80
3.7. Descripción general del flujo metodológico	83
3.7.1. Metodología de implementación de los modelos XGBoost y Random Forest para predicción de cobertura LTE	83
3.7.2. Metodología de la implementación del sistema de predicción de cobertura	91
3.8. Metodología de la implementación del sistema de visualización	107
CAPÍTULO IV: RESULTADOS Y ANÁLISIS	117
4.1. Condiciones y parámetros del entorno de evaluación	117
4.2. Evaluación de los modelos predictivos	119
4.2.1. Comparación por variable objetivo	120
4.2.2. Comparación entre valores reales y predichos	121
4.2.3. Distribución de errores	122

4.2.4.	Importancia de las variables predictoras	123
4.2.5.	Análisis del error en función de la distancia	124
4.2.6.	Análisis de sobreajuste	126
4.3.	Validación del sistema con mediciones reales de campo	127
4.3.1.	Comparación entre valores medidos y estimados.....	129
4.4.	Análisis comparativo.....	133
4.4.1.	Análisis gráfico del error de predicción	133
4.4.2.	Comparación con investigaciones relacionadas.....	135
4.4.3.	Comparación con herramientas de planificación radioeléctrica	138
4.5.	Propuesta de mejora para la planificación de la red LTE de CNT.....	140
4.5.1.	Caso 1: sector 2	140
4.5.2.	Caso 2: sector 1	142
4.5.3.	Caso 3: sector 3	144
4.6.	Discusión de resultados.....	146
CONCLUSIONES		149
RECOMENDACIONES		150
REFERENCIAS		151
ANEXOS		158
Anexo A. Instalación del entorno de desarrollo.....		158
Instalación de Python		158
Instalación de Visual Studio Code		160
Anexo B. Instalación de librerías del proyecto		163
Anexo C. Repositorio de código fuente.....		165
Anexo D. Diagrama de flujo del sistema desarrollado		167

ÍNDICE DE TABLAS

Tabla 1. Recomendaciones de intensidad de señal móvil en redes LTE.	38
Tabla 2. Comparativa de métodos para predecir la cobertura LTE.	50
Tabla 3. Comparativa entre software para la predicción de la cobertura LTE.	58
Tabla 4. Comparación de lenguajes y herramientas de programación aplicables al desarrollo de sistemas de predicción de cobertura LTE.	60
Tabla 5. Parámetros técnicos de la estación base LTE utilizada en el estudio.	68
Tabla 6. Origen de los parámetros empleados en las mediciones.	72
Tabla 7. Muestra de los datos utilizados para el entrenamiento de los modelos predictivos.	74
Tabla 8. Muestra de las mediciones de campo obtenidas mediante G-NetTrack Pro. ..	75
Tabla 9. Variables predictoras (entradas) utilizadas en los modelos predictivos de cobertura LTE.	78
Tabla 10. Variables objetivo y variables de contexto e identificación registradas durante la campaña.	79
Tabla 11. Librerías utilizadas en el desarrollo del sistema de predicción de cobertura LTE.	81
Tabla 12. Submódulos utilizados en el desarrollo del sistema de predicción de cobertura LTE.	82
Tabla 13. Parámetros del entorno de evaluación.	118
Tabla 14. Resumen comparativo del desempeño global de los modelos predictivos. ..	119
Tabla 15. Comparación del MAE obtenido por los modelos Random Forest y XGBoost para las variables objetivo evaluadas.	120
Tabla 16. Error absoluto medio por rango de distancia.	125
Tabla 17. Ubicación de los puntos de validación.	130
Tabla 18. Comparación entre mediciones reales y valores estimados por el sistema.	130
Tabla 19. Validación de los parámetros de calidad de señal (RSRQ, SINR y CQI). ..	131
Tabla 20. Validación de los parámetros de rendimiento (DL y UL Throughput).	132
Tabla 21. Comparación entre investigaciones relacionadas y la propuesta desarrollada.	136
Tabla 22. Comparación entre herramientas de planificación radioeléctrica y el sistema desarrollado.	138

Tabla 23. Comparación entre el escenario original y el escenario optimizado.	141
Tabla 24. Parámetros de configuración evaluados para ambos escenarios.	143
Tabla 25. Comparación entre escenarios de configuración del sector 3.....	145
Tabla 26. Librerías de Python utilizadas en el proyecto.....	163
Tabla 27. Estructura de archivos del proyecto.....	166

ÍNDICE DE FIGURAS

Figura 1. Diagrama de representación del desarrollo del proyecto.	35
Figura 2. Componentes principales y flujo de datos.	37
Figura 3. Desarrollo del aprendizaje automático. Parte 1.	41
Figura 4. Desarrollo del aprendizaje automático. Parte 2.	41
Figura 5. Proceso CRISP-DM para el modelado predictivo de redes LTE.....	43
Figura 6. Analizador selectivo de radiación electromagnética Narda SRM-3006.	44
Figura 7. Representación gráfica del funcionamiento de la regresión lineal.	45
Figura 8. Representación gráfica del funcionamiento de las máquinas de vectores de soporte.	46
Figura 9. Representación gráfica del funcionamiento del algoritmo Random Forest...	47
Figura 10. Representación gráfica del funcionamiento del algoritmo XGBoost.	47
Figura 11. Representación gráfica del funcionamiento de una red neuronal artificial..	48
Figura 12. Representación gráfica del funcionamiento de un proceso gaussiano.	49
Figura 13. Representación de información geoespacial en ArcGIS.....	52
Figura 14. Mapa de calor de densidad urbana elaborado en QGIS.....	52
Figura 15. Arquitectura por capas del sistema de predicción de cobertura LTE.	65
Figura 16. Ubicación geográfica de la estación base de CNT en el cantón La Libertad.	69
Figura 17. Vista de la estación base de CNT y su torre de telecomunicaciones desde la vía pública.	70
Figura 18. Vista frontal del edificio que aloja la estación base de CNT en el área de estudio.....	70
Figura 19. Detalle de la torre y las antenas sectoriales de la estación base LTE analizada.	71
Figura 20. Flujograma del proceso de entrenamiento del modelo XGBoost para la predicción de cobertura LTE.	84
Figura 21. Flujograma del proceso de entrenamiento del modelo Random Forest para la predicción de cobertura LTE.	85
Figura 22. Proceso de inicialización del sistema.....	92
Figura 23. Módulo de configuración del sistema.	95
Figura 24. Módulo de procesamiento geoespacial y radioeléctrico.	97

Figura 25. Construcción del vector de características.	102
Figura 26. Motor predictivo del sistema.	103
Figura 27. Proceso de selección del sector dominante y estimación de métricas LTE.	104
Figura 28. Panel de controles interactivos del sistema de predicción.	109
Figura 29. Módulo de simulación de mapa de calor LTE.	109
Figura 30. Mapa de calor interactivo de cobertura LTE.	112
Figura 31. Panel de visualización del sistema.	112
Figura 32. Panel informativo dinámico del sector.	113
Figura 33. Monitor de rendimiento y velocidad del enlace.....	113
Figura 34. Presupuesto de enlace del sector.....	114
Figura 35. Flujo de trabajo del comparador de escenarios de planificación.	115
Figura 36. Interfaz del comparador de escenarios de planificación.	115
Figura 37. Comparación del MAE obtenido por los modelos Random Forest y XGBoost para las variables objetivo evaluadas.....	121
Figura 38. Comparación entre los valores reales y predichos de RSRP para los modelos Random Forest y XGBoost.	122
Figura 39. Distribución de errores residuales para la predicción del RSRP.	122
Figura 40. Importancia relativa de las variables predictoras para Random Forest y XGBoost.....	124
Figura 41. Variación del MAE según la distancia entre la estación base y el usuario.	126
Figura 42. Comparación del coeficiente de determinación entre los conjuntos de entrenamiento y prueba.	126
Figura 43. Medición de campo con el equipo Narda SRM-3006 y sonda isotrópica en el área de estudio.	128
Figura 44. Caracterización espectral de la banda LTE 2120 MHz obtenida con el Narda SRM-3006.	128
Figura 45. Proceso de validación del sistema mediante mediciones reales de campo.	129
Figura 46. Comparación entre valores reales y estimados de RSRP.....	134
Figura 47. Error absoluto por punto de validación.....	134
Figura 48. Diagrama de dispersión entre los valores medidos y estimados de RSRP.	135

Figura 49. Comparación del mapa de cobertura entre el escenario original y el escenario optimizado.	141
Figura 50. Comparación de la cobertura LTE antes y después del incremento de la potencia de transmisión.	142
Figura 51. Comparación de escenarios de cobertura LTE antes y después de la optimización.	144
Figura 52. Descarga oficial de Python desde la web.....	158
Figura 53. Instalador de Python en Windows.	159
Figura 54. Instalación de Python completada en Windows.	159
Figura 55. Descarga de Visual Studio Code.....	160
Figura 56. Aceptación de términos de licencia en VS Code.	160
Figura 57. Selección de carpeta de destino en VS Code.	161
Figura 58. Selección de carpeta en el Menú Inicio de VS Code.	161
Figura 59. Selección de tareas adicionales en VS Code.....	162
Figura 60. Confirmación final antes de instalar VS Code.....	162
Figura 61. Instalación de Visual Studio Code completada.....	163
Figura 62. Diagrama de flujo del módulo de visualización, simulación de cobertura y comparador de escenarios del sistema desarrollado.	167

ÍNDICE DE ECUACIONES

Ecuación 1. Modelo Okumura-Hata.	53
Ecuación 2. Modelo COST-231 Hata.	54
Ecuación 3. Modelo de sombreado log-normal.	54
Ecuación 4. Error absoluto medio.	61
Ecuación 5. Raíz del error cuadrático medio.	62
Ecuación 6. Coeficiente de determinación.	63
Ecuación 7. Representación de la variable predictora en forma matemática.	87
Ecuación 8. Representación de la división del dataset.	87
Ecuación 9. Predicción del modelo Random Forest.	88
Ecuación 10. Predicción del modelo XGBoost.	89
Ecuación 11. Representación matemática de la evaluación de línea de vista.	97
Ecuación 12. Cálculo de la longitud de onda.	98
Ecuación 13. Radio de la primera zona de Fresnel.	98
Ecuación 14. Factor de corrección.	99
Ecuación 15. Pérdida horizontal de la antena.	100
Ecuación 16. Estimación del DL Throughput.	100
Ecuación 17. Estimación del UL Throughput.	100
Ecuación 18. Estimación del RSRP.	105
Ecuación 19. Estimación del RSRQ.	106
Ecuación 20. Estimación del RSSI.	106
Ecuación 21. Estimación del SINR.	106
Ecuación 22. Estimación del CQI.	107
Ecuación 23. Número de puntos de la malla radial.	110
Ecuación 24. Intensidad normalizada utilizada en el mapa de calor.	110
Ecuación 25. Estimación de la latitud.	111
Ecuación 26. Estimación de la longitud.	111
Ecuación 27. Importancia de la variable x_i	123
Ecuación 28. Brecha de sobreajuste.	126

ABREVIATURAS

ANN: Red Neuronal Artificial.

AWS: Servicios Inalámbricos Avanzados, banda 1700/2100 MHz.

CNN: Red Neuronal Convolucional.

CNT: Corporación Nacional de Telecomunicaciones.

COST-231: Modelo de Propagación del European Cooperation in Science and Technology.

CQI: Indicador de Calidad del Canal.

CRISP-DM: Proceso Estándar Interindustrial para la Minería de Datos.

CSV: Valores Separados por Comas.

dB: Decibelio.

dBm: Decibelio Referido a un Milivatio.

DEM: Modelo Digital de Elevación.

DL: Enlace Descendente.

EARFCN: Número Absoluto de Canal de Radiofrecuencia E-UTRA.

eNodeB: Estación Base Evolucionada de la Red LTE.

EPC: Núcleo de Paquetes Evolucionado.

E-UTRAN: Red de Acceso de Radio Terrestre Universal Evolucionada.

FDD: Duplexación por División de Frecuencia.

GPR: Regresión por Procesos Gaussianos.

GPS: Sistema de Posicionamiento Global.

GSM: Sistema Global para Comunicaciones Móviles.

HSS: Servidor de Suscriptores Locales.

IA: Inteligencia Artificial.

IP: Protocolo de Internet.

ITU-R: Unión Internacional de Telecomunicaciones – Sector de Radiocomunicaciones.

LTE: Evolución a Largo Plazo.

MAE: Error Absoluto Medio.

ML: Aprendizaje Automático.

MME: Entidad de Gestión de Movilidad.

MSE: Error Cuadrático Medio.

OFDM: Multiplexación por División de Frecuencia Ortogonal.

OFDMA: Acceso Múltiple por División de Frecuencias Ortogonales.

OSM: Proyecto Colaborativo de Cartografía Libre.

PCI: Identidad Física de Celda.

PGW: Pasarela de Red de Datos por Paquetes.

QGIS: Sistema de Información Geográfica Quantum.

R²: Coeficiente de Determinación.

RF: Random Forest (Bosques Aleatorios).

RMSE: Raíz del Error Cuadrático Medio.

RNN: Red Neuronal Recurrente.

RSRP: Potencia Recibida de la Señal de Referencia.

RSRQ: Calidad Recibida de la Señal de Referencia.

RSSI: Indicador de Intensidad de Señal Recibida.

SC-FDMA: Acceso Múltiple por División de Frecuencia de Portadora Única.

SDR: Radio Definida por Software.

SGW: Pasarela de Servicio.

SINR: Relación Señal a Interferencia más Ruido.

SVM: Máquina de Vectores de Soporte.

TDD: Duplexación por División de Tiempo.

TIFF: Formato de Archivo de Imagen Etiquetada.

UE: Equipo de Usuario.

UL: Enlace Ascendente.

UMTS: Sistema Universal de Telecomunicaciones Móviles.

XGBoost: Potenciación Extrema de Gradiente.

INTRODUCCIÓN

En Ecuador, el acceso a Internet ha experimentado un incremento continuo, hasta convertirse en un elemento esencial en las actividades cotidianas de trabajo y de estudio para gran parte de la población. Según cifras publicadas por DataReportal en su página oficial, en 2024 un 83,6 % de los ecuatorianos contaba con acceso a internet, un incremento del 3,9 % con respecto al año anterior [1]. El crecimiento se vio motivado, en gran medida, por la pandemia de COVID-19 entre 2020 y 2021, cuando numerosos hogares contrataron los servicios fijos o móviles de internet para sostener las clases virtuales y el teletrabajo. Una vez finalizadas las restricciones derivadas de la emergencia sanitaria, la mayoría de los usuarios mantuvo este servicio, consolidando su importancia en las actividades diarias de la población ecuatoriana.

En comparación con las tecnologías 2G y 3G, el LTE ofrece ventajas significativas en velocidad y cobertura. Sin embargo, su desempeño puede verse afectado por las condiciones del entorno y las características de propagación presentes en cada zona. Entre los principales factores que influyen en su desempeño se encuentran la distancia entre el usuario y el eNodeB, la interferencia electromagnética, el número de usuarios conectados simultáneamente y la presencia de edificios o vegetación. Estos factores pueden influir en los indicadores empleados para evaluar el comportamiento de la red, entre ellos el RSRP, el RSRQ, el RSSI y el SINR.

Obtener una estimación de cobertura suficientemente precisa continúa siendo un desafío, especialmente en zonas urbanas con alta densidad poblacional y elevada demanda de servicios de datos. Para estimar las pérdidas de propagación se han desarrollado diversos modelos empíricos, como Okumura-Hata y COST-231 Hata, los cuales fueron formulados para determinados rangos de frecuencia y condiciones de propagación; por ello, su precisión puede disminuir cuando las características del entorno de estudio difieren de aquellas para las cuales fueron diseñados. Por otro lado, las campañas de medición en campo permiten obtener información real sobre el comportamiento de la red, aunque requieren recursos económicos, humanos y de tiempo, además de que sus resultados pueden perder vigencia cuando las condiciones de la red o del entorno cambian. En este contexto, el aprendizaje automático constituye una alternativa intermedia, debido a que permite aprovechar datos obtenidos mediante

mediciones de campo para entrenar modelos capaces de identificar relaciones entre las características del entorno y el comportamiento de la señal en una zona determinada.

El presente estudio se desarrolla sobre una estación base LTE de CNT ubicada en la zona urbana del cantón La Libertad, provincia de Santa Elena, donde la demanda de servicios de datos móviles se incrementa paralelamente al crecimiento y desarrollo urbano. Entre mayo y junio de 2026 se realizaron recorridos en diferentes rutas alrededor del sitio de estudio, durante los cuales se registraron 2 822 muestras mediante la aplicación G-NetTrack Pro. De manera complementaria, se verificaron las condiciones de la portadora en la banda analizada mediante el analizador Narda SRM-3006. Debido a que la configuración de la estación base no presentó modificaciones durante la campaña de medición, se amplió el conjunto de datos mediante la generación de escenarios aumentados con variaciones de inclinación eléctrica de la antena (tilt), la altura y la potencia, calibrados a partir de un balance de enlace. De esta manera, se obtuvo un conjunto de datos compuesto por 4 962 registros. A partir de estos datos se entrenaron y evaluaron dos algoritmos de aprendizaje automático: Random Forest, basado en un conjunto de árboles de decisión, y XGBoost, basado en un método de potenciación de gradiente. El procesamiento, entrenamiento y evaluación de los modelos se realizaron en Python mediante librerías especializadas como Pandas, NumPy y Scikit-learn, mientras que Matplotlib fue empleada para la representación gráfica de los resultados.

Los resultados obtenidos permiten identificar los sectores del cantón La Libertad donde los niveles de señal se encuentran por debajo del umbral establecido, proporcionando información útil para la toma de decisiones relacionadas con la instalación de nuevas estaciones base, así como con la reorientación y optimización de las estaciones base ya instaladas. De esta manera, se reduce la necesidad de realizar campañas extensivas de medición para evaluar, punto por punto, el comportamiento de los principales parámetros LTE en toda el área de estudio. Además, el trabajo aporta información sobre el comportamiento radioeléctrico de un entorno urbano costero ecuatoriano, contribuyendo al desarrollo de futuras investigaciones orientadas a la planificación, la evaluación y la optimización de redes móviles mediante técnicas de aprendizaje automático.

CAPÍTULO I: EL PROBLEMA DE LA INVESTIGACIÓN

1.1. Antecedentes

En Ecuador se desarrolló el estudio titulado “Predicción de rendimiento de redes LTE con técnicas de Machine Learning” [2], en el cual se construyó una base de datos de una red móvil mediante una aplicación para dispositivos Android capaz de registrar parámetros de desempeño, entre ellos el RSRP, el RSRQ, la velocidad de descarga y la latencia. Para el desarrollo de la investigación, se entrenaron diferentes modelos siguiendo la metodología CRISP-DM, utilizando una partición de los datos de 70 % para entrenamiento y 30 % para evaluación. Posteriormente, se comparó el desempeño de los modelos mediante métricas estadísticas y se seleccionaron aquellos que presentaron un mejor ajuste para cada escenario. Aunque el estudio no empleó equipos especializados para la adquisición de las mediciones, constituye un referente para la aplicación de técnicas de aprendizaje automático sobre datos relacionados con el desempeño de redes móviles en Ecuador.

En la ciudad de Santa Elena se desarrolló un estudio sobre las interferencias electromagnéticas presentes en la banda no licenciada de 2,4 GHz [3], en el cual se empleó el analizador de espectro Narda SRM-3006. Los resultados evidenciaron que las interferencias de pulso, de barrido y de banda estrecha pueden afectar la estabilidad de las transmisiones Wi-Fi y, cuando alcanzan niveles perjudiciales, pueden provocar una disminución apreciable en la calidad del servicio. Además de sus resultados sobre el comportamiento de las interferencias, este trabajo constituye un antecedente relevante debido al empleo del analizador Narda SRM-3006 para la caracterización de señales electromagnéticas en condiciones reales. La utilización de este equipo también guarda relación con la presente investigación, en la cual este equipo se emplea como instrumento de apoyo para la verificación de las mediciones de parámetros asociados a la cobertura LTE.

En Santiago de Chile se desarrolló el estudio titulado “Aplicación de técnicas de aprendizaje automático y analítica predictiva para mejorar desempeño de redes 4G LTE” [4], orientado al análisis del desempeño de celdas 4G LTE y a la identificación de patrones de uso de las antenas. Para ello, se utilizaron datos del funcionamiento operativo de redes 4G LTE y se aplicaron modelos como XGBoost, Random Forest y LSTM para clasificar, predecir y pronosticar el comportamiento de las celdas. Asimismo, empleó el

algoritmo LDA para identificar patrones de uso de las antenas en diferentes zonas de Santiago. Los resultados destacaron el desempeño de XGBoost en las tareas de clasificación y permitieron identificar cinco grupos con patrones diferenciados. Este estudio guarda relación con el presente trabajo por el empleo de técnicas de aprendizaje automático en redes LTE. Sin embargo, se diferencia en el objeto de estudio, ya que se enfocó en el desempeño de las celdas y los patrones de uso de las antenas, mientras que la presente investigación se orienta a la predicción de cobertura LTE mediante mediciones obtenidas en campo y escenarios complementarios generados mediante simulación.

En Quito se implementó el estudio titulado "Modelos predictivos de zonas de handover en redes LTE con técnicas de Machine Learning" [5], centrado en el proceso de handover, analizado sobre dos escenarios. Para llevar a cabo el estudio, se emplearon variables como el RSRP, la velocidad del usuario y la distancia a la estación base, obtenidas mediante simulaciones de red y registros de mediciones. El objetivo principal de este estudio, “orientado a predecir zonas de handover en entornos simulados”, difiere del objetivo de la presente investigación, aunque ambos estudios recurren al aprendizaje automático para mejorar la eficiencia de redes LTE. La principal diferencia entre ambos trabajos radica en el propósito del análisis, las variables consideradas y el enfoque de predicción. El estudio referenciado se orienta a la identificación y predicción de zonas de handover y handover fallido mediante árboles de decisión, mientras que la presente investigación se centra en la predicción de parámetros asociados a la cobertura LTE mediante Random Forest y XGBoost, utilizando mediciones de campo complementadas con escenarios aumentados.

En Colombia se analizó la aplicación de modelos predictivos para la cobertura LTE mediante algoritmos de Machine Learning, considerando diversas dificultades asociadas con su implementación [6]. Entre los principales desafíos identificados se encuentran la recolección, limpieza y selección de los datos; la identificación de variables relevantes para el entrenamiento de los modelos; la prevención del sobreajuste; la integración de las soluciones predictivas con plataformas operativas existentes; y la disponibilidad de personal capacitado para interpretar los resultados. Estos aspectos guardan relación con la presente investigación, debido a que el desarrollo del modelo predictivo requiere abordar procesos similares. En particular, este trabajo considera la adquisición de datos mediante G-NetTrack Pro y la verificación de las condiciones de la portadora mediante

el Narda SRM-3006, así como la aplicación de técnicas de ingeniería de características orientadas a mejorar la calidad de las variables utilizadas durante el entrenamiento de los modelos de aprendizaje automático.

1.2. Descripción del proyecto

Una parte fundamental de la planificación y optimización de las redes LTE de operadoras como la Corporación Nacional de Telecomunicaciones (CNT) se apoya en modelos empíricos de propagación, entre ellos Okumura-Hata y COST-231 Hata. Sin embargo, estos modelos presentan limitaciones para estimar la cobertura con precisión en entornos urbanos específicos, ya que se basan en formulaciones empíricas desarrolladas bajo condiciones estandarizadas que no incorporan de forma directa las particularidades del entorno. Dichas limitaciones generan discrepancias entre los niveles de señal estimados y los observados en campo, lo que dificulta la identificación precisa de zonas con cobertura deficiente.

Esta problemática adquiere especial relevancia en zonas costeras como el cantón La Libertad, cuyo entorno presenta características complejas como la alta humedad, la salinidad, la brisa marina, una configuración urbana irregular y las variaciones en la densidad poblacional asociadas a la actividad turística. Estas condiciones influyen en los fenómenos de atenuación, reflexión, difracción y propagación multitrayecto, produciendo variaciones en la señal que los modelos empíricos no logran representar adecuadamente. Por ello, resulta pertinente complementar estos modelos mediante mediciones reales y técnicas de aprendizaje automático, capaces de identificar la relación entre las características del entorno y el comportamiento de la señal LTE.

En consecuencia, este proyecto desarrolló, mediante Python, una herramienta para la predicción de cobertura LTE basada en técnicas de aprendizaje automático. Para ello, se recopilaron mediciones en campo mediante la aplicación G-NetTrack Pro en distintos puntos geográficos del cantón La Libertad, complementadas con la caracterización espectral realizada mediante el analizador Narda SRM-3006 en un conjunto reducido de ubicaciones. Durante la campaña se registraron variables relacionadas con el comportamiento y la calidad de la red, entre ellas la potencia recibida de la señal de referencia (RSRP), la intensidad de la señal recibida (RSSI), la calidad de la señal de referencia recibida (RSRQ) y la relación señal-interferencia más ruido (SINR).

Antes del entrenamiento de los modelos, los datos se sometieron a un proceso de limpieza, depuración y selección de variables. En esta etapa se identificaron y trataron los valores faltantes, registros inválidos y posibles inconsistencias presentes en las mediciones. Posteriormente, se realizó la preparación de las variables de entrada de acuerdo con los requerimientos de los modelos utilizados. Estas etapas permitieron obtener un conjunto de datos estructurado y adecuado para el proceso de entrenamiento y evaluación.

Una vez depurados los datos, se entrenaron y evaluaron los modelos Random Forest y XGBoost. Random Forest es un algoritmo de aprendizaje supervisado que combina múltiples árboles de decisión y utiliza el conjunto de predicciones obtenidas para generar una estimación más robusta. Por su parte, XGBoost utiliza un enfoque de potenciación de gradiente en el que los árboles se construyen secuencialmente, incorporando en cada etapa la información asociada con los errores de los modelos anteriores. La utilización de ambos algoritmos permitió comparar su capacidad para predecir el comportamiento de las variables asociadas con la cobertura LTE.

Para el procesamiento de datos se emplearon librerías de Python como Pandas y NumPy, mientras que Scikit-learn proporcionó las herramientas de entrenamiento, validación y evaluación. La librería XGBoost se utilizó para la implementación del modelo de gradient boosting y Matplotlib para la representación gráfica. Asimismo, se emplearon herramientas de análisis geoespacial para mapear la distribución de la señal. Durante el proceso se aplicaron estrategias de validación cruzada y ajuste de hiperparámetros con el fin de optimizar el desempeño y mitigar el riesgo de sobreajuste.

La evaluación se planteó como un problema de regresión, dado que los modelos estiman valores continuos (como el RSRP) en lugar de clasificar datos en categorías discretas. El desempeño se midió a través del error absoluto medio (MAE) y el coeficiente de determinación (R^2), complementados con el análisis del sesgo promedio entre los valores predichos y los medidos en campo. El MAE cuantificó el error promedio en las unidades de la variable objetivo, mientras que el R^2 determinó la proporción de variabilidad explicada por el modelo. Adicionalmente, se comparó el rendimiento entre los conjuntos de entrenamiento y prueba para detectar posibles indicios de sobreajuste.

Finalmente, los resultados se representaron en mapas de cobertura y visualizaciones georreferenciadas que reflejan la distribución espacial de la señal en el cantón La

Libertad. Estas representaciones facilitaron la localización de sectores con niveles inferiores al umbral normativo y permitieron priorizar zonas de intervención. De este modo, la herramienta ofrece insumos valiosos para la planificación y optimización de redes LTE, reduciendo la necesidad de realizar campañas extensivas de medición en campo y sentando las bases para futuras investigaciones en redes móviles.

1.3. Objetivos

1.3.1. Objetivo general

Desarrollar un modelo de predicción de cobertura LTE para el cantón La Libertad, a partir de mediciones de señal de las antenas de CNT y mediante técnicas de aprendizaje automático, orientado a contribuir a la planificación y la optimización de las redes móviles de la zona.

1.3.2. Objetivos específicos

- **Realizar** un estudio sobre LTE y aprendizaje automático en Python, con el propósito de establecer la base teórica necesaria para el desarrollo de la investigación.
- **Recolectar** las señales de radio LTE de las antenas de CNT en el cantón La Libertad, utilizando medidores especializados y herramientas de análisis de datos, con el fin de construir una base de datos confiable para el entrenamiento del modelo.
- **Desarrollar** y entrenar un modelo de aprendizaje automático en Python para predecir la cobertura de señales LTE a partir de los datos recolectados.
- **Evaluar** los resultados de las predicciones mediante métricas de desempeño y análisis comparativos, con el propósito de generar información de apoyo para la planificación de las redes de radio de CNT en el cantón La Libertad.
- **Implementar** un sistema interactivo de visualización geoespacial para representar los resultados de cobertura LTE mediante mapas de calor, perfiles topográficos y elementos cartográficos que faciliten el análisis de la cobertura en el área de estudio.

1.4. Justificación

De acuerdo con el Instituto Nacional de Estadística y Censos (INEC), para julio de 2022 alrededor del 60,4 % de los ecuatorianos tenía acceso a internet. Un año después,

esta cifra había aumentado en 1,8 puntos porcentuales y, para julio de 2024, se registró un incremento adicional de 3,8 puntos porcentuales, hasta alcanzar aproximadamente el 66 % [7]. Este crecimiento se explica, en buena parte, por la pandemia de COVID-19, periodo en el que los hogares y las empresas se vieron obligados a contratar servicios de internet para desarrollar actividades educativas, laborales y trámites administrativos. Estas conexiones se mantuvieron después de la emergencia sanitaria, consolidando la importancia de los servicios de internet en las actividades cotidianas de la población. Estas cifras, obtenidas mediante encuestas de hogares a nivel nacional, difieren de las estimaciones internacionales citadas en la Introducción, que emplean una metodología distinta basada en el número de usuarios activos de internet.

En este escenario, la calidad del servicio depende del rendimiento y la cobertura de las redes LTE. Sin embargo, su planificación presenta limitaciones asociadas a la precisión y los elevados costos de recolección de mediciones en campo. Estas dificultades se evidenciaron durante la ejecución del proyecto, donde la caracterización de un único eNodeB en el cantón La Libertad requirió diversas jornadas de medición entre mayo y junio de 2026. En este contexto, el aprendizaje automático representó una alternativa eficaz para complementar los métodos convencionales, permitiendo estimar el comportamiento de la señal en puntos no medidos directamente y evaluar escenarios con distintas configuraciones en la estación base.

Este trabajo tuvo como objetivo principal desarrollar, mediante Python, un modelo predictivo capaz de estimar el nivel de la señal LTE representado por el RSRP en la zona urbana del cantón La Libertad, provincia de Santa Elena. Para ello, se implementaron algoritmos de aprendizaje automático como Random Forest y XGBoost, empleando librerías como Pandas, NumPy y Matplotlib para el procesamiento, análisis y visualización de datos. La evaluación de los modelos se realizó mediante la comparación entre las predicciones y las mediciones de campo, utilizando métricas como el error absoluto medio (MAE) y el coeficiente de determinación (R^2).

El aprendizaje automático ha demostrado resultados favorables en aplicaciones orientadas a la estimación de cobertura y al análisis de la calidad de señal en redes LTE. En la ciudad de Quito, un estudio basado en árboles de decisión alcanzó un 67 % de precisión al evaluar la intensidad de la señal en condiciones estáticas [8]. Asimismo, otra investigación desarrollada en la misma ciudad logró predecir fallos de handover con un

95,5 % de efectividad mediante técnicas de aprendizaje supervisado [5]. Estos antecedentes respaldan la aplicación de técnicas de aprendizaje automático en el análisis de redes móviles. En este sentido, la aplicación de estas técnicas para la predicción de cobertura LTE en el cantón La Libertad, permitió analizar el comportamiento de la cobertura en un entorno urbano costero y generar información que contribuya a la toma de decisiones relacionadas con la planificación y optimización de la infraestructura de la red.

1.5. Alcance del proyecto

El presente estudio se concentra en desarrollar y validar modelos predictivos basados en algoritmos de aprendizaje automático para estimar variables relacionadas con el comportamiento de la señal LTE en diferentes ubicaciones geográficas. La variable principal de predicción es la potencia de la señal de referencia recibida (RSRP), ya que constituye el principal indicador para determinar el nivel de cobertura. De manera complementaria, se entrenaron modelos independientes para estimar la calidad de la señal de referencia recibida (RSRQ), la relación señal-interferencia más ruido (SINR), el indicador de calidad de canal (CQI) y las velocidades de transmisión de descarga y de subida (DL y UL Throughput), utilizados para caracterizar la calidad y el rendimiento del enlace.

El estudio se delimitó a una estación base LTE de CNT, con sus tres sectores, ubicada en la zona urbana del cantón La Libertad, provincia de Santa Elena. Las mediciones se realizaron entre mayo y junio de 2026 dentro del radio de servicio de dicha estación. El análisis de estaciones base vecinas y de los procesos de traspaso entre celdas quedó fuera del alcance de este trabajo, aunque las muestras relacionadas con estos elementos se conservaron para estudios posteriores.

El desarrollo del estudio se realizó en Python, con el apoyo de librerías como Scikit-learn, Pandas, NumPy y Matplotlib para el procesamiento, análisis y visualización de los datos, así como para el entrenamiento, validación y evaluación de los modelos. Se emplearon los algoritmos de aprendizaje automático Random Forest y XGBoost, utilizando conjuntos de datos que integran información técnica, geoespacial y ambiental relacionada con el entorno de estudio, complementados con escenarios generados mediante un balance de enlace calibrado con las propias mediciones.

El despliegue físico de infraestructura de telecomunicaciones queda excluido de este trabajo, debido a que la investigación se concentra en el procesamiento de los datos recolectados, el desarrollo de los modelos predictivos y la generación de escenarios complementarios para analizar el comportamiento de la señal LTE bajo diferentes condiciones de los parámetros de la estación base. Del mismo modo, las configuraciones evaluadas mediante simulación no se implementaron sobre la red en operación.

Como parte del desarrollo, se implementó un sistema interactivo de visualización geoespacial que integra mapas de calor, perfiles topográficos y paneles de indicadores radioeléctricos, orientado a representar espacialmente el comportamiento de la señal LTE en el área de estudio. Estas herramientas de visualización podrán utilizarse como apoyo para el análisis y la toma de decisiones relacionadas con la planificación y optimización de redes móviles. El alcance del proyecto se limita a redes LTE (4G); sin embargo, la metodología desarrollada puede constituir una base para futuras investigaciones orientadas a otras tecnologías de comunicaciones móviles.

1.6. Metodología

Para cumplir con los objetivos propuestos, este proyecto adopta un enfoque cuantitativo, con un alcance descriptivo y aplicado, y un componente experimental, mediante el empleo de técnicas de aprendizaje automático implementadas en Python. El estudio se sustenta en mediciones reales recopiladas mediante la aplicación G-NetTrack Pro y en la caracterización espectral de la banda analizada realizada con el analizador Narda SRM-3006, las cuales se complementan con escenarios generados mediante simulación para ampliar el volumen de datos en el entrenamiento y evaluación de los modelos.

1.6.1. Metodología cuantitativa

La investigación se fundamenta en la recopilación y el análisis de datos cuantitativos relacionados con el comportamiento y desempeño de la señal LTE. El conjunto de datos integra parámetros técnicos, geoespaciales y ambientales, entre los que se incluyen el RSRP, el RSRQ, el RSSI, el SINR, el CQI, el DL Throughput y el UL Throughput. De estas variables, los modelos de aprendizaje automático se entrenan para predecir seis variables objetivo: RSRP, RSRQ, SINR, CQI, DL Throughput y UL Throughput. El RSSI, aunque está presente en el conjunto de datos, no se considera una variable objetivo

de los modelos y, por lo tanto, no forma parte de la evaluación comparativa de su desempeño.

1.6.2. Metodología descriptiva

El estudio caracteriza las condiciones de cobertura LTE en diferentes sectores del cantón La Libertad mediante la identificación de patrones en el comportamiento de la señal, representados en mapas de cobertura y visualizaciones georreferenciadas. Este diagnóstico describe el estado actual de la zona analizada y constituye la base estructural para el desarrollo y validación de los modelos predictivos.

1.6.3. Metodología aplicada

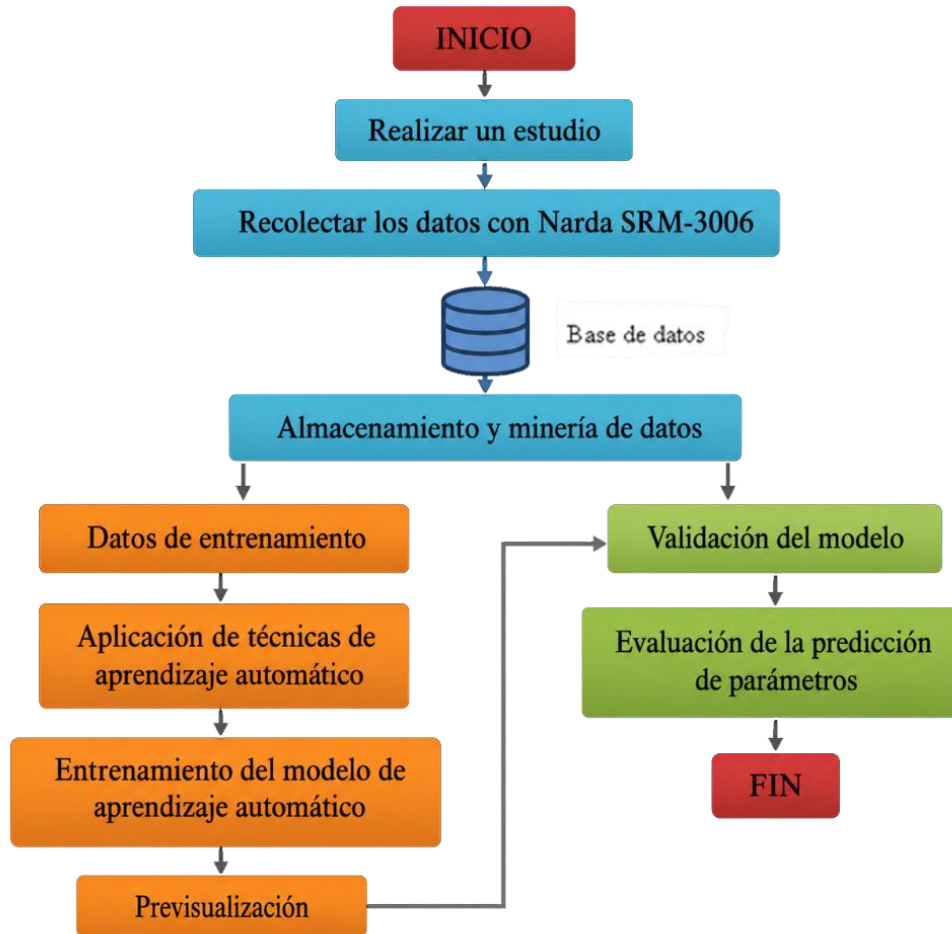
Por su carácter aplicado, la investigación aporta una solución orientada a una necesidad práctica del sector de las telecomunicaciones: la optimización del análisis y estimación de cobertura en redes móviles. Los resultados sirven como insumo técnico para la planificación de infraestructura, facilitando la identificación de zonas con niveles de señal inferiores a los umbrales normativos.

1.6.4. Metodología experimental

La fase experimental contempla el procesamiento de los datos, así como el entrenamiento y evaluación de los modelos de aprendizaje automático Random Forest y XGBoost. Los algoritmos se nutren de la información recopilada en campo y de las simulaciones complementarias. Los modelos se entrenan para predecir las seis variables objetivo definidas en el estudio: RSRP, RSRQ, SINR, CQI, DL Throughput y UL Throughput. Su desempeño se evalúa mediante las métricas de error absoluto medio (MAE) y coeficiente de determinación (R^2), complementadas con el análisis de la diferencia entre el desempeño observado en los conjuntos de entrenamiento y prueba. Esto permite comparar el comportamiento de ambos modelos y determinar cuál presenta el mejor desempeño bajo las condiciones establecidas en la investigación.

La **Figura 1** se representa la estructura que adoptó el desarrollo del proyecto, basada en los enfoques y tipos de investigación descritos.

Figura 1. Diagrama de representación del desarrollo del proyecto.



Nota. Elaboración propia.

CAPÍTULO II: MARCO TEÓRICO

Este capítulo presenta los fundamentos teóricos que sustentan la investigación. Se abordan la arquitectura y el funcionamiento de la tecnología LTE, los principales parámetros utilizados para evaluar la calidad de la señal, las técnicas de aprendizaje automático aplicables a la predicción de cobertura y los modelos empíricos de propagación. Asimismo, se describen las herramientas y métricas empleadas para el desarrollo y la evaluación de los modelos predictivos.

2.1. Fundamentos de la tecnología LTE

2.1.1. Arquitectura de la red LTE

La arquitectura de una red LTE se divide en dos bloques funcionales principales: la red de acceso radioeléctrico evolucionada (E-UTRAN) y el núcleo de paquetes evolucionado (EPC). Junto con el equipo de usuario (UE), estos elementos conforman la estructura básica de una red LTE [9].

Las estaciones base que componen la E-UTRAN se denominan eNodeB (Evolved Node B) y son los responsables de la comunicación de datos entre la red y los dispositivos móviles mediante la interfaz aérea. Además, cada eNodeB controla una o más celdas y participa en la gestión de la movilidad y de la calidad de servicio (QoS). Asimismo, puede comunicarse con los eNodeB vecinos mediante la interfaz X2, lo que permite la coordinación entre celdas y facilita la ejecución del traspaso de usuarios (handover).

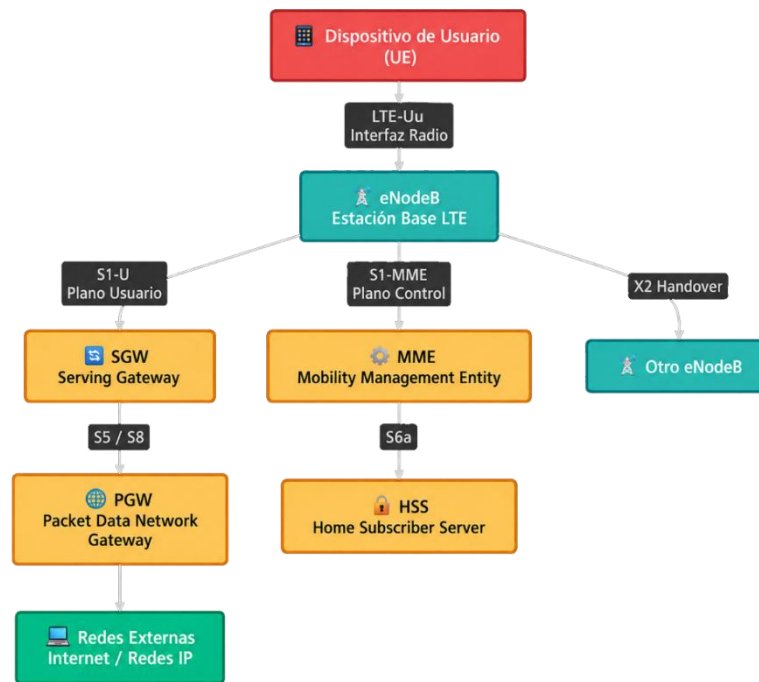
El EPC constituye el núcleo de la red y se encarga de funciones relacionadas con la gestión de la movilidad, la autenticación de los usuarios y la conexión con redes externas. Sus principales componentes son los siguientes:

- **La entidad de gestión de movilidad (MME)** administra la movilidad del usuario, la autenticación y la señalización de control.
- **La pasarela de servicio (SGW)** enruta y reenvía los datos entre el eNodeB y el PGW, y actúa como punto de anclaje del tráfico de datos.
- **La pasarela de red de datos por paquetes (PGW)** proporciona acceso a redes externas (como internet), asigna la dirección IP y aplica las políticas de calidad de servicio.
- **El servidor de suscriptores locales (HSS)** almacena la información de suscripción, incluidos los perfiles de usuario y los datos de autenticación.

La comunicación entre estas entidades ocurre a través de interfaces normalizadas identificadas como S1 y S5/S8. La arquitectura LTE está basada en el protocolo IP, lo que permite una integración eficiente con redes de datos y facilita la prestación de servicios de comunicación de alta velocidad [2].

La **Figura 2** muestra la arquitectura de una red LTE y sus principales componentes, entre los que se encuentran el equipo de usuario (UE), las estaciones base eNodeB y las entidades que conforman el núcleo de paquetes evolucionado (EPC), como el MME, el SGW, el PGW y el HSS. Además, se ilustran las interfaces que interconectan estos elementos, entre ellas LTE-Uu, S1-U, S1-MME, S5/S8, S6a y X2, así como la conexión entre estaciones base para la realización del handover.

Figura 2. Componentes principales y flujo de datos.



Nota. Elaboración propia con base en [9].

2.1.2. Principios de la transmisión LTE

El esquema de acceso múltiple empleado por LTE varía según el sentido de la comunicación. En el enlace descendente se utiliza OFDMA (Orthogonal Frequency Division Multiple Access), técnica que permite distribuir las subportadoras disponibles entre diferentes usuarios y favorece el uso eficiente del espectro. En el enlace ascendente se emplea SC-FDMA (Single Carrier Frequency Division Multiple Access), cuya principal ventaja es presentar una menor relación entre la potencia pico y la potencia

media (PAPR) en comparación con OFDMA, lo que reduce las exigencias del amplificador de potencia y favorece un menor consumo energético en los terminales.

En cuanto al esquema de duplexación, LTE admite dos modos de operación: FDD (Frequency Division Duplex), en el que la transmisión y la recepción se realizan simultáneamente utilizando bandas de frecuencia diferentes; y TDD (Time Division Duplex), en el que ambas direcciones utilizan la misma banda de frecuencia, pero se alternan en diferentes intervalos de tiempo [6].

2.1.3. Parámetros clave de la señal LTE

El desempeño de una red LTE puede evaluarse mediante diferentes indicadores relacionados con el nivel y la calidad de la señal recibida. En esta investigación se consideran los siguientes parámetros:

- **El RSRP (potencia de la señal de referencia recibida)** representa la potencia de las señales de referencia recibidas y constituye uno de los principales indicadores utilizados para evaluar el nivel de cobertura.
- **El RSRQ (calidad de la señal de referencia recibida)** permite evaluar las condiciones de recepción a partir de la relación entre la potencia de referencia y la potencia total recibida en el canal.
- **El RSSI (indicador de intensidad de la señal recibida)** representa la potencia total recibida dentro del ancho de banda considerado, incluyendo la señal útil, las señales interferentes y el ruido presente en el canal.
- **El SINR (relación señal-interferencia más ruido)** representa la relación entre la potencia de la señal útil y la suma de la interferencia y el ruido. Este indicador permite evaluar la calidad de la recepción y las condiciones radioeléctricas del enlace.

Tabla 1. Recomendaciones de intensidad de señal móvil en redes LTE.

Intensidad de señal	RSRP (dBm)	RSRQ (dB)	SINR (dB)	RSSI (dBm)
Excelente	≥ -80	≥ -10	≥ 20	≥ -65
Muy buena	-80 a -90	-10 a -15	13 a 20	-65 a -75
Regular	-90 a -100	-15 a -20	0 a 13	-75 a -85
Mala	≤ -100	≤ -20	≤ 0	≤ -85

Nota. Tomado de [10].

En los casos en que un valor coincide exactamente con el límite entre dos categorías (por ejemplo, -80 dBm o 20 dB), se asigna a la categoría de mayor calidad, siguiendo el criterio $\geq$ indicado en la fuente original.

La **Tabla 1** resume los umbrales de referencia correspondientes a las principales métricas empleadas para evaluar la calidad de la señal en redes LTE. En particular, la intensidad de la señal se clasifica a partir del RSRP, expresado en decibelios-milivatio (dBm), en diferentes niveles de calidad. El nivel excelente (≥ -80 dBm) corresponde a una señal fuerte y estable, mientras que el nivel muy bueno (-80 a -90 dBm) representa condiciones favorables para la prestación de servicios de datos. El nivel regular (-90 a -100 dBm) permite mantener la conexión, aunque puede presentar una reducción en la calidad y en la velocidad del servicio. Finalmente, los valores inferiores a -100 dBm corresponden a niveles de señal no utilizables y presentan una mayor probabilidad de pérdida de conectividad. Esta clasificación permite identificar zonas con niveles de cobertura deficientes y constituye una referencia para el análisis de la cobertura y la planificación de la infraestructura de red [10].

2.1.4. Cobertura y factores que la afectan en LTE

La cobertura de una red LTE depende de la propagación de las ondas de radio y está condicionada por diversos factores relacionados con las características de la red y del entorno. Entre los principales se encuentran:

- **La frecuencia de operación** influye en el alcance y las características de propagación de la señal. En condiciones comparables, las frecuencias más altas presentan mayores pérdidas de propagación y, por lo tanto, pueden requerir una mayor densidad de estaciones base para proporcionar una cobertura equivalente.
- **Los obstáculos físicos**, como edificios, árboles y estructuras metálicas pueden bloquear, absorber, reflejar o difractar las ondas de radio, afectando el nivel y la calidad de la señal recibida.
- **La topografía del terreno**, incluida la presencia de montañas, colinas, depresiones y otras irregularidades del terreno puede modificar las condiciones de propagación y afectar la existencia de una línea de vista entre el transmisor y el receptor.
- **Las condiciones climáticas** pueden producir efectos sobre la propagación de las ondas de radio. Su influencia depende de la frecuencia de operación, la distancia y las condiciones específicas del enlace.

Para estimar la cobertura y las pérdidas de propagación en diferentes entornos, se emplean modelos como Okumura-Hata y COST-231 Hata. Estos modelos permiten estimar la pérdida de propagación a partir de variables relacionadas con la frecuencia, la distancia entre el transmisor y el receptor, la altura de la antena y las características generales del entorno [5].

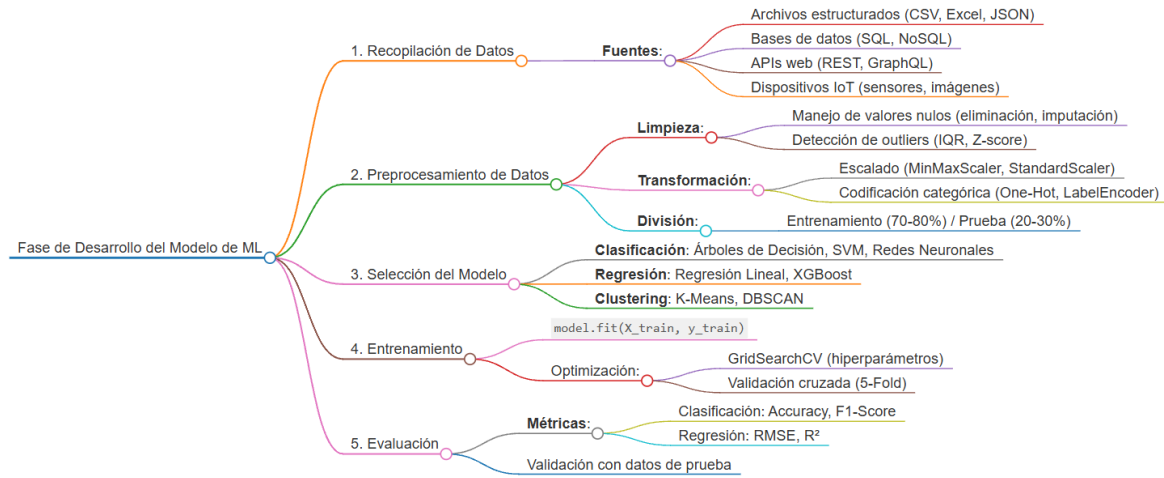
2.2. Aplicación del aprendizaje automático

El aprendizaje automático (Machine Learning) es una rama de la inteligencia artificial que permite a los sistemas aprender a partir de los datos e identificar patrones sin depender de reglas programadas explícitamente. En el ámbito de las redes LTE, estas técnicas se emplean para analizar grandes volúmenes de información generada por los dispositivos y las estaciones base, con el propósito de caracterizar el comportamiento de la señal, identificar patrones y apoyar a la optimización de los recursos de la red.

Técnicas como Random Forest, las máquinas de vectores de soporte (SVM) y XGBoost han demostrado ser de gran utilidad en tareas relacionadas con la clasificación de niveles de cobertura y la estimación de variables como el RSRP en diferentes ubicaciones geográficas. Los resultados de estos modelos pueden contribuir a identificar zonas con niveles deficientes de señal o condiciones desfavorables de propagación, proporcionando información de apoyo para la planificación y optimización de la infraestructura de telecomunicaciones [6].

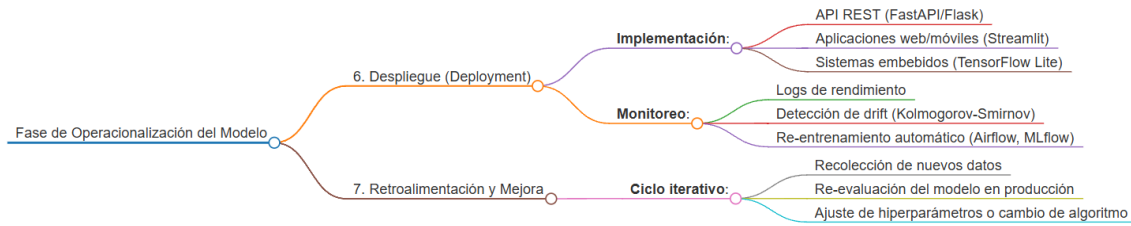
Entre las herramientas disponibles para el desarrollo de modelos de aprendizaje automático en Python se encuentran Scikit-learn y XGBoost. Scikit-learn proporciona funciones para el procesamiento de datos, el entrenamiento, la validación y la evaluación de modelos, mientras que XGBoost permite implementar algoritmos de aumento de gradiente basados en árboles de decisión. El uso de estas herramientas resulta especialmente útil en el análisis de redes LTE, donde las condiciones de propagación y desempeño pueden variar como consecuencia de cambios en el entorno urbano, el desplazamiento de los usuarios y las variaciones en los niveles de interferencia. En este contexto, los modelos predictivos pueden ser reentrenados con nuevos datos para mantener o mejorar su capacidad de generalización frente a las condiciones del entorno.

Figura 3. Desarrollo del aprendizaje automático. Parte 1.



Nota. Elaboración propia con base en [6].

Figura 4. Desarrollo del aprendizaje automático. Parte 2.



Nota. Elaboración propia con base en [6].

Las **Figuras 3 y 4** presentan, mediante un flujo iterativo, las principales etapas del proceso de aprendizaje automático, entre las que se encuentra la recopilación y preparación de los datos, la selección y el entrenamiento de los modelos, su evaluación y posterior aplicación, así como la actualización del modelo a partir de nuevos datos.

2.3. Predicción del rendimiento de las redes LTE mediante el uso de aprendizaje automático

El uso de técnicas de aprendizaje automático en el análisis de redes LTE permite identificar patrones en los datos de desempeño y detectar zonas con niveles deficientes de señal. Esto proporciona información que puede utilizarse como apoyo para la planificación y optimización de la red.

Indicadores como el RSRP, el RSRQ y el SINR permiten evaluar diferentes aspectos de la señal recibida por el usuario. Al combinar estas variables con información geográfica, características propias del entorno y otros parámetros obtenidos mediante

mediciones de campo, es posible emplear algoritmos como Random Forest y XGBoost para desarrollar modelos de clasificación y regresión orientados al análisis del comportamiento de redes móviles [5].

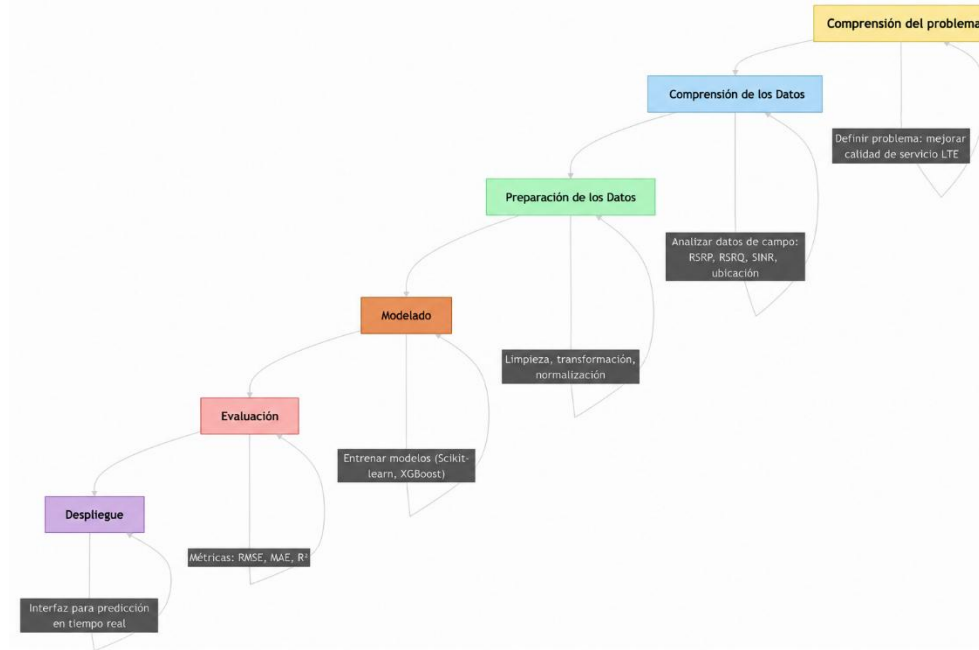
2.4. Proceso de modelado predictivo

El desarrollo de modelos predictivos mediante técnicas de aprendizaje automático puede estructurarse mediante la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), compuesta por seis etapas principales:

- **La comprensión del problema** define los objetivos y requisitos de la investigación. En esta etapa se establecen los objetivos que orientan el desarrollo del modelo predictivo.
- **La comprensión de los datos** incluye la exploración inicial de las mediciones recopiladas, considerando las variables disponibles y sus características, así como su distribución y posibles relaciones.
- **La preparación de los datos** consiste en limpiar, transformar y organizar la información, con el propósito de obtener un conjunto de datos adecuado para el entrenamiento de los modelos.
- **El modelado** comprende la selección y el entrenamiento de algoritmos de aprendizaje automático. Entre los algoritmos utilizados en tareas de predicción y análisis de datos se encuentran Random Forest y XGBoost, los cuales pueden implementarse mediante herramientas disponibles en Python.
- **La evaluación** consiste en analizar el desempeño de los modelos mediante métricas que permitan determinar la precisión de las predicciones, comparar sus resultados e identificar el modelo con mejor desempeño.
- **El despliegue** corresponde a la aplicación de los resultados del modelo en un entorno que permita utilizar y visualizar las predicciones obtenidas.

La **Figura 5** presenta el flujo general del proceso CRISP-DM aplicado al desarrollo del modelo predictivo, desde la adquisición y preparación de los datos hasta la evaluación y la representación de los resultados.

Figura 5. Proceso CRISP-DM para el modelado predictivo de redes LTE.



Nota. Elaboración propia.

2.5. Python

Python es un lenguaje de programación de alto nivel, interpretado y multiparadigma, ampliamente utilizado en áreas como el análisis de datos, la inteligencia artificial y el aprendizaje automático. Su sintaxis y disponibilidad de numerosas librerías facilitan el procesamiento de datos y el desarrollo de modelos predictivos, por lo que resulta adecuado para el análisis de información asociada con redes LTE.

Entre las principales librerías se encuentran Pandas y NumPy, utilizadas para el procesamiento y análisis de datos; Scikit-learn, empleada para el desarrollo, entrenamiento y evaluación de modelos de aprendizaje automático; XGBoost, orientada a la implementación de algoritmos de aumento de gradiente basados en árboles de decisión; y Matplotlib, utilizada para la representación gráfica de los resultados. En conjunto, estas herramientas proporcionan recursos para el procesamiento de datos, el desarrollo de modelos predictivos y la representación gráfica de los resultados obtenidos [11].

2.6. Instrumentación para la recolección de datos LTE

2.6.1. Narda SRM-3006

El Narda SRM-3006 es un analizador selectivo de radiación diseñado para la medición y caracterización de campos electromagnéticos en un amplio rango de frecuencias. El equipo permite analizar señales de radiofrecuencia y evaluar diferentes fuentes de radiación electromagnética en entornos urbanos y rurales, por lo que resulta adecuado para estudios de medición y caracterización del espectro radioeléctrico [12].

Figura 6. *Analizador selectivo de radiación electromagnética Narda SRM-3006.*



Nota. Tomado de [12].

La **Figura 6** muestra el equipo empleado como instrumento de apoyo para la caracterización de las señales radioeléctricas presentes en el área de estudio. Las mediciones obtenidas con este instrumento se utilizaron como referencia complementaria para contrastar los niveles de potencia registrados mediante G-NetTrack Pro y aportar datos adicionales para el análisis de las condiciones de señal durante la campaña de recolección de datos.

2.6.2. G-NetTrack Pro

G-NetTrack Pro es una aplicación móvil utilizada para monitorear y registrar parámetros de redes celulares. La herramienta permite recopilar información relacionada con el nivel y la calidad de la señal en diferentes tecnologías de comunicación móvil, entre ellas GSM, UMTS, LTE y 5G. Entre los parámetros que pueden registrarse se encuentran el nivel de la señal, los identificadores de celda, la frecuencia de operación y la ubicación geográfica.

Las mediciones pueden exportarse para su posterior procesamiento y análisis, lo que permite utilizar la información recopilada en estudios de cobertura, caracterización y optimización de redes móviles [13].

2.7. Técnicas de aprendizaje automático para la predicción de niveles de cobertura y propagación en redes LTE

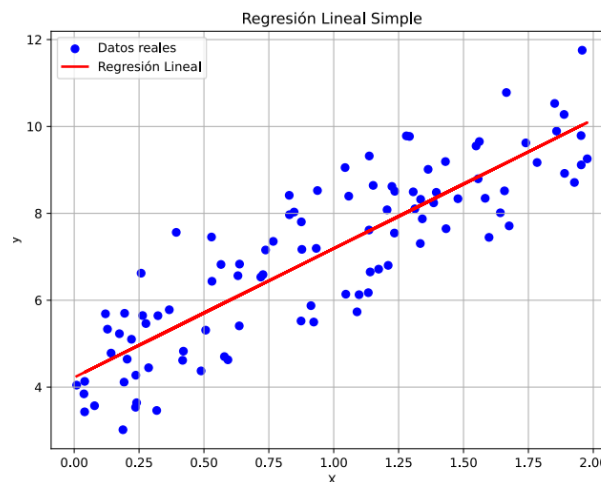
La estimación de la cobertura constituye un aspecto relevante en la planificación y optimización de redes LTE. Tradicionalmente, esta tarea se ha abordado mediante modelos empíricos de propagación. Como alternativa, las técnicas de aprendizaje automático permiten aprovechar datos obtenidos en campo para desarrollar modelos capaces de estimar parámetros relacionados con el nivel y la calidad de la señal a partir de variables geográficas, ambientales y características de la red. Entre las técnicas que pueden aplicarse a este tipo de problemas se encuentran las siguientes:

2.7.1. Regresión lineal (Linear Regression)

La regresión lineal es un modelo estadístico y de aprendizaje supervisado que establece una relación entre una variable dependiente y una o más variables independientes. Su principal ventaja radica en su simplicidad y facilidad de interpretación. Sin embargo, su capacidad para representar relaciones no lineales complejas es limitada, lo que puede afectar su desempeño cuando las variables presentan interacciones o comportamientos no lineales [14].

La **Figura 7** muestra el funcionamiento de un modelo de regresión lineal simple, desde la representación de los datos hasta la obtención de la línea utilizada para realizar las predicciones.

Figura 7. Representación gráfica del funcionamiento de la regresión lineal.



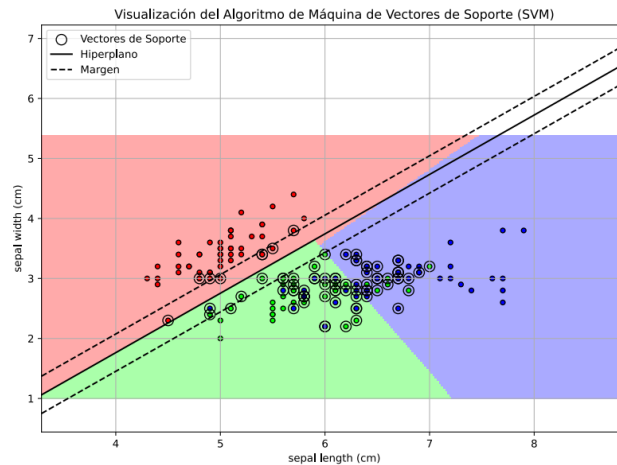
Nota. Elaboración propia.

2.7.2. Máquinas de vectores de soporte (Support Vector Machines)

Las máquinas de vectores de soporte (SVM) son algoritmos de aprendizaje supervisado que pueden emplearse tanto en tareas de clasificación como de regresión. En el ámbito de las redes móviles, pueden utilizarse para clasificar niveles de cobertura o estimar variables relacionadas con la señal. Su capacidad para trabajar con espacios de alta dimensión las convierte en una alternativa para problemas en los que intervienen múltiples variables [15].

La **Figura 8** describe el funcionamiento de una SVM aplicada a un problema de clasificación binaria, en el que se establece un límite de decisión entre dos grupos de datos.

Figura 8. Representación gráfica del funcionamiento de las máquinas de vectores de soporte.



Nota. Elaboración propia.

2.7.3. Random Forest y XGBoost

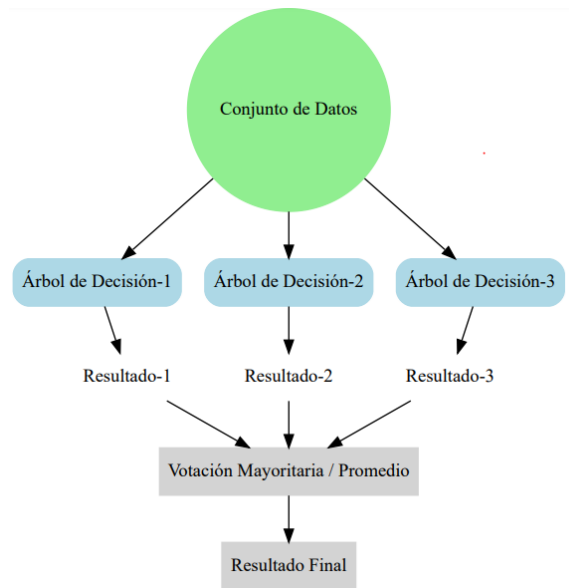
Los árboles de decisión son modelos de aprendizaje supervisado que utilizan una estructura jerárquica para realizar predicciones a partir de diferentes condiciones sobre las variables de entrada. A partir de estos modelos se desarrollaron técnicas de aprendizaje en conjunto (ensemble learning), entre las que se encuentran Random Forest y XGBoost.

Random Forest utiliza múltiples árboles de decisión entrenados sobre diferentes muestras de los datos y combina sus predicciones para obtener un resultado final más robusto y estable. Este procedimiento corresponde a un enfoque de *bagging*. Por otra parte, XGBoost emplea un enfoque de *boosting*, mediante el cual los árboles se construyen secuencialmente, incorporando en cada etapa la información asociada con los errores de las predicciones anteriores.

Ambos algoritmos pueden utilizarse en problemas de clasificación y regresión y presentan ventajas para trabajar con conjuntos de datos que contienen múltiples variables y relaciones no lineales. Por estas características, resultan adecuados para el desarrollo de modelos predictivos aplicados al análisis de redes LTE [16].

La **Figura 9** representa el funcionamiento de Random Forest, donde las predicciones obtenidas por diferentes árboles de decisión se combinan para generar una predicción final.

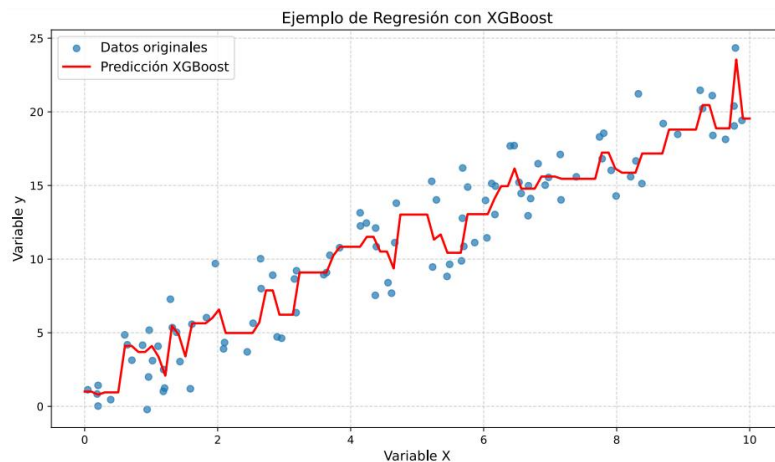
Figura 9. Representación gráfica del funcionamiento del algoritmo Random Forest.



Nota. Elaboración propia.

La **Figura 10** muestra el funcionamiento de XGBoost, que se basa en la construcción secuencial de árboles de decisión para mejorar progresivamente las predicciones.

Figura 10. Representación gráfica del funcionamiento del algoritmo XGBoost.



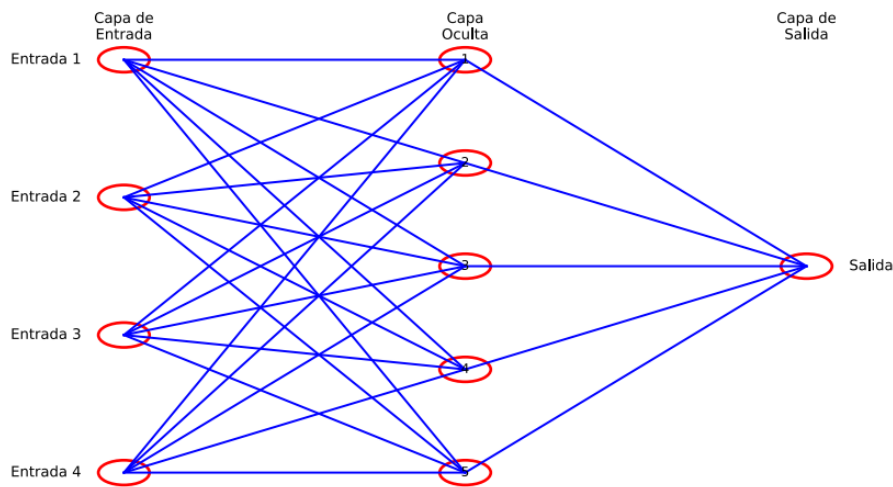
Nota. Elaboración propia.

2.7.4. Redes neuronales artificiales (Artificial Neural Networks, ANN)

Las redes neuronales artificiales son modelos de aprendizaje automático capaces de representar relaciones complejas entre diferentes variables mediante la conexión de neuronas organizadas en capas de entrada, ocultas y salida. Estas técnicas pueden aplicarse a problemas de clasificación y regresión, incluyendo la estimación de la intensidad de la señal en diferentes entornos [17].

La **Figura 11** representa una estructura básica de una red neuronal artificial, compuesta por capas de entrada, ocultas y salida.

Figura 11. Representación gráfica del funcionamiento de una red neuronal artificial.



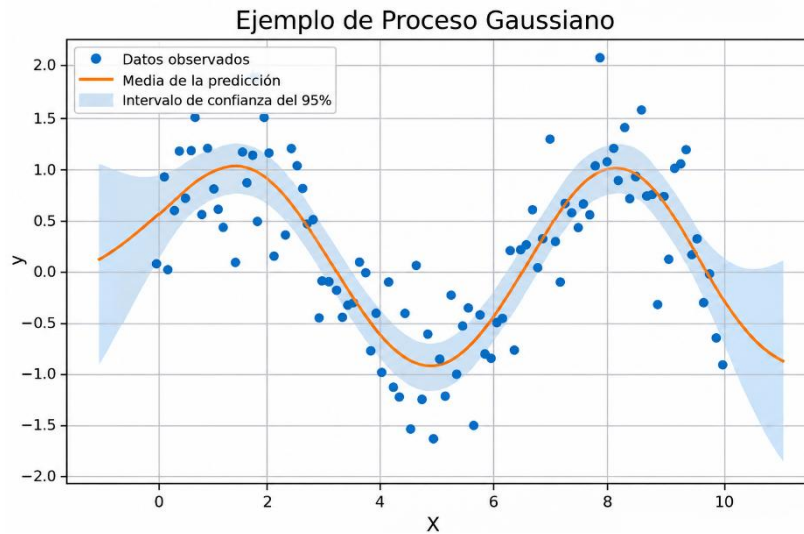
Nota. Elaboración propia.

2.7.5. Procesos gaussianos (Gaussian Process Regression, GPR)

Los procesos gaussianos constituyen un método probabilístico y no paramétrico utilizado principalmente en problemas de regresión. Además de generar una estimación de la variable objetivo, permiten cuantificar la incertidumbre asociada con las predicciones, característica que puede resultar útil en aplicaciones de planificación y análisis de redes móviles. Su utilización puede ser especialmente conveniente cuando se dispone de conjuntos de datos de tamaño reducido [18].

La **Figura 12** muestra una representación de un proceso gaussiano aplicado a un problema de regresión, en el que se obtiene una predicción acompañada de un intervalo de incertidumbre.

Figura 12. Representación gráfica del funcionamiento de un proceso gaussiano.



Nota. Elaboración propia.

2.8. Análisis comparativo de métodos de predicción utilizados en redes LTE

La selección del algoritmo de aprendizaje automático depende de las características de los datos, las variables disponibles y el objetivo de predicción. En el contexto de las redes LTE pueden emplearse diferentes enfoques, entre ellos modelos basados en árboles de decisión, redes neuronales y métodos de interpolación espacial.

2.8.1. Random Forest

Random Forest es un algoritmo de aprendizaje automático basado en la combinación de múltiples árboles de decisión. Su estructura permite trabajar con conjuntos de datos que contienen variables numéricas y categóricas, así como relaciones no lineales entre las características de entrada. En aplicaciones relacionadas con redes LTE puede utilizar variables como el RSRP, el RSRQ, la latitud, la longitud, la distancia y las características de la celda para realizar tareas de clasificación o regresión [8].

2.8.2. XGBoost

XGBoost es un algoritmo de aprendizaje automático basado en árboles de decisión que utiliza una estrategia de boosting. A diferencia de Random Forest, los árboles se construyen de forma secuencial, de manera que cada nuevo árbol busca reducir los errores de las predicciones anteriores. Esta característica le permite modelar relaciones complejas entre las variables y alcanzar un alto desempeño en diferentes problemas de regresión y clasificación. Además, incorpora mecanismos para manejar valores faltantes y permite analizar la importancia de las variables utilizadas en el modelo [8].

2.8.3. Redes neuronales (ANN, RNN y CNN)

Las redes neuronales artificiales (ANN) son modelos de aprendizaje automático capaces de representar relaciones no lineales complejas mediante diferentes capas de procesamiento. Dentro de esta familia se encuentran las redes neuronales recurrentes (RNN), orientadas principalmente al análisis de datos secuenciales, y las redes neuronales convolucionales (CNN), utilizadas con frecuencia en el procesamiento de información espacial o matricial, como imágenes y mapas de cobertura. Estas técnicas comparten una limitación frente a modelos más simples, relacionada con los requerimientos de datos, de capacidad computacional y de ajuste de hiperparámetros necesarios para obtener un buen desempeño [17].

2.8.4. Interpolación espacial mediante Kriging

El Kriging es un método geoestadístico de interpolación espacial que permite estimar valores en ubicaciones donde no existen mediciones directas, utilizando la correlación espacial entre los puntos disponibles. En estudios de cobertura puede emplearse para estimar niveles de señal entre puntos de medición. Sin embargo, su capacidad para representar cambios asociados con variables técnicas, ambientales o dinámicas de la red depende de las variables incorporadas y de la estructura espacial de los datos disponibles. La **Tabla 2** resume las características de los métodos descritos en esta sección [18].

Tabla 2. Comparativa de métodos para predecir la cobertura LTE.

Técnica	Precisión	Tiempo de entrenamiento	Adaptabilidad	Uso común
Regresión lineal	Baja / media	Muy bajo	Media	Alto
Máquinas de vectores de soporte	Media / alta	Medio / alto	Media	Medio
XGBoost / Random Forest	Alta	Medio	Muy alta	Medio
Redes neuronales	Muy alta	Alto / Muy alto	Alta	Medio
Procesos gaussianos	Alta	Alto	Media / alta	Alto

Kriging	Alta	Bajo	Baja	Alto
----------------	------	------	------	------

Nota. Elaboración propia a partir de [8], [17] y [18].

La selección de Random Forest y XGBoost para este proyecto se basa en su capacidad para trabajar con múltiples variables de entrada y modelar relaciones no lineales presentes en los datos de cobertura LTE. Ambos algoritmos permiten abordar problemas de regresión y analizar la importancia de las variables utilizadas en la predicción.

Kriging, en cambio, se fundamenta principalmente en la correlación espacial entre las mediciones, mientras que las redes neuronales pueden requerir mayores volúmenes de datos y un ajuste más complejo de sus hiperparámetros. Por estas características, Random Forest y XGBoost resultan adecuados para comparar la capacidad de estimación de ambos algoritmos a partir de las variables técnicas, geográficas y ambientales consideradas en este estudio.

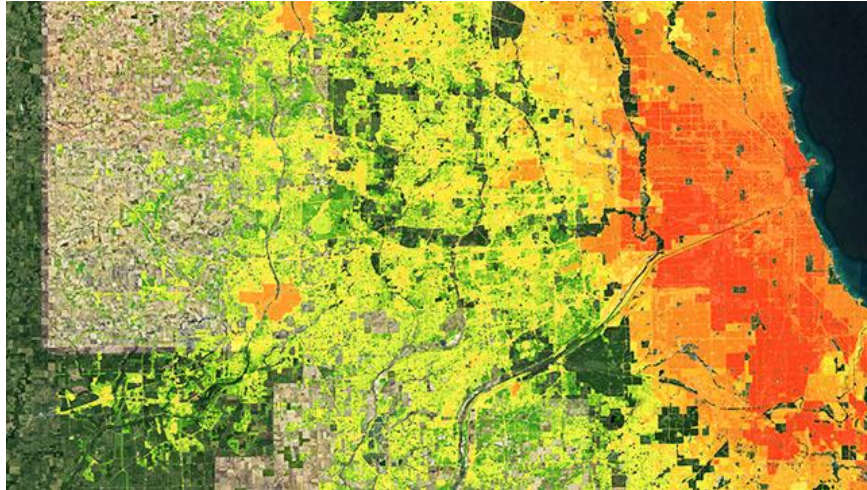
2.9. Ubicación geográfica y herramientas para la representación espacial de datos

La ubicación geográfica constituye un elemento fundamental para analizar la distribución espacial de los parámetros de señal LTE y representar los resultados mediante mapas de cobertura. Para este propósito, se emplean herramientas de posicionamiento y sistemas de información geográfica que permiten asociar los datos de señal con coordenadas geográficas y facilitar su posterior representación espacial:

- **Los sistemas GPS** proporcionan coordenadas geográficas que permiten determinar la ubicación de puntos de interés y asociar información de señal con una posición específica, facilitando su posterior representación espacial.
- **Google Earth Pro** permite importar y visualizar información geográfica mediante formatos como KML y KMZ. Esta herramienta facilita la inspección visual de puntos geográficos y su ubicación sobre imágenes satelitales [19].
- **QGIS y ArcGIS** son sistemas de información geográfica (SIG) que permiten importar, gestionar, analizar y representar datos geoespaciales. Estas herramientas admiten diferentes formatos de información geográfica y facilitan la elaboración de mapas temáticos para analizar la distribución espacial de los datos [20] [21].

La **Figura 13** muestra un proyecto desarrollado en ArcGIS, en el que se representa información geoespacial sobre una imagen satelital mediante diferentes elementos gráficos.

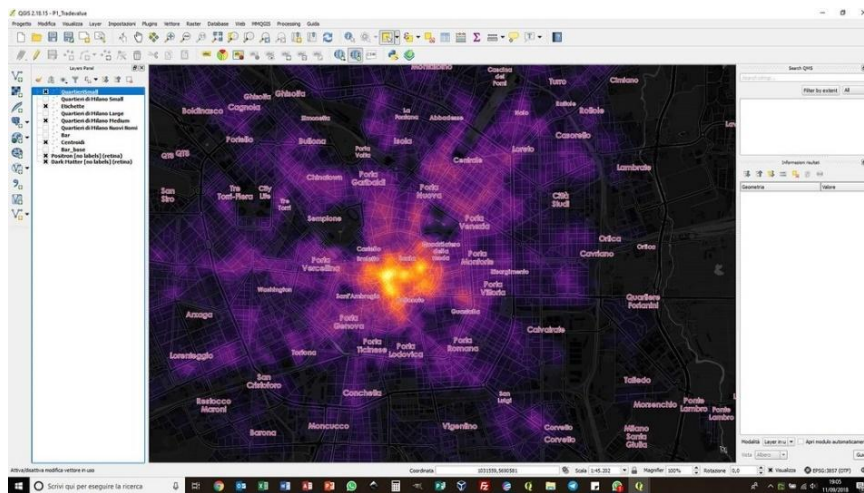
Figura 13. Representación de información geoespacial en ArcGIS.



Nota. Tomado de [20].

La **Figura 14** presenta un mapa de calor de densidad urbana elaborado en QGIS, en el que la información espacial se representa mediante una escala de colores sobre un mapa base.

Figura 14. Mapa de calor de densidad urbana elaborado en QGIS.



Nota. Tomado de [21].

2.10. Modelos de propagación empíricos

Los modelos de propagación empíricos son formulaciones matemáticas desarrolladas a partir de mediciones experimentales y utilizadas para estimar la pérdida de trayectoria de una señal de radio (path loss). Estos modelos constituyen una referencia para el diseño

y análisis de redes inalámbricas en diferentes entornos, debido a que consideran variables como la frecuencia de operación, la distancia entre el transmisor y el receptor y la altura de las antenas. A continuación, se presentan algunos de los modelos empleados en el análisis de redes móviles.

2.10.1. Okumura-Hata

El modelo Okumura-Hata es uno de los modelos empíricos más utilizados para estimar la pérdida de trayectoria en entornos urbanos, suburbanos y rurales. Su formulación se basa en los resultados de las mediciones realizadas por Yoshihisa Okumura y la posterior formulación matemática desarrollada por Hata. El modelo es aplicable, en su formulación original, a frecuencias comprendidas aproximadamente entre 150 y 1500 MHz y distancias de 1 a 20 km [22].

Su formulación general se expresa mediante:

$$L_{50}(dB) = L_F + A_{mu}(f, d) - G(h_{te}) - G(h_{re}) - G_{AREA}$$

Ecuación 1. Modelo Okumura-Hata.

Donde:

- L_{50} : pérdida media de propagación, expresada en dB.
- L_F : pérdida de propagación en el espacio libre.
- A_{mu} : atenuación media relativa respecto a la propagación en el espacio libre.
- $G(h_{te})$: factor de ganancia asociado con la altura de la antena transmisora.
- $G(h_{re})$: factor de ganancia asociado con la altura de la antena receptora.
- G_{AREA} : factor de corrección asociado con las características del entorno.

2.10.2. Modelo COST-231 Hata

El modelo COST-231 Hata constituye una extensión del modelo Okumura-Hata desarrollada para ampliar su aplicación hacia frecuencias superiores. Su formulación permite trabajar hasta aproximadamente 2000 MHz y con distancias de hasta 20 km, por lo que resulta aplicable al análisis de determinados escenarios urbanos [23].

Cabe señalar que, en el presente estudio, el modelo se aplica a una frecuencia de 2120 MHz y a distancias inferiores a 1 km, condiciones que exceden ligeramente el límite superior de frecuencia y se ubican por debajo del límite inferior de distancia para los cuales fue formulado originalmente. Esta aplicación fuera del rango nominal se detalla y justifica en la **sección 3.3.1**.

La formulación se expresa de la siguiente manera:

$$L = 46.3 + 33.9 \log(f) - 13.82 \log(h_{te}) - a(h_{re}) + [44.9 - 6.55 \log(h_{te})] \log(d) + C_m$$

Ecuación 2. Modelo COST-231 Hata.

Donde:

- **f**: frecuencia de operación, expresada en MHz.
- **h_{te}**: altura de la estación base, expresada en metros.
- **h_{re}**: altura del receptor, expresada en metros.
- **d**: distancia entre la estación base y el receptor, expresada en kilómetros.
- **a(h_{re})**: factor de corrección asociado con la altura de la antena receptora.
- **C_m**: factor de corrección relacionado con el tipo de entorno (urbano, suburbano o rural), cuyo valor varía según las características del escenario estudiado.

2.10.3. Modelo de sombreado log-normal (Log-Normal Shadowing Model)

El modelo de sombreado log-normal es un modelo estadístico utilizado para representar las variaciones de la pérdida de trayectoria que no pueden explicarse únicamente mediante la distancia entre el transmisor y el receptor. Estas variaciones pueden estar asociadas con obstáculos y cambios en las condiciones de propagación del entorno, lo que permite que el modelo represente de manera más realista las fluctuaciones de la señal [24].

Su formulación es:

$$PL(d) = PL_0 + 10\alpha \log 10 \left(\frac{d}{d_0} \right) + X\sigma$$

Ecuación 3. Modelo de sombreado log-normal.

Donde:

- **PL(d)**: pérdida de trayectoria a una distancia (d).
- **PL₀**: pérdida de trayectoria a la distancia de referencia d₀, que puede obtenerse mediante un modelo de propagación en espacio libre o mediante mediciones de campo.
- **α**: exponente de pérdida de trayectoria, que representa la rapidez con la que disminuye el nivel de la señal a medida que aumenta la distancia.
- **Xσ**: variable aleatoria asociada con el sombreado, caracterizada mediante una desviación estándar y expresada en dB.

2.11. Limitaciones de los modelos empíricos en zonas costeras

Los modelos de propagación convencionales presentan limitaciones para representar las condiciones particulares de las zonas costeras. En estos entornos, la propagación de las ondas electromagnéticas empleadas por las redes móviles puede verse influenciada por factores geográficos, ambientales y urbanos propios del área de estudio [25].

2.11.1. Influencia de la humedad relativa y la salinidad

La elevada humedad relativa propia de algunas zonas costeras puede alterar las propiedades electromagnéticas del entorno y modificar las condiciones de propagación. Sin embargo, la magnitud de este efecto depende de la frecuencia de operación y de las condiciones específicas del sitio donde se produce la transmisión. Por ello, su influencia debe analizarse considerando tanto la banda de frecuencia utilizada como las condiciones ambientales presentes en la zona de estudio [26].

De manera similar, la salinidad no suele incorporarse de forma explícita en los modelos de propagación tradicionales empleados para redes LTE. No obstante, los entornos marinos presentan condiciones particulares, como la presencia de superficies húmedas y agua salina, capaces de alterar el comportamiento de la señal en su trayectoria. Estos efectos pueden contribuir a variaciones en parámetros de calidad como el SINR, especialmente en zonas próximas a la línea costera [27].

2.11.2. Impacto de multitrayecto y superficie oceánica

La proximidad al océano Pacífico favorece la aparición de reflexiones sobre superficies como el agua y las estructuras cercanas al litoral. Como consecuencia, una misma señal puede llegar al receptor a través de diferentes trayectorias y con distintas fases, dando lugar al fenómeno conocido como multitrayecto. Dependiendo de la relación entre las señales recibidas, pueden producirse interferencias constructivas o destructivas, ocasionando variaciones en el nivel y la calidad de la señal.

Este fenómeno adquiere especial importancia en zonas costeras donde coexisten superficies reflectantes, como el mar, con edificaciones y otras estructuras próximas al litoral. Estas condiciones pueden generar fluctuaciones en la señal recibida y afectar el comportamiento de parámetros utilizados para evaluar la cobertura y la calidad del servicio LTE [28].

2.11.3. Restricciones de los modelos tradicionales y razonamiento del aprendizaje automático

Los modelos empíricos convencionales se basan principalmente en relaciones matemáticas obtenidas a partir de mediciones realizadas bajo determinadas condiciones. Por esta razón, pueden presentar dificultades para representar todas las variaciones presentes en entornos reales, especialmente cuando existen cambios en las características urbanas, ambientales y geográficas de la zona de estudio.

El uso de mediciones obtenidas directamente en campo permite contrastar las estimaciones de los modelos de propagación con el comportamiento real de la red y proporcionar información adicional para el desarrollo de modelos predictivos [29]. En este contexto, el aprendizaje automático permite establecer relaciones entre diferentes variables de entrada, como la distancia, la ubicación geográfica, las características de la estación base y los parámetros de señal, con el propósito de estimar el comportamiento de la cobertura LTE.

De esta manera, los modelos de aprendizaje automático pueden complementar de manera efectiva a los modelos de propagación tradicionales, ya que permiten aprovechar información obtenida directamente del entorno de estudio y adaptarse a patrones que no necesariamente son representados por las formulaciones empíricas.

2.12. Comparativa de software para predicción de cobertura LTE

En esta sección se describen diferentes herramientas empleadas en la planificación, simulación y medición de redes móviles, y se comparan sus principales características y funcionalidades para identificar sus posibilidades de aplicación en el análisis y predicción de cobertura LTE.

2.12.1. ICS Telecom

ICS Telecom (HTZ Communications) es una herramienta de planificación y modelado de redes inalámbricas que admite diferentes tecnologías, entre ellas GSM, UMTS, LTE y TETRA. También incorpora diversos modelos de propagación, como Okumura-Hata, Walfisch-Ikegami e ITU-R P.1812. La herramienta permite representar escenarios urbanos, rurales y mixtos mediante información geográfica y modelos digitales del terreno. Además, ofrece funciones para analizar la cobertura, las interferencias y otros parámetros relacionados con la planificación de redes [30].

2.12.2. Atoll

Atoll es una plataforma de planificación y optimización de redes inalámbricas que permite trabajar con diferentes tecnologías móviles, entre ellas LTE y 5G. La herramienta integra modelos de propagación, información geográfica y parámetros de red para realizar estudios de cobertura, capacidad e interferencia. También permite analizar diferentes escenarios de planificación y optimización de celdas, por lo que constituye una herramienta utilizada en procesos profesionales de diseño y optimización de redes móviles [31].

2.12.3. QualiPoc Android

A diferencia de las herramientas orientadas principalmente a la planificación y simulación, QualiPoc Android está enfocada en la realización de mediciones de campo. La aplicación permite recopilar información relacionada con el comportamiento de las redes móviles, incluyendo parámetros como el RSRP, el RSRQ y el SINR, además de otros indicadores de desempeño. Los datos obtenidos pueden utilizarse para analizar la cobertura y la calidad de la red, así como para contrastar las mediciones reales con resultados provenientes de modelos de simulación o predicción [32].

2.12.4. ZedBoard + Python

Las plataformas basadas en sistemas de radio definida por software (SDR), como ZedBoard, pueden utilizarse junto con módulos de adquisición y procesamiento de señales para desarrollar prototipos de comunicación y análisis de señales de radiofrecuencia. Estos sistemas pueden complementarse con Python y librerías como Pandas, NumPy y Scikit-learn para el procesamiento de datos y el desarrollo de modelos de aprendizaje automático. Esta combinación proporciona flexibilidad para implementar soluciones experimentales y adaptar el procesamiento de acuerdo con los requerimientos específicos de cada investigación.

La **Tabla 3** sintetiza las características de las herramientas analizadas [33].

Tabla 3. Comparativa entre software para la predicción de la cobertura LTE.

Software	Tipo de uso	Entrada requerida	Salida/ resultados	Licencia	Facilidad de uso
ICS Telecom	Planificación de red.	Datos geográficos, parámetros LTE.	Mapas de cobertura, análisis interferencias.	Comercial	Media
Atoll	Diseño y simulación.	Datos de red, KPIs.	Cobertura, KPIs, handover, interferencias.	Comercial	Alta
QualiPoc	Medición de campo.	Datos de equipos móviles.	Indicadores LTE (RSRP, RSRQ, SINR, etc.).	Comercial	Alta
ZedBoard + Python	Prototipo de predicción.	Datos CSV, modelo ML.	Predicciones de señal.	Libre	Baja

Nota. Elaboración propia con base en [30], [31], [32] y [33].

En comparación con herramientas comerciales como Atoll e ICS Telecom, que están orientadas principalmente a la planificación y optimización profesional de redes, la solución desarrollada en este proyecto se basa en herramientas de programación y análisis de datos, lo que proporciona mayor flexibilidad para adaptar el procesamiento a las características del área de estudio.

Una diferencia importante radica en el enfoque utilizado. Las herramientas de planificación emplean modelos de propagación, información geográfica y parámetros de red para realizar simulaciones y análisis de cobertura. En cambio, la propuesta desarrollada en este proyecto utiliza mediciones obtenidas directamente en el área de estudio para entrenar modelos de aprendizaje automático capaces de estimar los parámetros de cobertura LTE.

De esta manera, la propuesta no busca sustituir a las herramientas profesionales de planificación, sino ofrecer una alternativa académica y experimental que permita analizar

la cobertura a partir de datos reales y generar predicciones adaptadas a las características del entorno estudiado.

2.13. Comparación de lenguajes y entornos de programación para el aprendizaje automático

En esta sección se comparan diferentes lenguajes y entornos de programación que pueden emplearse en el desarrollo de aplicaciones relacionadas con el análisis de datos, las telecomunicaciones y el aprendizaje automático. La comparación permite identificar las características y posibilidades de aplicación de cada alternativa, entre las cuales se encuentra Python como una opción orientada al procesamiento de datos y desarrollo de modelos de aprendizaje automático.

2.13.1. Python

Python es un lenguaje de programación de alto nivel que se caracteriza por su sintaxis sencilla y por disponer de un amplio conjunto de librerías orientadas al análisis de datos, la visualización y el aprendizaje automático. Entre ellas se encuentran Scikit-learn, utilizada para la implementación de algoritmos de aprendizaje automático; Matplotlib, empleada para la representación gráfica de datos; y Pandas, orientada a la manipulación y gestión de datos estructurados. Asimismo, librerías como GeoPandas permiten procesar y representar información geoespacial, lo que facilita su aplicación en estudios relacionados con la cobertura de redes móviles [11].

2.13.2. MATLAB

MATLAB es un entorno de programación y cálculo numérico que dispone de herramientas especializadas para el análisis y la simulación de sistemas de telecomunicaciones. Entre sus funcionalidades se encuentran la generación y el análisis de señales, la simulación de sistemas de comunicación y el estudio de diferentes escenarios de propagación. Una de sus principales limitaciones para proyectos académicos está relacionada con el costo de las licencias y de algunas herramientas especializadas [34].

2.13.3. C++

C++ es un lenguaje de programación ampliamente utilizado en aplicaciones que requieren un alto rendimiento computacional y un uso eficiente de los recursos del sistema. Estas características lo hacen adecuado para el desarrollo de sistemas embebidos y aplicaciones relacionadas con radio definida por software (SDR), donde puede ser

necesario interactuar directamente con el hardware y procesar señales de manera eficiente. Además, herramientas como GNU Radio permiten diseñar flujos de procesamiento de señales mediante diferentes bloques funcionales. Sin embargo, su complejidad de desarrollo puede ser mayor en comparación con lenguajes de alto nivel como Python [35].

2.13.4. Java

Java es un lenguaje de programación utilizado en diferentes tipos de aplicaciones, entre ellas el desarrollo de aplicaciones móviles y sistemas con interfaces gráficas. En el contexto de las redes LTE, puede emplearse para desarrollar aplicaciones de monitoreo, interfaces y sistemas destinados a la presentación de resultados. Sin embargo, para el análisis de datos, el aprendizaje automático y el procesamiento geoespacial, Python dispone de un conjunto de librerías directamente relacionado con las necesidades de esta investigación.

La **Tabla 4** resume la comparación entre los lenguajes considerados [36].

Tabla 4. Comparación de lenguajes y herramientas de programación aplicables al desarrollo de sistemas de predicción de cobertura LTE.

Lenguaje / Herramienta	Propósito principal	Bibliotecas útiles	Comunidad	Flexibilidad
Python	Aprendizaje automático y análisis de datos.	Pandas, Scikit-learn, XGBoost, Matplotlib, GeoPandas.	Alta	Muy alta
MATLAB	Cálculo numérico, simulación y procesamiento de señales.	LTE Toolbox, Statistics and Machine Learning Toolbox, Curve Fitting Toolbox.	Alta	Alta
C++	Sistemas de alto rendimiento y	GNU Radio, IT++.	Media	Alta

	aplicaciones embebidas.			
Java	Aplicaciones móviles y sistemas multiplataforma.	Android SDK y bibliotecas de Java.	Alta	Media

Nota. Elaboración propia con base en [11], [34], [35] y [36].

Se seleccionó Python como lenguaje principal debido a la disponibilidad de librerías especializadas para el procesamiento de datos, el aprendizaje automático y la representación geoespacial. Entre las herramientas empleadas se encuentran Pandas para la gestión de datos, Scikit-learn y XGBoost para el desarrollo de los modelos predictivos, Matplotlib para la visualización de resultados y las herramientas geoespaciales utilizadas para representar la cobertura.

Además, Python permite integrar en un mismo entorno las diferentes etapas de la investigación, desde el procesamiento y depuración de las mediciones hasta el entrenamiento, evaluación y visualización de los modelos. Al tratarse de un ecosistema ampliamente utilizado y con numerosas herramientas de código abierto, facilita la modificación y adaptación del sistema de acuerdo con los requerimientos del proyecto.

2.14. Métricas de evaluación para modelos de predicción

Las métricas estadísticas permiten evaluar el desempeño de los modelos predictivos y determinar el grado de correspondencia entre los valores estimados y los valores observados. Entre las principales métricas utilizadas para este propósito se encuentran el error absoluto medio (MAE), la raíz del error cuadrático medio (RMSE) y el coeficiente de determinación (R^2).

2.14.1. Error absoluto medio (MAE)

El error absoluto medio (MAE) representa el promedio de las diferencias absolutas entre los valores estimados por el modelo y los valores observados. Esta métrica permite cuantificar el error promedio de las predicciones y se expresa en las mismas unidades que la variable objetivo [37]. Su ecuación se expresa de la siguiente manera:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Ecuación 4. Error absoluto medio.

Donde:

- **MAE**: error absoluto medio.
- **n**: número total de mediciones evaluadas.
- **y_i**: valor de RSRP medido en el punto (i).
- **ŷ_i**: valor de RSRP predicho por el modelo para el punto (i).

2.14.2. Raíz del error cuadrático medio (RMSE)

La raíz del error cuadrático medio (RMSE) permite evaluar la magnitud de los errores entre los valores predichos y los valores observados. A diferencia del MAE, esta métrica eleva al cuadrado las diferencias antes de calcular su promedio y posteriormente obtiene la raíz cuadrada del resultado. Debido a este procedimiento, los errores de mayor magnitud tienen una influencia más significativa sobre el valor final de la métrica [37]. Su ecuación se expresa de la siguiente manera:

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

Ecuación 5. Raíz del error cuadrático medio.

Donde:

- **RMSE**: raíz del error cuadrático medio, expresada en las mismas unidades que la variable objetivo.
- **n**: número total de mediciones evaluadas.
- **y_i**: valor de RSRP medido en el punto (i).
- **ŷ_i**: valor de RSRP predicho por el modelo para el punto (i).

2.14.3. Coeficiente de determinación (R²)

El coeficiente de determinación indica la proporción de la variabilidad de la variable objetivo que puede ser explicada por el modelo predictivo. Su valor permite evaluar el grado de ajuste entre las predicciones y los valores observados. En términos generales, un valor de R² cercano a 1 indica que el modelo explica una mayor proporción de la variabilidad de los datos [37].

No obstante, esta métrica debe interpretarse conjuntamente con otras medidas de error, como el MAE y el RMSE, para obtener una evaluación más completa del desempeño del modelo.

Su ecuación se expresa de la siguiente manera:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Ecuación 6. Coeficiente de determinación.

Donde:

- **R²**: coeficiente de determinación, una medida adimensional.
- **SS_{res}**: suma de los cuadrados de los residuos $\hat{y}_i$, $\sum (y_i - \hat{y}_i)^2$, representa el error del modelo.
- **SS_{tot}**: suma total de cuadrados $\sum (y_i - \bar{y})^2$, representa la variabilidad total de los datos respecto a su media $\bar{y}$.
- **n**: número total de mediciones evaluadas.
- **y_i**: valor de RSRP medido en el punto i.
- **ŷ_i**: valor de RSRP predicho por el modelo para el punto i.
- **ȳ**: valor medio de los valores RSRP observados.

CAPÍTULO III: DESARROLLO DE LA PROPUESTA TÉCNICA

El componente técnico de este trabajo se centra en el desarrollo de un modelo predictivo capaz de estimar la cobertura de una red móvil LTE. Para ello, se relacionan variables físicas, geográficas y radioeléctricas mediante técnicas computacionales y matemáticas.

El desarrollo adopta un enfoque experimental, debido a que los modelos de aprendizaje automático se aplican y evalúan utilizando datos reales obtenidos en campo. Esta estrategia permite analizar el comportamiento de la señal en el entorno estudiado y desarrollar una solución sustentada en evidencia empírica.

El problema radica en que los modelos empíricos de propagación tradicionales, aunque son útiles para realizar estimaciones generales, presentan limitaciones para representar la complejidad de los entornos urbanos, donde factores como la densidad de edificaciones, las condiciones ambientales y la orientación de las antenas influyen en la propagación de la señal. Para abordar esta limitación, se propone un enfoque híbrido que combina un modelo físico-empírico con técnicas de aprendizaje automático, con el propósito de mejorar la estimación de la intensidad de la señal recibida.

3.1. Diseño de la solución

La propuesta consiste en un sistema de predicción basado en un modelo híbrido que integra el modelo de propagación COST-231 Hata con algoritmos de aprendizaje automático supervisado.

A partir de los parámetros físicos del sistema, el modelo COST-231 Hata proporciona una primera estimación de la pérdida de trayectoria. Sin embargo, al tratarse de un modelo empírico, no contempla completamente las características particulares del entorno local, lo que puede llegar a limitar su precisión. Para complementar esta estimación, se incorporan Random Forest y XGBoost, ambos capaces de identificar relaciones no lineales a partir de los datos disponibles. Random Forest emplea un conjunto de árboles de decisión mediante la técnica de bagging, mientras que XGBoost utiliza boosting para construir los árboles de manera secuencial y reducir los errores de las predicciones anteriores.

3.2. Arquitectura del diseño

La **Figura 15** presenta la arquitectura general del sistema, organizada en capas que agrupan las principales funciones del proceso.

Figura 15. Arquitectura por capas del sistema de predicción de cobertura LTE.



Nota. Elaboración propia.

El sistema se estructura en cinco capas funcionales:

- **La capa de adquisición** registra los indicadores de nivel y calidad de la señal LTE, entre ellos el RSRP, el RSRQ y el RSSI, mediante la aplicación G-NetTrack Pro instalada en el terminal móvil. De manera complementaria, incorpora la caracterización espectral de la banda analizada obtenida con el analizador Narda SRM-3006.
- **La capa de preprocesamiento** realiza la limpieza y depuración de los datos, trata los valores nulos y filtra las muestras correspondientes a la estación base y al sector de interés. Para ello, cada registro se asigna al sector correspondiente según su PCI. Las muestras provenientes de estaciones base vecinas se conservan en un conjunto independiente para su posterior análisis de handover. Además, se calculan variables como la distancia geográfica, la orientación relativa respecto al sector y la altura de los obstáculos obtenida mediante una API. Finalmente, se genera el conjunto aumentado mediante escenarios que incorporan variaciones de la configuración de la antena, complementando las mediciones reales obtenidas en campo.
- **La capa de procesamiento** comprende la preparación de los datos y el entrenamiento de los modelos de aprendizaje automático Random Forest y XGBoost. Para el procesamiento se utiliza una partición de 80 % para el

entrenamiento y 20 %. La división se realiza agrupando las muestras correspondientes a un mismo punto físico de medición, con el propósito de reducir el riesgo de fuga de información entre ambos conjuntos. El proceso se complementa con una validación cruzada agrupada.

- **La capa de evaluación** analiza el desempeño de los modelos mediante métricas estadísticas como el error absoluto medio (MAE), la raíz del error cuadrático medio (RMSE) y el coeficiente de determinación (R^2). Una vez seleccionados los modelos con mejor desempeño, estos se exportan en formato .pkl, lo que facilita su reutilización posterior.
- **La capa de visualización e interfaz** genera mapas interactivos para representar los resultados mediante herramientas como ipyleaflet y folium. Además, incorpora un panel de control desarrollado con ipywidgets, que permite al usuario interactuar con el sistema y consultar las predicciones de cobertura LTE.

3.3. Infraestructura del hardware

Para la adquisición de los datos utilizados en el entrenamiento y la evaluación de los modelos, se empleó un dispositivo móvil con capacidad de medición LTE y la aplicación especializada G-NetTrack Pro para el monitoreo de parámetros radioeléctricos. El dispositivo se configuró en modo LTE para evitar el cambio automático hacia tecnologías 2G o 3G. De esta manera, las mediciones correspondieron exclusivamente a la red LTE objeto de estudio.

Entre los principales parámetros registrados se encuentran el RSRP, el RSRQ, el SINR, el RSSI, el PCI, el EARFCN y las coordenadas geográficas de cada punto de medición. Esta información fue obtenida mediante el módem del dispositivo y complementada con los datos de posicionamiento proporcionados por su sistema GPS. Esto permitió asociar cada muestra con su ubicación geográfica y facilitar su posterior análisis y representación espacial.

3.3.1. Configuración de la estación base

La estación base objeto de estudio presenta una arquitectura sectorizada de tres sectores, con azimuts de 355°, 161° y 240°, una altura de antena aproximada de 30 m y un tilt eléctrico de 6°. Cada sector fue identificado mediante los valores de PCI registrados durante la campaña de medición, de modo que los PCI 87, 88 y 89 correspondieron a los tres sectores de la estación base analizada.

La frecuencia de operación se determinó a partir del parámetro EARFCN = 2050, correspondiente a la banda LTE 4 (Band 4). De acuerdo con la fórmula de conversión establecida en el estándar 3GPP TS 36.101 para la frecuencia de enlace descendente de esta banda, se tiene: $F_{DL} = 2110 + 0,1 \times (EARFCN - 1950)$. Por lo tanto, la frecuencia de enlace descendente correspondiente al EARFCN 2050 es de 2120 MHz. Este valor se utilizó como frecuencia de entrada para el modelo de propagación COST-231 Hata.

Es importante señalar que la frecuencia utilizada en el presente estudio (2120 MHz), así como las distancias registradas durante la campaña de medición, comprendidas entre 4,7 m y 520 m, se encuentran fuera del rango de validez formal del modelo COST-231 Hata, formulado originalmente para frecuencias de hasta 2000 MHz y distancias de 1 a 20 km [23]. Por esta razón, en el presente trabajo, el modelo no se emplea como una estimación físicamente precisa de la pérdida de trayectoria, sino como una referencia físico-empírica cuya salida se incorpora como una variable predictora adicional (cost231_path_loss) al vector de características utilizado por los modelos de aprendizaje automático. De esta manera, las predicciones de Random Forest y XGBoost se realizan a partir de datos reales obtenidos durante la campaña de medición y de las demás variables consideradas en el estudio, lo que permite que los modelos aprendan las relaciones entre estas características y los valores observados de RSRP. Así, el modelo COST-231 Hata se utiliza como una fuente de información físico-empírica para el proceso de ingeniería de características, y no como un modelo autónomo de predicción de cobertura.

3.3.2. Parámetros de la estación LTE analizada

La estación base analizada presenta una configuración sectorizada mediante tres antenas direccionales. Sus parámetros técnicos fueron determinados a partir de las mediciones realizadas durante la campaña de campo y contrastados con información de referencia obtenida mediante simulaciones en Atoll para estaciones LTE de CNT en el sector de La Libertad. La **Tabla 5** resume los principales parámetros considerados en el estudio.

Asimismo, la configuración radioeléctrica de referencia, correspondiente a variables como el azimut, el tilt, la altura y la potencia por sector, fue contrastada con la parametrización de estaciones LTE de CNT en la provincia de Santa Elena reportada en el trabajo de Muñoz Rodríguez [13].

El modelo de antena sectorial considerado es el Tongyu TDJH-182018DE-65F. De acuerdo con la hoja de datos del fabricante, este modelo opera en un rango de frecuencias de 1710 a 2170 MHz, utiliza polarización cruzada de $\pm 45^\circ$, presenta un ancho de haz horizontal de 65° , una ganancia máxima de 18 dBi y un valor típico de 17,8 dBi en el rango de 1920 a 2170 MHz. Además, dispone de un tilt eléctrico ajustable de 0° a 10° y una potencia máxima de entrada de 300 W [38].

Tabla 5. *Parámetros técnicos de la estación base LTE utilizada en el estudio.*

Parámetro	Valor
Tecnología	LTE
Banda de operación	Band 4 (AWS)
EARFCN	2050
Frecuencia DL aproximada	2120 MHz
Altura de antena	30 m
Tilt eléctrico	6°
Número de sectores	3
Azimut sector 1	355°
Azimut sector 2	161°
Azimut sector 3	240°
Ubicación aproximada BTS	La Libertad
Potencia estimada	43–49 dBm
Modelo de antena sectorial	Tongyu TDJH-182018DE-65F
Ganancia de antena	17,8 dBi (máx. 18 dBi)
Ancho de haz horizontal	65°
Potencia máxima de entrada	300 W

Nota. Elaboración propia con base en [13] y [38].

3.4. Área de estudio

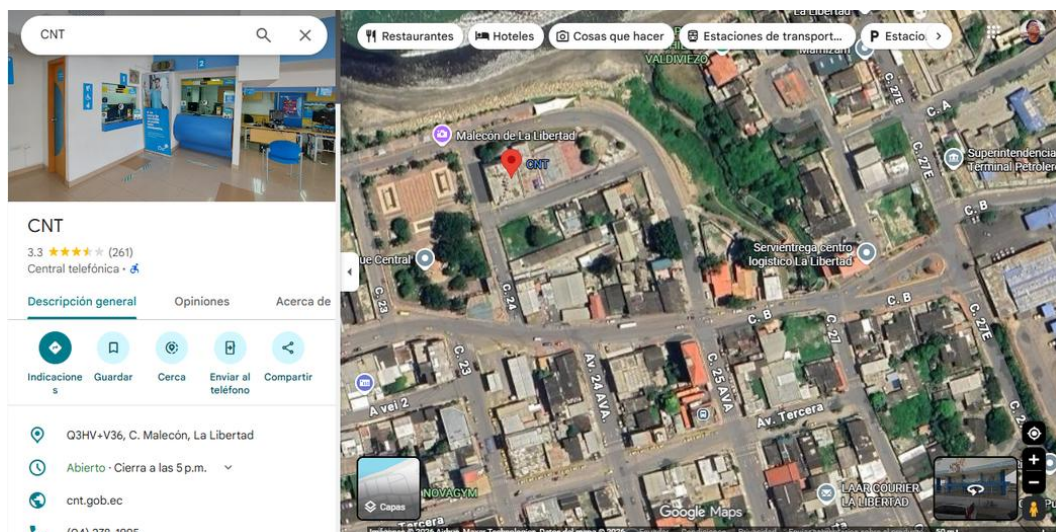
La investigación se centra en analizar y predecir la cobertura LTE dentro de un área geográfica específica del cantón La Libertad, considerando que la propagación de la señal depende tanto de los parámetros técnicos de la red como de las características físicas y geográficas del entorno.

Con el propósito de aproximar el análisis a las condiciones reales de propagación, se incorporaron variables relevantes para el comportamiento de la señal, entre ellas la distancia entre el usuario y la estación base, la frecuencia de operación, la altura de la antena y del receptor, así como la elevación del terreno obtenida mediante modelos digitales de elevación (DEM). La inclusión de estas variables permite considerar la influencia de las características del terreno sobre la propagación de la señal.

La investigación se desarrolló en el entorno urbano del cantón La Libertad, específicamente alrededor de la estación base de CNT seleccionada para la investigación. La **Figura 16** muestra su ubicación geográfica en Google Maps. Las **Figuras 17 y 18** presentan, respectivamente, vistas de la estación base y de la torre de telecomunicaciones desde la vía pública.

Finalmente, la **Figura 19** muestra un detalle de las antenas sectoriales de la estación LTE analizada.

Figura 16. Ubicación geográfica de la estación base de CNT en el cantón La Libertad.



Nota. Elaboración propia con base en [19].

Figura 17. Vista de la estación base de CNT y su torre de telecomunicaciones desde la vía pública.



Nota. Elaboración propia.

Figura 18. Vista frontal del edificio que aloja la estación base de CNT en el área de estudio.



Nota. Elaboración propia.

Figura 19. *Detalle de la torre y las antenas sectoriales de la estación base LTE analizada.*



Nota. Elaboración propia.

3.5. Recolección de las muestras y datos

La recolección de los datos se realizó mediante campañas de medición en campo orientadas a caracterizar la cobertura LTE. Para ello, se empleó un dispositivo móvil con la aplicación G-NetTrack Pro, que permitió registrar parámetros de la red y la ubicación geográfica de cada muestra durante los recorridos. De manera complementaria, se utilizó el analizador Narda SRM-3006 para contrastar los niveles de potencia registrados durante las mediciones.

Es importante precisar el origen de los parámetros utilizados en el estudio. Los indicadores de red y las coordenadas geográficas fueron obtenidos mediante G-NetTrack Pro, que accede a la información proporcionada por el módem del terminal y su sistema de posicionamiento. De manera complementaria, se empleó el analizador Narda SRM-3006 en un conjunto reducido de puntos dentro del área de estudio, con el propósito de verificar la presencia de la portadora en la banda analizada y contrastar cualitativamente el nivel de potencia observado. Por razones de disponibilidad y de traslado del equipo, esta caracterización espectral no se extendió a la totalidad de los recorridos, por lo que el

conjunto de datos utilizado para el entrenamiento y la evaluación de los modelos se construyó íntegramente a partir de los registros de G-NetTrack Pro.

Los equipos empleados durante la campaña no permitieron determinar directamente algunos parámetros correspondientes a la configuración radioeléctrica de la estación base, como el azimut, el tilt, la altura de la antena y la potencia de transmisión. Por esta razón, estos valores se tomaron de la parametrización de estaciones LTE de CNT reportada en el informe técnico de referencia [13]. En particular, se consideró un tilt eléctrico de 6°, valor utilizado como referencia para la configuración de la estación analizada, debido a que la instrumentación disponible no permitía medirlo. La **Tabla 6** detalla el origen de cada uno de los parámetros empleados en el estudio.

Tabla 6. Origen de los parámetros empleados en las mediciones.

Parámetro	Fuente	Observación
RSRP, RSRQ, SINR, RSSI, CQI	G-NetTrack Pro.	KPIs de red leídos del módem del terminal durante la campaña.
PCI, EARFCN, Band, Cell_ID	G-NetTrack Pro.	Identificación de celda y de portadora.
Latitud y longitud	G-NetTrack Pro.	Georreferenciación de cada muestra.
Nivel de potencia recibida medido	Narda SRM-3006.	Verificación de la presencia y del nivel de potencia de la portadora en la banda de 2120 MHz.
Azimut por sector (355°, 161°, 240°)	Informe técnico de referencia [13].	Configuración fija de la estación.
Tilt (6°)	Informe técnico de referencia [13].	No medible con los equipos disponibles; se tomó de la parametrización de CNT.
Altura de antena (30 m)	Informe técnico de referencia [13].	Configuración fija del sitio.
Potencia de transmisión (43 a 49 dBm)	Informe técnico de referencia [13].	Configuración fija del sitio.

Parámetro	Fuente	Observación
Ganancia (17,8 dBi) y ancho de haz horizontal (65°)	Hoja de datos Tongyu Communication [38].	Especificación del fabricante.

Nota. Elaboración propia con base en [13] y [38].

Durante las mediciones se registró el RSRP junto con otras variables técnicas y contextuales, entre ellas el RSRQ, el RSSI, el SINR, el CQI, el PCI, el EARFCN, la frecuencia de operación y la ubicación geográfica. A partir de estos registros se calcularon variables adicionales, como la distancia respecto de la estación base y la orientación relativa al sector. Asimismo, se incorporaron características del entorno, entre ellas la presencia de edificaciones y la elevación del terreno, con el propósito de representar el comportamiento de la señal considerado en el modelo.

Las rutas de medición se planificaron de manera que abarcaran diferentes condiciones de cobertura y variaciones en las condiciones de propagación. Los datos obtenidos fueron almacenados y posteriormente consolidados en un conjunto estructurado para su procesamiento y análisis en Python.

La **Tabla 7** presenta una muestra del conjunto de datos utilizado para el entrenamiento, mientras que la **Tabla 8** muestra una parte de las mediciones de campo obtenidas mediante G-NetTrack Pro.

Tabla 7. Muestra de los datos utilizados para el entrenamiento de los modelos predictivos.

rsrp	distancia_ km	dif_ azimuth	tilt	altura_ antena	num_ edificios	altura_max _edificio	humedad	potencia _mw	cost231_ path_loss	obs_ terreno_m	L_ant
-73	0,0047	144,98	6,0	30,0	13	41,5	70,0	39811,0	71,200	19,5	25,000
-70	0,0047	75,02	6,0	30,0	13	13,9	70,0	19953,0	71,200	19,4	15,985
-92	0,0104	11,13	6,0	30,0	11	21,5	70,0	19953,0	71,801	24,9	0,352
-73	0,1241	163,72	6,0	30,0	20	13,6	70,0	39811,0	109,725	28,0	25,000
-80	0,1321	117,26	6,0	30,0	4	22,5	70,0	19953,0	110,680	18,8	25,000
-84	0,1765	113,63	6,0	30,0	16	13,2	70,0	39811,0	115,121	26,7	25,000
-91	0,2209	107,67	6,0	30,0	23	3,9	70,0	19953,0	118,552	28,7	25,000
-89	0,4296	162,14	6,0	30,0	9	42,9	70,0	39811,0	128,724	28,4	25,000
-95	0,4358	127,71	6,0	30,0	20	22,8	70,0	19953,0	128,943	9,5	25,000
-103	0,5183	50,50	6,0	30,0	24	27,8	70,0	19953,0	131,595	3,2	7,243

Nota. Elaboración propia.

Tabla 8. Muestra de las mediciones de campo obtenidas mediante G-NetTrack Pro.

Time	Latitude	Longitude	RSRQ (dB)	SINR (dB)	RSSI (dBm)	CQI	DL_Throughput (Mbps)	UL_Throughput (Mbps)	PCI	EARFCN	Band	Cell_ID
02/06/2026 07:54	-2.220203	-80.907490	-8,0	3,0	-45,0	5	0,45	0,13	88	2050	4	51625
02/06/2026 07:54	-2.220203	-80.907490	-7,0	18,0	-43,0	11	10,48	3,15	89	2050	4	51625
20/05/2026 08:00	-2.220140	-80.907450	-10,0	9,0	-62,0	8	24,30	7,29	87	2050	4	51625
16/06/2026 07:38	-2.220355	-80.908570	-6,0	23,0	-47,0	13	5,42	1,63	88	2050	4	51625
25/05/2026 08:37	-2.220683	-80.906360	-8,0	12,0	-52,0	9	2,00	0,60	87	2050	4	51625
02/06/2026 07:58	-2.221555	-80.908340	-12,0	11,0	-52,0	8	1,62	0,49	88	2050	4	51625
25/05/2026 08:35	-2.221571	-80.905990	-8,0	8,0	-63,0	7	1,07	0,32	89	2050	4	51625
17/06/2026 08:22	-2.220761	-80.911290	-11,0	-5,0	-58,0	2	0,10	0,05	88	2050	4	51625
26/05/2026 07:28	-2.222351	-80.904160	-14,0	-3,0	-61,0	3	0,10	0,05	87	2050	4	51625
02/06/2026 08:06	-2.224830	-80.908230	-8,0	12,0	-75,0	9	14,03	4,21	89	2050	4	51625

Nota. Elaboración propia.

Las **Tablas 7 y 8** presentan muestras representativas de los conjuntos de datos utilizados durante el desarrollo de los modelos Random Forest y XGBoost. Los registros contienen variables relacionadas con el nivel y la calidad de la señal LTE, parámetros de configuración radioeléctrica, características geográficas y topográficas, condiciones ambientales, información de identificación de la red y métricas de desempeño.

3.5.1. Preparación de los datos

Una vez obtenidas las mediciones, se aplicó un proceso de preprocesamiento orientado a construir un conjunto de datos consistente para el entrenamiento de los modelos. En esta etapa se depuraron los registros vacíos o inválidos, se interpolaron las pérdidas breves de la señal GPS y se identificaron y trataron los valores cero considerados anómalos. Asimismo, se realizó una depuración por estación base y sector, conservando únicamente las muestras servidas por el eNodeB objetivo, identificado mediante su Cell_ID, y asignando cada registro al sector correspondiente según su PCI, sin realizar reasignaciones geométricas.

Las muestras provenientes de estaciones base vecinas no fueron descartadas, sino almacenadas en un conjunto independiente para su posterior utilización en el análisis de handover. Finalmente, se eliminaron los puntos cuya distancia respecto de la estación base superaba el radio de servicio considerado, debido a que podían estar asociados con errores de posicionamiento GPS o con registros ajenos al área de estudio.

Durante la campaña de medición, la configuración física de la estación base se mantuvo constante. Por esta razón, variables como el tilt, la altura de la antena y la potencia de transmisión no presentaban variación suficiente en las mediciones de campo para que el modelo aprendiera de manera directa su influencia sobre la señal. En consecuencia, el entrenamiento basado exclusivamente en las mediciones reales permitiría estimar el comportamiento de la red bajo la configuración observada durante la campaña, pero limitaría la evaluación de escenarios con diferentes configuraciones de la estación base.

Para incorporar sensibilidad frente a estos parámetros sin realizar una campaña de medición para cada configuración, se aplicó un procedimiento de aumento de datos basado en el modelo físico. A lo largo del documento, los registros obtenidos mediante este procedimiento se denominan escenarios aumentados. Para cada punto medido se calibró un balance de enlace que consideró la potencia por elemento de recurso, la

ganancia de la antena, la pérdida de trayectoria calculada mediante COST-231 Hata y los patrones de radiación horizontal y vertical. A partir de la diferencia entre la estimación física y la medición real se obtuvo un residuo por punto, utilizado para representar las variaciones no explicadas por el modelo de propagación y asociadas al entorno observado.

Con base en este procedimiento se generaron los escenarios aumentados mediante variaciones de los parámetros de configuración de la estación base. El tilt se consideró en un rango de 0° a 10° , la altura de la antena entre 25 m y 35 m y la potencia de transmisión entre 43 dBm y 49 dBm. Para cada configuración generada se recalcularon las variables radioeléctricas correspondientes.

Cada registro del conjunto aumentado incorpora una variable de origen que permite distinguir entre las muestras obtenidas directamente en campo y las generadas mediante simulación. El conjunto final está compuesto por 4 962 registros, de los cuales 2 822 corresponden a mediciones de campo y 2 140 a registros generados mediante el modelo físico.

Los escenarios aumentados fueron generados y seleccionados mediante un procedimiento de muestreo por configuración, de modo que las diferentes combinaciones consideradas estuvieran representadas en el conjunto resultante sin necesidad de replicar todas las configuraciones para cada punto de medición. Las mediciones reales constituyeron la base de calibración, mientras que los escenarios aumentados permitieron ampliar el rango de condiciones observadas por los modelos. La variable de origen permitió, además, diferenciar ambos grupos de datos durante el análisis de los resultados.

La humedad se mantuvo como una variable de contexto asociada a cada jornada de medición y no se incorporó como parámetro de variación de los escenarios aumentados. Esta decisión se fundamentó en la baja influencia esperada de la absorción atmosférica en la frecuencia de operación considerada, de acuerdo con la referencia utilizada [25].

Finalmente, el conjunto de datos se dividió en 80 % para entrenamiento y 20 % para prueba mediante GroupShuffleSplit. La agrupación se realizó utilizando un identificador correspondiente a cada punto físico de medición, de manera que todas las observaciones asociadas a un mismo punto, incluidos sus escenarios aumentados, permanecieran en un único conjunto. Este procedimiento evita la fuga de información espacial y permite evaluar la capacidad de generalización del modelo sobre puntos no utilizados durante el

entrenamiento. Los resultados se analizaron de forma diferenciada para las mediciones obtenidas en campo y para los escenarios aumentados.

3.6. Definición de variables del sistema

La variable objetivo principal del modelo es el RSRP, que representa la potencia recibida de las señales de referencia LTE y constituye uno de los principales indicadores utilizados para evaluar el nivel de cobertura de una red móvil. De manera complementaria, se entrenaron modelos independientes para estimar otras cinco variables objetivo (RSRQ, SINR, CQI, DL Throughput y UL Throughput), de modo que el sistema entrega en total seis parámetros estimados por punto. Estas seis variables constituyen las salidas del sistema y en ningún caso forman parte del vector de características de entrada.

3.6.1. Variables predictoras (entradas del modelo)

Las variables independientes —es decir, las que conforman el vector de características de entrada de los modelos— se seleccionaron considerando su relación con la propagación de la señal y las características del entorno. La **Tabla 9** presenta estas variables predictoras, mientras que la **Tabla 10** recoge las variables objetivo y las variables de contexto e identificación registradas durante la campaña, que no se utilizan como entradas de los modelos.

Tabla 9. Variables predictoras (entradas) utilizadas en los modelos predictivos de cobertura LTE.

Variable	Significado	Tipo
distancia_km	Distancia entre la antena y el usuario.	Geográfica
dif_azimuth	Diferencia angular entre la dirección de la antena y el usuario.	Radioeléctrica
tilt	Inclinación vertical de la antena.	Configuración RF
altura_antena	Altura de la estación base.	Infraestructura
num_edificios	Cantidad de edificios entre antena y usuario.	Obstáculos urbanos
altura_max_edificio	Altura del edificio más alto en el trayecto.	Obstáculos urbanos
humedad	Humedad ambiental.	Climática
potencia_mw	Potencia transmitida en miliwatts.	RF

cost231_path_loss	Pérdida de propagación calculada con COST-231 Hata.	Modelo empírico
obs_terreno_m	Obstrucción del terreno/elevación.	Topografía
L_ant	Pérdidas de antena o feeder.	Sistema RF

Nota. Elaboración propia.

3.6.2. Variables objetivo y variables de contexto

La Tabla 10 agrupa las variables que se registraron durante la campaña pero que no forman parte del vector de características de entrada de los modelos. Estas se dividen en dos grupos. El primero corresponde a las variables objetivo, es decir, las que el sistema estima: RSRQ, SINR, CQI, DL Throughput y UL Throughput, además del RSRP ya indicado. El segundo corresponde a las variables de contexto e identificación —Time, Latitude, Longitude, RSSI, PCI, EARFCN, Band y Cell_ID—, empleadas para georreferenciar cada muestra, asignarla al sector correspondiente y verificar que pertenezca a la estación base analizada. El RSSI se registró como referencia de nivel, pero no se utilizó como variable objetivo ni como variable predictora.

Para cada una de las seis variables objetivo se entrenó un modelo independiente sobre el mismo conjunto de datos y con el mismo vector de características descrito en la sección anterior. De esta manera, el RSRQ, el SINR, el CQI, el DL Throughput y el UL Throughput no se derivan del valor estimado de RSRP, sino que se predicen directamente a partir de las condiciones del enlace. El RSRP se considera la variable principal del estudio, debido a que constituye el indicador utilizado para determinar el nivel de cobertura y para seleccionar el sector dominante; las cinco variables restantes se emplean como indicadores complementarios de calidad y rendimiento del enlace.

Tabla 10. Variables objetivo y variables de contexto e identificación registradas durante la campaña.

Variable	Significado	Tipo
Time	Marca temporal de la medición LTE.	Temporal
Latitude	Latitud GPS del usuario/equipo móvil.	Geoespacial
Longitude	Longitud GPS del usuario/equipo móvil.	Geoespacial
RSRQ (dB)	Calidad de señal LTE.	RF / Calidad

SINR (dB)	Relación señal/interferencia/ruido.	RF / Calidad
RSSI (dBm)	Potencia total recibida (señal + ruido + interferencia).	RF
CQI	Indicador de calidad de canal LTE.	Recursos radioeléctricos
DL_Throughput (Mbps)	Velocidad de descarga estimada.	Rendimiento LTE
UL_Throughput (Mbps)	Velocidad de subida estimada.	Rendimiento LTE
PCI	Identidad física de celda LTE (PCI).	Identificación LTE
EARFCN	Número de canal LTE (frecuencia LTE).	Frecuencia LTE
Band	Banda LTE utilizada.	Banda RF
Cell_ID	Identificador único de la celda.	Identificación LTE

Nota. Elaboración propia.

Es importante identificar el origen de las variables utilizadas en el modelo. Los parámetros de radiofrecuencia y la información de identificación de la red fueron obtenidos a partir de las mediciones realizadas en campo mediante G-NetTrack Pro y complementados con la caracterización espectral realizada mediante el Narda SRM-3006 en un conjunto reducido de puntos. Los valores de tilt, altura de antena y potencia de transmisión se tomaron de la configuración de referencia utilizada para la estación y posteriormente se variaron en los escenarios aumentados. La humedad se mantuvo como variable de contexto asociada a cada jornada de medición. Las variables adicionales derivadas del modelo físico y del procesamiento geoespacial fueron calculadas durante la etapa de preparación de los datos. Asimismo, el conjunto aumentado incorpora una variable de origen que permite identificar si cada registro corresponde a una medición real o a un escenario aumentado.

3.6.3. Entornos de desarrollo y librerías

El sistema se desarrolló en Python 3.14.6, seleccionado por la disponibilidad de herramientas para el procesamiento de datos, el aprendizaje automático, el análisis geoespacial y la visualización de resultados.

La **Tabla 11** presenta las principales librerías utilizadas y la función que desempeñan dentro del sistema, mientras que la **Tabla 12** detalla los submódulos empleados.

Tabla 11. *Librerías utilizadas en el desarrollo del sistema de predicción de cobertura LTE.*

Librería	Descripción	Función dentro del sistema
Pandas	Biblioteca para análisis y manipulación de datos.	Construcción y almacenamiento del dataset LTE mediante DataFrames.
NumPy	Librería para cálculo matemático y científico.	Generación de datos aleatorios y operaciones matriciales.
scikit-learn	Biblioteca de aprendizaje automático y análisis de datos en Python.	División train/test y métricas de evaluación como el MAE y el R^2 .
XGBoost	Biblioteca de aprendizaje automático basada en potenciación de gradiente (gradient boosting).	Entrenamiento del modelo de potenciación de gradiente para la predicción de cobertura LTE.
Joblib	Herramienta de serialización eficiente.	Guardado y carga de modelos entrenados y sus configuraciones.
OSMnx	Librería para descargar datos de OpenStreetMap.	Obtención de edificios y geometrías urbanas.
GeoPandas	Extensión geoespacial de Pandas.	Manipulación de capas y geometrías espaciales.
Shapely	Biblioteca de geometría computacional.	Cálculo de intersecciones y trayectorias de propagación RF.
Geopy	Librería para cálculos geográficos.	Cálculo de distancias reales entre la antena y el usuario.
Rasterio	Librería para lectura de datos raster geoespaciales.	Obtención de elevación del terreno desde archivos DEM/GeoTIFF.
ipyleaflet	Librería para mapas interactivos en JupyterLab.	Visualización dinámica de cobertura LTE y mapas RF.

ipywidgets	Biblioteca de widgets interactivos.	Creación de sliders, botones y controles interactivos en la simulación.
Folium	Biblioteca para insertar mapas interactivos basada en Leaflet.	Visualización geográfica de cobertura y puntos LTE.
IPython	Entorno interactivo para notebooks JupyterLab.	Visualización de widgets y mapas, y actualización dinámica de interfaces.
Math	Módulo matemático estándar de Python.	Operaciones trigonométricas y cálculos geométricos RF.
Datetime	Librería estándar para manejo temporal.	Generación de marcas temporales para los escenarios aumentados.
Warnings	Módulo estándar de control de advertencias.	Supresión de mensajes innecesarios durante ejecución.
OS	Módulo estándar para interacción con el sistema operativo.	Gestión de rutas, archivos y directorios del sistema.
Matplotlib	Biblioteca para generación de gráficos estáticos en Python.	Elaboración de gráficos estadísticos y comparativos de desempeño de los modelos.

Nota. Elaboración propia.

Tabla 12. Submódulos utilizados en el desarrollo del sistema de predicción de cobertura LTE.

Submódulo	Función
sklearn.model_selection.GroupShuffleSplit	División del dataset entrenamiento/prueba.
sklearn.metrics.mean_absolute_error	Cálculo del error absoluto medio.
sklearn.metrics.r2_score	Cálculo del coeficiente R ² .
rasterio.transform.rowcol	Conversión de coordenadas geográficas a índices de fila y columna del raster.
shapely.geometry.LineString	Construcción de trayectorias geométricas.

folium.plugins.MousePosition	Mostrar coordenadas sobre mapa.
ipyleaflet.Map	Creación de mapas interactivos.
ipyleaflet.Marker	Marcadores geográficos.
ipyleaflet.Heatmap	Mapas de calor RF.
IPython.display.display	Visualización de elementos dinámicos.
IPython.display.clear_output	Actualización en tiempo real.

Nota. Elaboración propia.

3.7. Descripción general del flujo metodológico

La implementación se desarrolló en Python mediante bibliotecas especializadas en procesamiento de datos y aprendizaje automático. Los modelos XGBoost y Random Forest comparten un flujo general que comprende la carga y validación de los datos, la preparación del conjunto de entrenamiento y prueba, el entrenamiento, la evaluación y el almacenamiento de los modelos. La principal diferencia entre ambos radica en el procedimiento mediante el cual se construye el conjunto de árboles utilizados para generar las predicciones.

3.7.1. Metodología de implementación de los modelos XGBoost y Random Forest para predicción de cobertura LTE

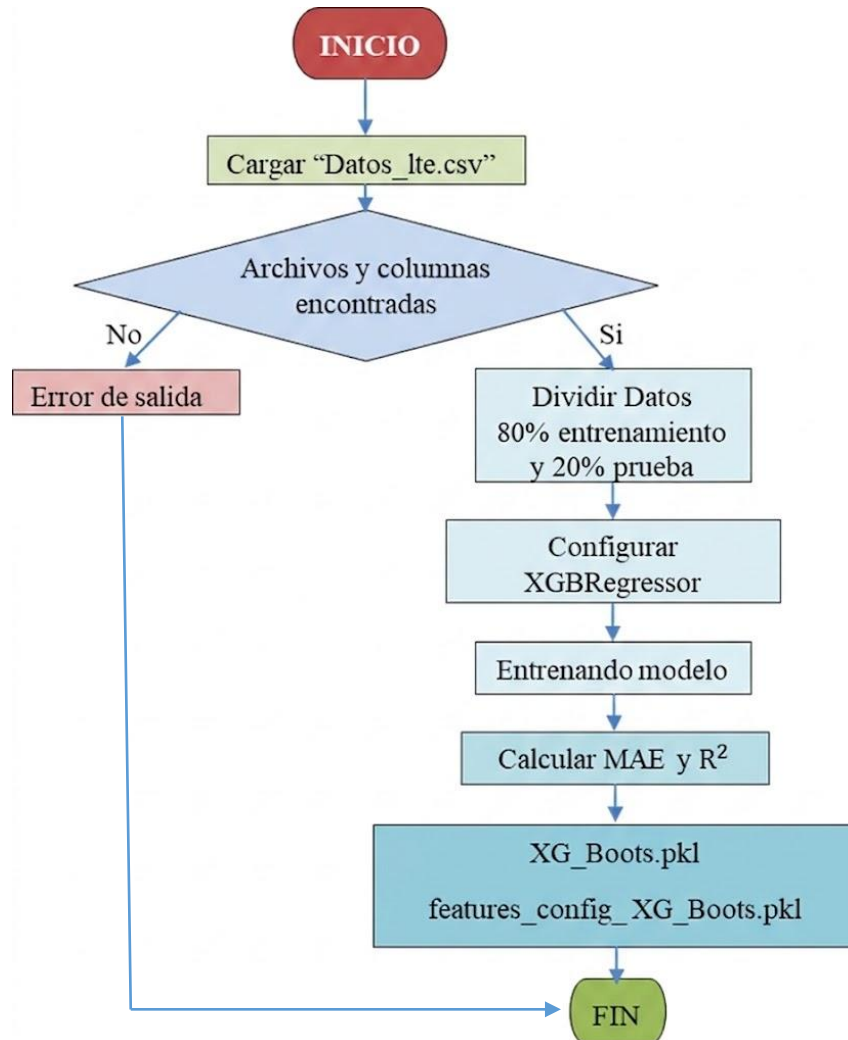
El proceso se inicia mediante las funciones principales `entrenar_xgboost ()` y `entrenar_random_forest ()`, encargadas de coordinar las diferentes etapas del entrenamiento y evaluación de cada modelo.

La metodología se divide en las siguientes fases:

1. Inicialización del proceso.
2. Carga del conjunto de datos.
3. Validación estructural del dataset.
4. Verificación de parámetros constantes.
5. Separación de variables predictoras y variable objetivo.
6. División de datos en conjuntos de entrenamiento y prueba.
7. Entrenamiento del modelo XGBoost y Random Forest.
8. Evaluación estadística de su desempeño.
9. Almacenamiento del modelo entrenado.
10. Fin.

Diagrama de flujo del entrenamiento del modelo de aprendizaje automático XGBoost.

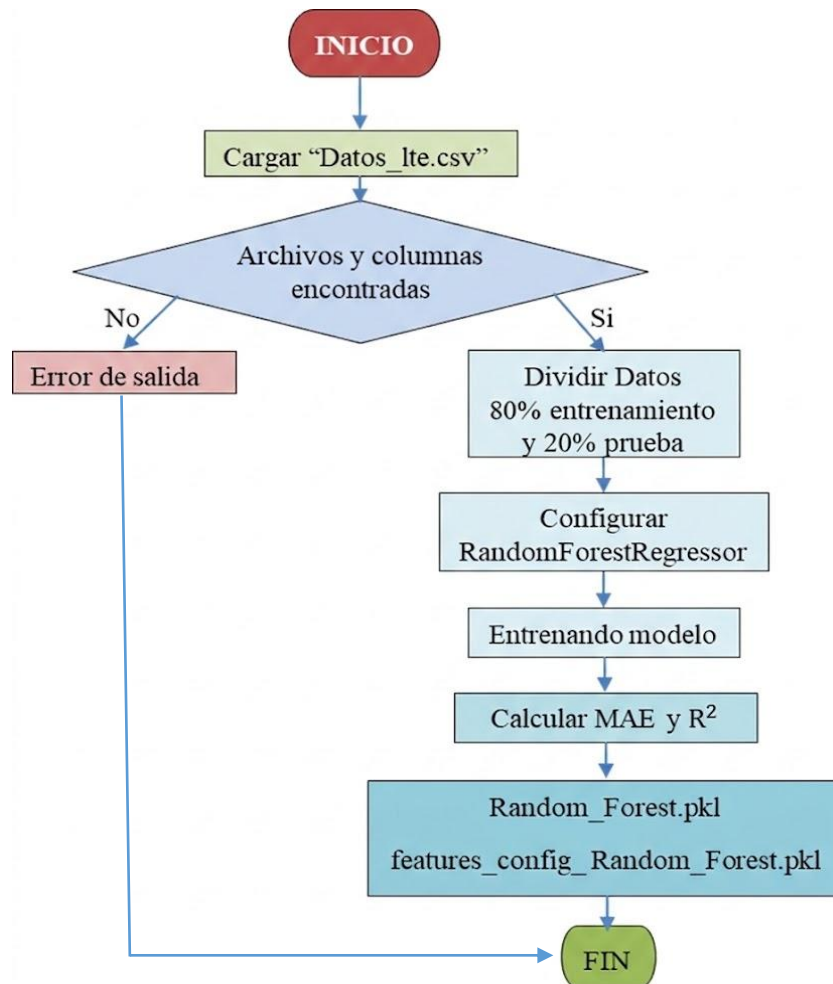
Figura 20. Flujograma del proceso de entrenamiento del modelo XGBoost para la predicción de cobertura LTE.



Nota. Elaboración propia.

Diagrama de flujo del entrenamiento del modelo de aprendizaje automático Random Forest.

Figura 21. Flujograma del proceso de entrenamiento del modelo Random Forest para la predicción de cobertura LTE.



Nota. Elaboración propia.

La **Figura 20** presenta el proceso de entrenamiento del modelo XGBoost, mientras que la **Figura 21** muestra el proceso correspondiente al modelo Random Forest. Ambos procedimientos comparten las mismas etapas generales: la carga y validación del conjunto de datos LTE; la separación de las muestras entre los conjuntos de entrenamiento y prueba; el entrenamiento del modelo; la evaluación mediante métricas como el MAE y R^2 ; y el almacenamiento del modelo entrenado junto con la configuración de las variables utilizadas. Estos archivos permiten reutilizar posteriormente los modelos para la predicción y visualización de la cobertura LTE.

El proceso de entrenamiento se desarrolla mediante las siguientes fases:

- **La inicialización del entrenamiento** comienza con la ejecución de las funciones `entrenar_xgboost ()` y `entrenar_random_forest ()`, encargadas de coordinar las diferentes etapas de entrenamiento y prueba. Durante esta fase, el sistema muestra información relacionada con el modelo seleccionado, el conjunto de datos, las variables predictoras y la variable objetivo, permitiendo comprobar la configuración antes de iniciar el entrenamiento.
- **La carga del conjunto de datos** consiste en la lectura del archivo CSV mediante la biblioteca Pandas y la función `pd.read_csv(filename)`. El conjunto de datos contiene registros de cobertura LTE y variables relacionadas con las condiciones radioeléctricas, geográficas y del entorno. Entre las variables consideradas se encuentran la distancia entre el transmisor y el receptor, la elevación del terreno, la altura de la antena, el tilt eléctrico, las coordenadas geográficas, la potencia recibida y la pérdida de propagación.
- **La verificación de la existencia del archivo** consiste en comprobar, una vez indicada la ruta del conjunto de datos, que el archivo esté disponible. Si el archivo se encuentra disponible, el proceso continúa con la validación de las columnas requeridas. En caso contrario, se genera un mensaje de advertencia y el entrenamiento se detiene.
- **La validación de las columnas requeridas** consiste en comprobar que el conjunto de datos contenga todas las variables necesarias para el entrenamiento. Estas corresponden a `FEATURES_RF`, `FEATURES_XGB` y `TARGET`.

Donde:

- **FEATURES_RF:** conjunto de variables de entrada utilizadas por el modelo Random Forest.
- **FEATURES_XGB:** conjunto de variables de entrada utilizadas por el modelo XGBoost.
- **TARGET:** variable objetivo que debe ser estimada por el modelo.

Código 1. Lectura y validación del conjunto de datos

```
data = pd.read_csv(filename)
missing_cols = [c for c in FEATURES_XGB + [TARGET]
                 if c not in data.columns]
if missing_cols:
```

```
print(f'Columnas faltantes: {missing_cols}')
```

```
return
```

El código realiza la lectura del archivo CSV y verifica que el conjunto de datos contenga las variables requeridas para el entrenamiento del modelo.

La separación de variables predictoras y objetivo se realiza una vez validado el conjunto de datos y consiste en distinguir las variables de entrada de la variable que se desea predecir, se representa de la siguiente manera:

$$X = \text{data}[\text{FEATURES_XGB}]$$

Estas variables contienen la información utilizada por el modelo para generar las predicciones. Matemáticamente, pueden representarse como:

$$X = \{x_1, x_2, x_3, \dots, x_n\}$$

***Ecuación 7.** Representación de la variable predictora en forma matemática.*

Donde:

- **X**: vector de variables predictoras (features), que alimentan al modelo.
- **x1, x2, x3, ..., xn**: cada una de las variables predictoras (distancia, diferencia de azimut, tilt, altura de antena, pérdida COST-231 Hata, entre otras).
- **n**: número total de variables predictoras del modelo.

La variable objetivo se representa mediante $y = \text{data}[\text{TARGET}]$. Corresponde al parámetro que el modelo debe estimar. En el presente estudio la variable objetivo es el RSRP.

La división del dataset consiste en repartir el conjunto total en 80 % para entrenamiento y 20 % para prueba, agrupando las observaciones según el punto físico de medición. Matemáticamente se representa de la siguiente manera:

$$D = D_{\text{train}} + D_{\text{test}}$$

***Ecuación 8.** Representación de la división del dataset.*

Donde:

- **D_{train}**: subconjunto de entrenamiento, correspondiente al 80 % de los datos.
- **D_{test}**: subconjunto de prueba, correspondiente al 20 % de los datos.
- **D**: conjunto total de datos disponible, compuesto por 4 962 registros.

La división tuvo como propósito evaluar la capacidad de generalización de los modelos sobre datos no utilizados durante el entrenamiento. Debido a que el conjunto aumentado contenía varias observaciones asociadas a un mismo punto físico (incluidos

sus escenarios aumentados), la partición se realizó agrupando las observaciones por latitud, longitud y sector. De esta manera, todos los registros correspondientes a una misma ubicación permanecieron en un único conjunto, lo que permitió evitar la fuga de información entre los conjuntos de entrenamiento y prueba, reducir el sesgo en la evaluación y prevenir la obtención de métricas artificialmente elevadas.

Código 2. División de muestras para entrenamiento y validación

```
grupo = lat.astype(str) + "_" + lon.astype(str) + "_" + sector.astype(str)
gss = GroupShuffleSplit(
    n_splits=1, test_size=0.20,
    random_state=42
)
idx_train, idx_test = next(
    gss.split(X, y, groups=grupo)
)
```

El conjunto de datos se divide mediante GroupShuffleSplit, perteneciente a la biblioteca scikit-learn. Previamente, se construye un identificador de grupo a partir de la latitud, longitud y sector. Así, las mediciones repetidas de un mismo punto y sus escenarios aumentados permanecen en el mismo subconjunto. Este procedimiento permite evaluar el desempeño sobre puntos no observados durante el entrenamiento y evita la fuga de información espacial.

El entrenamiento del modelo Random Forest se lleva a cabo mediante RandomForestRegressor(), algoritmo de aprendizaje supervisado basado en un conjunto de árboles de decisión.

- **El algoritmo funciona** construyendo múltiples árboles de manera independiente, utilizando diferentes muestras y subconjuntos de variables. Para una tarea de regresión, la predicción final se obtiene mediante el promedio de las predicciones individuales de los árboles [39].

La predicción puede expresarse de la siguiente manera:

$$\hat{y}_i = \frac{1}{N} \sum_{i=1}^N T_i(x)$$

Ecuación 9. Predicción del modelo Random Forest.

Donde:

- **N**: representa el número total de árboles de decisión.
- **T_i(x)**: es la predicción del árbol (i).
- **ŷ_i**: es la predicción final.

Los parámetros utilizados para configurar el modelo fueron los siguientes:

- **n_estimators = 200**: define el número de árboles en el bosque aleatorio del modelo.
- **max_depth = 15**: define la complejidad máxima de cada árbol.
- **min_samples_leaf = 2**: muestras mínimas por hoja; este parámetro contribuye a limitar la complejidad del modelo y puede favorecer su capacidad de generalización.

Código 3. Configuración del modelo Random Forest

```
model_rf = RandomForestRegressor(  
    n_estimators=200,  
    max_depth=15,  
    min_samples_leaf=2,  
    random_state=42  
)  
model_rf.fit(X_train, y_train)
```

El entrenamiento del modelo XGBoost se lleva a cabo mediante `XGBRegressor()`, perteneciente a la familia de algoritmos de gradient boosting basados en árboles de decisión.

- **El funcionamiento del algoritmo XGBoost**, a diferencia de Random Forest, consiste en construir los árboles de manera secuencial. Cada nuevo árbol se incorpora para reducir los errores generados por los modelos anteriores, permitiendo mejorar progresivamente la predicción [40].

Una representación simplificada del modelo puede expresarse como:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i)$$

Ecuación 10. Predicción del modelo XGBoost.

Donde:

- **f_k**: representa cada árbol de decisión.

- **k**: es el número total de árboles.
- **x_i**: es el vector de variables predictoras de la muestra i.
- **y_i**: predicción final del modelo para la muestra i, obtenida como la suma de las salidas de los K árboles.
- **K**: número total de árboles del ensamble.

Los **parámetros utilizados** para configurar el modelo XGBoost fueron los siguientes:

- **n_estimators = 300**: define la cantidad de árboles generados durante el entrenamiento.
- **max_depth = 10**: define la complejidad del modelo.
- **learning_rate = 0.05**: regula cuánto aporta cada árbol nuevo a la solución final.
- **subsample = 0.8**: reduce el sobreajuste haciendo uso de subconjuntos de datos.
- **colsample_bytree = 0.8**: selecciona aleatoriamente variables de cada árbol.
- **random_state = 42**: fija la semilla de aleatoriedad del algoritmo para asegurar la reproducibilidad de los resultados entre distintas ejecuciones.

Código 4. Configuración del modelo XGBoost

```

model_xgb.fit(X_train, y_train)
model_xgb = XGBRegressor(
n_estimators=300,
max_depth=10,
learning_rate=0.05,
subsample=0.8,
colsample_bytree=0.8,
random_state=42
)
model_xgb.fit(X_train, y_train)

```

La **evaluación del modelo** se lleva a cabo, una vez finalizado su entrenamiento, mediante métricas estadísticas de regresión.

- **El error absoluto medio (MAE)** cuantifica el error absoluto promedio entre los valores observados y los valores predichos por el modelo. Una representación simplificada del modelo puede expresarse como:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Valores pequeños indican mayor precisión.

- **El coeficiente de determinación (R^2)** permite determinar la proporción de la variabilidad de la variable objetivo explicada por el modelo. Una representación simplificada del modelo puede expresarse como:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Un valor de R^2 cercano a 1 indica que el modelo explica una mayor proporción de la variabilidad observada. Esta métrica debe interpretarse conjuntamente con el MAE para obtener una evaluación más completa del desempeño.

Una vez finalizada la evaluación, se realiza el **almacenamiento de los modelos** entrenados mediante `joblib.dump()`. Los archivos generados permiten conservar los objetos entrenados y reutilizarlos posteriormente sin repetir el proceso de aprendizaje. Dependiendo de la configuración utilizada, estos archivos pueden contener el modelo entrenado y sus parámetros internos. Entre los elementos almacenados se encuentran:

- El modelo entrenado.
- Los parámetros del modelo.
- La estructura necesaria para realizar posteriormente las predicciones.

Código 5. Evaluación y almacenamiento del modelo

```
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
joblib.dump(model_xgb, "XG_Boost.pkl")
```

El código calcula las métricas de desempeño a partir de las predicciones obtenidas sobre el conjunto de prueba y posteriormente almacena el modelo XGBoost entrenado para su utilización en la etapa de predicción.

Una vez completadas las etapas de entrenamiento, evaluación y almacenamiento, el sistema informa la finalización del proceso y deja el modelo disponible para su posterior utilización en tareas de inferencia y predicción de cobertura LTE.

3.7.2. Metodología de la implementación del sistema de predicción de cobertura

Esta sección describe la metodología utilizada para desarrollar el sistema interactivo de predicción y visualización de la cobertura LTE en el área de estudio. El sistema integra módulos de procesamiento radioeléctrico, análisis geoespacial, aprendizaje automático y

visualización cartográfica, que trabajan de manera articulada para generar los resultados de cobertura. El flujo completo de funcionamiento del sistema se presenta en el Anexo D correspondiente (ver la **Figura 62**).

Entre las principales funciones implementadas se encuentran:

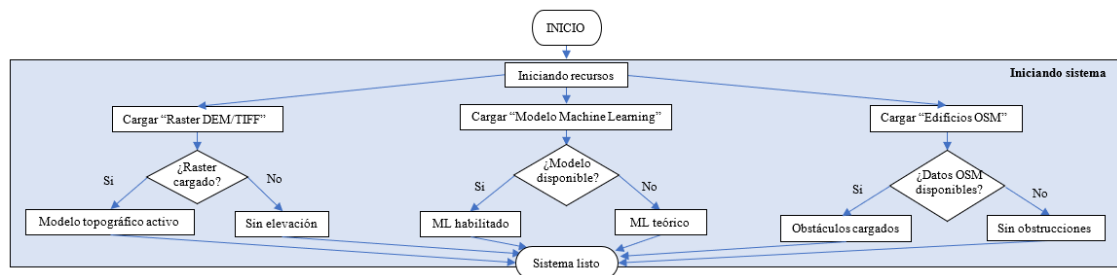
1. Calcular niveles de señal LTE.
2. Estimar las pérdidas de propagación.
3. Evaluar obstrucciones producidas por edificaciones y características del terreno.
4. Visualizar zonas de cobertura.
5. Estimar parámetros de calidad de la señal LTE.
6. Representar perfiles topográficos y zona de Fresnel.

La arquitectura del sistema se divide en varios módulos funcionales encargados de procesar la información geográfica, radioeléctrica y predictiva necesaria para la estimación de cobertura LTE.

Inicialización del sistema

La **Figura 22** muestra la etapa de inicialización del sistema, durante la cual se cargan los recursos necesarios para ejecutar los procesos radioeléctricos y geoespaciales. En esta fase se incorporan el Modelo Digital de Elevación (DEM), el modelo de aprendizaje automático previamente entrenado y la información geográfica obtenida de OpenStreetMap. Estos elementos constituyen los insumos utilizados en las etapas posteriores de análisis y predicción de cobertura LTE.

Figura 22. Proceso de inicialización del sistema.



Nota. Elaboración propia.

- **El procesamiento de elevación** utiliza archivos raster en formato GeoTIFF para obtener información georreferenciada sobre la elevación del terreno. La carga del archivo se realiza mediante la función `rasterio.open(path)`. Esta información se utiliza en diferentes procesos del sistema, entre ellos:

- La generación del perfil topográfico.
- El cálculo de la línea de vista (LOS).
- El cálculo de la zona de Fresnel.
- La identificación de posibles obstrucciones del terreno.

La elevación correspondiente a cada coordenada geográfica se obtiene mediante la función `elevacion(lat, lon)`. A partir de esta función, las coordenadas geográficas se transforman en índices de fila y columna del raster mediante la función `rowcol(transform, lon, lat)`. Posteriormente, el sistema recupera el valor de elevación correspondiente al punto analizado. Este procedimiento permite obtener el perfil del terreno a lo largo del trayecto entre la estación base y el punto de interés.

Código 6. Carga del modelo digital de elevación

```
dem = rasterio.open(dem_path)
def elevacion(lat, lon):
    row, col = rowcol(dem.transform, lon, lat)
    return dem.read(1)[row, col]
```

El código abre el archivo GeoTIFF y define la función `elevacion()`, que recibe las coordenadas geográficas de un punto, las transforma en índices del raster y devuelve el valor correspondiente a la elevación del terreno.

- **La carga del modelo de aprendizaje automático** consiste en recuperar el modelo previamente entrenado mediante `joblib.load(archivo.pkl)`. En la etapa de inferencia se utiliza el modelo predictivo almacenado para estimar el RSRP, definido como la variable objetivo principal del sistema, a partir del vector de características cuya construcción se detalla en la sección "Construcción del vector de características".

La predicción puede representarse mediante la siguiente relación:

$$\text{RSRP} = f(d, \Delta\theta, \text{tilt}, h_b, N_{\text{edif}}, H_{\text{edif}}, H, P_t, \text{PL}, O_{\text{terr}}, L_{\text{ant}})$$

Donde:

- **d**: distancia entre la estación base y el punto de interés.
- **$\Delta\theta$** : diferencia angular entre el azimut del sector y la dirección del punto de interés.
- **tilt**: inclinación eléctrica de la antena.
- **h_b** : altura de la antena de la estación base.
- **N_{edif}** : número de edificaciones presentes en el trayecto.

- H_{edif} : altura máxima de la edificación considerada en el trayecto, expresada en metros.
- H : humedad ambiental.
- P_t : potencia de transmisión.
- PL : pérdida de trayectoria calculada mediante el modelo de propagación.
- O_{terr} : obstrucción del terreno a lo largo del trayecto, expresada en metros.
- L_{ant} : pérdidas asociadas con el sistema de antena.

Código 7. Carga del modelo predictivo

```
modelo = joblib.load("modelo_xgboost.pkl")
```

El código carga el modelo previamente entrenado y lo deja disponible para realizar predicciones durante la etapa de inferencia.

- **El procesamiento de edificios OSM** obtiene información de las edificaciones presentes en el entorno de estudio mediante OpenStreetMap, utilizando herramientas de la biblioteca OSMnx. Esta información se emplea para identificar posibles obstrucciones en el trayecto entre la estación base y el punto de interés, así como para calcular variables asociadas con las características del entorno urbano.

Las edificaciones obtenidas se utilizan principalmente para:

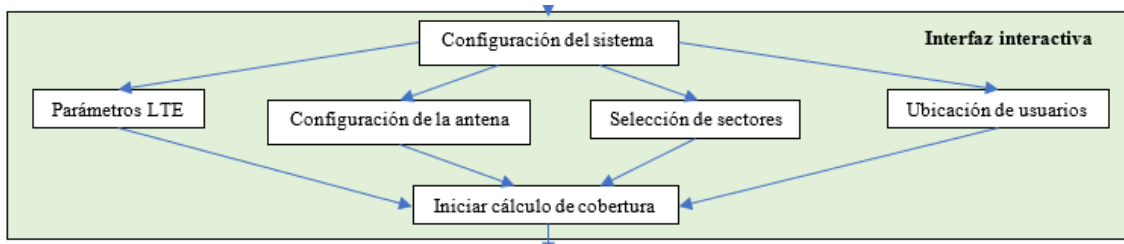
- Determinar posibles obstrucciones en la línea de vista (LOS).
- Incorporar información sobre obstáculos urbanos en el análisis del trayecto.
- Generar perfiles del enlace entre la estación base y el punto de interés.

Cuando una edificación no dispone de información sobre su altura, el sistema utiliza un procedimiento de estimación basado en características disponibles de la construcción. De esta manera, se obtiene una aproximación de la altura necesaria para el análisis de los obstáculos presentes en el entorno urbano.

Configuración e interacción del usuario

La **Figura 23** presenta el módulo de configuración del sistema, encargado de gestionar los parámetros utilizados durante la simulación y predicción de cobertura LTE.

Figura 23. Módulo de configuración del sistema.



Nota. Elaboración propia.

Desde esta interfaz, el usuario puede modificar diferentes parámetros asociados con la estación base, las antenas y el entorno de análisis. Entre ellos se encuentran:

- La frecuencia de operación de la portadora.
- La potencia de transmisión.
- La altura de la antena.
- El azimut de los sectores.
- El tilt eléctrico.
- La activación o desactivación de los sectores.
- La posición geográfica de la antena y el usuario.

Las modificaciones realizadas en la interfaz actualizan los parámetros utilizados por el sistema y permiten recalcular las variables correspondientes para generar una nueva estimación de cobertura.

- **La configuración global del sistema LTE** comienza con la definición de los parámetros generales asociados con la estación base LTE, entre ellos ANT_LAT y ANT_LON, que representan las coordenadas geográficas de la antena transmisora. También se establecen parámetros relacionados con las características de la antena, de la siguiente manera:

$$\text{GANANCIA_DBI} = 17.8$$

$$\text{ANCHO_HAZ} = 65$$

$$\text{L_ANT_MAX} = 25$$

Donde:

- **GANANCIA_DBI**: representa la ganancia de la antena, expresada en dBi.
- **ANCHO_HAZ**: corresponde al ancho de haz horizontal del lóbulo principal, expresado en grados.

- **L_ANT_MAX**: representa el límite de pérdida de radiación establecido en la implementación del sistema.

La ganancia de 17,8 dBi y el ancho de haz horizontal de 65° corresponden a las características consideradas para la antena sectorial Tongyu TDJH-182018DE-65F. El fabricante especifica un rango de frecuencia de operación de 1710 a 2170 MHz, un ancho de haz horizontal de 65°, un tilt eléctrico ajustable de 0° a 10° y una potencia máxima de entrada de 300 W, de acuerdo con la información técnica utilizada en el estudio [38].

Código 8. Configuración de parámetros LTE

```
tilt_slider = FloatSlider(value=6)
potencia_slider = FloatSlider(value=43)
altura_slider = FloatSlider(value=30)
tilt_slider.observe(actualizar)
```

Los controles interactivos permiten modificar los parámetros de configuración utilizados en los escenarios de simulación y ejecutar automáticamente la actualización de los resultados.

- **La sectorización LTE del sistema** se basa en una arquitectura celular dividida en tres sectores. Cada sector dispone de parámetros propios de configuración, entre ellos el azimuth, el tilt eléctrico, la potencia de transmisión y el estado de activación de los sectores. Estos parámetros pueden modificarse desde la interfaz gráfica para evaluar diferentes escenarios de cobertura.

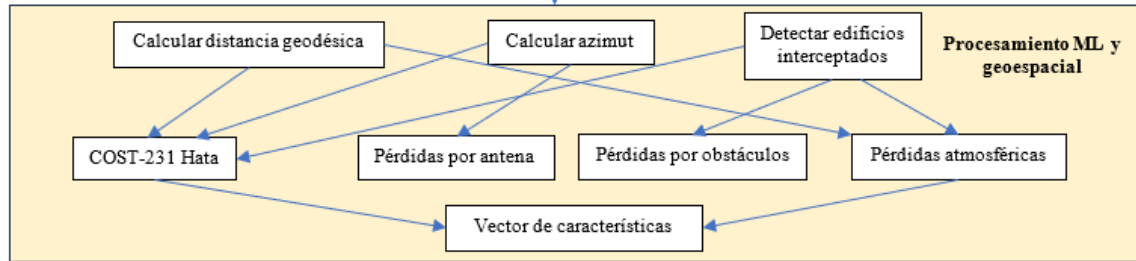
Representado de la siguiente manera:

$$\text{Sector}_i = \{\text{Azimut, Tilt, Potencia, Estado}\}$$

Procesamiento geoespacial y radioeléctrico

La **Figura 24** muestra el módulo de procesamiento geoespacial y radioeléctrico, encargado de integrar la información geográfica con los parámetros de propagación LTE para generar las variables de entrada utilizadas durante la predicción. Entre sus principales operaciones se encuentran el cálculo de distancias, la determinación de la orientación del enlace, el análisis de obstáculos, la extracción de información del terreno y el cálculo de pérdidas de propagación.

Figura 24. Módulo de procesamiento geoespacial y radioeléctrico.



Nota. Elaboración propia.

A partir de la información contenida en el Modelo Digital de Elevación (DEM), el sistema obtiene el perfil topográfico entre la estación base y el punto de interés mediante la función `elevacion(lat, lon)`. Esta función transforma las coordenadas geográficas en índices del raster utilizando `rowcol(transform, lon, lat)` y recupera el valor de elevación correspondiente. La información obtenida permite construir el perfil del terreno y analizar la posible presencia de obstrucciones en el trayecto.

- **La evaluación de línea de vista** determina la altura de la trayectoria directa mediante una interpolación lineal entre las alturas del transmisor y el receptor. Para un punto situado a una distancia (x) dentro del trayecto, dicha altura puede expresarse como:

$$LOS = y_1 + \frac{(y_2 - y_1)}{(x_2 - x_1)}(x - x_1)$$

Ecuación 11. Representación matemática de la evaluación de línea de vista.

Donde:

- **LOS:** altura de la trayectoria directa en el punto (x), expresada en metros.
- **x :** distancia horizontal del punto evaluado dentro del perfil.
- **x_1, x_2 :** posiciones horizontales correspondientes al transmisor y al receptor.
- **y_1, y_2 :** alturas de los extremos del enlace respecto al sistema de referencia utilizado.

Este cálculo permite establecer la altura teórica de la trayectoria directa entre el transmisor y el receptor, y posteriormente compararla con el perfil del terreno para identificar posibles obstrucciones.

Código 9. Obtención de distancia geodésica

```

distancia = geodesic(
(lat_ant, lon_ant),

```

(lat_usr, lon_usr)
).km

El código calcula la distancia geodésica entre la estación base y el usuario a partir de sus coordenadas geográficas y expresa el resultado en kilómetros.

- **El cálculo de la longitud de onda** se realiza a partir de la velocidad de propagación de la onda y de la frecuencia de operación, se representa de la siguiente manera:

$$\lambda = \frac{c}{f}$$

Ecuación 12. Cálculo de la longitud de onda.

Donde:

- **λ** : longitud de onda, expresada en metros.
- **c** : velocidad de propagación de la luz en el vacío, aproximadamente (3×10^8 m/s).
- **f** : frecuencia de operación, expresada en Hz.

La longitud de onda se utiliza posteriormente para calcular el radio de la primera zona de Fresnel.

- **El radio de la primera zona de Fresnel** se calcula mediante la siguiente ecuación:

$$F_1 = \sqrt{\frac{\lambda d_1 * d_2}{d_1 + d_2}}$$

Ecuación 13. Radio de la primera zona de Fresnel.

Donde:

- **F_1** : radio de la primera zona de Fresnel en el punto analizado, expresado en metros.
- **λ** : longitud de onda de la señal, expresada en metros.
- **d_1** : distancia entre el transmisor y el punto analizado, expresada en metros.
- **d_2** : distancia entre el punto analizado y el receptor, expresada en metros.

El cálculo del radio de Fresnel permite determinar el volumen espacial alrededor de la trayectoria principal que debe mantenerse suficientemente libre de obstáculos para reducir pérdidas adicionales asociadas con la difracción. Esta información se utiliza

conjuntamente con el perfil topográfico y la ubicación de las edificaciones para evaluar las condiciones de propagación del enlace.

Código 10. Cálculo del radio de Fresnel

```
fresnel = np.sqrt(  
    (lambda_ * d1 * d2) /  
    (d1 + d2)  
)
```

El radio de la primera zona de Fresnel se calcula mediante operaciones vectorizadas de NumPy utilizando la función `np.sqrt()`. El resultado corresponde al radio de la zona de Fresnel en el punto analizado, expresado en metros. Para evitar resultados negativos derivados de operaciones previas, durante el cálculo se puede recurrir a la función `np.maximum(0, ...)`, con el fin de mantener valores no negativos y contribuir a la estabilidad numérica.

- **El modelo de propagación COST-231 Hata**, implementado por el sistema y presentado en la sección 2.10.2, permite estimar la pérdida de trayectoria en entornos urbanos a partir de la frecuencia, la altura de la estación base, la altura del receptor y la distancia. En esta sección se detalla específicamente el factor de corrección $a(hm)$ asociado con la altura del terminal móvil, utilizado como parte de dicho modelo, se representa de la siguiente manera:

$$a(hm) = (1.1\log_{10}(f) - 0.7)hm - (1.56\log_{10}(f) - 0.8)$$

Ecuación 14. Factor de corrección.

Donde:

- **f**: corresponde a la frecuencia de operación, expresada en MHz.
- **hm**: altura del terminal móvil, expresada en metros.

Las constantes incluidas en la formulación corresponden a los parámetros empíricos del modelo. El factor (C_m) depende del escenario de propagación considerado. Las variables PL, hb, d y C, correspondientes a la ecuación principal del modelo, se definieron previamente en la sección 2.10.2.

El modelo COST-231 Hata se utiliza como referencia físico-empírica para estimar la pérdida de trayectoria a partir de variables como la frecuencia de operación, la altura de la estación base, la altura del receptor y la distancia entre ambos. La estimación obtenida

se incorpora posteriormente al procesamiento de las variables utilizadas por el sistema predictivo.

- **El cálculo de pérdidas por antena** considera la atenuación horizontal asociada con el patrón de radiación de la antena mediante una función dependiente de la diferencia angular respecto al azimut del sector, que se calcula de la siguiente manera:

$$L_{\text{hor}} = \min \left(12 \left(\frac{\Delta\theta}{\theta_{3\text{dB}}} \right)^2, L_{\text{max}} \right)$$

Ecuación 15. Pérdida horizontal de la antena.

Donde:

- **L_{hor}** : pérdida horizontal de la antena, expresada en dB.
- **$\Delta\theta$** : diferencia angular entre la dirección del sector y la dirección hacia el punto de interés.
- **$\theta_{3\text{dB}}$** : ancho de haz horizontal de la antena.
- **L_{max}** : pérdida máxima considerada en el modelo.

En el sistema se establece un valor máximo de pérdida de $L_{\text{max}} = 25\text{dB}$.

Este valor limita la atenuación calculada por el patrón horizontal y evita que el modelo produzca pérdidas excesivamente elevadas fuera del lóbulo principal.

- **El monitor de rendimiento y velocidad** incorpora un módulo para estimar velocidades de transmisión a partir del CQI. Estas estimaciones se utilizan como indicadores complementarios del desempeño del enlace LTE.

Para el enlace descendente se utiliza la siguiente representación matemática:

$$DL = CQI \times 9.5$$

Ecuación 16. Estimación del DL Throughput.

Donde:

- **DL**: velocidad estimada de descarga, expresada en Mbps.
- **CQI**: indicador de calidad del canal (0–15).
- **9.5**: factor de escala utilizado por el modelo implementado para aproximar la velocidad descendente.

Para el enlace ascendente se utiliza la siguiente representación matemática:

$$UL = CQI \times 3.1$$

Ecuación 17. Estimación del UL Throughput.

Donde:

- **UL**: velocidad estimada de subida, expresada en Mbps.
- **CQI**: indicador de calidad del canal (0–15).
- **3.1**: factor de escala utilizado para aproximar la velocidad ascendente.

Estos valores constituyen aproximaciones empleadas por el sistema y no representan directamente el Throughput real medido por la red, debido a que la velocidad efectiva depende también de factores como el ancho de banda, la modulación, la asignación de recursos, la carga de la celda y las condiciones de interferencia.

Código 11. Representación gráfica en el sistema

```
ax.fill_between(
    xs,
    los_y - fresnel_r,
    los_y + fresnel_r
)
```

El área sombreada representa la sección bidimensional del volumen asociado con la primera zona de Fresnel a lo largo del trayecto. Esta representación permite identificar posibles obstáculos del terreno o de las edificaciones que puedan afectar las condiciones de propagación.

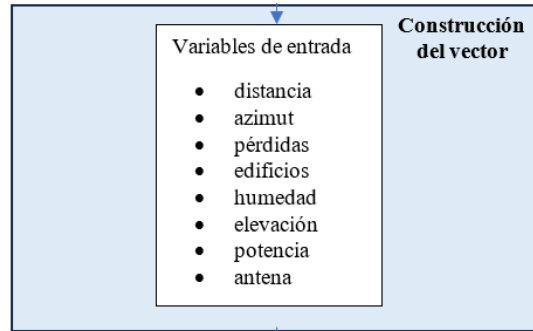
Entre los elementos que pueden visualizarse se encuentran:

- Las obstrucciones estructurales.
- Los bloqueos del terreno.
- La interferencia estructural.
- Las pérdidas por edificios.

Construcción del vector de características

La **Figura 25** presenta el proceso de construcción del vector de características, en el que el sistema integra las variables geoespaciales, radioeléctricas, ambientales y de propagación calculadas previamente. Este conjunto de variables constituye la entrada utilizada por los modelos Random Forest y XGBoost para realizar la estimación de la cobertura LTE.

Figura 25. Construcción del vector de características.



Nota. Elaboración propia.

Las variables empleadas pueden representarse mediante el siguiente vector, se representa de la siguiente manera:

$$X = [d, \Delta\theta, \text{tilt}, h_b, N_{\text{edif}}, H_{\text{edif}}, H, P_t, PL, O_{\text{terr}}, L_{\text{ant}}]$$

Donde:

- ***d***: distancia entre la estación base y el usuario.
- **$\Delta\theta$** : diferencia angular respecto al sector seleccionado.
- ***tilt***: inclinación eléctrica de la antena.
- ***h_b***: altura de la estación base.
- ***N_{edif}***: número de edificaciones consideradas en el trayecto.
- ***H_{edif}***: altura máxima de la edificación considerada en el trayecto.
- ***H***: humedad ambiental.
- ***P_t***: potencia de transmisión.
- ***PL***: pérdida de trayectoria (path loss), calculada mediante la fórmula COST-231 Hata.
- ***O_{terr}***: obstrucción del terreno a lo largo del trayecto.
- ***L_{ant}***: pérdidas de la antena.

Código 12. Construcción del vector de características

```
features = np.array([[
    distancia_km,
    dif_azimuth,
    tilt,
    altura_antena,
    num_edificios,
```

```

altura_max_edificio,
humedad,
potencia_mw,
cost231_path_loss,
obs_terreno_m,
L_ant

```

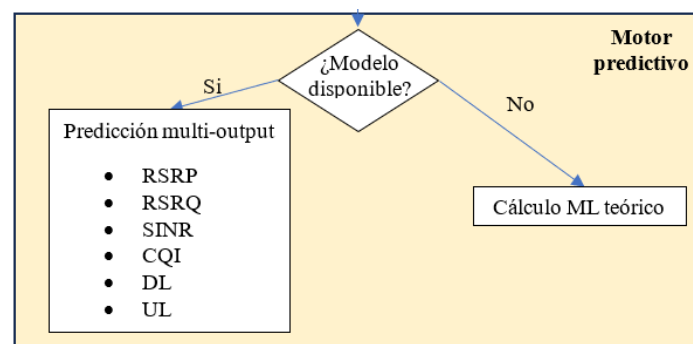
]])

El código construye un DataFrame con las mismas variables utilizadas durante el entrenamiento del modelo, reordenando las columnas de acuerdo con la secuencia definida en `features_nombres` para garantizar la correspondencia exacta con el vector usado en la etapa de entrenamiento. Una vez calculadas, las variables se agrupan en un único vector de características que representa las condiciones del enlace utilizadas por el modelo predictivo. Cada escenario generado por el sistema produce un nuevo conjunto de valores de entrada, que posteriormente se proporciona al modelo para obtener la estimación correspondiente.

Motor predictivo basado en aprendizaje automático

La **Figura 26** muestra el funcionamiento del motor predictivo. Una vez construido el vector de características, el sistema verifica la disponibilidad del modelo previamente entrenado. Cuando el modelo está disponible, se ejecuta la predicción a partir de las variables de entrada. En caso contrario, el sistema utiliza un mecanismo alternativo basado en los modelos físicos de propagación disponibles, permitiendo continuar con la simulación.

Figura 26. *Motor predictivo del sistema.*



Nota. Elaboración propia.

El núcleo del procesamiento corresponde a la función `motor_calculo_ml()`, que integra la información calculada mediante los modelos físicos de propagación con las variables geoespaciales y las características utilizadas por el modelo de aprendizaje automático.

Entre los principales elementos que intervienen se encuentran:

- Modelos físicos de propagación.
- Modelo de aprendizaje automático entrenado.
- Variables geoespaciales.
- Características y pérdidas asociadas con el entorno urbano.

Código 13. Inferencia del modelo

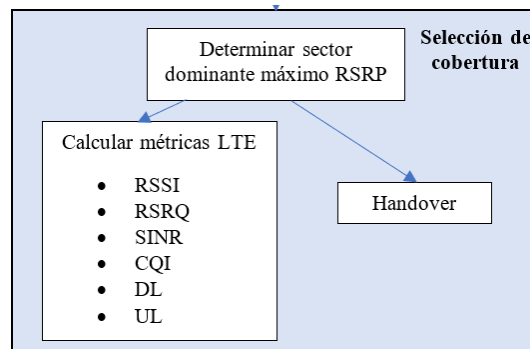
```
pred = modelo.predict(features)
```

La instrucción ejecuta la predicción utilizando el vector de características previamente construido y el modelo cargado en el sistema. El resultado se utiliza posteriormente para representar la estimación de cobertura LTE en la interfaz de visualización.

Selección de cobertura y métricas LTE

La **Figura 27** presenta el proceso de selección de cobertura del sistema, en el cual se identifica el sector dominante a partir del mayor valor de RSRP estimado entre los sectores activos.

Figura 27. Proceso de selección del sector dominante y estimación de métricas LTE.



Nota. Elaboración propia.

Una vez seleccionado el sector dominante, el sistema obtiene los indicadores complementarios de calidad y el rendimiento del enlace, como RSRQ, RSSI, SINR, CQI, DL Throughput y UL Throughput. Para ello dispone de dos rutas de cálculo. La ruta principal, utilizada durante toda la evaluación presentada en el Capítulo IV, consiste en la inferencia mediante los modelos de aprendizaje automático entrenados para cada

variable objetivo. Cada modelo recibe el mismo vector de características construido previamente y devuelve su estimación de manera independiente. La ruta alternativa se activa únicamente cuando los archivos de los modelos entrenados no se encuentran disponibles en el entorno de ejecución, condición representada en el flujograma del sistema. En ese caso, el sistema recurre a las relaciones físico-empíricas implementadas y expresadas en las ecuaciones siguientes, que permiten mantener operativa la simulación y la visualización de escenarios, aunque con menor precisión. Estas expresiones se documentan por corresponder al comportamiento implementado en el sistema, pero no fueron la fuente de los valores estimados que se reportan en el Capítulo IV. Aun así, los indicadores obtenidos mediante ellas permiten complementar la estimación de cobertura y analizar las condiciones del enlace radioeléctrico.

Código 14. Selección del sector dominante

```
sector_dominante = np.argmax(rsrp_sectores)
rsrp_final = max(rsrp_sectores)
```

Cuando el modelo de aprendizaje automático no se encuentra disponible, el sistema evalúa los sectores activos mediante el modelo físico de propagación y selecciona automáticamente el sector dominante con base en el mayor valor estimado de RSRP.

La estimación del RSRP mediante el modelo físico se expresa como:

$$RSRP = (P_{tx} - 10\log_{10}(12 \times N_{RB})) + G_{ant} - PL - L_{ant} - L_{edificios}$$

Ecuación 18. Estimación del RSRP.

Donde:

- P_{tx} : potencia total de transmisión del sector, expresada en dBm.
- N_{RB} : número de bloques de recursos disponibles según el ancho de banda del canal, con 12 subportadoras por RB.
- G_{ant} : ganancia de la antena, expresada en dBi.
- PL : pérdida de trayectoria, expresada en dB.
- L_{ant} : pérdidas asociadas con el patrón de radiación de la antena, expresadas en dB.
- $L_{edificios}$: pérdidas adicionales asociadas con las obstrucciones de las edificaciones, expresadas en dB.

Dado que el RSRP se define como la potencia recibida por elemento de recurso, mientras que P_{tx} representa la potencia total transmitida por el sector, la ecuación incorpora el término de conversión $10\log_{10}(12 \times N_{RB})$. El valor de N_{RB} se determinó a partir del ancho de banda de canal configurado para la banda E-UTRA Band 4 de la red LTE de CNT en la provincia de Santa Elena, reportado en 10 MHz (50 bloques de recursos) por el informe técnico de [13], estudio que documenta la misma infraestructura de red utilizada como referencia en la presente investigación. Con este valor, la conversión aplicada equivale a $10\log_{10}(12 \times 50) \approx 27,8$ dB.

El RSRQ se estima mediante la siguiente formulación matemática:

$$\mathbf{RSRQ} = \mathbf{0.18} \times \mathbf{RSRP} + \mathbf{2}$$

Ecuación 19. Estimación del RSRQ.

Donde:

- **RSRQ**: calidad de la señal de referencia recibida, expresada en dB.
- **RSRP**: potencia de la señal de referencia recibida, expresada en dBm.

La expresión corresponde a una relación empírica implementada en el sistema, por lo que debe interpretarse como una aproximación y no como la definición estándar del RSRQ.

El RSSI se calcula mediante la relación entre el RSRP, el RSRQ y el número de bloques de recursos utilizados, se representa de la siguiente manera:

$$\mathbf{RSSI} = \mathbf{RSRP} - \mathbf{RSRQ} + \mathbf{10\log_{10}(N)}$$

Ecuación 20. Estimación del RSSI.

Donde:

- **RSSI**: potencia total recibida, expresada en dBm.
- **RSRP**: potencia de la señal de referencia recibida, expresada en dBm.
- **RSRQ**: calidad de la señal de referencia recibida, expresada en dB.
- **N**: número de bloques de recursos utilizados en el ancho de banda considerado.
- **$10\log_{10}$** : término logarítmico que pondera la contribución de N al cálculo del RSSI.

El SINR se estima mediante una aproximación basada en el RSRP y la distancia entre la estación base y el punto de análisis, se representa de la siguiente manera:

$$\mathbf{SINR} = \mathbf{RSRP} + \mathbf{100} - \mathbf{3d}$$

Ecuación 21. Estimación del SINR.

Donde:

- **SINR**: relación señal-interferencia más ruido, expresada en dB.
- **RSRP**: potencia de la señal de referencia recibida, expresada en dBm.
- **d**: distancia entre la estación base y el punto de análisis, expresada en kilómetros.
- **100 y 3**: constantes utilizadas por la aproximación implementada para relacionar el nivel de RSRP con la condición estimada de interferencia.

Esta expresión constituye una aproximación específica de la implementación y no corresponde a una ecuación estándar de 3GPP para el cálculo directo de SINR.

El sistema obtiene un valor estimado de CQI a partir del SINR mediante la siguiente relación:

$$CQI = \frac{(SINR + 6)}{31} \times 15$$

Ecuación 22. Estimación del CQI.

Donde:

- **CQI**: con valores de 0 a 15.
- **SINR**: relación señal-interferencia más ruido, expresada en dB.
- **6 y 31**: constantes empleadas por la aproximación implementada para transformar el SINR al rango considerado.
- **15**: valor máximo utilizado en la escala de CQI.

El resultado obtenido mediante la aproximación se restringe al intervalo de 0 a 15, correspondiente al rango utilizado para el CQI en LTE.

El CQI se utiliza como indicador de calidad del canal y puede asociarse con diferentes esquemas de modulación y codificación, entre ellos QPSK, 16QAM, 64QAM y, en determinadas configuraciones LTE, 256QAM. Estas estimaciones permiten complementar la representación de la cobertura mediante indicadores asociados con la calidad y el rendimiento del enlace. En síntesis, las expresiones anteriores corresponden al modo de respaldo del sistema. Todos los valores estimados que se presentan y validan en el Capítulo IV provienen de los modelos de aprendizaje automático entrenados para cada variable objetivo.

3.8. Metodología de la implementación del sistema de visualización

El sistema incorpora un módulo de visualización interactiva destinado a representar gráficamente los resultados del procesamiento radioeléctrico y de la predicción de

cobertura LTE. Este módulo integra parámetros relacionados con la calidad de la señal, las condiciones del canal y las características geográficas del entorno, facilitando la interpretación de los resultados.

Cuando el usuario modifica parámetros de configuración, como la frecuencia de operación, la potencia de transmisión, el tilt, la altura de la antena, la humedad, el estado de los sectores o la posición de los marcadores geográficos, el sistema actualiza los cálculos y las representaciones correspondientes.

El módulo de visualización no participa en el entrenamiento de los modelos. Su función consiste en cargar, en cada sesión, el modelo previamente entrenado y almacenado que se seleccione para la inferencia (Random Forest o XGBoost), junto con la configuración de las variables de entrada. Ese modelo se utiliza posteriormente para realizar las predicciones durante la interacción con el sistema. Posteriormente, se aplican los ajustes físicos definidos en la implementación, entre ellos la corrección asociada con la humedad y la penalización fuera del lóbulo principal del sector activo. Esta última permite representar una disminución progresiva de la señal cuando el punto de análisis se aleja de la dirección de máxima radiación del sector.

Por otra parte, la detección de handover utiliza el conjunto de datos que incluye las mediciones correspondientes a estaciones base vecinas. Cuando se identifica una condición de cambio entre sitios, el sistema ejecuta la función principal de actualización, recalcula los parámetros correspondientes y sincroniza los componentes de la interfaz.

Código 15. Actualización dinámica de la interfaz

```
def actualizar(change=None):
```

```
    actualizar_mapa()
```

```
    actualizar_perfil()
```

```
    actualizar_metricas()
```

```
    actualizar_heatmap()
```

La función actualizar() coordina la actualización de los principales componentes visuales cada vez que se modifica un parámetro de simulación o la posición geográfica del análisis.

La **Figura 28** muestra el panel de controles del sistema, uno de los elementos que componen la interfaz desarrollada. En conjunto, la interfaz está compuesta por los siguientes elementos:

1. Panel de controles.
2. Mapa geográfico de cobertura.
3. Panel informativo dinámico.
4. Monitor de rendimiento y velocidad.
5. Perfil topográfico del enlace.
6. Comparador de resultados.

Figura 28. Panel de controles interactivos del sistema de predicción.



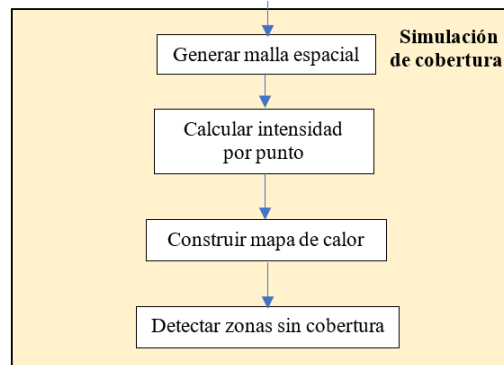
Nota. Elaboración propia.

- **La generación del mapa de calor** incorpora una representación espacial de la cobertura desarrollada con la biblioteca ipyleaflet. Esta herramienta permite visualizar la ubicación de la estación base, la posición del usuario y la distribución espacial estimada de los niveles de señal LTE.

El mapa se actualiza a partir de las predicciones generadas para los puntos considerados en el área de análisis, permitiendo identificar visualmente las zonas con diferentes niveles de cobertura y comparar el comportamiento de la señal entre los sectores de la estación base.

La **Figura 29** presenta el módulo de simulación de cobertura, en el cual se genera una malla espacial alrededor de la estación base y se estima el nivel de la señal en cada punto mediante el modelo predictivo. Los resultados obtenidos se representan posteriormente mediante un mapa de calor, lo que permite visualizar la distribución espacial de la cobertura LTE.

Figura 29. Módulo de simulación de mapa de calor LTE.



Nota. Elaboración propia.

El sistema genera una malla distribuida radialmente alrededor de la estación base. El número total de puntos de evaluación se determina mediante la siguiente relación matemática:

$$N = 10 \times 36$$

Ecuación 23. Número de puntos de la malla radial.

Donde:

- **N**: número total de puntos que conforman la malla radial de evaluación.
- **10**: número de distancias radiales consideradas.
- **36**: número de divisiones angulares.

Código 16. Generación de puntos radiales

for r in radios:

 for theta in angulos:

 lat = ANT_LAT + (r/111.32)*np.cos(theta)

 lon = ANT_LON + (r/(111.32*np.cos(np.radians(ANT_LAT))))*np.sin(theta)

Para cada punto de evaluación se estima el nivel de RSRP y posteriormente se normaliza para facilitar su representación mediante un gradiente de colores. La normalización utilizada por el sistema se expresa como:

$$I = \frac{RSRP + 110}{45}$$

Ecuación 24. Intensidad normalizada utilizada en el mapa de calor.

Donde:

- **I**: intensidad normalizada utilizada para la representación gráfica.
- **RSRP**: potencia de la señal de referencia recibida en el punto evaluado, expresada en dBm.
- **110**: valor utilizado para trasladar el límite inferior de la escala considerada hacia un rango positivo.
- **45**: amplitud del intervalo utilizado para normalizar los valores de RSRP.

La normalización se basa en el intervalo de referencia: $-110 \text{ dBm} \leq \text{RSRP} \leq -65 \text{ dBm}$.

Código 17. Generación del mapa de calor LTE

Heatmap(

 locations=heatmap_data,

radius=25,
blur=15
)

Para evitar valores extremos que puedan afectar la representación visual, el sistema limita la intensidad normalizada al intervalo $0.1 \leq I \leq 1$.

- **Las coordenadas geográficas de cada punto de la malla** se calculan mediante las siguientes expresiones matemáticas:

$$\text{Lat}_p = \text{Lat} + \frac{r}{111.32 \times \cos(\theta)}$$

***Ecuación 25.** Estimación de la latitud.*

$$\text{Lon}_p = \text{Lon} + \frac{r}{111.32 \cos(\text{Lat})} \times \sin(\theta)$$

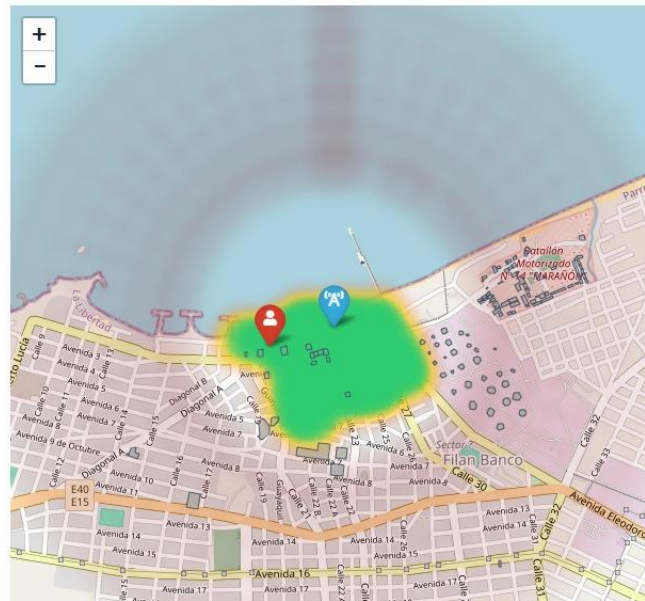
***Ecuación 26.** Estimación de la longitud.*

Donde:

- **Lat_p**: latitud del punto de la malla, expresada en grados decimales.
- **Lon_p**: longitud del punto de la malla, expresada en grados decimales.
- **Lat**: latitud de la estación base, expresada en grados decimales.
- **Lon**: longitud de la estación base, expresada en grados decimales.
- **θ**: ángulo utilizado para determinar la posición del punto dentro de la malla radial, expresado en radianes.
- **111.32 km**: aproximación de la cantidad de kilómetros correspondientes a un grado de latitud.

Para el cálculo de la longitud, el término del coseno debe utilizar la latitud expresada en radianes cuando se implemente mediante funciones trigonométricas de Python. El algoritmo genera los puntos de la malla alrededor de la estación base y evalúa las condiciones de propagación en cada ubicación. Para esto, se calculan parámetros como la pérdida de trayectoria, la atenuación angular, las pérdidas asociadas con el patrón vertical de la antena y otros factores considerados por el modelo, presentados en la **Figura 30**. Los resultados se representan mediante una escala de colores que permite diferenciar visualmente los niveles de cobertura LTE de acuerdo con el RSRP estimado.

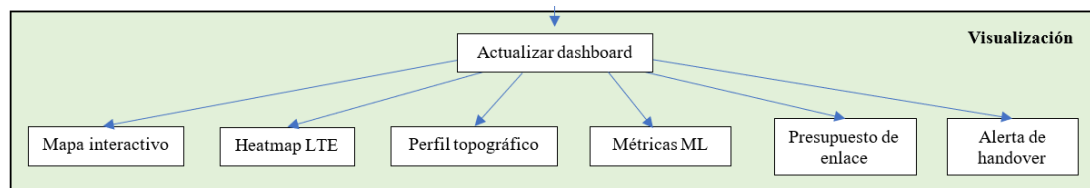
Figura 30. Mapa de calor interactivo de cobertura LTE.



Nota. Elaboración propia.

Visualización de resultados

Figura 31. Panel de visualización del sistema.



Nota. Elaboración propia.

La **Figura 31** presenta el módulo principal de visualización del sistema, encargado de representar los resultados mediante mapas interactivos, mapas de calor, perfiles topográficos, presupuestos de enlace y métricas asociadas con la predicción de cobertura LTE. Las **Figuras 32, 33 y 34** muestran diferentes componentes del panel informativo desarrollado.

Figura 32. Panel informativo dinámico del sector.



Nota. Elaboración propia.

Figura 33. Monitor de rendimiento y velocidad del enlace.



Nota. Elaboración propia.

Código 18. Actualización del panel de visualización

```
display(
    VBox([
        mapa,
        panel_metricas,
        perfil_topografico
    ])
)
```

Los componentes de la interfaz se actualizan mediante la función actualizar(), que coordina la renovación de los elementos visuales a partir de los cambios realizados en los parámetros de configuración. Esta sincronización permite mantener correspondencia entre los cálculos realizados por el sistema y su representación gráfica.

Figura 34. Presupuesto de enlace del sector.

PRESUPUESTO DE ENLACE - Sector 2		
Potencia TX	+43.0	TRANSMISOR dBm
Ganancia Antena	+17.6	dBd
EIRP	+60.6	dBm
Path Loss COST231	-124.6	PÉRDIDAS dB
L. Azimut	-0.0	dB
L. Edificios	-2.0	dB
L. Humedad	-0.3	dB
Total Pérdidas	-126.8	dB
RSRP Estimado	-84.3	RESULTADO dBm
Longitud de onda λ	0.142 m	1ª ZONA DE FRESNEL
Radio máximo (r_1)	3.41 m	
60% recomendado	2.05 m	
Distancia enlace	329 m	

Nota. Elaboración propia.

Comparador de escenarios

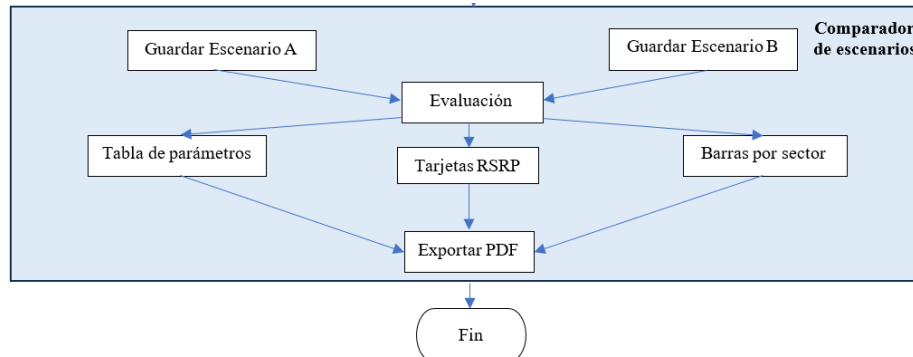
El sistema incorpora un módulo de comparación de escenarios que permite contrastar dos configuraciones diferentes de la red LTE, denominadas Escenario A y Escenario B, antes de aplicar cambios en los parámetros de la estación base. Esta herramienta permite analizar las variaciones estimadas en el nivel de RSRP y otros parámetros de cobertura para una misma ubicación, facilitando la evaluación de diferentes alternativas de configuración.

El comparador funciona mediante cuatro etapas principales. Primero, el usuario configura y guarda el Escenario A, que representa la condición de referencia. Posteriormente, modifica los parámetros seleccionados y guarda el Escenario B, correspondiente al ajuste alternativo. A continuación, el sistema evalúa ambas configuraciones en la ubicación del usuario y compara los resultados obtenidos. Finalmente, genera los elementos visuales y el reporte correspondiente.

La **Figura 35** representa este flujo de trabajo. Los bloques “Guardar Escenario A” y “Guardar Escenario B” almacenan el estado de los controles del sistema. El bloque “Evaluación” aplica temporalmente cada configuración y ejecuta el motor de predicción. A partir de los resultados se generan la tabla comparativa de parámetros, las tarjetas de

RSRP y el gráfico de barras por sector, que posteriormente pueden integrarse en un reporte PDF.

Figura 35. Flujo de trabajo del comparador de escenarios de planificación.



Nota. Elaboración propia.

- **La captura de un escenario** consiste en representarlo internamente mediante un diccionario que almacena la frecuencia de operación, la altura de la antena y los parámetros del azimut, el tilt, la potencia y el estado de activación de cada sector, tal como se aprecia en la **Figura 36**.

Figura 36. Interfaz del comparador de escenarios de planificación.



Nota. Elaboración propia.

Al guardar un escenario, la función correspondiente captura el estado actual de los controles sin modificar el estado activo del sistema. Durante la evaluación, cada escenario se aplica temporalmente para calcular el RSRP en la posición del usuario. Una vez finalizado el cálculo, el sistema restaura la configuración original para evitar modificaciones no deseadas en el resto de la interfaz.

Código 19. Captura y evaluación de un escenario

```

def _capturar_escenario():
    return {
        'frecuencia_mhz': frec_s.value,
        'altura_antena_m': alt_ant_s.value,
        'sectores': {k: dict(v) for k, v in sectores.items()}
    }

```

```
def _evaluar_escenario(cfg):
```

```
    # Aplica cfg, calcula RSRP por sector con motor_calculo_ml
```

```
    # y restaura el estado original en el bloque finally
```

Una vez obtenido el RSRP de ambos escenarios, el sistema genera tres elementos principales. El primero es una tabla comparativa que presenta, por sector, los valores de azimut, tilt y potencia correspondientes a los escenarios A y B. El segundo corresponde a las tarjetas de RSRP, que muestran el nivel estimado para cada escenario en la ubicación del usuario y la diferencia entre ambos resultados. El tercero es un gráfico de barras que permite comparar el RSRP estimado para cada sector activo.

De manera complementaria, el sistema genera mapas de calor para ambos escenarios utilizando el mismo procedimiento empleado en el mapa de cobertura principal. Esto permite observar espacialmente las variaciones producidas por cambios en parámetros como el azimut, el tilt o la potencia de transmisión.

El comparador también puede generar sugerencias automáticas basadas en las diferencias observadas entre los escenarios evaluados. Estas recomendaciones permiten identificar posibles efectos asociados con los cambios realizados, como variaciones en la distribución espacial de la señal o modificaciones en el nivel de cobertura de determinados sectores.

Finalmente, el botón “Guardar PDF” permite exportar los resultados de la comparación a un documento de cuatro páginas, compuesto por:

1. La tabla consolidada de parámetros y de valores de RSRP.
2. El gráfico comparativo por sector.
3. El mapa de calor correspondiente al escenario A.
4. El mapa de calor correspondiente al escenario B.

De esta manera, el comparador de escenarios permite analizar y documentar diferentes ajustes de la red antes de considerar su aplicación en la infraestructura real del cantón La Libertad.

CAPÍTULO IV: RESULTADOS Y ANÁLISIS

En este capítulo se presentan los resultados obtenidos mediante el sistema de predicción de cobertura LTE. Se analiza el desempeño de los modelos Random Forest y XGBoost en la estimación del RSRP a partir de variables geográficas, ambientales, radioeléctricas y de propagación. Para evaluar su desempeño se consideran principalmente el error absoluto medio (MAE) y el coeficiente de determinación (R^2), mediante la comparación entre los valores predichos y las mediciones obtenidas en campo.

Asimismo, se presentan los resultados del módulo de visualización interactiva mediante mapas de calor, perfiles topográficos y paneles de monitoreo, con el propósito de analizar la distribución espacial de la cobertura y valorar la utilidad de la herramienta como apoyo para la planificación y optimización de la red LTE de CNT en el cantón La Libertad.

4.1. Condiciones y parámetros del entorno de evaluación

Esta sección describe las condiciones bajo las cuales se evaluaron los modelos predictivos y el sistema de visualización. El conjunto de datos utilizado está compuesto por 4 962 registros, de los cuales 2 822 corresponden a mediciones LTE obtenidas en campo y 2 140 a escenarios generados mediante el procedimiento de aumento de datos. Las observaciones se asociaron con la estación base analizada, configurada en la banda 4 de LTE (Band 4), con un EARFCN de 2050 y una frecuencia de enlace descendente de 2120 MHz. Para el entrenamiento y prueba de los modelos se utilizó una partición de 80 % y 20 %, respectivamente, con agrupación por punto físico de medición. La **Tabla 13** resume estas condiciones.

Los parámetros radioeléctricos de referencia utilizados durante la evaluación, entre ellos la frecuencia, la altura de la antena, el azimut, el tilt y la potencia de transmisión, corresponden a los valores descritos en la **Tabla 5** y se sustentan en la parametrización de estaciones LTE de CNT en la provincia de Santa Elena reportada en [13]. Por su parte, la ganancia de la antena, el ancho de haz horizontal y las demás características del modelo Tongyu TDJH-182018DE-65F se obtuvieron de la hoja de datos del fabricante [38].

Las 2 822 mediciones reales corresponden a los puntos registrados durante las campañas de campo descritas en la sección 3.5, una vez que se eliminaron los registros

vacíos o inválidos, los puntos afectados por errores de posicionamiento y las muestras que no correspondían al Cell_ID de la estación base objetivo.

Los 2 140 escenarios aumentados se generaron mediante el procedimiento de aumento de datos descrito en la sección 3.5.1. Para ello, se modificaron los parámetros de tilt, la altura de antena y la potencia de transmisión dentro de los rangos establecidos, generando escenarios complementarios a partir de las mediciones reales y del modelo físico de propagación. Posteriormente, se aplicó un muestreo por configuración para evitar que los escenarios aumentados tuvieran una representación desproporcionada dentro del conjunto final.

La cantidad de mediciones reales estuvo condicionada por el área de servicio analizada y por el hecho de que el estudio se concentró en los tres sectores de una única estación base. La incorporación de estaciones base vecinas habría introducido diferentes configuraciones radioeléctricas y condiciones de servicio, mientras que la repetición de mediciones en los mismos puntos no habría aportado nuevas ubicaciones al análisis espacial.

El conjunto de datos utilizado permitió disponer de suficientes observaciones para realizar la separación entre entrenamiento y prueba, así como para evaluar el desempeño de los modelos sobre puntos físicos no utilizados durante el aprendizaje. Debido a que las observaciones presentan una estructura agrupada espacialmente, la cantidad efectiva de unidades independientes para evaluar la generalización está determinada por los puntos físicos y no únicamente por el número total de registros.

Tabla 13. *Parámetros del entorno de evaluación.*

Parámetro	Valor
Tecnología	LTE
Banda	Band 4
Frecuencia DL	2120 MHz
Muestras	4 962
División	80 % entrenamiento / 20 % prueba
Modelos	Random Forest y XGBoost

Nota. Elaboración propia.

4.2. Evaluación de los modelos predictivos

En esta sección se presentan los resultados obtenidos por los modelos Random Forest y XGBoost sobre el conjunto de prueba. La evaluación se realiza mediante el error absoluto medio (MAE) y el coeficiente de determinación (R^2), complementados con el análisis de la diferencia entre el desempeño observado en los conjuntos de entrenamiento y prueba.

La **Tabla 14** resume los resultados obtenidos para la predicción del RSRP. En esta variable, Random Forest presentó el mejor desempeño, con un MAE inferior y un R^2 superior a los obtenidos mediante XGBoost.

Tabla 14. Resumen comparativo del desempeño global de los modelos predictivos.

Métrica	Random Forest	XGBoost
MAE (RSRP)	1,82 dB	2,55 dB
R^2 (RSRP)	0,964	0,940
Variables objetivo con menor MAE	6/6	0/6
Variables objetivo con mayor R^2	5/6	1/6
Brecha de sobreajuste (train–test)	0,0722	0,127

Nota. Elaboración propia.

Los resultados muestran que Random Forest obtuvo un menor error absoluto medio en la estimación del RSRP, con un valor de 1,82 dB, frente a 2,55 dB obtenido por XGBoost. Asimismo, el coeficiente de determinación alcanzó 0,964 para Random Forest y 0,940 para XGBoost, lo que indica un mayor nivel de ajuste del primer modelo sobre los datos evaluados.

En cuanto a la generalización, la brecha entre el desempeño observado durante el entrenamiento y el obtenido sobre el conjunto de prueba —definida formalmente en la sección 4.2.6 (**Ecuación 28**)— fue de 0,0722 para Random Forest y 0,127 para XGBoost. La menor diferencia observada en Random Forest indica un comportamiento más estable entre ambos conjuntos, mientras que la mayor brecha de XGBoost sugiere una mayor diferencia entre su ajuste durante el entrenamiento y su desempeño sobre datos no utilizados previamente.

En consecuencia, bajo las condiciones y el conjunto de datos utilizados en esta investigación, Random Forest presentó el mejor desempeño global para la estimación de las variables analizadas, al obtener un menor MAE en las seis variables objetivo, un

mayor R^2 en cinco de ellas y una menor brecha entre entrenamiento y prueba. La excepción corresponde al RSRQ, para el cual XGBoost presentó un R^2 ligeramente superior al obtenido por Random Forest.

4.2.1. Comparación por variable objetivo

Para evaluar el comportamiento de los modelos en diferentes indicadores de calidad y rendimiento de la red LTE, se analizaron las variables RSRP, RSRQ, SINR, CQI, velocidad de descarga y velocidad de subida. Los resultados se presentan en la **Tabla 15**.

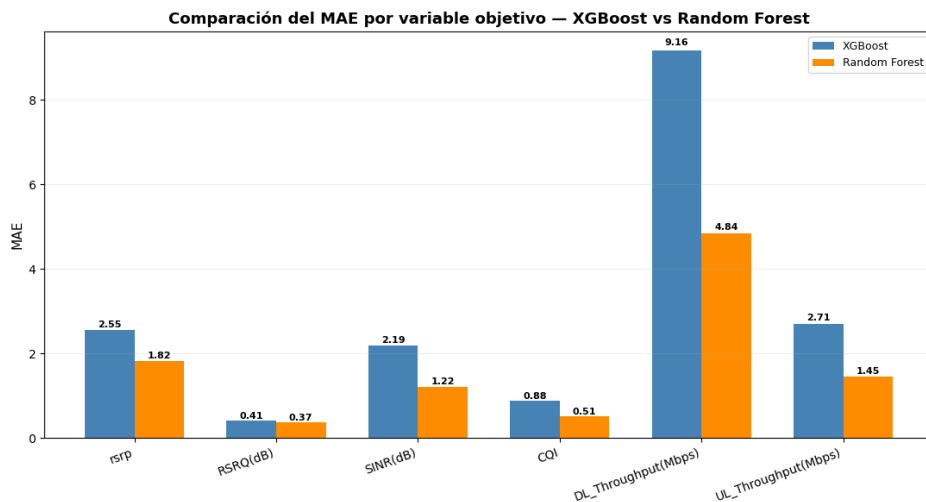
Tabla 15. Comparación del MAE obtenido por los modelos Random Forest y XGBoost para las variables objetivo evaluadas.

Variable objetivo	MAE RF	R^2 RF	MAE XGB	R^2 XGB	Mejor modelo
RSRP (dBm)	1,82	0,964	2,55	0,940	RF
RSRQ (dB)	0,37	0,863	0,41	0,871	RF*
SINR (dB)	1,22	0,962	2,19	0,900	RF
CQI	0,51	0,958	0,88	0,890	RF
DL Throughput (Mbps)	4,84	0,966	9,16	0,895	RF
UL Throughput (Mbps)	1,45	0,966	2,71	0,899	RF

Nota. Elaboración propia.

La columna "Mejor modelo" se determinó según el menor MAE, por corresponder a la métrica principal de evaluación. (*) En el RSRQ, Random Forest obtuvo el menor MAE (0,37 dB frente a 0,41 dB), mientras que XGBoost alcanzó un R^2 ligeramente superior (0,871 frente a 0,863); ambas diferencias son marginales, por lo que el desempeño de los dos modelos en esta variable puede considerarse equivalente.

Figura 37. Comparación del MAE obtenido por los modelos Random Forest y XGBoost para las variables objetivo evaluadas.



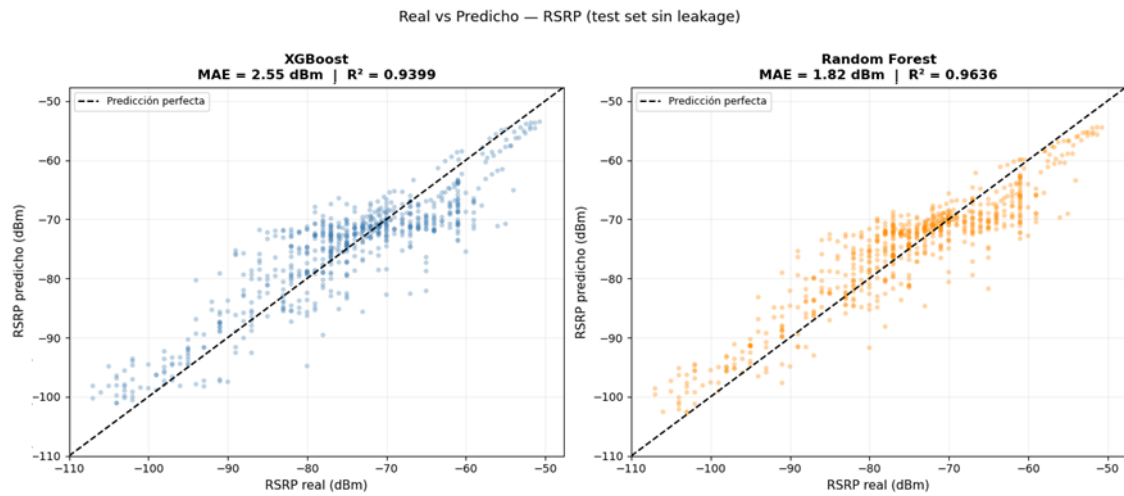
Nota. Elaboración propia.

Tanto la **Tabla 15** como la **Figura 37** muestran que Random Forest presentó el menor MAE en las seis variables evaluadas. Asimismo, obtuvo el mayor R^2 en cinco de ellas. La única excepción corresponde al RSRQ, para el cual XGBoost alcanzó un R^2 ligeramente superior, con 0,871 frente a 0,863 de Random Forest; sin embargo, Random Forest presentó un MAE menor, de 0,37 dB frente a 0,41 dB.

4.2.2. Comparación entre valores reales y predichos

Al comparar los valores reales con los predichos, las diferencias más marcadas se observan en el RSRP y en las velocidades de transmisión. En ambos casos, Random Forest presenta valores de MAE inferiores a los obtenidos por XGBoost. En consecuencia, considerando conjuntamente el MAE y el R^2 , Random Forest muestra un desempeño superior en la mayoría de las variables evaluadas, tal como se aprecia en la **Figura 38**.

Figura 38. Comparación entre los valores reales y predichos de RSRP para los modelos Random Forest y XGBoost.



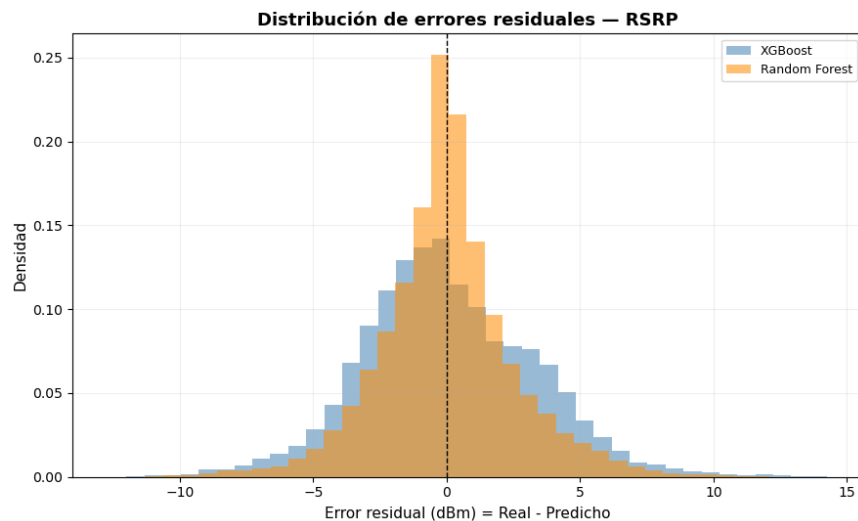
Nota. Elaboración propia.

4.2.3. Distribución de errores

La **Figura 39** presenta la distribución de los residuos obtenidos en la predicción del RSRP para ambos modelos. La concentración de los residuos alrededor de cero permite evaluar la presencia de errores sistemáticos, mientras que su dispersión refleja la variabilidad de las diferencias entre los valores observados y predichos.

Random Forest presenta una distribución de residuos más concentrada en torno al valor central, lo que es consistente con su menor MAE y su mayor estabilidad en la predicción. XGBoost muestra una dispersión mayor, en concordancia con el error observado en las métricas de evaluación.

Figura 39. Distribución de errores residuales para la predicción del RSRP.



Nota. Elaboración propia.

4.2.4. Importancia de las variables predictoras

La **Figura 40** presenta la importancia relativa de las variables utilizadas para estimar el RSRP. Entre las variables con mayor contribución se encuentran la diferencia de azimut, la pérdida de propagación calculada mediante COST-231 Hata, la distancia entre la estación base y el usuario y las pérdidas asociadas con el sistema de antena.

Los valores de importancia que se presentan en la **Figura 40** no se asignan de forma manual o subjetiva, sino que se obtienen directamente de los modelos ya entrenados. En Random Forest se utiliza el atributo `feature_importances_`, basado en la reducción media de impureza, mientras que en XGBoost se utiliza la importancia basada en la ganancia (gain). Los valores se normalizan para facilitar su representación y comparación dentro de cada modelo.

En Random Forest, la importancia de una variable se determina a partir de la reducción de impureza producida por las divisiones de los nodos en los que interviene. Para un problema de regresión, esta reducción se relaciona con la disminución del error cuadrático utilizado como criterio de división. El resultado se acumula en los diferentes nodos, se pondera de acuerdo con las muestras que alcanzan cada división y posteriormente se promedia entre los árboles del bosque.

Una representación simplificada puede expresarse mediante:

$$Imp(x_i) = \left(\frac{1}{T}\right) \times \sum_{t \in Nt(x_i)} \left(\frac{mn}{m}\right) \times \Delta MSE_n$$

Ecuación 27. Importancia de la variable x_i .

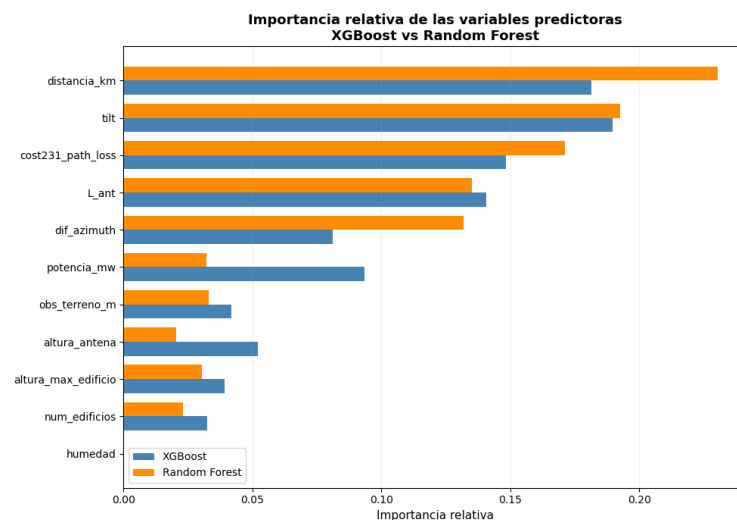
Donde:

- **Imp(x_i)**: importancia de la variable (x_i).
- **T**: número total de árboles del bosque.
- **Nt(x_i)**: conjunto de nodos del árbol (t) en los que interviene la variable x_i .
- **mn**: número de muestras que alcanzan el nodo (n).
- **m**: número total de muestras utilizadas para el entrenamiento.
- **ΔMSE_n** : reducción de impureza producida por la división del nodo (n).

Debido a que Random Forest y XGBoost utilizan criterios diferentes para calcular la importancia de las variables, sus valores individuales no deben interpretarse como directamente equivalentes. La comparación debe realizarse principalmente en términos del orden relativo de importancia dentro de cada modelo.

La coincidencia de las principales variables identificadas por ambos algoritmos proporciona evidencia de que las características relacionadas con la geometría del enlace, la distancia y las pérdidas de propagación tienen una influencia relevante en las predicciones. Sin embargo, esta coincidencia no implica que ambos modelos presenten el mismo desempeño predictivo. Como se detalló en la sección 4.2, Random Forest obtuvo un menor error y un mayor coeficiente de determinación que XGBoost en la estimación del RSRP. Esto sugiere que, si bien ambos algoritmos aprovechan características similares relacionadas con la geometría del enlace y la pérdida de propagación, Random Forest logra combinarlas con mayor precisión en el conjunto de datos evaluado.

Figura 40. Importancia relativa de las variables predictoras para Random Forest y XGBoost.



Nota. Elaboración propia.

4.2.5. Análisis del error en función de la distancia

Para analizar el comportamiento espacial de los modelos, se evaluó el MAE en diferentes rangos de distancia entre la estación base y el punto de medición. A diferencia de la evaluación de desempeño general presentada en la sección 4.2, que se calcula exclusivamente sobre el conjunto de prueba, este análisis se realizó sobre la totalidad de las 4 962 muestras del conjunto de datos (entrenamiento y prueba), con el propósito de caracterizar la distribución espacial del error a lo largo de toda el área de estudio y contar con una cantidad suficiente de observaciones en cada rango de distancia. Los resultados se presentan en la **Tabla 16**.

Tabla 16. Error absoluto medio por rango de distancia.

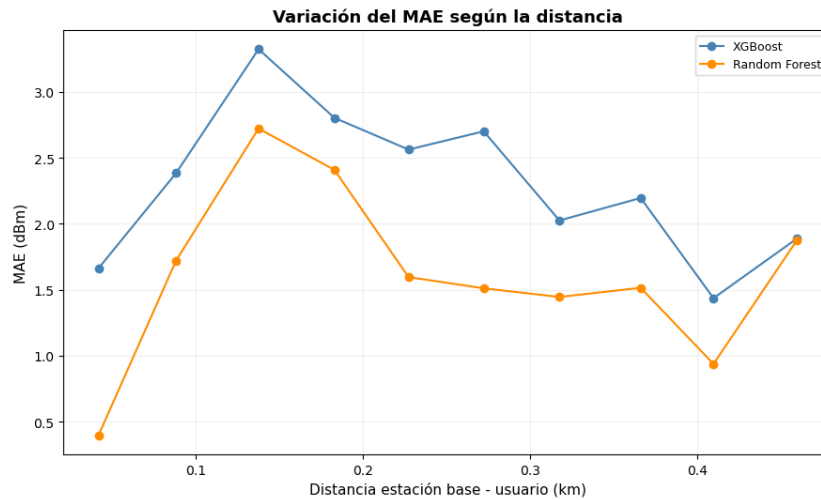
Rango de distancia (km)	Muestras	MAE RF (dB)	MAE XGB (dB)
0,00–0,07	492	0,40	1,66
0,07–0,11	932	1,72	2,38
0,11–0,16	889	2,72	3,32
0,16–0,20	932	2,41	2,80
0,20–0,25	561	1,60	2,56
0,25–0,30	440	1,51	2,70
0,30–0,34	319	1,44	2,02
0,34–0,39	242	1,51	2,20
0,39–0,43	104	0,94	1,44
0,43–0,52	51	1,87	1,89

Nota. Elaboración propia.

El MAE se calculó sobre el conjunto completo de datos (entrenamiento y prueba); por lo tanto, estos valores no deben interpretarse como una medida de la capacidad de generalización del modelo, la cual se evalúa exclusivamente en la sección 4.2 sobre el conjunto de prueba.

La **Figura 41** muestra que ambos modelos presentan variaciones en el error según la distancia respecto de la estación base. El mayor MAE de ambos modelos se observa en el intervalo de 0,11 a 0,16 km, mientras que en los rangos más cercanos los errores son menores. Este comportamiento puede estar relacionado con las características particulares de propagación presentes en cada ubicación, como los obstáculos urbanos, la geometría del enlace y las variaciones de la señal.

Figura 41. Variación del MAE según la distancia entre la estación base y el usuario.

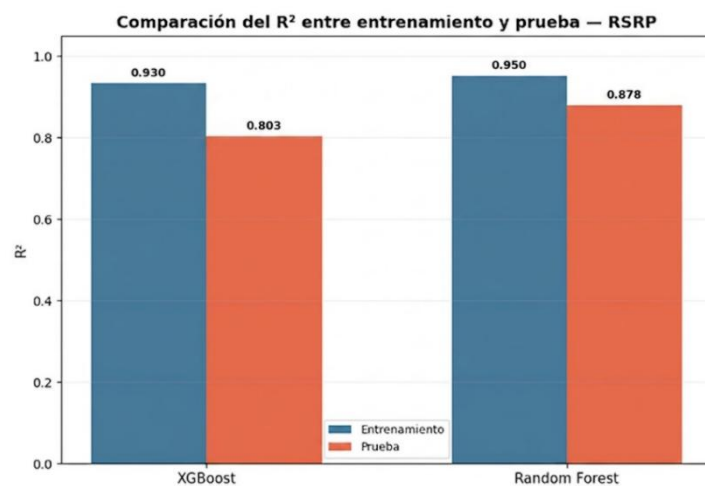


Nota. Elaboración propia.

4.2.6. Análisis de sobreajuste

La **Figura 42** compara el coeficiente de determinación obtenido en los conjuntos de entrenamiento y prueba. Para cuantificar la diferencia entre ambos resultados se define la brecha de sobreajuste como la diferencia absoluta entre los valores de R^2 .

Figura 42. Comparación del coeficiente de determinación entre los conjuntos de entrenamiento y prueba.



Nota. Elaboración propia.

$$\text{Gap}_{\text{overfitting}} = |R^2_{\text{train}} - R^2_{\text{test}}|$$

Ecuación 28. Brecha de sobreajuste.

Donde:

- **Gap_{overfitting}**: brecha de sobreajuste, expresada como una medida adimensional.

- R^2_{train} : coeficiente de determinación obtenido sobre el conjunto de entrenamiento, correspondiente al 80 % de los datos.
- R^2_{test} : coeficiente de determinación obtenido sobre el conjunto de prueba, correspondiente al 20 % de los datos.

Random Forest presentó una brecha de R^2 de 0,0722, mientras que XGBoost registró una brecha de 0,127. La menor diferencia observada en Random Forest indica una mayor estabilidad entre el desempeño obtenido durante el entrenamiento y el alcanzado sobre el conjunto de prueba. En contraste, la mayor brecha de XGBoost evidencia una diferencia más marcada entre ambos conjuntos, compatible con una mayor tendencia al sobreajuste.

Considerando conjuntamente el MAE, el R^2 , la distribución de los residuos y la brecha entre los conjuntos de entrenamiento y prueba, Random Forest fue seleccionado como el modelo con mejor desempeño para el sistema de predicción de cobertura LTE desarrollado en esta investigación.

4.3. Validación del sistema con mediciones reales de campo

Para verificar el funcionamiento del sistema, se realizó una validación utilizando mediciones reales obtenidas mediante G-NetTrack Pro en diferentes ubicaciones del cantón La Libertad. Durante la campaña se registraron parámetros como el RSRP, el RSRQ, el SINR, el PCI y las coordenadas geográficas. Estas ubicaciones se utilizaron como puntos de entrada para generar las estimaciones del sistema, cuyos resultados fueron posteriormente comparados con las mediciones obtenidas en campo mediante métricas de error y análisis visual.

Los diez puntos utilizados en esta validación se registraron en una jornada de medición posterior al entrenamiento de los modelos, por lo que ninguna de estas observaciones formó parte del conjunto de datos empleado para su ajuste. Las mediciones se realizaron en ubicaciones previamente recorridas durante la campaña, de modo que esta validación evalúa la correspondencia del sistema frente a mediciones nuevas obtenidas bajo condiciones de red distintas a las registradas originalmente. La evaluación sobre ubicaciones no observadas durante el entrenamiento se presenta en la sección 4.2, realizada sobre el conjunto de prueba mediante una partición agrupada por punto físico de medición.

De manera complementaria, se empleó el analizador Narda SRM-3006 con sonda isotrópica para caracterizar las señales radioeléctricas presentes en el área de estudio. Las

mediciones se realizaron alrededor de la frecuencia de operación considerada de 2120 MHz, con el propósito de contrastar las condiciones de potencia observadas durante la campaña de medición.

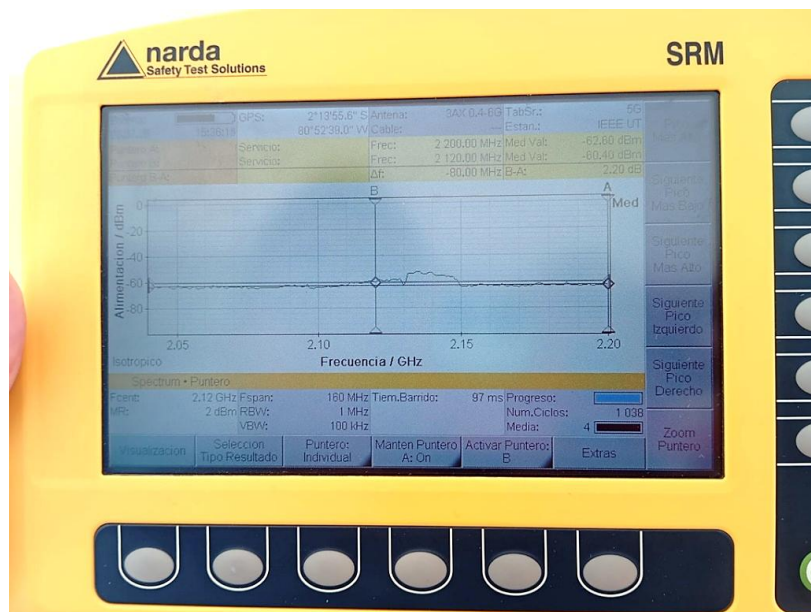
La **Figura 43** muestra el procedimiento de medición en campo con el analizador Narda SRM-3006 y la sonda isotrópica, mientras que la **Figura 44** presenta un ejemplo de la caracterización espectral obtenida durante la campaña.

Figura 43. Medición de campo con el equipo Narda SRM-3006 y sonda isotrópica en el área de estudio.



Nota. Elaboración propia.

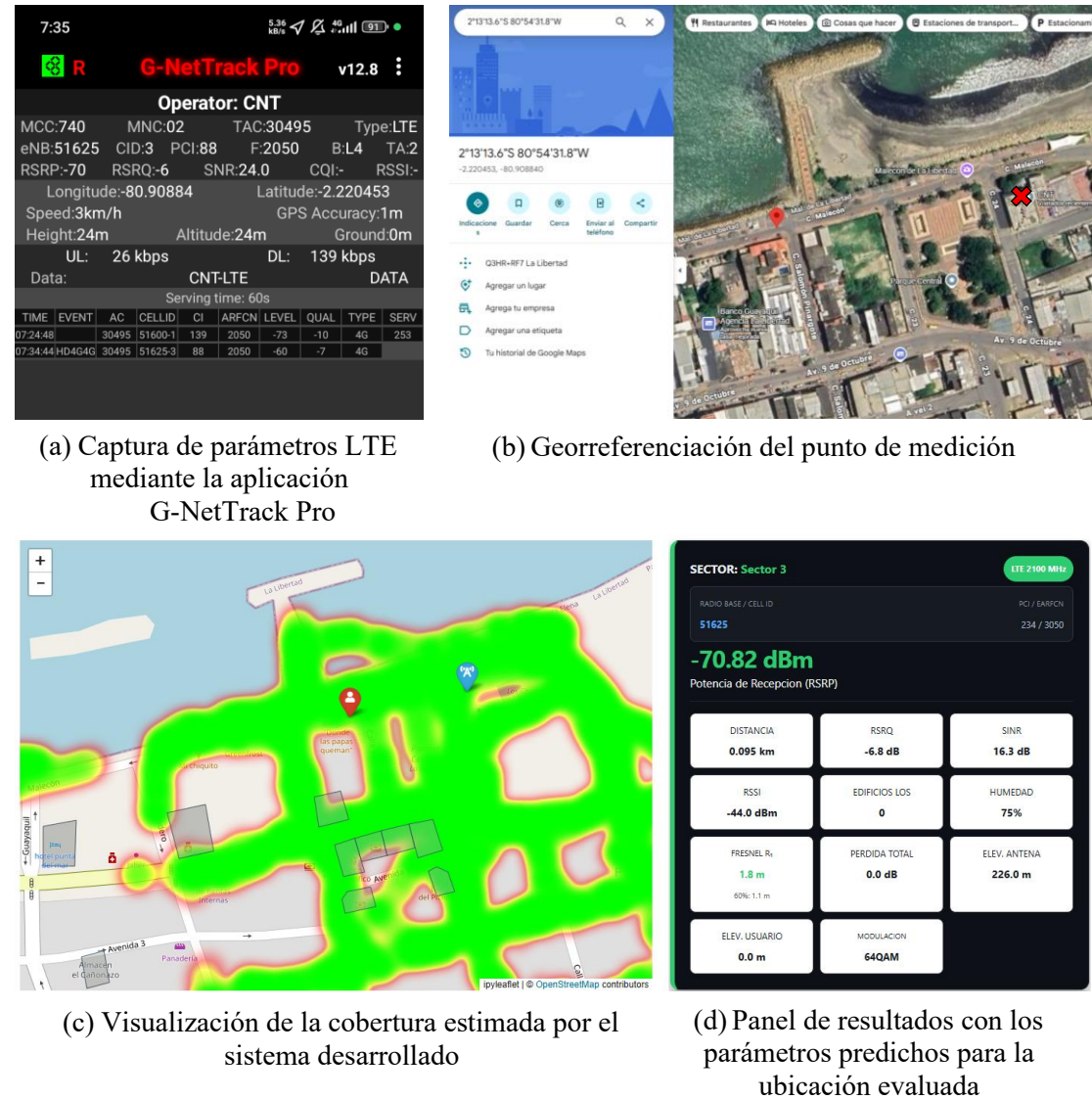
Figura 44. Caracterización espectral de la banda LTE 2120 MHz obtenida con el Narda SRM-3006.



Nota. Elaboración propia.

La **Figura 45** sintetiza el proceso de validación. Primero se obtuvieron las mediciones reales mediante G-NetTrack Pro, incluyendo el RSRP, el RSRQ, el SINR, el PCI y las coordenadas geográficas. Posteriormente, se verificó la información espacial de los puntos registrados. Una vez establecida la ubicación de cada muestra, el sistema generó las estimaciones correspondientes y las representó mediante un mapa de calor interactivo y un panel de visualización.

Figura 45. Proceso de validación del sistema mediante mediciones reales de campo.



Nota. Elaboración propia.

4.3.1. Comparación entre valores medidos y estimados

La **Tabla 17** presenta las coordenadas correspondientes a los diez puntos de validación, distribuidos dentro del área de estudio con el propósito de evaluar el desempeño del modelo bajo diferentes condiciones de propagación.

Tabla 17. Ubicación de los puntos de validación.

Punto	Latitud	Longitud
P01	-2.221823	-80.90841
P02	-2.219721	-80.90711
P03	-2.220453	-80.90884
P04	-2.221035	-80.90791
P05	-2.220455	-80.90644
P06	-2.220765	-80.90970
P07	-2.221321	-80.90959
P08	-2.220613	-80.90962
P09	-2.220644	-80.90971
P10	-2.220766	-80.91025

Nota. Elaboración propia.

La **Tabla 18** presenta la comparación entre los valores de RSRP medidos en campo y los valores estimados por el sistema para los diez puntos de evaluación. En general, se observa una buena correspondencia entre ambos conjuntos de valores. La mayoría de los puntos presenta errores absolutos inferiores a 3 dB, mientras que el mayor error se registró en P01, con un valor de 3,29 dB.

Tabla 18. Comparación entre mediciones reales y valores estimados por el sistema.

Punto	RSRP real (dBm)	RSRP estimado (dBm)	Error absoluto (dB)
P01	-64	-67,29	3,29
P02	-83	-80,65	2,35
P03	-70	-70,82	0,82
P04	-67	-67,55	0,55
P05	-79	-77,25	1,75
P06	-83	-82,75	0,25
P07	-84	-83,88	0,12
P08	-63	-64,76	1,76
P09	-68	-70,90	2,90
P10	-72	-73,35	1,35

Nota. Elaboración propia.

En el punto P03, la medición de campo fue de $-70,00$ dBm, mientras que la estimación alcanzó $-70,82$ dBm, lo que representa una diferencia de $0,82$ dB. Asimismo, los puntos P06 y P07 presentan diferencias de apenas $0,25$ dB y $0,12$ dB, respectivamente, evidenciando una elevada correspondencia entre los valores medidos y estimados. Por otra parte, los mayores errores se registraron en P01, con $3,29$ dB, y P09, con $2,90$ dB. Estas diferencias pueden asociarse a condiciones locales del entorno, como la presencia de edificaciones, efectos de multitrayecto y variaciones propias del canal radioeléctrico que no son completamente representadas por las variables de entrada del modelo.

En conjunto, los resultados muestran una correspondencia adecuada entre los valores reales y estimados de RSRP en los puntos de validación, lo que evidencia la capacidad del sistema para representar de manera consistente las condiciones de cobertura LTE en el área evaluada.

A continuación, se valida el desempeño del sistema para los parámetros de calidad de señal RSRQ, SINR y CQI, así como para los parámetros de rendimiento DL Throughput y UL Throughput, utilizando los mismos puntos de evaluación. Los resultados se presentan en las **Tablas 19** y **20**. Los valores estimados que se presentan a continuación fueron obtenidos mediante los modelos entrenados para cada variable objetivo, evaluados sobre los mismos diez puntos utilizados en la validación del RSRP.

Tabla 19. Validación de los parámetros de calidad de señal (RSRQ, SINR y CQI).

Punto	RSRQ real (dB)	RSRQ estimado (dB)	SINR real (dB)	SINR estimado (dB)	CQI real	CQI estimado
P01	-8	-8,2	16	17,1	10	11
P02	-10	-9	9	14,7	9	10
P03	-6	-6,8	24	16,3	11	12
P04	-8	-8,8	12	12,6	10	9
P05	-7	-7,2	14	15	11	11
P06	-8	-8,2	10	11,4	9	9
P07	-9	-8,2	6	8	10	11
P08	-9	-7,5	6	7,4	12	13
P09	-7	-8,2	7	8,7	9	8
P10	-7	-8,5	8	6,7	7	7

Nota. Elaboración propia.

Los resultados de la **Tabla 19** muestran una correspondencia generalmente adecuada en el parámetro RSRQ, con diferencias relativamente pequeñas entre los valores medidos y estimados. En el caso del SINR, se observan diferencias más pronunciadas en determinados puntos, especialmente en P02 y P03. En P03, por ejemplo, el valor real fue de 24,0 dB, mientras que el estimado fue de 16,3 dB. Esta diferencia puede estar relacionada con la naturaleza dinámica del SINR, debido a que este parámetro depende no solo de la potencia de la señal recibida, sino también de las condiciones de interferencia y ruido presentes durante la medición.

En cuanto al CQI, los valores estimados mantienen una tendencia similar a los registrados en campo, con diferencias de entre cero y dos unidades en los puntos analizados. Esto indica que el modelo logra representar de manera aproximada la calidad de enlace asociada a las condiciones de propagación observadas.

Tabla 20. Validación de los parámetros de rendimiento (DL y UL Throughput).

Punto	DL real (Mbps)	DL estimado (Mbps)	UL real (Mbps)	UL estimado (Mbps)
P01	37,10	29,98	11,13	8,99
P02	14,05	26,45	4,22	7,93
P03	5,40	4,77	1,62	1,43
P04	2,02	2,89	0,60	0,87
P05	3,61	2,82	1,08	0,85
P06	1,34	0,81	0,40	0,24
P07	0,36	0,76	0,11	0,23
P08	1,47	1,00	0,44	0,30
P09	1,07	0,89	0,32	0,27
P10	1,32	1,00	0,40	0,30

Nota. Elaboración propia.

En la **Tabla 20** se observa que los valores estimados de Throughput presentan una tendencia general acorde con las mediciones reales, aunque se identifican diferencias en determinados puntos. Por ejemplo, en P02 el Throughput de descarga medido fue de 14,05 Mbps, mientras que el valor estimado alcanzó 26,45 Mbps. De manera similar, en P07 se obtuvo un valor real de 0,36 Mbps frente a una estimación de 0,76 Mbps.

Estas diferencias pueden atribuirse a que el Throughput no depende exclusivamente de las condiciones radioeléctricas, sino también de factores adicionales relacionados con la operación de la red, como la carga de la celda, la disponibilidad de recursos radio, el tráfico de usuarios, las condiciones de modulación y codificación y las variaciones temporales de la red. Por ello, las estimaciones de DL y UL Throughput deben interpretarse como una aproximación del rendimiento esperado bajo las condiciones de señal modeladas.

4.4. Análisis comparativo

Esta sección contrasta los resultados obtenidos con investigaciones previas relacionadas con la predicción de cobertura LTE mediante técnicas de aprendizaje automático, considerando las variables utilizadas, las métricas de evaluación, los modelos implementados y los métodos empleados para la adquisición de datos.

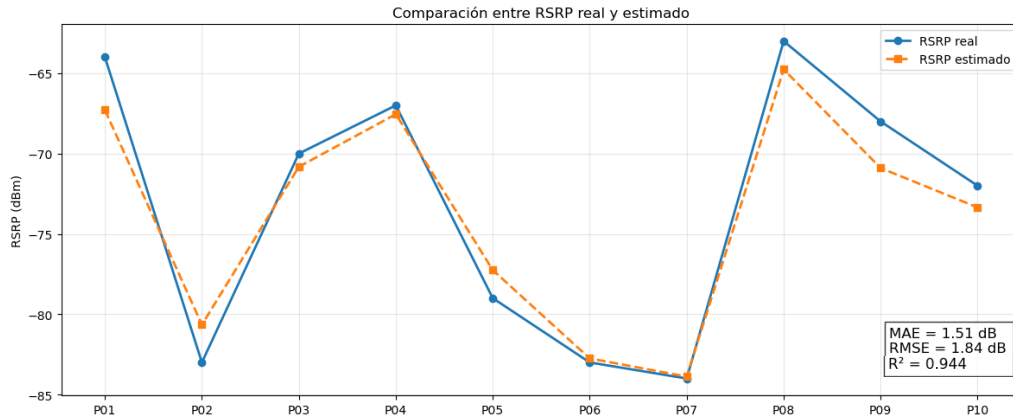
La incorporación de variables geoespaciales, ambientales y de propagación permitió al modelo considerar diferentes condiciones asociadas al comportamiento de la señal dentro del área de estudio. Esta combinación de variables contribuyó a representar las características particulares del entorno y permitió obtener las estimaciones de cobertura presentadas en esta investigación.

A diferencia de los modelos de propagación convencionales, que se basan principalmente en relaciones matemáticas establecidas entre la distancia, la frecuencia y las características generales del entorno, el enfoque desarrollado incorpora información obtenida mediante mediciones reales y variables adicionales relacionadas con las condiciones locales. Esto permite adaptar la estimación a las características específicas del área de estudio y complementar los modelos tradicionales de propagación.

4.4.1. Análisis gráfico del error de predicción

La **Figura 46** presenta la comparación entre los valores de RSRP medidos en campo y los valores estimados por el sistema para los diez puntos de validación. Se observa que las predicciones mantienen una tendencia similar a las mediciones reales. Los puntos P03, P04, P06 y P07 presentan las menores diferencias, con errores absolutos inferiores a 1 dB, lo que evidencia una elevada correspondencia entre ambas variables.

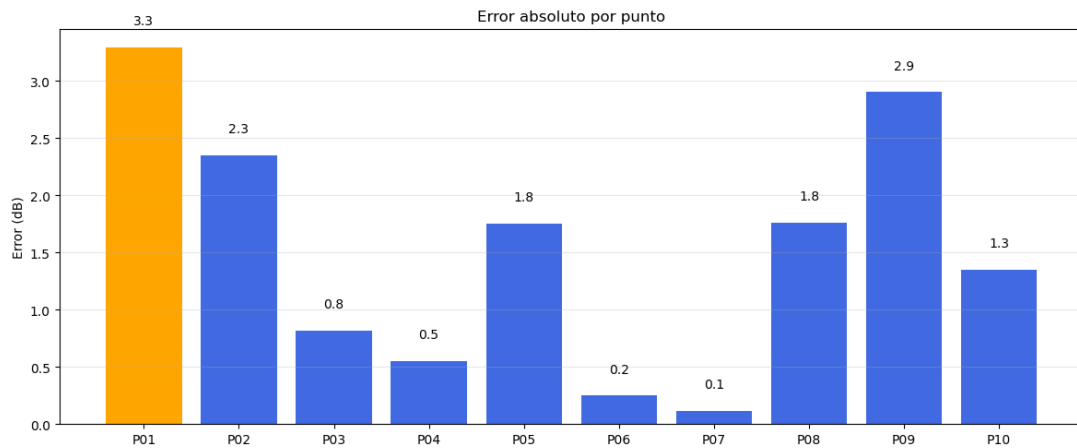
Figura 46. Comparación entre valores reales y estimados de RSRP.



Nota. Elaboración propia.

La **Figura 47** muestra el error absoluto obtenido en cada punto de validación. El mayor error corresponde a P01, con 3,29 dB, seguido de P09, con 2,90 dB. Estas diferencias pueden estar relacionadas con condiciones locales de propagación, como la presencia de edificaciones, efectos de multitrayecto, variaciones del canal radioeléctrico y otros factores que no son completamente representados mediante las variables de entrada utilizadas por el modelo. En contraste, los puntos P06 y P07 presentan los menores errores, con valores de 0,25 dB y 0,12 dB, respectivamente.

Figura 47. Error absoluto por punto de validación.



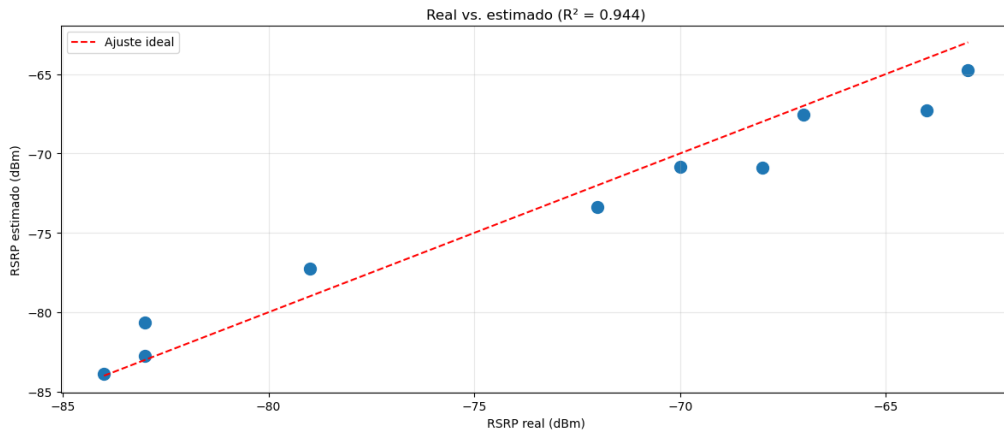
Nota. Elaboración propia.

Considerando los diez puntos de validación, se obtuvo un error absoluto medio (MAE) de 1,51 dB y una raíz del error cuadrático medio (RMSE) de 1,84 dB. Estos resultados indican que, en promedio, las estimaciones de RSRP se encuentran próximas a los valores registrados en campo, evidenciando un desempeño adecuado del modelo en los puntos evaluados.

El error obtenido en esta validación es inferior al registrado sobre el conjunto de prueba, que fue de 1,82 dB. Este comportamiento resulta coherente con el hecho de que las ubicaciones evaluadas corresponden a puntos ya recorridos durante la campaña, cuyas condiciones de propagación fueron representadas en el conjunto de entrenamiento, y no debe interpretarse como una mejora del desempeño de los modelos.

La **Figura 48** presenta el diagrama de dispersión entre los valores de RSRP medidos y estimados. La mayoría de los puntos se encuentra próxima a la diagonal de referencia, lo que evidencia una correspondencia elevada entre ambas variables. El coeficiente de determinación obtenido fue de $R^2 = 0,944$, lo que indica que aproximadamente el 94,4 % de la variabilidad observada en los valores medidos de RSRP es explicada por las estimaciones del modelo para este conjunto de validación.

Figura 48. Diagrama de dispersión entre los valores medidos y estimados de RSRP.



Nota. Elaboración propia.

En términos generales, los resultados obtenidos evidencian un desempeño satisfactorio del sistema en la estimación de RSRP para los puntos evaluados. Aunque se presentan algunas desviaciones puntuales, particularmente en P01 y P09, el comportamiento global muestra una elevada correspondencia entre las mediciones de campo y las predicciones. Por lo tanto, la propuesta desarrollada presenta condiciones adecuadas para apoyar el análisis de cobertura y las actividades de planificación radioeléctrica en el área de estudio.

4.4.2. Comparación con investigaciones relacionadas

Con el propósito de identificar los principales aportes de esta investigación frente a trabajos previos, se realizó una comparación considerando la metodología empleada, las fuentes de información, los modelos de aprendizaje automático, las variables utilizadas,

los parámetros estimados y los mecanismos de validación. Los resultados de esta comparación se presentan en la **Tabla 21**.

Tabla 21. Comparación entre investigaciones relacionadas y la propuesta desarrollada.

Criterio	Carrión (2022)	Gutiérrez (2021)	Párraga (2023)	Propuesta desarrollada
Datos reales de campo	✓	✗	✓	✓
Aprendizaje automático	✓	✓	✓	✓
Random Forest	✓	✓	✗	✓
XGBoost	✓	✓	✗	✓
Variables geográficas	Parcial	✓	✓	✓
Variables ambientales	✗	Parcial	✗	✓
Visualización interactiva	✗	✗	✗	✓
Predicción de RSRP	✓	✓	✓	✓
Predicción de RSRQ	✗	✗	✗	✓
Predicción de SINR	✗	✗	✗	✓
Predicción de CQI	✗	✗	✗	✓
Predicción de Throughput	✓	Parcial	✗	✓
Validación con mediciones propias	Parcial	✗	✗	✓

Nota. Elaboración propia.

La comparación evidencia que las investigaciones analizadas presentan diferentes enfoques para la predicción de cobertura y calidad de señal en redes móviles. En general, los trabajos revisados emplean técnicas de aprendizaje automático para estimar principalmente el comportamiento del RSRP y, en algunos casos, parámetros relacionados con el rendimiento de la red. Sin embargo, existen diferencias en cuanto a las fuentes de información, las variables utilizadas y los procedimientos de validación.

Una de las principales diferencias de la propuesta desarrollada corresponde al uso de mediciones reales de campo como fuente de información para la construcción y

validación del modelo. Estas mediciones fueron obtenidas en el área de estudio y complementadas con variables geoespaciales, ambientales y de propagación, permitiendo incorporar características particulares del entorno urbano y costero del cantón La Libertad. De esta manera, el modelo no se limita únicamente a variables asociadas con la distancia o a valores derivados de modelos teóricos de propagación, sino que considera información relacionada con las condiciones observadas en la red LTE evaluada.

En cuanto a los modelos de aprendizaje automático, la propuesta implementa Random Forest y XGBoost, lo que permite comparar el desempeño de ambos algoritmos y determinar cuál obtiene mejores resultados en la predicción de las variables analizadas. Esta comparación permite identificar el modelo que presenta un mejor desempeño con el conjunto de datos obtenido en el área de estudio.

Otro aspecto diferenciador corresponde al número de parámetros considerados. Mientras que las investigaciones revisadas se concentran principalmente en la predicción del RSRP, la propuesta desarrollada extiende el análisis hacia otros indicadores de calidad y rendimiento de la red, incluyendo RSRP, RSRQ, SINR, CQI, DL Throughput y UL Throughput. Esta integración permite obtener una caracterización más amplia de las condiciones radioeléctricas y del rendimiento esperado de la red LTE.

Asimismo, el sistema incorpora un módulo de visualización interactiva, mediante el cual los resultados pueden representarse espacialmente a través de mapas de cobertura, mapas de calor, perfiles topográficos y diferentes indicadores de desempeño. Esta funcionalidad facilita la interpretación de los resultados y permite identificar sectores con condiciones de señal desfavorables que podrían requerir análisis o acciones de optimización.

Finalmente, la propuesta incluye una etapa de validación mediante mediciones propias de campo, utilizando puntos independientes de evaluación para comparar los valores medidos con las estimaciones generadas por el sistema. Esta característica permite evaluar el comportamiento del modelo bajo condiciones reales de propagación y proporciona evidencia adicional sobre su capacidad para representar las condiciones de cobertura del área estudiada.

En conjunto, la principal contribución de la propuesta no radica únicamente en la utilización de técnicas de aprendizaje automático, sino en la integración de datos reales de campo, variables geoespaciales, ambientales y de propagación, modelos de

aprendizaje automático, estimación de múltiples parámetros y herramientas de visualización interactiva dentro de un mismo sistema orientado al análisis de cobertura LTE. Por lo tanto, el sistema desarrollado constituye una alternativa complementaria para apoyar el análisis y la planificación de redes en el área de estudio.

4.4.3. Comparación con herramientas de planificación radioeléctrica

Además de comparar la propuesta con investigaciones académicas previas, resulta pertinente contrastarla con herramientas utilizadas en el mismo ámbito, como Atoll. Estas plataformas permiten realizar análisis de propagación y representación geográfica de la cobertura mediante modelos de propagación, información cartográfica y parámetros de configuración de los sistemas de comunicaciones. La comparación se presenta en la **Tabla 22**.

Tabla 22. Comparación entre herramientas de planificación radioeléctrica y el sistema desarrollado.

Característica	Atoll	Sistema desarrollado
Licencia comercial	Sí	No
Mapa de calor	✓	✓
Visualización geográfica	✓	✓
Modelos aprendizaje automático	✗	✓
Predicción automática de RSRP	✓	✓
Predicción de RSRQ	✓	✓
Predicción de SINR	✓	✓
Predicción de CQI	✗	✓
Predicción de DL Throughput	✗	✓
Predicción de UL Throughput	✗	✓
Integración con datos de campo	Parcial	✓
Dashboard interactivo	✗	✓
Código modificable	✗	✓

Nota. Elaboración propia.

Los resultados muestran que ambas alternativas permiten representar espacialmente la cobertura mediante herramientas de visualización geográfica y mapas de intensidad de

señal. Sin embargo, existen diferencias importantes en cuanto al enfoque utilizado y al propósito de cada solución.

Atoll constituye una plataforma especializada para la planificación y optimización de redes de comunicaciones inalámbricas, por lo que dispone de un conjunto de funcionalidades más amplio para el diseño de redes, configuración de estaciones, análisis de cobertura y evaluación de diferentes escenarios de propagación. Su principal ventaja se encuentra en el enfoque integral de planificación radioeléctrica y en las herramientas especializadas que incorpora para el diseño de redes.

El sistema desarrollado presenta un enfoque diferente, debido a que está orientado principalmente a la predicción mediante técnicas de aprendizaje automático a partir de datos obtenidos de mediciones reales. En particular, incorpora los modelos Random Forest y XGBoost para procesar las variables disponibles y generar estimaciones relacionadas con la calidad y el rendimiento de la red LTE.

Otra diferencia corresponde a la cantidad de indicadores integrados en el proceso de análisis. Mientras que las herramientas tradicionales de planificación se orientan principalmente a la estimación y representación de parámetros radioeléctricos asociados con la cobertura, el sistema permite obtener estimaciones del RSRP, el RSRQ, el SINR, el CQI, el DL Throughput y el UL Throughput, proporcionando una visión complementaria de la calidad y el rendimiento esperado de la red.

Asimismo, el sistema incorpora un dashboard interactivo que permite visualizar los resultados de manera dinámica mediante mapas de calor, indicadores y representaciones geográficas. Esta característica facilita la interpretación de los resultados obtenidos por los modelos y permite relacionarlos con la ubicación espacial de los puntos analizados.

Una ventaja adicional de la propuesta corresponde a su carácter modificable y adaptable, debido a que el código y los modelos pueden ser ajustados para incorporar nuevos datos, variables o algoritmos de aprendizaje automático. Esto permite adaptar el sistema a diferentes escenarios de estudio sin depender exclusivamente de una configuración predeterminada.

No obstante, es importante señalar que la propuesta desarrollada no pretende sustituir herramientas profesionales de planificación radioeléctrica como Atoll, ya que estas plataformas cuentan con funcionalidades especializadas y un alcance mayor para el diseño integral de redes. La propuesta se orienta principalmente al análisis predictivo

basado en aprendizaje automático y puede considerarse una herramienta complementaria para estudiar el comportamiento de la cobertura a partir de mediciones reales.

En este sentido, la principal ventaja de la propuesta consiste en combinar datos reales de campo, variables geoespaciales, ambientales y radioeléctricas, algoritmos de aprendizaje automático y visualización interactiva en un único entorno de análisis. Esto permite complementar los métodos tradicionales de planificación y proporcionar una perspectiva basada en datos sobre las condiciones de cobertura y rendimiento de la red LTE en el cantón La Libertad.

4.5. Propuesta de mejora para la planificación de la red LTE de CNT

A partir de los resultados obtenidos mediante los modelos predictivos y el sistema de visualización, se plantea una propuesta de mejora orientada a apoyar la planificación y optimización de la red LTE de CNT en el cantón La Libertad. La propuesta consiste en utilizar el sistema desarrollado como una herramienta de apoyo para evaluar diferentes escenarios de configuración radioeléctrica antes de realizar modificaciones sobre la infraestructura existente.

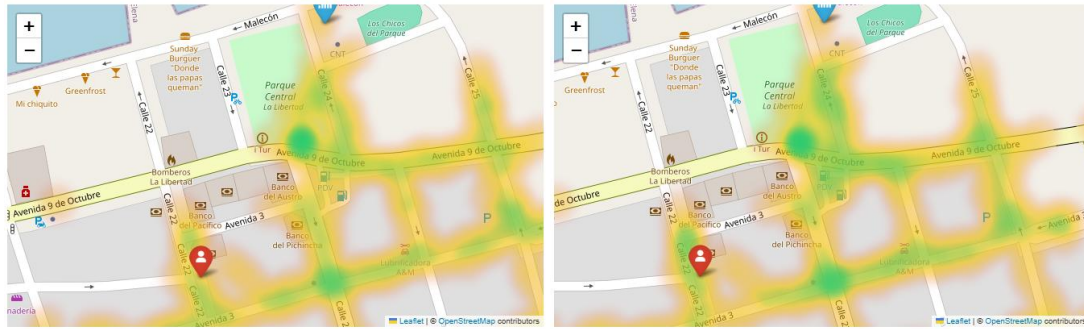
El sistema permite simular diferentes configuraciones de los sectores de una estación base mediante la variación de parámetros como el azimut, el tilt y la potencia de transmisión. Para cada escenario, se estima el RSRP y se representa espacialmente la distribución de la cobertura mediante mapas de calor. Esto permite comparar diferentes alternativas de configuración e identificar aquellas que presentan mejores condiciones de señal en sectores específicos del área de estudio.

De esta manera, el sistema puede utilizarse como una herramienta complementaria para la toma de decisiones, permitiendo evaluar previamente los posibles efectos de una modificación en la configuración de la estación base y reducir la necesidad de realizar múltiples pruebas directamente sobre la infraestructura.

4.5.1. Caso 1: sector 2

Para demostrar esta funcionalidad, se compararon dos escenarios correspondientes al sector 2, considerando un usuario ubicado en las coordenadas -2.221921 , -80.908288 . El escenario A corresponde a la configuración inicial de la antena, mientras que el escenario B incorpora modificaciones en el tilt y la potencia de transmisión, manteniendo constante el azimut. Los resultados se presentan en la **Figura 49**.

Figura 49. Comparación del mapa de cobertura entre el escenario original y el escenario optimizado.



a) Escenario A

b) Escenario B

Nota. Elaboración propia.

En el escenario inicial, el sector 2 presentó un RSRP estimado de $-82,7$ dBm en el punto de evaluación. En el escenario optimizado, el ajuste de los parámetros modificó la distribución espacial de la cobertura y permitió obtener un valor estimado de $-77,8$ dBm. La representación mediante el mapa de calor permite observar una mayor intensidad de señal en el área correspondiente al punto de evaluación. La **Tabla 23** recoge los valores cuantitativos.

Tabla 23. Comparación entre el escenario original y el escenario optimizado.

Parámetro	Escenario A	Escenario B
Azimut sector 2	161°	161°
Tilt sector 2	6°	8°
Potencia	20 W (43 dBm)	50 W (47 dBm)
RSRP estimado	$-82,7$ dBm	$-77,8$ dBm
Mejora obtenida	—	+4,9 dB

Nota. Elaboración propia.

El ajuste del tilt de 6° a 8° y el incremento de la potencia de transmisión de 20 W (43 dBm) a 50 W (47 dBm), manteniendo el azimut en 161° , produjo un aumento del RSRP estimado de $-82,7$ dBm a $-77,8$ dBm. Esto representa una mejora de 4,9 dB en el punto de evaluación.

La mejora observada es resultado de la modificación simultánea de dos parámetros de configuración. Por una parte, el incremento de la potencia de transmisión aumenta el nivel de potencia disponible en el receptor; por otra, la variación del tilt altera la distribución vertical del haz y, en consecuencia, la zona de cobertura del sector. La

combinación de ambos ajustes permite obtener un nivel de señal superior en el punto analizado.

De acuerdo con los umbrales de referencia utilizados en el sistema, el valor obtenido de $-77,8$ dBm se encuentra dentro del rango considerado como cobertura favorable. Sin embargo, esta clasificación debe interpretarse en función de los criterios establecidos para el estudio y no implica, por sí sola, una garantía de mejora en todos los indicadores de calidad o rendimiento de la red.

Los sectores que no proporcionan servicio principal al punto de evaluación presentan valores de RSRP cercanos a -145 dBm y no experimentan modificaciones relevantes entre los escenarios, debido a que sus parámetros de configuración permanecen constantes.

4.5.2. Caso 2: sector 1

Con el propósito de analizar de manera independiente el efecto de la potencia de transmisión, se simuló una modificación de este parámetro manteniendo constantes el azimut y el tilt. La **Figura 50** presenta la comparación entre dos escenarios para un usuario ubicado en las coordenadas $-2.221900, -80.908382$.

Figura 50. Comparación de la cobertura LTE antes y después del incremento de la potencia de transmisión.



Nota. Elaboración propia.

En el escenario A se mantuvo la configuración inicial, mientras que en el escenario B se incrementó la potencia de transmisión del sector 1 de 20 W (43 dBm) a 40 W (46 dBm), manteniendo el azimut en 355° y el tilt en 6° . Los resultados cuantitativos se presentan en la **Tabla 24**.

Tabla 24. *Parámetros de configuración evaluados para ambos escenarios.*

Parámetro	Escenario A	Escenario B
Azimut sector 1	355°	355°
Tilt sector 1	6°	6°
Potencia	20 W (43 dBm)	40 W (46 dBm)
RSRP estimado	−82,7 dBm	−79,3 dBm
Mejora obtenida	—	+3,4 dB

Nota. Elaboración propia.

El incremento de la potencia de transmisión produjo un aumento del RSRP estimado de −82,7 dBm a −79,3 dBm, equivalente a una mejora de 3,4 dB en el punto de evaluación.

Cabe señalar que los Casos 1 y 2 parten del mismo valor de RSRP estimado (−82,7 dBm) pese a corresponder a sectores con azimuts distintos. Esto se debe a que ambos puntos de evaluación están separados entre sí por aproximadamente 11 m y se ubican a una distancia prácticamente idéntica de la estación base (unos 210 m). En consecuencia, comparten las mismas condiciones de distancia, obstrucción del terreno y edificaciones en el trayecto, que son las variables de mayor peso en la estimación. Por esta razón, ambos casos deben interpretarse como la evaluación de dos sectores sobre una misma zona y no como escenarios geográficamente independientes. El Caso 3, en cambio, corresponde a una ubicación distinta, ya que su punto de evaluación se sitúa en el borde de celda.

El resultado es consistente con el comportamiento esperado al incrementar la potencia de transmisión. En condiciones ideales y manteniendo constantes los demás parámetros, una duplicación de la potencia corresponde a un incremento teórico de aproximadamente 3 dB en el nivel de potencia recibido. La diferencia de 3,4 dB obtenida en la simulación puede estar relacionada con las características específicas del modelo utilizado y con la forma en que se representan las variables de propagación.

Con el valor estimado de −79,3 dBm, el punto de evaluación alcanza un nivel de señal superior a −80 dBm, de acuerdo con el umbral de referencia empleado en el sistema. Esto evidencia que una modificación relativamente sencilla de la potencia puede producir una mejora apreciable en el nivel de la señal en determinadas ubicaciones, aunque su

aplicación real debería considerar también aspectos como la interferencia, el consumo energético, los límites de potencia y la planificación de frecuencias.

Las pequeñas variaciones observadas en otros sectores corresponden a cambios derivados del comportamiento del modelo de simulación y no representan una modificación intencional de su configuración, debido a que los parámetros de dichos sectores permanecen constantes entre ambos escenarios.

4.5.3. Caso 3: sector 3

Para evaluar una estrategia de optimización combinada, se modificaron simultáneamente el tilt y la potencia de transmisión del sector 3, manteniendo constante el azimut. El usuario se ubicó en una posición próxima al borde de la celda, correspondiente a las coordenadas $-2.220774, -80.910712$. La **Figura 51** presenta la comparación entre el escenario inicial y el escenario propuesto.

En el escenario A, el punto de evaluación presentó un RSRP estimado de $-87,8$ dBm. En el escenario B, la modificación de los parámetros permitió obtener un valor estimado de $-79,9$ dBm. En consecuencia, se obtuvo una mejora de $7,9$ dB en el punto analizado.

Figura 51. Comparación de escenarios de cobertura LTE antes y después de la optimización.



a) Escenario A

b) Escenario B

Nota. Elaboración propia.

La propuesta consistió en reducir el tilt de 6° a 2° e incrementar la potencia de transmisión de 20 W (43 dBm) a 80 W (49 dBm), manteniendo el azimut en 240° . Los resultados cuantitativos se presentan en la **Tabla 25**.

Tabla 25. Comparación entre escenarios de configuración del sector 3.

Parámetro	Escenario A	Escenario B
Azimut sector 3	240°	240°
Tilt sector 3	6°	2°
Potencia	20 W (43 dBm)	80 W (49 dBm)
RSRP estimado	-87,8 dBm	-79,9 dBm
Mejora obtenida	—	+7,9 dB

Nota. Elaboración propia.

El ajuste del tilt permite alterar la distribución vertical del haz y favorecer la cobertura en una zona más alejada de la estación, mientras que el incremento de potencia contribuye a elevar el nivel de la señal recibida.

La combinación de ambos parámetros produjo un incremento del RSRP de 7,9 dB, pasando de -87,8 dBm a -79,9 dBm. De acuerdo con los umbrales de referencia empleados en el sistema, el valor final se encuentra ligeramente por encima del límite de -80 dBm establecido para el nivel de cobertura favorable considerado en este estudio. Este resultado demuestra que la modificación conjunta de parámetros puede generar un efecto más significativo que la modificación individual de un único parámetro en determinadas condiciones de propagación.

No obstante, este escenario debe interpretarse como una simulación de apoyo a la toma de decisiones y no como una recomendación inmediata de implementación en la infraestructura real. Un incremento de potencia hasta 80 W (49 dBm) y un cambio significativo del tilt deben ser evaluados considerando aspectos adicionales, como la interferencia entre sectores, el solapamiento de cobertura, la capacidad de la estación para atender la demanda de tráfico, el consumo energético, la normativa aplicable y las condiciones de operación de la red.

En definitiva, los casos analizados demuestran que el sistema no se limita a estimar parámetros de cobertura, sino que también puede utilizarse como una herramienta de apoyo para evaluar diferentes alternativas de configuración radioeléctrica. La posibilidad de ajustar virtualmente parámetros como el azimut, el tilt y la potencia permite comparar escenarios antes de realizar intervenciones sobre la infraestructura física.

Este enfoque puede contribuir a reducir el número de pruebas necesarias en campo, facilitar la identificación de configuraciones potencialmente favorables y apoyar la toma

de decisiones durante procesos de planificación y optimización de la red LTE. Sin embargo, los resultados obtenidos mediante simulación deben complementarse con mediciones de campo y análisis de ingeniería antes de implementar cualquier cambio en una estación base real.

4.6. Discusión de resultados

Los resultados obtenidos muestran que Random Forest presentó un mejor desempeño que XGBoost en la estimación de los parámetros evaluados, al registrar menores errores y mejores indicadores de ajuste para el conjunto de datos analizado. En el caso del RSRP, considerado el parámetro principal para el análisis de cobertura, Random Forest alcanzó un MAE de 1,82 dB y un coeficiente de determinación de $R^2 = 0,964$, mientras que XGBoost obtuvo un MAE de 2,55 dB y un R^2 de 0,940. Estos resultados indican que Random Forest logró representar con mayor precisión la relación entre las variables de entrada y el comportamiento de la señal en el área de estudio.

La comparación del desempeño entre el entrenamiento y la prueba también mostró una menor brecha de sobreajuste para Random Forest. Este comportamiento sugiere una mayor estabilidad del modelo frente a datos no utilizados durante su entrenamiento. Esta característica resulta especialmente relevante en el ámbito de la planificación radioeléctrica, debido a que un modelo destinado a apoyar la toma de decisiones debe mantener un comportamiento consistente cuando se aplica sobre ubicaciones o condiciones diferentes a aquellas utilizadas durante su construcción.

La validación mediante mediciones reales de campo permitió complementar la evaluación realizada sobre el conjunto de datos. En los diez puntos de validación se obtuvo un MAE de 1,51 dB y un RMSE de 1,84 dB para el RSRP, mientras que el coeficiente de determinación alcanzó un $R^2 = 0,944$. Estos valores corresponden al conjunto específico de puntos utilizado para la validación final y no deben confundirse con las métricas obtenidas durante la evaluación comparativa de los modelos. La mayoría de los puntos presentó diferencias inferiores a 3 dB, aunque se identificaron algunas desviaciones puntuales asociadas a las características particulares del entorno. Esto evidencia que el modelo mantiene una correspondencia adecuada con las mediciones realizadas en condiciones reales dentro del área evaluada.

Las diferencias observadas entre las mediciones y las estimaciones pueden explicarse, en parte, por fenómenos propios de la propagación radioeléctrica, como el multitrayecto,

la presencia de edificaciones y otras obstrucciones, así como por variaciones del canal que no siempre pueden ser representadas completamente mediante las variables utilizadas por el modelo. Estos resultados ponen de manifiesto que, aunque las técnicas de aprendizaje automático permiten establecer relaciones complejas entre diferentes variables, su precisión depende de la calidad, cantidad y representatividad de los datos utilizados durante el entrenamiento.

Los resultados obtenidos son consistentes con el enfoque planteado en investigaciones previas sobre la aplicación de técnicas de aprendizaje automático a la predicción de cobertura en redes móviles. Sin embargo, la propuesta desarrollada amplía este enfoque mediante la incorporación conjunta de variables geoespaciales, ambientales y de propagación, obtenidas o derivadas a partir de información del área de estudio y mediciones de campo. Además, el sistema integra herramientas de visualización geoespacial que permiten representar mapas de calor, perfiles topográficos y diferentes parámetros radioeléctricos. Esta integración facilita la interpretación espacial de los resultados y proporciona información adicional para analizar las condiciones de cobertura.

Desde la perspectiva de la ingeniería de telecomunicaciones, uno de los principales aportes de la propuesta es la posibilidad de evaluar virtualmente diferentes escenarios de configuración radioeléctrica antes de realizar modificaciones sobre la infraestructura existente. Las simulaciones realizadas demostraron que la modificación conjunta del tilt y la potencia de transmisión puede producir mejoras significativas en el RSRP en determinados puntos. Por ejemplo, en uno de los escenarios analizados se obtuvo una mejora de 4,9 dB, al pasar de $-82,7$ dBm a $-77,8$ dBm. Este resultado demuestra la utilidad del sistema para comparar alternativas de configuración y seleccionar escenarios potencialmente favorables antes de realizar pruebas directamente sobre la red.

No obstante, los resultados también permiten identificar algunas limitaciones. La precisión del sistema depende en gran medida de la calidad y representatividad del conjunto de datos utilizado para entrenar los modelos. Por esta razón, la aplicación del sistema en zonas con características geográficas, urbanísticas o radioeléctricas significativamente diferentes podría requerir la incorporación de nuevos datos y un proceso adicional de entrenamiento o ajuste del modelo. Asimismo, determinados factores dinámicos, como las variaciones temporales del tráfico, los niveles de

interferencia, las condiciones de operación de la red y los cambios en el entorno físico, pueden generar diferencias entre las estimaciones y las condiciones observadas en un momento específico. Por otra parte, aunque la salinidad fue identificada en el marco teórico como un factor relevante en entornos costeros, su efecto no se incorporó como variable explícita en el modelo desarrollado, quedando representado de manera indirecta a través de la humedad y de las mediciones reales de campo realizadas en la zona de estudio.

Asimismo, la caracterización espectral mediante el analizador Narda SRM-3006 se realizó en un número limitado de ubicaciones, debido a las condiciones de disponibilidad y traslado del equipo. Su función en el estudio fue la de verificar las condiciones de la portadora en la banda analizada, y no la de constituir la fuente de los valores de referencia utilizados para el entrenamiento y la validación de los modelos.

Otra consideración importante corresponde a la estimación de parámetros de rendimiento como el DL y el UL Throughput. A diferencia del RSRP, estos parámetros no dependen únicamente de las condiciones de propagación, sino también de factores relacionados con la operación de la red, tales como la carga de la celda, la asignación de recursos radio, el tráfico de usuarios y las condiciones de modulación y codificación. Por ello, las estimaciones de Throughput deben interpretarse como una aproximación del rendimiento esperado y no como una representación exacta del comportamiento instantáneo de la red.

En conjunto, los resultados permiten establecer que la propuesta desarrollada constituye una herramienta viable de apoyo para la planificación y optimización de la red LTE de CNT en el cantón La Libertad. La integración de modelos de aprendizaje automático con variables geoespaciales, información radioeléctrica y herramientas de visualización permite analizar las condiciones de cobertura desde una perspectiva espacial y predictiva. Asimismo, la posibilidad de evaluar diferentes configuraciones antes de su implementación facilita la identificación de alternativas potencialmente favorables y puede contribuir a reducir el número de pruebas requeridas en campo. Por lo tanto, la propuesta proporciona información objetiva que puede servir como apoyo a los procesos de análisis, planificación y optimización radioeléctrica, sin sustituir la validación técnica y las mediciones necesarias antes de implementar cambios en la infraestructura real.

CONCLUSIONES

- Se estableció la base teórica necesaria para el desarrollo de la investigación, mediante el estudio de la arquitectura y los parámetros de la red LTE, así como de las técnicas de aprendizaje automático aplicables a la predicción de cobertura, lo que permitió sustentar el diseño del sistema desarrollado.
- Se recopiló y procesó información radioeléctrica, geoespacial y ambiental mediante campañas de medición en campo, conformando un conjunto de 4 962 muestras (2 822 mediciones reales y 2 140 escenarios aumentados), utilizado para el entrenamiento y evaluación de los modelos de aprendizaje automático.
- Se entrenaron y evaluaron dos modelos de aprendizaje automático, Random Forest y XGBoost, para la estimación de parámetros asociados a la cobertura LTE. Random Forest presentó el mejor desempeño para la predicción del RSRP, con un MAE de 1,82 dB y un $R^2 = 0,964$, frente a un MAE de 2,55 dB y un $R^2 = 0,940$ obtenidos por XGBoost. Además, Random Forest presentó una menor brecha entre los valores de R^2 de entrenamiento y prueba, lo que indica un comportamiento más estable entre ambos conjuntos.
- Se evaluaron los resultados de las predicciones mediante métricas de desempeño y análisis comparativos. La validación final mediante diez puntos de medición de campo permitió comprobar la correspondencia entre los valores de RSRP medidos y estimados, obteniéndose un MAE de 1,51 dB, un RMSE de 1,84 dB y un R^2 de 0,944. Asimismo, las simulaciones de optimización demostraron que la modificación conjunta de parámetros radioeléctricos puede producir mejoras significativas en el nivel de la señal, alcanzando incrementos de 4,9 dB y hasta 7,9 dB en los escenarios analizados. Estos resultados muestran un comportamiento adecuado del sistema en los puntos de validación evaluados y proporcionan información de apoyo para la planificación y optimización de la red LTE de CNT en el cantón La Libertad.
- Se implementó un sistema interactivo de visualización geoespacial que integra mapas de calor, perfiles topográficos, presupuestos de enlace y paneles de indicadores radioeléctricos, permitiendo representar espacialmente los resultados obtenidos por los modelos. Esta integración facilita la interpretación de las condiciones de cobertura y proporciona una herramienta de apoyo para el análisis de diferentes escenarios dentro del área de estudio.

RECOMENDACIONES

- Ampliar las campañas de medición a diferentes épocas del año y bajo distintas condiciones de operación de la red, con el propósito de incrementar la diversidad y representatividad del conjunto de datos y fortalecer la capacidad de generalización de los modelos.
- Incorporar nuevas variables de estudio, como la carga de tráfico, la interferencia entre celdas, las condiciones meteorológicas en tiempo real y otros parámetros de operación de la red, con el objetivo de mejorar la representación de los factores que influyen en la calidad de la señal y el rendimiento de la red.
- Evaluar el sistema en redes de nueva generación, como el 5G, para determinar su comportamiento y realizar los ajustes necesarios en los modelos, variables y parámetros de propagación requeridos para estos nuevos escenarios.
- Utilizar el sistema como una herramienta de apoyo en los procesos de planificación y optimización de la red LTE de CNT, permitiendo evaluar previamente diferentes escenarios de configuración, como variaciones del azimut, el tilt y la potencia de transmisión, antes de realizar modificaciones sobre la infraestructura existente.
- Integrar el sistema desarrollado con plataformas profesionales de planificación radioeléctrica, como Atoll, con el propósito de complementar los métodos tradicionales de planificación con modelos de aprendizaje automático y datos obtenidos mediante mediciones reales, facilitando así el análisis comparativo de escenarios y la toma de decisiones por parte de los ingenieros responsables de la planificación de la red.
- Desarrollar y evaluar el módulo de detección de handover incorporado en el sistema, actualmente reservado para las mediciones de estaciones base vecinas, con el propósito de analizar su desempeño y extender el alcance de la herramienta hacia la gestión de traspasos entre celdas.

REFERENCIAS

- [1] DataReportal, «Digital 2024,» 23 02 2024. [En línea]. Available: <https://datareportal.com/reports/digital-2024-ecuador>. [Último acceso: 21 07 2025].
- [2] M. d. I. Á. Carrión Herrera, «Predicción de rendimiento de redes LTE con técnicas de Machine Learning,» 2022. [En línea]. Available: <https://bibliotecautpl.utpl.edu.ec/cgi-bin/abnetclwo?ACC=DOSEARCH&xsqf99=130434.TITN..> [Último acceso: 27 03 2025].
- [3] C. A. Muñoz Pérez, «Estudio de las interferencias en el campo electromagnético que se producen en las diferentes bandas de frecuencia no licenciadas, para determinar cómo se ven afectados las señales si se transmiten en una misma frecuencia o en un mismo canal,» 17 05 2023. [En línea]. Available: <https://repositorio.upse.edu.ec/handle/46000/9606>. [Último acceso: 15 04 2025].
- [4] J. J. Gutiérrez Terraza, «Aplicación de técnicas de aprendizaje automático y analítica predictiva para mejorar desempeño de redes 4G LTE,» 2021. [En línea]. Available: <https://repositorio.uchile.cl/handle/2250/184380>. [Último acceso: 27 03 2025].
- [5] V. R. C. N. H. & L.-M. P. Párraga Villamar, «Modelos Predictivos de Zonas de Handover en Redes LTE con Base a Mediciones de Campo y Árboles de Decisión (Caso de Estudio Ciudad de Quito),» 2023. [En línea]. Available:

- https://revistapolitecnica.epn.edu.ec/ojs2/index.php/revista_politecnica2/article/view/1414/627. [Último acceso: 27 03 2025].
- [6] J. E. Herrera Rubio, G. A. Portilla González y A. M. Osorio Velazco, «Predicción experimental de cobertura en una red LTE mediante el uso de algoritmos de Machine Learning,» 25 11 2024. [En línea]. Available: <https://doi.org/10.17081/invinno.12.2.7215>. [Último acceso: 03 04 2025].
- [7] Instituto Nacional de Estadística y Censos, «INEC,» 2023. [En línea]. Available: https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Sociales/TIC/2024/202407_Tecnologia_de_la_Informacion_y_Comunicacion-TICs.pdf. [Último acceso: 01 04 2025].
- [8] M. Curipallo, G. Pozo, P. Lupera-Morillo y V. Párraga, «Modelo de aprendizaje automático para la predicción de calidad en modo estático-inactivo y modo móvil-conectado en redes LTE,» 18 08 2022. [En línea]. Available: <https://doi.org/10.36825/RITI.10.21.010>. [Último acceso: 15 04 2025].
- [9] C. A. Serra Jimenez y F. R. Marante Rizo, «Arquitectura general del sistema LTE,» 2013. [En línea]. Available: <https://hdl.handle.net/20.500.14330/PER01000404238>. [Último acceso: 15 08 2025].
- [10] Teltonika, «Teltonika,» [En línea]. Available: https://wiki.teltonika-networks.com/view/Mobile_Signal_Strength_Recommendations. [Último acceso: 25 05 2025].
- [11] Python Software Foundation, «Python,» 2024. [En línea]. Available: <https://www.python.org>. [Último acceso: 21 05 2025].

- [12 Narda Safety Test Solutions, «SRM-3006 Selective Radiation Meter,» [En línea]. Available: <https://www.narda-sts.com/en/>. [Último acceso: 20 05 2025].
- [13 J. M. Muñoz Rodríguez, «Optimization of an LTE network in the province of Santa Elena using the Atoll Simulation Tool for the study of packet scheduling strategies: Round robin and Proportional Fair,» 22 12 2024. [En línea]. Available: <https://doi.org/10.22201/icat.24486736e.2024.22.4.2408>. [Último acceso: 18 06 2025].
- [14 Google for Developers, «Regresión lineal,» [En línea]. Available: <https://developers.google.com/machine-learning/crash-course/linear-regression?hl=es-419>. [Último acceso: 29 05 2025].
- [15 N. Moraitis, L. Tsipi y D. Vouyioukas, «Machine Learning-Based Methods for Path Loss Prediction in Urban Environment for LTE Networks,» 2020. [En línea]. Available: <https://doi.org/10.1109/WiMob50308.2020.9253369>. [Último acceso: 29 05 2025].
- [16 E. Eyceyurt, «Machine Learning-Based Prediction of Optimal Neighbour Cells for LTE Handover in Dense Urban Areas: A Case Study from San Francisco.,» 2025. [En línea]. Available: <https://doi.org/10.22399/ijcesen.3438>. [Último acceso: 13 06 2025].
- [17 G. Eiber A, P. Jairo A y F.-G. Juan C, «Estudio comparativo entre máquinas de soporte vectorial multiclase, redes neuronales artificiales y sistema de inferencia neuro-difuso auto organizado para problemas de clasificación,» 02 2020. [En línea]. Available: <http://dx.doi.org/10.4067/S0718-07642020000100273>. [Último acceso: 28 05 2025].

- [18 J. Horak y A. Sabek, «Fixed Rank Kriging for Cellular Coverage Analysis,» 2015.
[En línea]. Available: <https://arxiv.org/pdf/1505.07062>. [Último acceso: 12 06 2025].
- [19 Google Earth Pro, «Importar datos de sistemas de información geográfica (GIS) en Google Earth para ordenadores,» [En línea]. Available: https://www.google.com/intl/es_es/earth/outreach/learn/importing-geographic-information-systems-gis-data-in-google-earth/. [Último acceso: 30 05 2025].
- [20 ArcGIS, «What is ArcGIS?,» [En línea]. Available: <https://www.esri.com/en-us/arcgis/geospatial-platform/overview>. [Último acceso: 31 05 2025].
- [21 QGIS, «QGIS overview,» [En línea]. Available: <https://qgis.org/project/overview/>. [Último acceso: 31 05 2025].
- [22 A. P. Quistial Correa, «Modelo matemático adaptado de pérdidas de trayectoria mediante el análisis de propagación en la banda de 850 MHz para una zona urbana de Quito,» 11 11 2016. [En línea]. Available: <http://bibdigital.epn.edu.ec/handle/15000/16924>.
- [23 Y. . B. Logro Lema y G. A. Mayanquer Méndez, «Simulación y análisis de los modelos de propagación para un canal implementado bajo el estándar IEEE 802.11p,» 30 01 2020. [En línea]. Available: <http://bibdigital.epn.edu.ec/handle/15000/20711>.
- [24 T. S. Rappaport, «Wireless communications : principles and practice,» 2010. [En línea]. Available: https://archive.org/details/wirelesscommunic0000rapp_x1m2/mode/2up. [Último acceso: 01 08 2025].

- [25 Recommendation ITU-R P.676-13, «Attenuation by atmospheric gases and related effects,» 08 2022. [En línea]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.676-13-202208-I!!PDF-E.pdf. [Último acceso: 12 07 2025].
- [26 J. Ribadeneira Ramírez, M. Suarez, E. Sarmiento y F. Santacruz, «Propagación Indoor de 10 a 20 Ghz con Correcciones por Humedad y Concentración de Oxígeno para la Región Andina,» 29 08 2022. [En línea]. Available: https://perspectivas.esPOCH.edu.ec/RCP_ESPOCH/es/article/view/154/131. [Último acceso: 25 07 2025].
- [27 Union Internacional de las Telecomunicaciones, «Datos de propagación y métodos de predicción necesarios para el diseño de sistemas terrenales con visibilidad directa,» 09 2025. [En línea]. Available: <https://www.itu.int/rec/R-REC-P.530-19-202509-I/es>. [Último acceso: 12 08 2025].
- [28 G. Crespín Angulo, C. Bermeo Reyes y C. V. Chico Vallejo, «Modelos de propagación aplicados a tecnologías LTE en zonas residenciales de Riobamba – Ecuador,» 29 07 2022. [En línea]. Available: <https://doi.org/10.33262/concienciadigital.v5i3.1.2263>.
- [29 L. J. Sefla Urquizo, J. A. Aguilar Bonilla, D. A. Basantes Cherrez y E. A. Tintin Sánchez, «Modelos de propagación aplicados a áreas abiertas con tecnología LTE y ajuste de mediciones,» 16 08 2023. [En línea]. Available: <https://doi.org/10.47187/perspectivas.5.2.197>. [Último acceso: 12 07 2025].
- [30 D. E. Barrera Salamea, «Simulación y análisis de cobertura para tecnología LTE en el Centro Histórico de la ciudad de Cuenca,» 2015. [En línea]. Available:

- <http://dspace.ucuenca.edu.ec/handle/123456789/21661>. [Último acceso: 12 06 2025].
- [31 Forsk, «4G/5G,» [En línea]. Available: <https://www.forsk.com/4g5g>. [Último acceso: 12 06 2025].
- [32 S. Kazmierczyk, «Implementación de una sistema de análisis de medidas de cobertura en LTE mediante la herramienta QUALIPOC,» 2020. [En línea]. Available: https://oa.upm.es/57836/1/PFC_SYLWIA_KAZMIERCZYK_20.pdf. [Último acceso: 12 06 2025].
- [33 N. Porras Garzón y H. A. Gómez Pérez, «Implementación de un sistema de predicción de medidas LTE utilizando la tarjeta de desarrollo Zedboard y Machine Learning,» 2023. [En línea]. Available: <http://hdl.handle.net/11349/41055>. [Último acceso: 12 06 2025].
- [34 The MathWorks, «Acerca de MathWorks,» [En línea]. Available: https://la.mathworks.com/company.html?s_tid=nav_company_about_us. [Último acceso: 12 06 2025].
- [35 Microsoft, «Desarrolle aplicaciones con C y C++,» [En línea]. Available: <https://visualstudio.microsoft.com/es/vs/features/cplusplus/>. [Último acceso: 12 06 2025].
- [36 Java, «¿Qué es la tecnología Java y por qué la necesito?,» [En línea]. Available: https://www.java.com/es/download/help/whatis_java.html. [Último acceso: 12 06 2025].

- [37 scikit-learn developers, «Metrics and scoring: quantifying the quality of predictions,» [En línea]. Available: https://scikit-learn.org/stable/modules/model_evaluation.html. [Último acceso: 01 08 2025].
- [38 Tongyu Communication, «Antenna Specifications GSM TDJH-182018DE-65F DCS,» [En línea]. Available: <https://fccid.io/m/eba908908b4ab38e4a7148375563855dd9ec06049a6aff612161a8a145a465fa.pdf>. [Último acceso: 10 07 2026].
- [39 L. Breiman, «Random Forests,» 01 2001. [En línea]. Available: <https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf>. [Último acceso: 01 08 2025].
- [40 T. Chen y C. Guestrin, «XGBoost: A Scalable Tree Boosting System,» 19 3 2016. [En línea]. Available: <https://arxiv.org/abs/1603.02754>. [Último acceso: 02 08 2025].

ANEXOS

Con el propósito de facilitar la reproducibilidad del sistema desarrollado y su uso por parte de terceros, se documenta el proceso de instalación del entorno de desarrollo, las librerías empleadas y el acceso al repositorio que contiene el código fuente completo del proyecto.

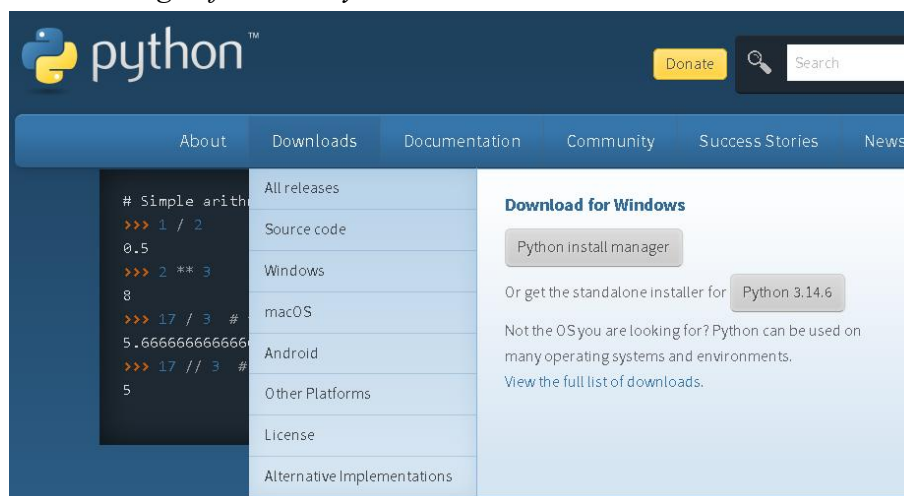
Anexo A. Instalación del entorno de desarrollo

El presente anexo describe el proceso de instalación de Python 3.14.6 y Visual Studio Code (VS Code), herramientas utilizadas durante el desarrollo, entrenamiento de los modelos de aprendizaje automático y ejecución del sistema de visualización descrito en el Capítulo 3.

Instalación de Python

Como se muestra en la **Figura 52**, la página oficial de Python dispone de la sección de descargas, desde la cual es posible seleccionar la versión correspondiente al sistema operativo utilizado. Para el desarrollo de este proyecto se seleccionó Python 3.14.6.

Figura 52. Descarga oficial de Python desde la web.



Nota. Elaboración propia.

Como se observa en la **Figura 53**, durante la instalación de Python 3.14.6 en Windows se seleccionó la opción “Install Now”. Asimismo, se habilitó la opción “Add python.exe to PATH”, con el propósito de agregar el intérprete de Python a las variables de entorno del sistema y permitir su ejecución directamente desde la línea de comandos. Las opciones relacionadas con privilegios administrativos se mantuvieron de acuerdo con los requerimientos del sistema.

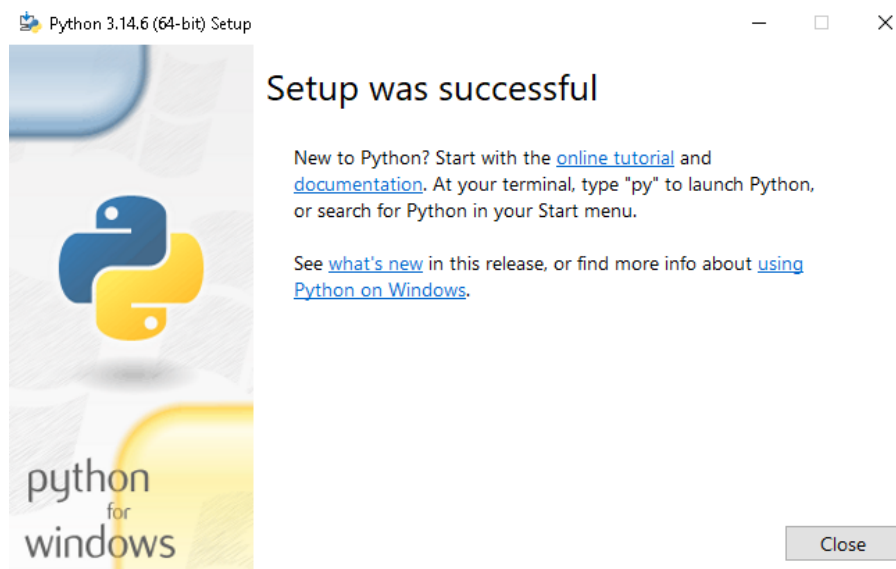
Figura 53. Instalador de Python en Windows.



Nota. Elaboración propia.

Como se aprecia en la **Figura 54**, el instalador muestra un mensaje de confirmación una vez finalizado correctamente el proceso. Esto indica que Python y las herramientas básicas asociadas quedaron disponibles para su utilización en el entorno de desarrollo.

Figura 54. Instalación de Python completada en Windows.

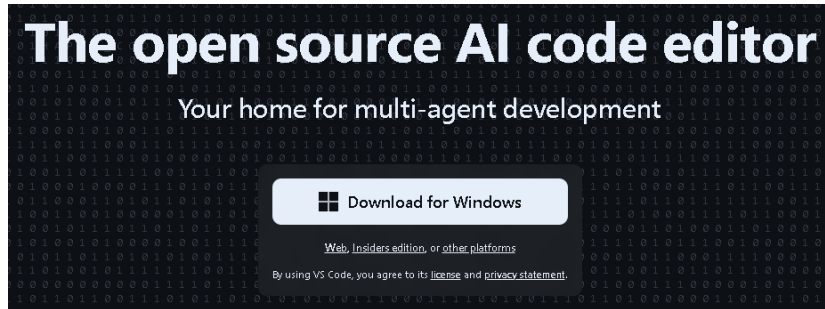


Nota. Elaboración propia.

Instalación de Visual Studio Code

Como se muestra en la **Figura 55**, la página oficial de Visual Studio Code permite descargar el instalador correspondiente al sistema operativo. En este proyecto se seleccionó la versión destinada a Windows.

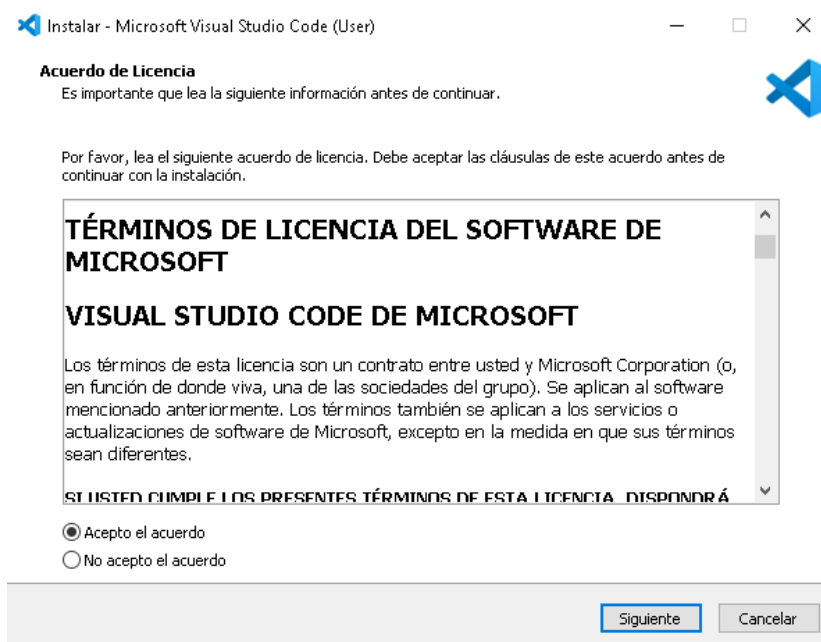
Figura 55. Descarga de Visual Studio Code.



Nota. Elaboración propia.

Como se observa en la **Figura 56**, durante el proceso de instalación se presenta el acuerdo de licencia del software. Para continuar con la instalación es necesario aceptar los términos establecidos.

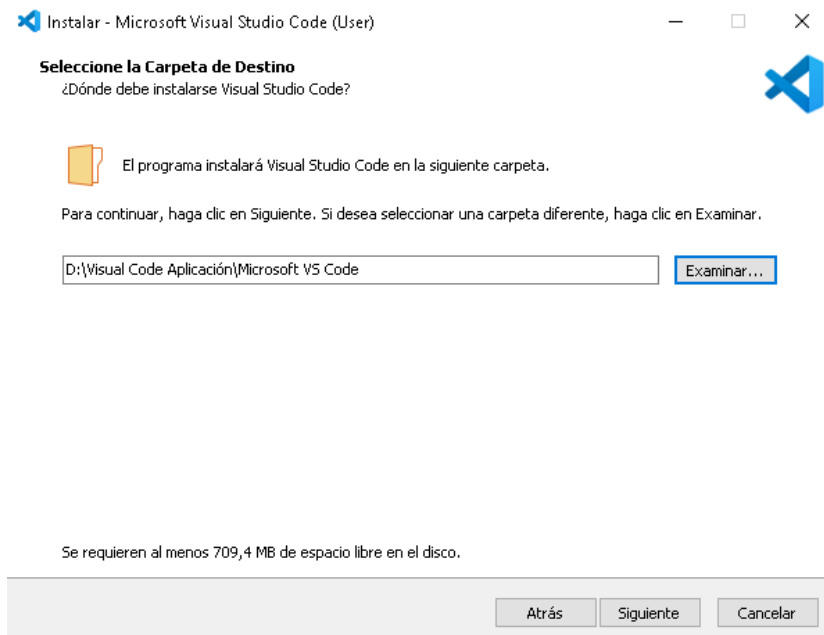
Figura 56. Aceptación de términos de licencia en VS Code.



Nota. Elaboración propia.

Como se muestra en la **Figura 57**, el instalador permite seleccionar la carpeta de destino donde se instalará Visual Studio Code. En este caso se mantuvo la ubicación predeterminada propuesta por el instalador.

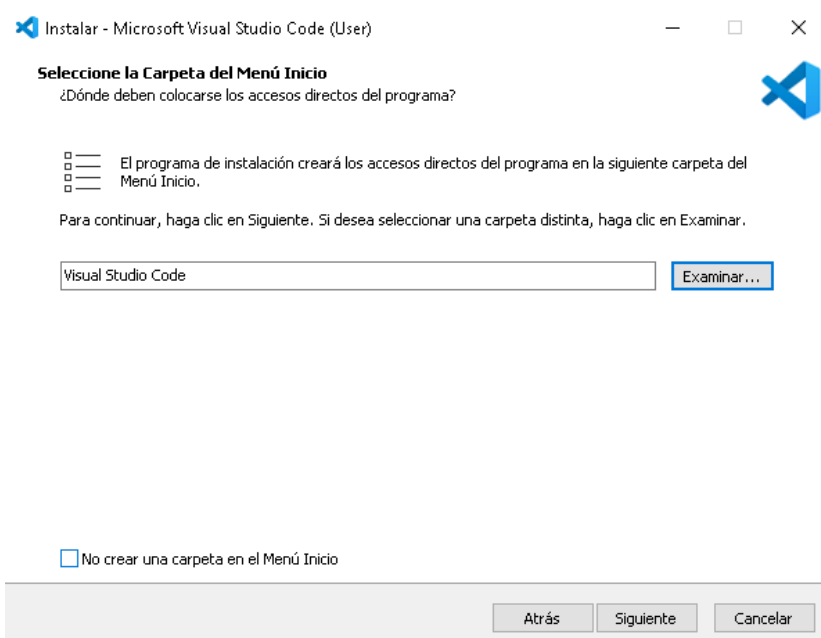
Figura 57. Selección de carpeta de destino en VS Code.



Nota. Elaboración propia.

Como se observa en la **Figura 58**, el instalador permite seleccionar la carpeta del Menú Inicio donde se crearán los accesos directos correspondientes al programa. Se mantuvo la configuración predeterminada.

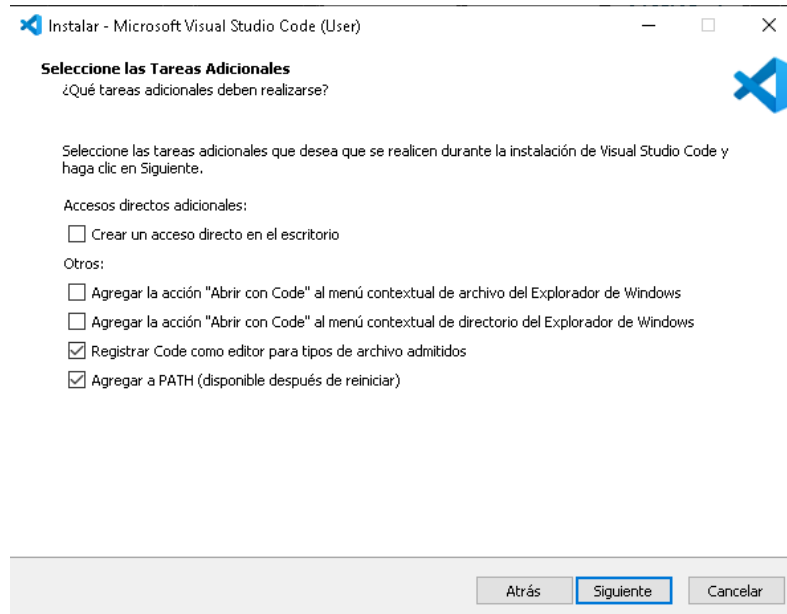
Figura 58. Selección de carpeta en el Menú Inicio de VS Code.



Nota. Elaboración propia.

Como se muestra en la **Figura 59**, el instalador permite seleccionar diferentes tareas adicionales para integrar Visual Studio Code con el sistema operativo. Para este proyecto se mantuvieron las opciones establecidas por defecto.

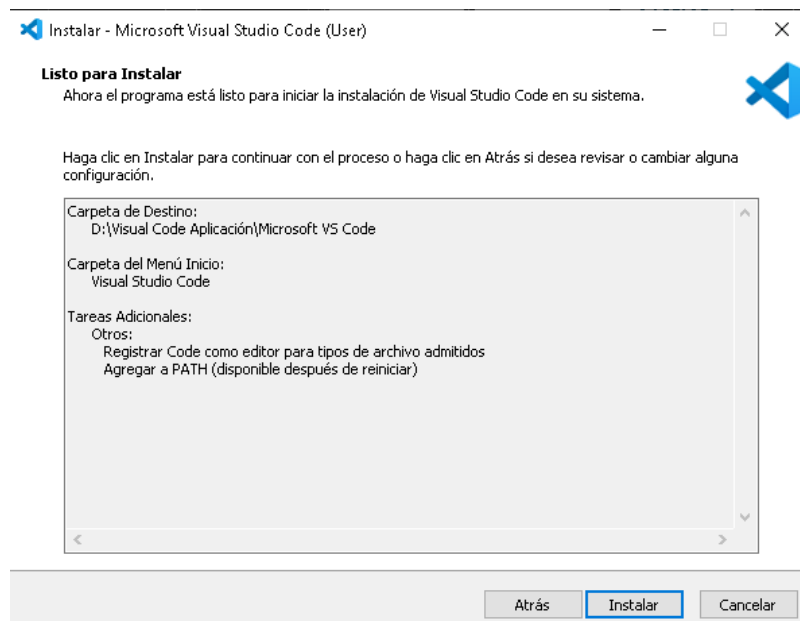
Figura 59. Selección de tareas adicionales en VS Code.



Nota. Elaboración propia.

Como se aprecia en la **Figura 60**, el asistente de instalación presenta la ventana de confirmación “Listo para instalar”. En esta etapa se seleccionó el botón “Instalar” para iniciar el proceso.

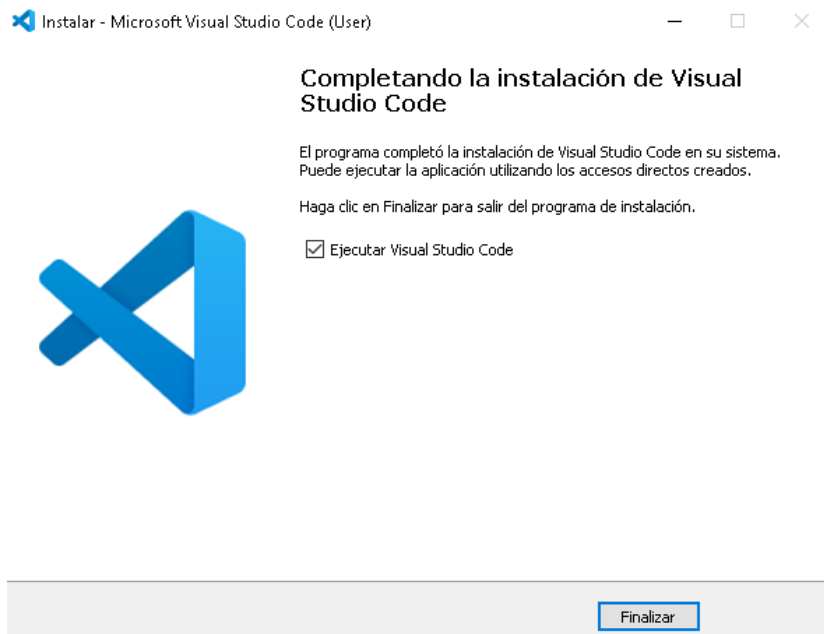
Figura 60. Confirmación final antes de instalar VS Code.



Nota. Elaboración propia.

Finalmente, como se observa en la **Figura 61**, el asistente confirma que la instalación de Visual Studio Code se completó correctamente.

Figura 61. *Instalación de Visual Studio Code completada.*



Nota. Elaboración propia.

Anexo B. Instalación de librerías del proyecto

Una vez instalado el entorno de desarrollo, es necesario incorporar las librerías de Python utilizadas para el procesamiento de datos, entrenamiento de los modelos de aprendizaje automático, análisis geoespacial y visualización de los resultados. La **Tabla 26** presenta las principales librerías utilizadas y su función dentro del proyecto.

Tabla 26. *Librerías de Python utilizadas en el proyecto.*

Librería	Propósito en el proyecto
pandas	Manipulación y análisis de los conjuntos de datos de mediciones LTE.
numpy	Operaciones numéricas y cálculos matriciales.
scikit-learn	Entrenamiento del modelo Random Forest y preprocesamiento de variables.
xgboost	Entrenamiento del modelo de potenciación de gradiente (XGBoost).
joblib	Serialización y carga de los modelos entrenados.
geopandas	Manejo de datos geoespaciales vectoriales.
rasterio	Lectura y procesamiento de datos ráster (elevación, topografía).

shapely	Operaciones geométricas sobre coordenadas y polígonos.
osmnx	Obtención de datos de infraestructura urbana desde OpenStreetMap.
geopy	Cálculo de distancias geográficas entre puntos.
folium / ipyleaflet	Generación de mapas interactivos y mapas de calor.
ipywidgets	Construcción de la interfaz interactiva del sistema.
matplotlib	Generación de gráficos estadísticos y comparativos.
Pillow (PIL)	Procesamiento de imágenes utilizadas en la interfaz.

Nota. Elaboración propia.

La instalación de estas dependencias se realiza mediante pip, gestor de paquetes utilizado por Python. A continuación, se describe el procedimiento recomendado para configurar el entorno de ejecución del proyecto.

Paso 1. Verificar la instalación de Python y pip: desde el Símbolo del sistema o PowerShell de Windows se comprueba que Python y el gestor de paquetes pip estén disponibles mediante los siguientes comandos:

```
python --version
pip --version
```

Estos comandos permiten verificar la versión instalada y confirmar que Python puede ser ejecutado desde la terminal.

Paso 2. Actualizar el gestor de paquetes: se recomienda actualizar pip antes de instalar las dependencias del proyecto, con el propósito de disponer de una versión actualizada del gestor:

```
python -m pip install --upgrade pip
```

Paso 3. Crear y activar un entorno virtual (opcional, recomendado): con el propósito de aislar las dependencias del proyecto y reducir posibles conflictos con otras instalaciones de Python, se recomienda crear un entorno virtual:

```
python -m venv
venv\Scripts\activate
```

Una vez activado, las librerías instaladas quedarán asociadas al entorno virtual creado para el proyecto.

Paso 4. Instalar las librerías del proyecto: las dependencias principales pueden instalarse mediante el siguiente comando, con las librerías descritas en la **Tabla 26**:

```
pip install pandas numpy scikit-learn xgboost joblib geopandas rasterio shapely osmnx  
geopy folium ipyleaflet ipywidgets matplotlib pillow
```

Durante la instalación, algunas librerías pueden incorporar dependencias adicionales necesarias para su funcionamiento. Estas serán gestionadas automáticamente por pip.

Paso 5. Verificar la instalación: una vez finalizado el proceso, se puede comprobar el listado de paquetes instalados mediante:

```
pip list
```

Si las librerías descritas en la **Tabla 26** aparecen en el listado, el entorno dispone de las principales dependencias necesarias para ejecutar los notebooks de entrenamiento y el sistema de predicción y visualización.

Nota. Para mejorar la reproducibilidad del proyecto, se recomienda generar un archivo de requerimientos después de configurar y verificar el entorno definitivo mediante el comando:

```
pip freeze > requirements.txt
```

Posteriormente, la instalación del entorno puede realizarse mediante:

```
pip install -r requirements.txt
```

De esta manera, se conservan las versiones específicas de las dependencias utilizadas durante el desarrollo, facilitando la reproducción del entorno en otro equipo.

Anexo C. Repositorio de código fuente

El código fuente y los archivos asociados al sistema desarrollado, incluidos los notebooks utilizados para el entrenamiento de los modelos Random Forest y XGBoost, el sistema interactivo de predicción y visualización, los modelos entrenados y los archivos de datos necesarios para su funcionamiento, se encuentran disponibles en la siguiente carpeta de almacenamiento:

Recurso del proyecto:

https://drive.google.com/drive/folders/1HdJxo95G8GJFWckf3Fs_pGOpta7ksaCX?usp=drive_link

La estructura de los principales archivos y carpetas disponibles se describe en la **Tabla 27**.

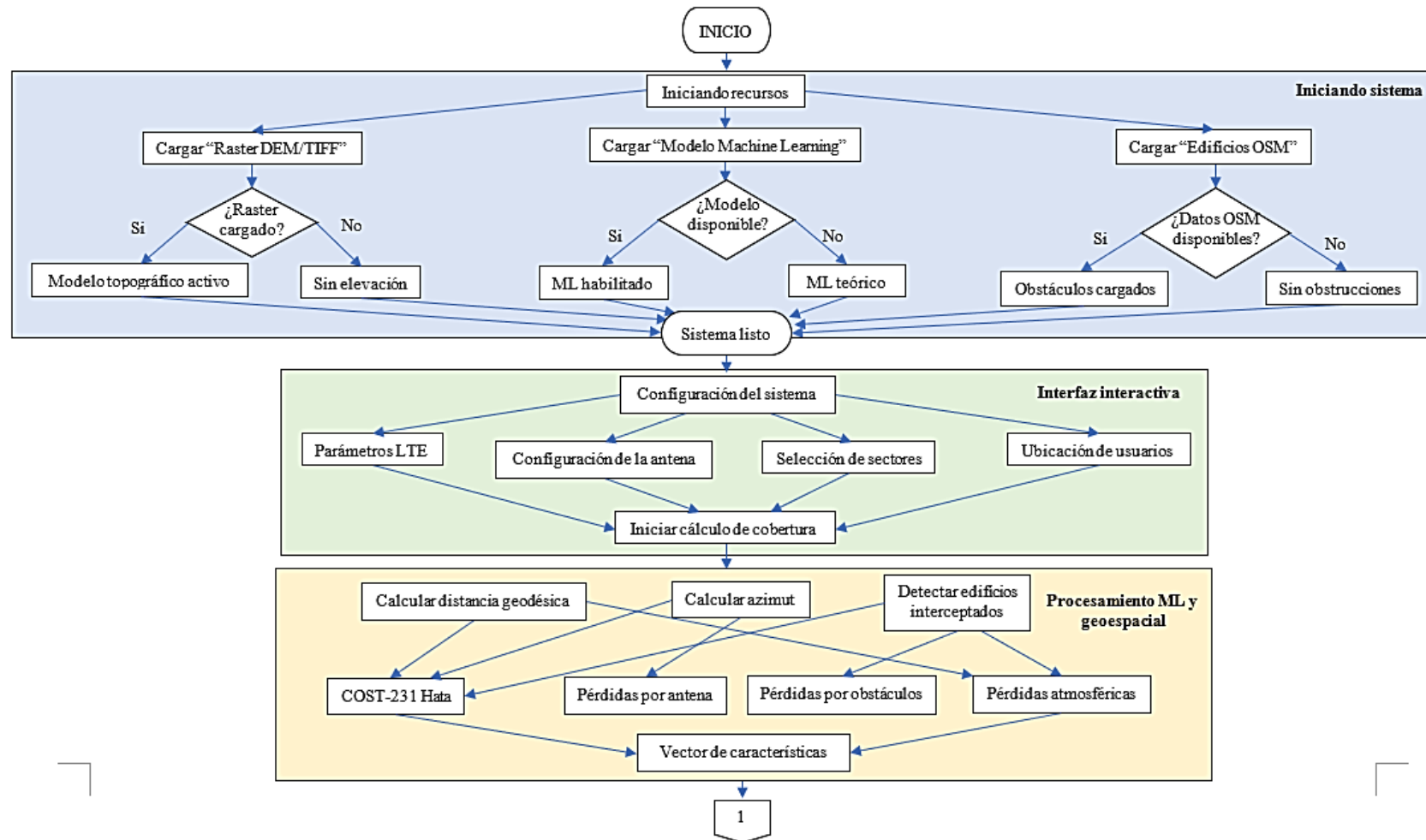
Tabla 27. Estructura de archivos del proyecto.

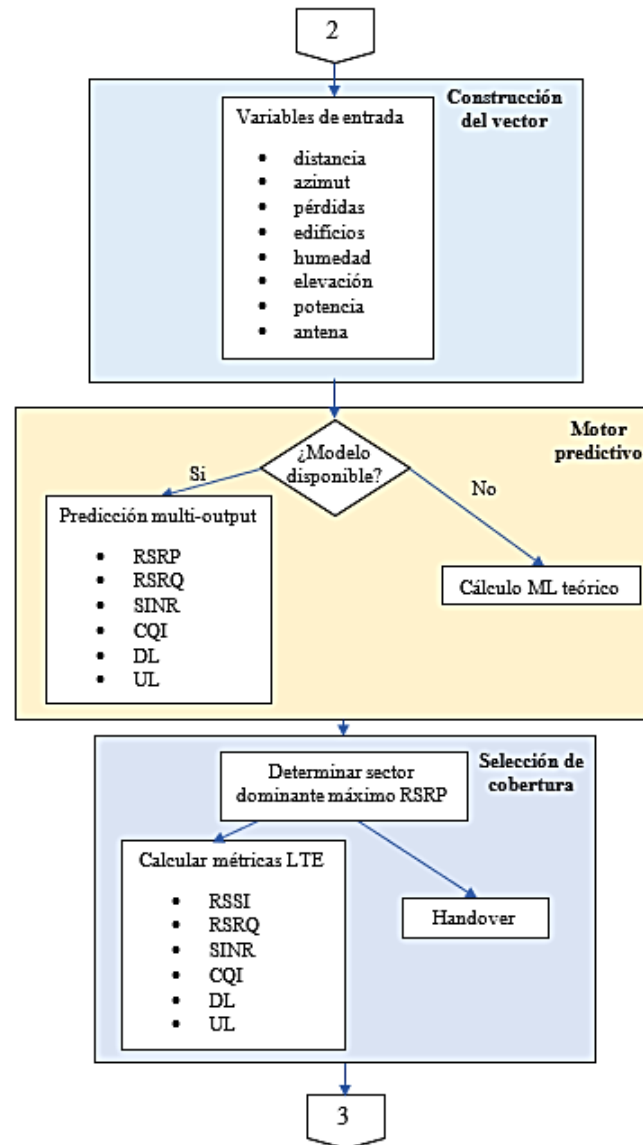
Archivo / Carpeta	Contenido
Entrenamiento_de_los_modelos_de_machine_learning.ipynb	Notebook de entrenamiento de los modelos XGBoost y Random Forest multi-salida.
/Archivo de entrenamiento de los modelos	Modelos entrenados serializados (formato.pkl).
/dataset	Conjunto de datos de mediciones LTE utilizado para el entrenamiento y validación.
GEOTIFF	Archivo requerido para que el sistema detecte la elevación del terreno.
Código de visualización.ipynb	Notebook de visualización para cargar el modelo. Sistema interactivo de predicción, visualización de cobertura y comparador de escenarios.

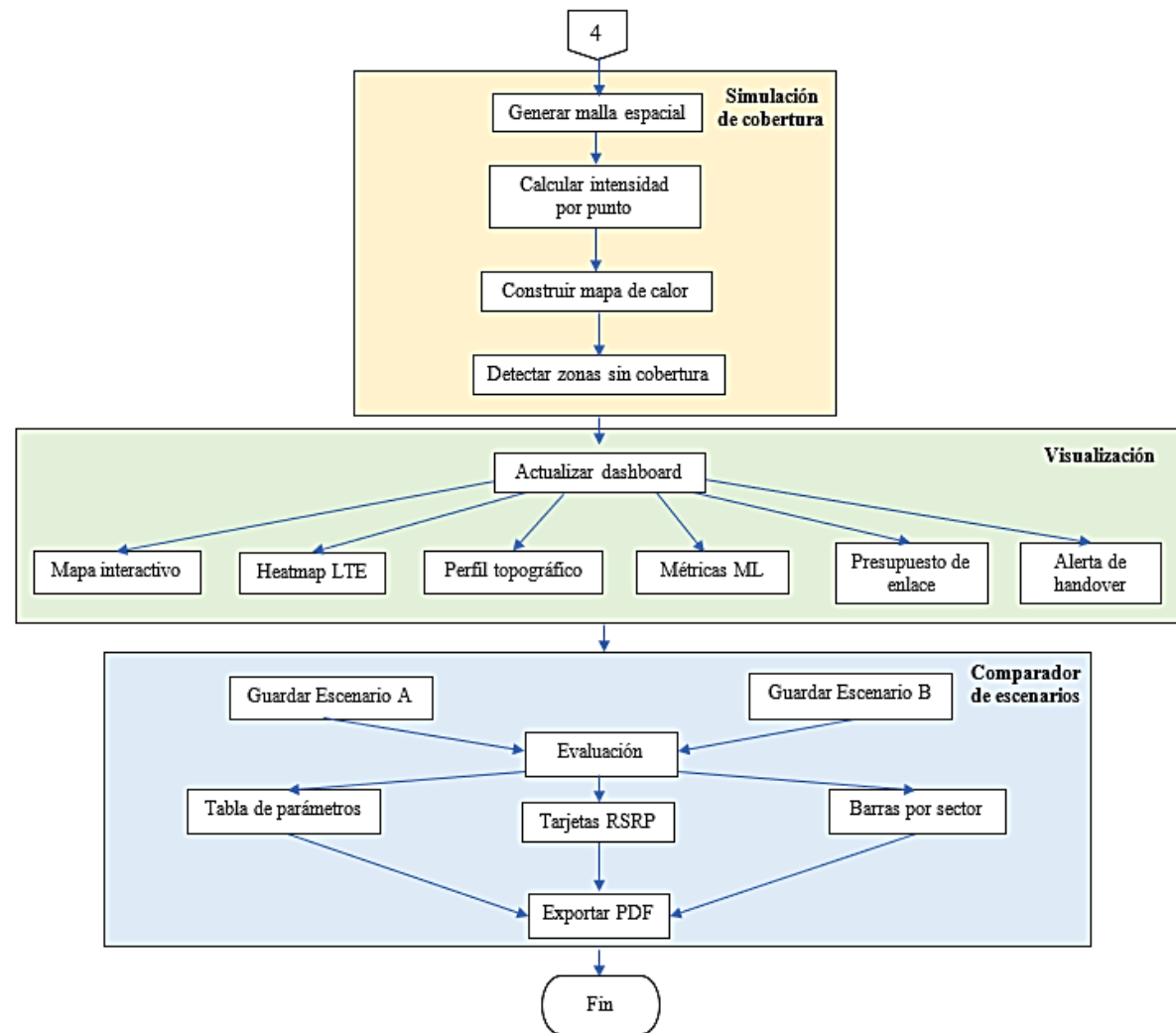
Nota. Elaboración propia.

Anexo D. Diagrama de flujo del sistema desarrollado

Figura 62. Diagrama de flujo del módulo de visualización, simulación de cobertura y comparador de escenarios del sistema desarrollado.







Nota. Elaboración propia.