



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

PREDICCIÓN DE LA CALIDAD DE SERVICIO DE INTERNET MÓVIL
BASADA EN MEDICIONES PASIVAS

TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN CIENCIAS, MENCIÓN COMPUTACIÓN

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL EN COMPUTACIÓN

DIEGO IGNACIO MADARIAGA ROMÁN

PROFESOR GUÍA:
JAVIER BUSTOS JIMÉNEZ

MIEMBROS DE LA COMISIÓN:
ALEJANDRO HEVIA ANGULO
LUIS MATEU BRULÉ
GONZALO ACUÑA LEIVA

ESTE TRABAJO HA SIDO PARCIALMENTE FINANCIADO POR NIC CHILE
RESEARCH LABS

SANTIAGO DE CHILE
2018

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE MAGÍSTER EN CIENCIAS, MENCIÓN COMPUTACIÓN
POR: DIEGO IGNACIO MADARIAGA ROMÁN
FECHA: 2018
PROF. GUÍA: JAVIER BUSTOS JIMÉNEZ

PREDICCIÓN DE LA CALIDAD DE SERVICIO DE INTERNET MÓVIL BASADA EN MEDICIONES PASIVAS

El continuo aumento a nivel global en el uso de Internet móvil, ha posicionado a este medio como la principal forma de acceder a Internet tanto en Chile como en el mundo. Lo anterior, ha generado gran interés en la realización de estudios que permitan determinar objetivamente la calidad de servicio recibida por los usuarios de Internet móvil, así como estudios que permitan estimar la calidad futura asociada a los servicios presentes en la red de telefonía celular.

Múltiples investigaciones han analizado los servicios de Internet móvil desde distintos puntos de vista e utilizando distintos tipos de indicadores de calidad, obtenidos por medio de la realización de mediciones pasivas y activas de la red. Dentro de estos parámetros, destaca la intensidad de señal recibida, ya que se ha demostrado su impacto en otros indicadores de calidad comúnmente utilizados, afectando de gran manera la calidad de servicio general de Internet móvil.

El presente trabajo propone la predicción de intensidad de señal recibida para un área geográfica determinada, y además, propone la incorporación de información externa en los modelos de predicción, respecto a factores ajenos a las redes móviles que presentan un efecto sobre ellas, tales como eventos meteorológicos y eventos masivos de personas.

Para lograr lo mencionado, este estudio hace uso de distintos modelos de predicción basados en técnicas estadísticas y basados en el aprendizaje de máquinas, incorporando información climática acerca de la temperatura, humedad y precipitaciones medidas en el área de interés.

Los resultados obtenidos muestran la factibilidad de predecir la intensidad de señal recibida por los usuarios de la red móvil utilizando datos agregados correspondientes a dispositivos móviles dentro de la comuna de Santiago. Además, los resultados muestran la mejora en el desempeño de los modelos predictivos al incorporar información externa acerca del clima presente en el área, obteniendo modelos de predicción semanal más precisos.

De acuerdo al estudio realizado y al conocimiento adquirido durante la realización de este trabajo, este es el primer trabajo de investigación enfocado en predecir calidad de servicio de Internet móvil utilizando mediciones pasivas, y además, el primero en incluir factores externos en la predicción de calidad de servicio de Internet móvil.

A mi hermosa familia

Agradecimientos

A Catalina, por tu amor, por ser mi amiga y compañera durante casi 9 años. Gracias por cada momento juntos, por dejarme ser parte de tu vida desde que eramos escolares, y por haber decidido ser parte de la mía. Gracias por tu apoyo permanente y por haberme acompañado (literalmente) al otro lado del mundo durante la última etapa de estos estudios de magíster. Espero que nuestros días juntos me alcancen para demostrarte todo mi amor y mis agradecimientos.

A Iván, por ser junto a tu mamá los mejores compañeros de aventuras, por brindarme tantas alegrías y momentos hermosos. Por cada abrazo, por cada beso, por cada vez que me buscaste para jugar o leerte tus libros favoritos. Gracias por alegrarme la vida cada día, eres sin duda alguna lo más precioso de mi vida.

A mis padres, por todas sus enseñanzas que me han hecho crecer, por toda su ayuda y su preocupación, y por haber estado siempre en los momentos importantes que han marcado y forjado mi vida. Gracias por todos los esfuerzos, entregas y sacrificios silenciosos que realizaron y siguen realizando por sus cinco hijos, esfuerzos que desde hace un poco más de dos años, creo entender mejor.

A mis cuatro hermanos, las personas más geniales y brillantes que conozco. Gracias por su apoyo incondicional frente a cada dificultad, por todas las risas y penas que hemos compartido. Los amo mucho.

A Javier Bustos, por haberme recibido como alumno cuando iba recién en tercer año de universidad, por la confianza puesta en mí y en mi trabajo, y por toda la ayuda entregada durante estos últimos tres años para formarme profesionalmente.

A todos mis compañeros y amigos de NIC Labs, lugar donde realicé esta tesis, con quienes he vivido un sinnúmero de excelentes momentos en los últimos años. Haber compartido con ustedes es sin duda una de las mejores cosas que me ha traído este trabajo de tesis. Especiales gracias a Martín, por ser un excelente amigo y compañero tanto dentro como fuera del laboratorio, y por toda tu ayuda para confeccionar este trabajo.

Al único que no varía, al Padre de las luces. Gracias por cada buena dádiva, por sostener nuestra familia y por darnos siempre la victoria por medio de nuestro Señor Jesucristo.

Tabla de Contenido

Introducción	1
1. Motivación y Antecedentes	4
1.1. Contexto	4
1.2. Problema a Resolver	5
1.2.1. Preguntas de Investigación	6
1.3. Hipótesis	6
1.4. Objetivos	6
1.4.1. Objetivo General	6
1.4.2. Objetivos Específicos	6
1.5. Metodología de Investigación	7
1.6. Resultados Esperados	7
1.7. Aportes de la Tesis	8
2. Estado del Arte	9
2.1. Calidad de Internet	9
2.1.1. Internet Fijo	9
2.1.2. Internet Móvil	10
2.2. Mediciones Activas y Pasivas	11
2.2.1. Mediciones Activas	12
2.2.2. Mediciones Pasivas	14
2.3. Análisis de Calidad de Servicio	16
2.3.1. Análisis Estadísticos	16
2.3.2. Análisis Predictivos	17
2.4. Origen de Datos	18
2.4.1. Datos Internos de la Red	18
2.4.2. Datos Externos	18
2.5. Resumen	19
3. Descripción de Datos	21
3.1. Adkintun Mobile	21
3.1.1. Telephony Monitor	21
3.1.2. Connectivity Monitor	22
3.1.3. Traffic Monitor	23
3.2. Privacidad de Datos	24
3.3. Procesamiento de Datos	24

4. Modelos de Predicción	27
4.1. ARIMA	27
4.2. Seasonal ARIMA	28
4.3. Redes Neuronales	28
4.3.1. LSTM	28
4.3.2. CNN-LSTM	29
4.3.3. Naïve Persistent Model	30
4.4. Evaluación de Modelos	30
4.4.1. Errores de Predicción	31
4.4.2. Distancia Entre Series de Tiempo	32
5. Predicción de Intensidad de Señal	33
5.1. Experimentos Realizados	33
5.2. Intensidad de Señal General	34
5.3. Intensidad de Señal 3G UMTS	36
5.4. Intensidad de Señal 4G LTE	38
5.5. Discusión de Resultados	40
6. Inclusión de Factores Externos	42
6.1. Condiciones Climáticas	42
6.2. Agrupaciones Masivas	45
6.3. Experimentos Realizados	46
6.4. Discusión de Resultados	52
Conclusión	54
Bibliografía	62
Anexos	67

Índice de Tablas

2.1. Estudios de calidad de servicio para Internet fijo.	19
2.2. Estudios de calidad de servicio para Internet móvil.	19
3.1. Registros de intensidad de señal recolectados por <i>Adkintun Mobile</i>	22
3.2. Registros de antenas generados en el servidor de <i>Adkintun Mobile</i>	22
3.3. Registros de estado de la conectividad de red recolectados <i>Adkintun Mobile</i> . .	23
3.4. Registros de tráfico de datos por aplicación recolectados por <i>Adkintun Mobile</i> . .	24
4.1. Niveles generales de intensidad de señal recibida.	31
5.1. Medidas de precisión para intensidad de señal general.	36
5.2. Medidas de precisión para intensidad de señal 3G UMTS.	38
5.3. Medidas de precisión para intensidad de señal 4G LTE.	40
6.1. Factores climáticos y su correlación de Pearson con la intensidad de señal recibida.	44
6.2. Agrupaciones masivas y su correlación de Pearson con la intensidad de señal recibida.	45
6.3. Medidas de precisión para intensidad de señal general considerando condiciones climáticas.	48
6.4. Medidas de precisión para intensidad de señal 3G UMTS considerando condiciones climáticas.	50
6.5. Medidas de precisión para intensidad de señal 4G LTE considerando condiciones climáticas.	52
6.6. Bandas de frecuencias para redes móviles utilizadas por los principales proveedores de Chile.	54

Índice de Ilustraciones

1.	Crecimiento global del tráfico de datos por dispositivos móviles, primer trimestre 2017.	1
2.1.	Número de países con despliegue de infraestructura de red móvil.	11
2.2.	Ejemplo de medición activa de la red.	12
2.3.	Ejemplo de medición pasiva de la red.	14
3.1.	Intensidad de señal recibida en la comuna de Santiago durante una semana. .	26
5.1.	Predicción semana 1 de intensidad de señal general.	34
5.2.	Predicción semana 2 de intensidad de señal general.	35
5.3.	Predicción semana 3 de intensidad de señal general.	35
5.4.	Predicción semana 1 de intensidad de señal 3G UMTS.	36
5.5.	Predicción semana 2 de intensidad de señal 3G UMTS.	37
5.6.	Predicción semana 3 de intensidad de señal 3G UMTS.	37
5.7.	Predicción semana 1 de intensidad de señal 4G LTE.	38
5.8.	Predicción semana 2 de intensidad de señal 4G LTE.	39
5.9.	Predicción semana 3 de intensidad de señal 4G LTE.	39
6.1.	Ubicación geográfica de estaciones base y estación meteorológica en la comuna de Santiago.	43
6.2.	Relación entre intensidad de señal recibida y porcentaje de humedad en la comuna de Santiago.	44
6.3.	Predicción semana 1 de intensidad de señal general considerando condiciones climáticas.	47
6.4.	Predicción semana 2 de intensidad de señal general considerando condiciones climáticas.	47
6.5.	Predicción semana 3 de intensidad de señal general considerando condiciones climáticas.	48
6.6.	Predicción semana 1 de intensidad de señal 3G UMTS considerando condiciones climáticas.	49
6.7.	Predicción semana 2 de intensidad de señal 3G UMTS considerando condiciones climáticas.	49
6.8.	Predicción semana 3 de intensidad de señal 3G UMTS considerando condiciones climáticas.	50
6.9.	Predicción semana 1 de intensidad de señal 4G LTE considerando condiciones climáticas.	51

6.10. Predicción semana 2 de intensidad de señal 4G LTE considerando condiciones climáticas.	51
6.11. Predicción semana 3 de intensidad de señal 4G LTE considerando condiciones climáticas.	52
6.12. Promedio de intensidad de señal para un ISP en Chile.	56
6.13. Áreas en Santiago, Chile agrupadas según sus patrones de intensidad de señal.	58

Introducción

Crecimiento de Internet Móvil

Durante los últimos años el uso del servicio de Internet móvil ha ido en aumento a nivel global, razón por la cual algunas empresas especializadas han destinado sus esfuerzos en evaluar la evolución que ha tenido este servicio durante el último período. Este es el caso de la empresa *Akamai*, quien realiza periódicamente reportes del estado de Internet (*State of the Internet Reports*), en donde informan, entre otras cosas, acerca del aumento en la velocidad de las conexiones a Internet y acerca de la evolución del tráfico de datos móviles. En la Figura 6.12, correspondiente al reporte de *Akamai* del primer trimestre de 2017 [2], se puede observar la tendencia al alza del uso de Internet móvil, la cual se presenta de forma continua y sostenida durante los últimos reportes.

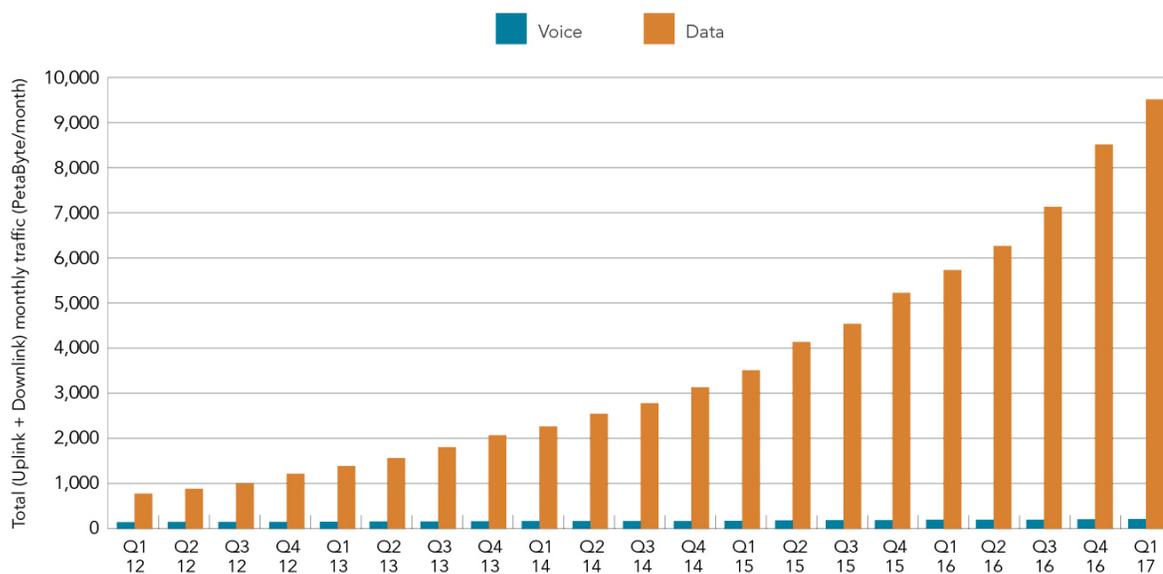


Figure 39: Total Monthly Mobile Voice and Data Traffic as Measured by Ericsson



Figura 1: Crecimiento global del tráfico de datos por dispositivos móviles, primer trimestre 2017.

El gran aumento en el uso de Internet móvil, ha implicado también que este servicio se haya convertido en la principal forma de acceder a los contenidos de la red, desplazando a otros medios como los servicios de Internet fijo. Este fenómeno se ve también reflejado en Chile en donde, según datos de la Subsecretaría de Telecomunicaciones (SUBTEL), durante el primer trimestre de 2017 el 75.8% del total de accesos a Internet fue realizado por medio de dispositivos móviles [62].

Calidad de Servicio

Dada la gran magnitud del uso actual de Internet móvil, es importante la realización de estudios que permitan determinar la calidad de este servicio entregado por las empresas proveedoras.

Actualmente, las mediciones de calidad de Internet que los usuarios realizan comúnmente, se basan en la aplicación de mediciones activas, es decir, mediciones en las cuales se introducen paquetes en la red, los que son transmitidos entre dos dispositivos o nodos de la red, midiendo tiempo de ida y vuelta de los paquetes (RTT¹), promedio de paquetes perdidos, ancho de banda de la conexión, entre otros parámetros. Una de las mayores dificultades en la aplicación de mediciones activas utilizando dispositivos móviles, es que al producir tráfico en la red se puede presentar un fuerte impacto en el consumo del plan de datos de los usuarios, en el caso de que se realicen pruebas constantes, como también una sobrecarga en la infraestructura, si muchas pruebas se realizan en paralelo. Como alternativa a esto, es posible realizar mediciones pasivas de la calidad de Internet móvil, las cuales difieren de las mediciones activas en que no introducen paquetes en la red, sino que observan el estado de la red en el dispositivo sin generar tráfico, entregando información acerca de un punto específico de la red y no del camino entre dos puntos de esta, como en el caso de las mediciones activas. Algunos parámetros que se obtienen al realizar mediciones pasivas son la cantidad de datos transmitidos y recibidos, estado de la conectividad a Internet, intensidad de señal recibida, entre otros.

Dentro de los parámetros de calidad en entornos móviles destaca el indicador de intensidad de señal, llamado también indicador de fuerza de la señal recibida o RSSI², el cual es una escala para medir la potencia de la señal recibida por los dispositivos dentro de una red inalámbrica. En el caso de la red de telefonía celular, se entiende como la potencia de señal que reciben los dispositivos celulares de las estaciones base de transmisión de la red, comúnmente conocidas como antenas. Estas estaciones de transmisión permiten a los dispositivos el acceso a la red móvil y a los múltiples servicios otorgados por esta, tales como llamadas telefónicas, mensajería instantánea o Internet móvil.

Durante mucho tiempo y desde los tiempos en que los teléfonos celulares solo proveían servicios básicos como llamadas telefónicas y mensajería instantánea, la intensidad de señal recibida ha sido uno de los indicadores de calidad de servicio más importantes y visibles por los usuarios, principalmente por el hecho de que esta información ha estado siempre presente en la pantalla de los celulares de forma gráfica como barras de señal, en donde a mayor

¹Por sus siglas en inglés: Round-Trip Time

²Por sus siglas en inglés: Received Signal Strength Indicator

cantidad de barras presentes, mayor es la intensidad de la señal recibida por el dispositivo.

Con la inclusión de nuevos servicios en la red móvil tales como Internet móvil, han aumentado también los indicadores de calidad de servicio a considerar. Entre los parámetros más utilizados y conocidos por los usuarios se encuentran la latencia, cantidad de paquetes perdidos y velocidad de transmisión de datos, cuyos valores suelen ser entregados por la ejecución de las llamadas *pruebas de velocidad*, realizadas generalmente en sitios populares como *SpeedTest*³ y *Fast*⁴. A pesar de la importancia de todos estos indicadores, la intensidad de señal recibida sigue teniendo una gran relevancia ya que estudios han demostrado que la intensidad de señal recibida incide en los valores de los indicadores mencionados anteriormente, impactando directamente la calidad general del Internet móvil [49].

La medición de la calidad de servicio de Internet móvil resulta importante tanto para las empresas prestadoras de servicios, con el fin de mantener los niveles de calidad establecidos y conocer fallas en sus sistemas, como para los usuarios que desean conocer las características reales de los servicios adquiridos y mediante esta información poder tomar decisiones al respecto. Dado lo anterior, también resulta importante el desarrollo de métodos predictivos de la calidad de servicio de Internet móvil, con el fin de que ambas contrapartes (empresas y clientes) puedan anticiparse a posibles bajas en la calidad de Internet móvil para un determinado tiempo y lugar geográfico.

El presente trabajo de tesis se enfoca en la implementación y comparación de distintos modelos de predicción del indicador pasivo de intensidad de señal recibida para un área específica, mediante la utilización de registros agregados de intensidad de señal provenientes de múltiples dispositivos móviles en la parte central de la ciudad de Santiago, Chile. Además, se considera la utilización de factores externos a la red móvil, específicamente acerca de factores climáticos, y se analiza la mejora en los resultados que produce su inclusión en los modelos predictivos.

De acuerdo al estudio realizado y al conocimiento adquirido durante la realización de este trabajo, este es el primer trabajo de investigación enfocado en predecir calidad de servicio de Internet móvil utilizando mediciones pasivas, y además, el primero en incluir factores externos en la predicción de calidad de servicio de Internet móvil.

³<http://www.speedtest.net>

⁴<https://fast.com>

Capítulo 1

Motivación y Antecedentes

1.1. Contexto

Los altos niveles de tráfico de datos móviles presentes en la actualidad, han provocado que investigadores, empresas proveedoras y clientes se interesen cada vez más en la calidad de los servicios móviles entregados. Lo anterior, se ha traducido en un incremento en la cantidad de estudios realizados y en el desarrollo de herramientas que tienen como objetivo principal establecer de forma objetiva los niveles de calidad de servicio presentes en el Internet móvil recibido por los usuarios.

Dentro de los parámetros de calidad de servicio de la red móvil comúnmente estudiados destaca el indicador de intensidad de señal recibida, el cual informa acerca del nivel de cobertura de red que reciben los dispositivos móviles. Este indicador tiene la característica de impactar la calidad general de Internet móvil, lo cual se refleja en la amplia cantidad de indicadores de calidad que se ven afectados por la intensidad de señal[49].

Pese a que las estaciones móviles monitorean constantemente su intensidad de señal, estas mediciones no suelen ser visibles para los usuarios de dispositivos móviles, quienes solo tienen acceso a los valores de intensidad de señal recibida por sus dispositivos, la cual puede variar en gran manera con respecto a la potencia que reporta la antena de telefonía celular, producto principalmente de factores externos a la red. De todas formas, para medir la calidad de servicio de la red móvil y, particularmente, para medir la intensidad de señal, resulta más coherente realizar las mediciones en los dispositivos móviles, quienes son los nodos finales de la red y por lo tanto, quienes pueden reportar efectivamente acerca de los niveles de intensidad de señal que se les ha entregado.

Utilizando métodos de agregación de datos es posible, a partir de múltiples mediciones individuales, realizar mediciones colaborativas vía procesos de *crowd-measuring* [23, 29]. De esta forma es posible agregar registros de intensidad de señal provenientes de usuarios ubicados en una misma área geográfica, con el fin de encontrar un valor representativo de la intensidad de señal que se recibe en dicha área. Así, al unir mediciones agregadas de intensidad de señal consecutivas en el tiempo, se genera el vector o serie temporal que describe el

comportamiento de la intensidad de señal recibida en el área a través del tiempo. Lo anterior permite la aplicación de técnicas propias del análisis de series temporales para realizar predicciones de la intensidad de señal recibida.

Por lo tanto, el trabajo de tesis presentado, busca comparar técnicas de aprendizaje de máquinas en relación con su operación con un gran volumen de datos, con el fin de crear un modelo de predicción de la calidad de servicio de Internet móvil, tomando como base mediciones pasivas realizadas constantemente por dispositivos móviles, capturando intensidad de la señal recibida, cantidad de datos transmitidos y recibidos, eventos de desconexión de la red e identificación de las antenas utilizadas para acceder a la red. Además, se propone la consideración de factores externos que pudiesen tener un efecto en la calidad de servicio de Internet móvil para una área geográfica específica, tales como la agrupación masiva de personas o fenómenos naturales locales.

1.2. Problema a Resolver

Este trabajo busca caracterizar la calidad de servicio de Internet móvil mediante el uso de mediciones pasivas (sin gasto de datos móviles), con el propósito de crear modelos predictivos que permitan estimar la calidad de servicio de Internet móvil para un lugar geográfico y tiempo determinado. El trabajo presentado toma la intensidad de señal recibida como principal indicador de calidad a estudiar, dada su importancia en la calidad de servicio de Internet móvil en términos generales.

Para lograr este propósito, se hará uso de la aplicación *Adkintun Mobile* [40], mediante la cual los dispositivos móviles toman mediciones de las antenas a las cuales se conectan, reportando la intensidad de señal y el tipo de red a la que están conectados (2G EDGE, 3G UMTS, 3G+ HSDPA, 4G LTE). Además, se mide continuamente el tráfico de Internet de cada aplicación instalada (bytes enviados y recibidos), pudiendo asociar de esta manera el tráfico de datos móviles con las antenas utilizadas para acceder a Internet en un momento determinado. También, dado que las mediciones de tráfico de datos se realizan con intervalo de pocos segundos, es posible obtener una estimación de la velocidad de carga y descarga de datos de cada usuario en un tiempo específico.

Otras de las funcionalidades de la aplicación es el monitoreo constante del estado de la conectividad a Internet reportada por el dispositivo móvil, con lo cual se puede saber en qué momentos los usuarios estuvieron conectados a Internet por medio de la red móvil y cuándo sufrieron eventos de desconexión. Esta última información también puede ser cruzada con la información de las antenas utilizadas por los dispositivos móviles, pudiendo mapear eventos de desconexión con la antena utilizada en ese instante, y por consiguiente, asociar a un punto geográfico. La aplicación cuenta además con otras funcionalidades enfocadas en los usuarios, con lo que se espera una mayor aceptación por parte de ellos para instalar la aplicación en sus dispositivos y seguir contribuyendo con la recolección de datos.

Con los datos disponibles, se pretende realizar modelos predictivos de intensidad de señal, basados en técnicas de análisis de series temporales presentes en el estado del arte. Además,

dada la literatura estudiada, se investigará acerca del efecto que tienen algunos factores externos sobre esta variable, tales como eventos naturales locales o la agrupación masiva de personas en un lugar geográfico (manifestaciones ciudadanas, recitales musicales o eventos deportivos).

1.2.1. Preguntas de Investigación

- ¿Es posible predecir la calidad de servicio de Internet móvil para un lugar geográfico y tiempo determinados?
- ¿Qué patrones temporales y geográficos sigue la calidad de servicio de Internet móvil?
- ¿Qué factores externos impactan en la calidad de servicio de Internet móvil? y ¿cuál es su efecto?

1.3. Hipótesis

Es posible predecir la calidad de servicio de Internet móvil para un lugar y tiempo determinados, utilizando mediciones pasivas.

1.4. Objetivos

1.4.1. Objetivo General

Crear modelos predictivos para la calidad de servicio de Internet móvil en un lugar geográfico y tiempo determinado, por medio de la realización de mediciones pasivas.

1.4.2. Objetivos Específicos

- Realizar un modelo predictivo para la intensidad de la señal recibida por los dispositivos móviles en un área geográfica determinada, basado en el análisis de series temporales.
- Comparar técnicas de aprendizaje de máquinas según su operación con grandes cantidades de datos procesados y los resultados obtenidos al realizar predicciones sobre ellos.
- Reconocer factores externos que pudiesen tener un efecto en la calidad de servicio de Internet móvil e identificar cuál es la magnitud de dichos efectos.

1.5. Metodología de Investigación

Para abordar el problema planteado, se hará uso de los datos recolectados por la aplicación *Adkintun Mobile* en la ciudad de Santiago a partir del mes de octubre del año 2016, en donde aproximadamente 300 dispositivos móviles pertenecientes a dicha localidad tenían instalada la aplicación. Esta porción de la base de datos real será manejada de forma independiente, obteniendo una copia local de los datos para trabajar con ellos.

Este set de datos debe ser previamente procesado, eliminando datos duplicados y datos inconsistentes con el fin de facilitar la realización de los análisis posteriores.

La información acerca de la intensidad de señal puede ser obtenida agrupando registros directamente de la base de datos de la aplicación y creando una serie de tiempo que identifique la intensidad de señal recibida en un área geográfica específica, la que será sometida a modelamiento para su predicción temporal. Es importante destacar que la intensidad de señal tiene gran importancia, ya que influye directamente en la calidad de servicio de Internet móvil, afectando de manera importante la latencia, cantidad de paquetes perdidos y *throughput* [49], cuyos valores no suelen ser obtenidos por medio de la realización de mediciones pasivas. Tomando como base la literatura estudiada, se utilizarán técnicas de análisis de series temporales para modelar la intensidad de señal recibida en un lugar geográfico, tales como modelos ARIMA¹, SARIMA² y el uso de redes neuronales. Las predicciones obtenidas por los modelos serán comparadas con los valores reales obtenidos para dichas fechas, presentes también en la base de datos. Esta comparación será realizada según valores de errores de predicción tales como MAE³, MAPE⁴ y RSSE⁵, y según medidas de distancias de series temporales tales como distancia de Fréchet y distancia de edición para secuencias reales o EDR⁶. Dichos errores de predicción y distancias de series temporales son definidos en detalle en la Sección 4.4.

Finalmente se investigará acerca de los posibles factores externos que tengan algún efecto en la calidad de servicio de Internet móvil y que generen una mejora en los resultados de las predicciones, evidenciada según las medidas de comparación mencionadas, tales como la agrupación masiva de personas o condiciones climáticas locales.

1.6. Resultados Esperados

Como resultado del presente trabajo de tesis se espera obtener modelos de predicción para la calidad de servicio de Internet móvil basados en mediciones pasivas aplicados en una porción central de la ciudad de Santiago, los cuales permitan predecir valores de intensidad de señal, obteniendo al menos, mejores resultados al ser comparados con un modelo “ingenuo”

¹Por sus siglas en inglés: *AutoRegressive Integrated Moving Average*

²Por sus siglas en inglés: *Seasonal ARIMA*

³Por sus siglas en inglés: *Mean Absolute Error*

⁴Por sus siglas en inglés: *Mean Absolute Percentage Error*

⁵Por sus siglas en inglés: *Root Mean Squared Error*

⁶Por sus siglas en inglés: *Edit Distance on Real sequence*

el cual repita como predicción para un cierto día los valores obtenidos una semana atrás. Además, se espera obtener mejoras en las predicciones realizadas, al incorporar en el análisis información acerca de factores externos que afecten la calidad de servicio de Internet móvil. Las comparaciones entre modelos se realizarán considerando los valores de errores de predicción y distancias de series temporales mencionados en la Sección 1.5.

1.7. Aportes de la Tesis

Los principales aportes del trabajo una vez finalizado son:

- Implementación de modelos especializados para la predicción geográfica y temporal de la calidad de servicio de Internet móvil, capaz de ser utilizado por una herramienta futura para este mismo fin.
- Comparación entre distintos modelos estadísticos y de aprendizaje de máquinas en relación con su aplicación en un problema real, identificando fortalezas y debilidades.
- Cuantificación del impacto de factores externos en la calidad de servicio de Internet móvil.
- Los métodos realizados y los resultados obtenidos conforman material para una publicación científica.

Capítulo 2

Estado del Arte

2.1. Calidad de Internet

Desde su aparición, y principalmente durante las últimas tres décadas, el servicio de Internet ha sufrido una permanente expansión, en donde solo durante los años 2000 y 2018 el uso de Internet tuvo un crecimiento de más del 1000 %, alcanzando una penetración de más del 50 % en la población total mundial [26]. Dado lo anterior, también ha aumentado la importancia de investigar acerca de la calidad de servicio de Internet.

2.1.1. Internet Fijo

A lo largo de mucho tiempo, el acceso a Internet estuvo limitado a conexiones de Internet fijo, utilizando cable coaxial, fibra óptica, entre otros, las que hasta el día de hoy corresponden a una importante porción del total de conexiones a Internet. Dado el rápido crecimiento y masificación del uso de Internet fijo, se han realizado de forma continua una gran cantidad de estudios con el fin de determinar la calidad del servicio entregado por las empresas proveedoras (ISP¹) a los clientes de Internet fijo.

De esta manera, los estudios se han enfocado en la toma de mediciones y en el análisis de los distintos indicadores de calidad obtenidos, permitiendo objetivizar la calidad de servicio entregada y utilizando esta información para diversos fines tales como:

- Permitir que usuarios puedan conocer la real calidad de servicio recibida a través de sus equipos [14], con el fin de verificar la calidad ofrecida por las empresas ISPs en sus servicios.
- Proponer nuevas formas de modelar el tráfico de Internet, tomando en cuenta los distintos servicios utilizados por los clientes y sus respectivos requerimientos de calidad de servicio [61].

¹Por sus siglas en inglés: Internet Service Provider

- Determinar herramientas de regulación de precios considerando los requerimientos de calidad de usuarios y su disposición a pagar y cumpliendo el principio de neutralidad de la red [58].
- Determinar acuerdos de nivel de servicio (SLA²) y regulaciones que permitan una comunicación efectiva entre clientes y proveedores [45].
- Entregar a las empresas proveedoras de Internet métodos que permitan conocer el estado de sus servicios, realizando comparaciones con estándares de calidad establecidos y alertando de posibles anomalías en la red [24, 32].

2.1.2. Internet Móvil

Con el posterior crecimiento de la red de telefonía celular y la aparición de servicios de Internet móvil a partir de los años 2000, el uso de Internet móvil ha adquirido continuamente más importancia, logrando en el año 2017, según el reporte de la empresa BrighthEdge [9], ser el responsable del 57% del tráfico total de Internet, por lo que también ha resultado importante el desarrollo de estudios con el fin de medir la calidad de servicio de Internet móvil. Algunos de estos estudios han utilizado en sus análisis técnicas similares a las utilizadas anteriormente para Internet fijo, replicando experimentos que permitiesen la obtención de los mismos indicadores de calidad, utilizados principalmente para la caracterización de los servicios entregados por las distintas empresas proveedoras de Internet móvil [11, 12, 19].

Otros estudios han puesto mayor atención en el hecho de que el Internet móvil debe lidiar con el constante movimiento de los usuarios del servicio, lo que introduce nuevas variables ya que dependiendo del lugar donde se encuentre un usuario puede recibir diferentes niveles de calidad de servicio [23]. Además, la regularidad de los patrones de actividades humanas en distintas áreas [25, 50] producen diferentes patrones en el uso de la red móvil [65], lo cual repercute en la calidad de servicio que recibirán los usuarios en cada sector, dada las limitantes físicas que representa el ancho de banda.

También, algunos estudios se han enfocado en el análisis de otros indicadores de calidad relacionados con la naturaleza de las redes móviles los cuales pueden sufrir cambios importantes a medida de que los dispositivos móviles cambien de celda en la red, como el tipo de tecnología que se recibe por parte de las estaciones de transmisión, analizando el desempeño de Internet en redes de tipo 3G UMTS [54], 3G HSPA [53], 3G+ HSDPA [39], 4G LTE [63], entre otras, o la cobertura de red recibida por medio del análisis del indicador de intensidad de señal recibida [23, 40]. Este último indicador, ha sido uno de los más importantes y más estudiados, principalmente dada la facilidad con la que los dispositivos realizan estas mediciones [60] y dada la influencia que tiene la intensidad de señal en la calidad de servicio global recibida en las redes móviles, lo que se refleja en el impacto que producen las variaciones de intensidad de señal en mediciones de desempeño de la red [31, 49, 60].

El trabajo presentado se enfoca en el análisis de la calidad de servicio de Internet móvil, principalmente por el hecho de que actualmente es la manera predominante de acceder a Internet (por sobre el uso de Internet fijo) y se proyecta un aumento aún mayor en el uso

²Por sus siglas en inglés: Service-Level Agreement

de Internet móvil en los próximos años. Además, como lo refleja la Figura 2.1 perteneciente al reporte global de Internet del año 2015 de la organización *Internet Society* [10], cada vez son más los países que se embarcan en proyectos de desarrollo y actualización de sus infraestructuras de red móvil, producto del hecho de que este servicio se encuentra en constante evolución tecnológica. Lo anterior sumado al hecho de estar *ad portas* de la llegada y masificación de los llamados servicios de 5G, indica que el análisis de calidad de Internet móvil es un área donde hay y seguirá existiendo la posibilidad de realizar trabajos de investigación y así aportar a la comunidad científica.

Number of countries with mobile network deployments using different technologies

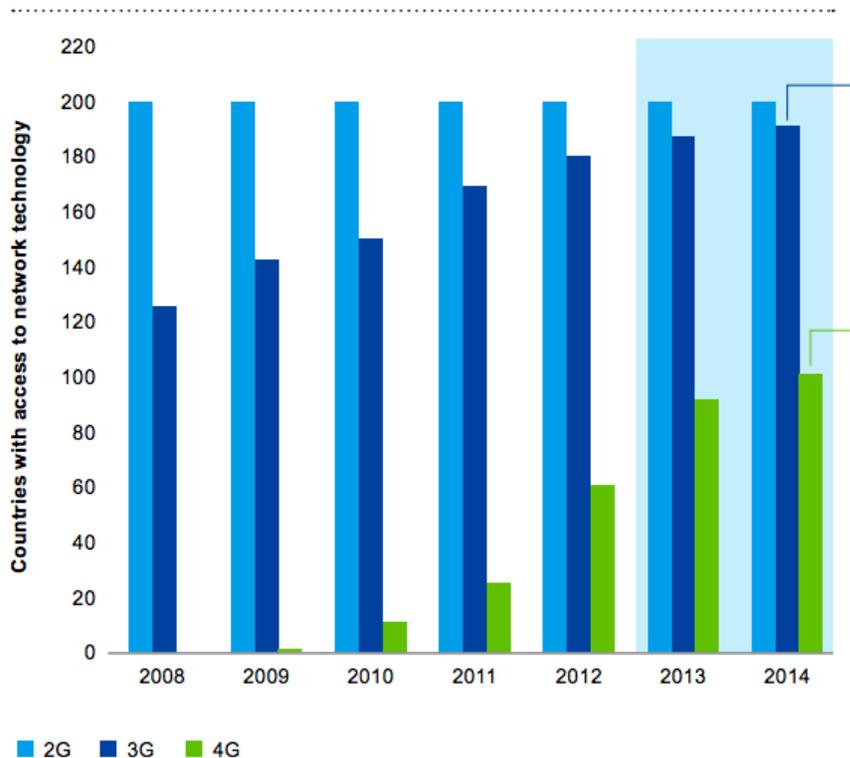


Figura 2.1: Número de países con despliegue de infraestructura de red móvil.

2.2. Mediciones Activas y Pasivas

Para el estudio de la calidad de servicio de Internet tanto fijo como móvil, se utilizan diferentes parámetros o indicadores de calidad, los que informan acerca de variados aspectos de la conexión a Internet establecida, pudiendo conocer la calidad de servicio recibida desde distintos puntos de vista. Estos indicadores son calculados a partir de mediciones tomadas por dispositivos o nodos de la red, las cuales son comúnmente clasificadas en mediciones activas y pasivas.

Es importante notar que a pesar de que las mediciones activas y pasivas han sido em-

pleadas en gran manera durante las últimas dos décadas, y se han desarrollado estudios para entender sus diferencias [47] y limitaciones [27], recién en el año 2016 la organización IETF³, encargada de establecer los estándares de Internet, entregó definiciones claras y precisas acerca de qué es una medición activa y una medición pasiva, cómo son definidos sus métodos y sus métricas, dando a conocer herramientas que permitan evaluar futuros métodos de mediciones de Internet desarrollados [48].

2.2.1. Mediciones Activas

Las mediciones activas son aquellas en las cuales el dispositivo tomador de las mediciones introduce paquetes a la red, estableciendo una comunicación con otro nodo presente en la red y por lo tanto, generando tráfico de datos y recopilando información de este intercambio de mensajes [48].

En la Figura 2.2, se ejemplifica un proceso de medición activa, en donde un nodo agente tomador de la medición envía una petición hacia un nodo servidor de la red, generando tráfico de red. Una vez recibido el mensaje, el servidor envía una respuesta al nodo agente, quien al recibirla puede medir el tiempo de respuesta del servidor.

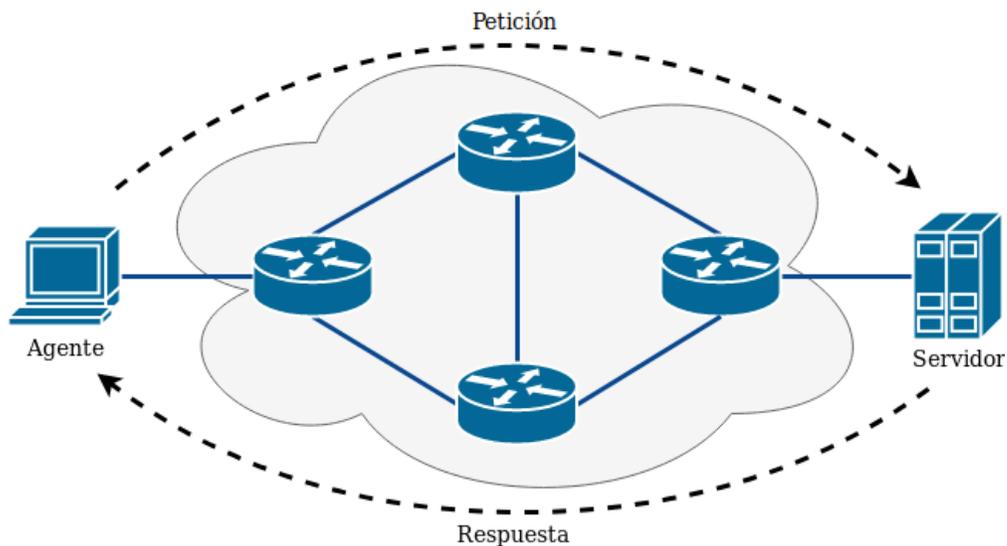


Figura 2.2: Ejemplo de medición activa de la red.

Dentro de los indicadores recolectados por mediciones activas más utilizados en estudios de calidad de servicio se encuentran aquellos que son recolectados por medio de la herramienta de software conocida como *ping*, la cual realiza sus mediciones enviando paquetes ICMP⁴ de tipo *echo request* y esperando la llegada de paquetes ICMP de tipo *echo reply* como respuesta. Los indicadores obtenidos por *ping* son:

³Por sus siglas en inglés: Internet Engineering Task Force

⁴Por sus siglas en inglés: Internet Control Message Protocol

- **Tiempo de ida y vuelta (RTT):** Es el tiempo que transcurre entre el envío del mensaje ICMP *echo request* y la llegada de su respuesta *echo reply*. Comúnmente es referido como **latencia**.
- **Jitter:** Calculado como la varianza entre mediciones consecutivas de RTT. Jitter es un valor no deseado, ya que en una comunicación óptima, los paquetes de respuesta debiesen llegar en forma periódica.
- **Porcentaje de paquetes perdidos:** Porcentaje de mensajes ICMP *echo request* que no recibieron respuesta con respecto al total de mensajes *echo request* enviados.

Estos parámetros suelen ser analizados en conjunto para determinar niveles de calidad de Internet fijo [44, 56] y móvil [11, 19, 49]. También, con estos valores se ha estudiado la **disponibilidad de conexión** (o qué tan probable es sufrir una desconexión de Internet) monitoreando constantemente la accesibilidad de algunos nodos en la red [12] y pudiendo relacionar desconexiones de la red con eventos meteorológicos severos como huracanes [28] y tormentas [56].

A pesar de su común uso, *ping* no reporta medidas precisas, ni refleja necesariamente la calidad de servicio recibida, principalmente debido a que se basa en el uso de paquetes ICMP (utilizados solo con fines de control y diagnóstico), y pueden ser afectados por diferentes mecanismos de control de tráfico que los paquetes TCP.

Otro tipo de mediciones activas común en análisis de calidad de Internet, es la velocidad de transmisión de datos, o **velocidad de subida** y **velocidad de bajada**, en donde el nodo que realiza estas pruebas, se conecta con un servidor en la red e inicia un intercambio de datos, generalmente descargando y subiendo un archivo al servidor. Durante este proceso, los paquetes son monitoreados con el fin de establecer la velocidad con la que se envían y reciben datos con respecto a ese servidor en particular [39, 54, 63].

Es importante notar que estas mediciones son calculadas en conexiones *end to end*, obteniendo solamente información respecto a la calidad de servicio presente en el camino entre los dos nodos comunicados (cliente y servidor), lo cual no representa necesariamente la calidad general del servicio recibido por el dispositivo que realiza las mediciones, debido a que las pruebas activas reportan el comportamiento de solo una de las múltiples conexiones que pueden presentarse de forma simultánea en el dispositivo. Además, dichas pruebas son realizadas en un momento específico y capturan la calidad recibida en ese preciso instante, por lo que se requiere mucha información para poder determinar cuándo es un buen momento para realizar las pruebas. Realizar mediciones activas de forma constante para registrar la calidad recibida durante un mayor tiempo resulta difícil, ya que esto implicaría una permanente sobrecarga del dispositivo debido a la constante generación de tráfico de datos, lo cual sería especialmente perjudicial en dispositivos móviles, pues los planes de datos de Internet se verían fuertemente afectados, debido a que estos suelen ser limitados en cuanto a la cantidad de datos enviados y recibidos permitidos.

2.2.2. Mediciones Pasivas

Las mediciones pasivas son aquellas que son tomadas sin crear o modificar el tráfico de red, sino que observando el estado de la red en un punto particular de esta. Este tipo de mediciones tiene la ventaja de que al no introducir paquetes en la red, no generan un impacto en el resto de conexiones de Internet activas, lo cual es muy importante para poder llevar a cabo mediciones de forma constante. Las pruebas pasivas no interfieren con el tráfico de red que es manejado por el dispositivo que toma las mediciones, ni genera más tráfico de datos que pudiera disminuir el desempeño de las conexiones activas. Por lo tanto, este tipo de mediciones permite monitorear simultáneamente todas las conexiones activas en un dispositivo minimizando el impacto sobre ellas. Además, ya que en este tipo de mediciones solo se observa el estado de la red en un punto, no es necesario que el dispositivo que toma las mediciones sea un dispositivo final de la red, sino que es posible que se trate de un nodo intermedio por el cual pasen distintos caminos entre nodos de la red, permitiendo así que estos nodos intermedios puedan tomar medidas pasivas de todo el tráfico que circula a través de ellos, permitiendo realizar análisis de grandes porciones de la red en tiempo real [52, 55].

En la Figura 2.3, se ejemplifica un proceso de medición pasiva, en donde el tráfico de red es pasivamente monitoreado duplicando el tráfico con la ayuda de un duplicador de red (*splitter*) o analizando los *buffers* presentes en un *router* monitoreado.

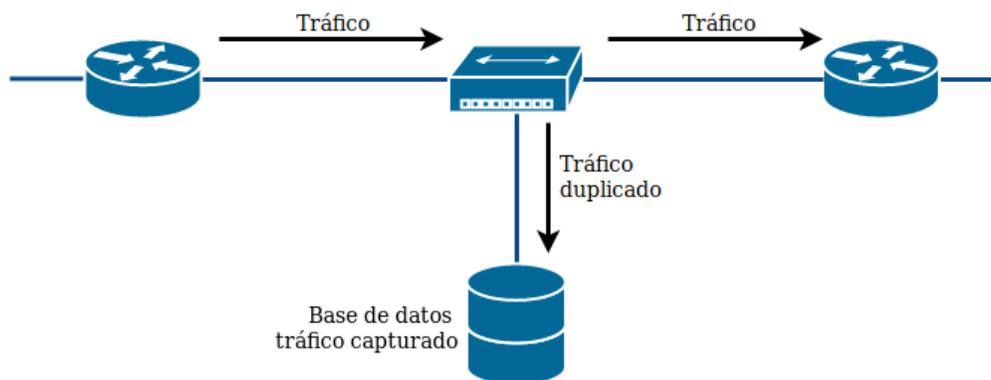


Figura 2.3: Ejemplo de medición pasiva de la red.

Dentro de los indicadores obtenidos por medio de mediciones pasivas utilizados para analizar el Internet fijo y móvil se encuentran:

- **Volumen de tráfico de datos:** Este indicador se obtiene sumando el tamaño en bytes de todos los paquetes circulantes por la red, para lo que comúnmente se utilizan sets de datos provenientes de ISPs, quienes cuentan con la información completa del tráfico de datos generada por todos sus usuarios [46, 65, 67], logrando análisis muy finos. También, se han desarrollado aplicaciones móviles que permiten capturar esta información observando el tráfico de datos circulante [15, 40], pudiendo diferenciar los flujos de datos de cada aplicación instalada en el dispositivo.
- **Latencia:** Esta información, comúnmente obtenida por medio de la herramienta de medición activa *ping*, puede ser también obtenida de forma pasiva, analizando los tiempos de salida y entrada de paquetes TCP por la interfaz de red [37] o por una máquina

intermedia por donde el dispositivo accede a Internet [43], o analizando la opción *timestamp* en las cabeceras TCP [35], siguiendo el mecanismo introducido en RFC 1323 [36]. A diferencia de la herramienta *ping*, estos métodos no introducen paquetes en la red, sino que analizan los paquetes TCP pertenecientes a las demás conexiones de Internet establecidas por el cliente.

- **Disponibilidad:** Esto ha sido estudiado en dispositivos móviles por medio del análisis del tipo de conexión (WiFi o Móvil) y analizando el flujo de datos transmitidos y recibidos (información entregada por el sistema operativo en dispositivos Android) [40].
- **Throughput:** Definida como la tasa real en que la información es transferida y calculada como la cantidad de bits (tamaño de los paquetes) que se reciben en el dispositivo por unidad de tiempo.
- **Goodput:** Es el símil de *throughput* pero a nivel de aplicación y se entiende como la tasa en que la información útil es transferida. A diferencia de *throughput*, esta medida no considera los bits correspondientes a *headers* de protocolos ni a paquetes retransmitidos [8].

También, en el estudio de conexiones inalámbricas como redes WLAN y redes móviles, se ha considerado en gran manera el estudio del indicador de **intensidad de señal recibida**, el cual cobra gran importancia ya que se ha probado su directa influencia en indicadores de calidad de servicio importantes. Se ha demostrado una clara correlación entre la intensidad de señal y la velocidad de transmisión de datos en conexiones TCP [31, 60], así como una correlación con indicadores importantes como latencia, cantidad de paquetes perdidos y velocidad de transmisión de datos [49]. En general, dichos estudios establecen que con intensidades de señales muy bajas, la calidad de Internet es baja ya que se sufren muchas desconexiones, mientras que a medida que aumenta la intensidad de señal, la calidad de servicio aumenta hasta un límite en el cual una mejor intensidad de señal no produce mejoras en la calidad de servicio de Internet. Otra ventaja en el uso de este indicador, es que resulta fácil su obtención, por lo que variados estudios han desarrollado herramientas para dispositivos móviles que capturan estos registros para su análisis [40, 60] o para, a partir de la intensidad de señal recibida, realizar estimaciones de otros indicadores de calidad [1, 15].

Una de las principales dificultades que tienen las mediciones pasivas, es que suelen presentar problemas de escalabilidad [47], ya que si se cuenta con varios puntos en la red que capturen mediciones pasivas de los paquetes, la cantidad de datos a almacenar resulta muy grande dada la magnitud del tráfico de datos de Internet, lo que dificulta a su vez la realización de análisis sobre los datos.

Tanto mediciones activas, pasivas e híbridas [48] han sido utilizadas en los estudios para obtener métricas de calidad que permitan objetivar la calidad de servicio recibida por los usuarios de Internet y así poder comparar los servicios entregados por las empresas proveedoras, tanto de Internet fijo [14, 32, 45] como móvil [12, 19, 64].

El trabajo de tesis presentado comprende el uso de mediciones pasivas para realizar análisis de calidad de Internet, aprovechando las grandes cantidades de datos que este tipo de mediciones genera para aplicar algoritmos estadísticos y de aprendizaje de máquinas. Los datos utilizados en este trabajo corresponden a mediciones de calidad de servicio de Internet móvil, provenientes de la base de datos de la aplicación chilena *Adkintun Mobile*, lo cual será

tratado en más detalle en el Capítulo 3.

2.3. Análisis de Calidad de Servicio

Con la gran cantidad de datos recopilados por las herramientas desarrolladas para tomar medidas de calidad de servicio de Internet, investigadores e ISPs han llevado a cabo estudios enfocados en el análisis de estas mediciones de calidad de servicio. Los análisis han sido realizados considerando las mediciones tomadas por un único dispositivo y obteniendo información respecto a ese dispositivo en particular [15, 37, 63], o agrupando las mediciones de múltiples dispositivos y obteniendo información acerca de la calidad percibida por un conjunto de usuarios, generalmente agrupados según la ubicación geográfica. Dada la naturaleza estática de las conexiones a Internet fijo, la geolocalización de estos equipos resulta fácil [14, 56]. Sin embargo, para el caso de conexiones de Internet móvil, se deben utilizar servicios de GPS [12, 23] o geolocalización de las antenas usadas por los dispositivos para acceder a Internet [40], y de esta manera poder estimar la posición geográfica de los dispositivos para su posterior agrupación.

Las investigaciones llevadas a cabo para analizar los registros de calidad de servicio de Internet, pueden enfocarse en la confección de reportes de la calidad de servicio observada en un intervalo de tiempo pasado o para realizar predicciones de la calidad de servicio que se recibirá en un tiempo futuro. Según lo anterior, es posible establecer la diferencia entre análisis estadísticos y análisis predictivos de la calidad de servicio de Internet.

2.3.1. Análisis Estadísticos

Los análisis estadísticos corresponden a estudios que a partir de una gran cantidad de datos recopilados realizan procesos para obtener valores representativos de los indicadores de calidad considerados. Dichos estudios tienen el fin de mostrar el estado de la calidad de servicio de Internet durante el período en el cual se tomaron las mediciones, teniendo la característica de presentar información acerca de un intervalo de tiempo pasado, el cual puede ser muy reciente en el caso de los llamados análisis en tiempo real.

Algunos estados de calidad de servicio son realizados para intervalos grandes de tiempo, del orden de semanas o meses, los cuales suelen tener como objetivo principal caracterizar la calidad de Internet en un área específica y comparar el desempeño de las distintas empresas proveedoras de Internet [12, 13, 42], en donde ciertos estudios utilizan como parámetro de comparación mediciones de calidad tomadas con respecto a los servicios de Internet más demandados [19, 64]. Otras investigaciones que desarrollan análisis estadísticos, se enfocan en realizar un estado de la calidad observada en un período corto o en un evento particular, con duración de uno o pocos días, en donde no se busca caracterizar la calidad general del área, sino aspectos atinentes al evento en particular. Entre estos análisis se encuentra el estudio de la accesibilidad a Internet durante eventos climáticos fuertes [28, 56] o el análisis de la calidad de servicio recibida durante eventos masivos [23, 57].

Un ejemplo notable de este tipo de análisis, corresponde a los estados de las redes móviles llevados a cabo por la empresa *OpenSignal* para una gran cantidad de países, utilizando los datos recopilados por más de 20 millones de usuarios de su aplicación móvil [51].

2.3.2. Análisis Predictivos

Los análisis predictivos de los servicios de Internet entregados corresponden a estudios que toman los datos de calidad recopilados e intentan, a partir de ellos, predecir valores futuros de la calidad de servicio que recibirán los usuarios de las redes. Para realizar esto, se suelen representar los indicadores de calidad en forma de series temporales, las que son entregadas como variable de entrada a modelos de predicción basados en técnicas estadísticas y de aprendizaje de máquinas para obtener pronósticos acerca de los siguientes valores de las series de tiempo, y por consiguiente, pronosticar acerca de los valores futuros de los indicadores de calidad de servicio.

Dentro de los estudios predictivos para la calidad de Internet fijo, destacan la predicción de RTT utilizando redes neuronales y regresiones lineales [66] y la predicción del tiempo de respuesta de servicios web utilizando modelos ARIMA⁵ y modelos GARCH⁶ [5].

Con respecto al Internet móvil, un trabajo recientemente publicado acerca de predicción de calidad de servicio se refiere a la predicción de la velocidad de descarga para red 4G LTE mediante la aplicación de mediciones activas y utilizando variaciones de modelos ARIMA [63].

El presente trabajo busca aportar precisamente en el área de predicción de calidad de servicio de Internet móvil, en donde, según el presente estudio del estado del arte, se ha publicado solamente un trabajo de investigación realizado, el cual data del año 2017. Además, según lo estudiado, aún no ha sido explorado el uso de modelos predictivos de calidad de Internet móvil que utilicen como base mediciones pasivas de la red.

Existen estudios que han implementado técnicas predictivas para otros parámetros presentes en la red móvil, los cuales se encuentran estrechamente relacionados con la calidad de servicio, lo que sustenta la idea de realizar pronósticos de la calidad de Internet móvil. Algunos de estos han implementado modelos predictivos para conocer la congestión de la red en un lugar geográfico específico, como la predicción del tráfico de llamadas utilizando modelos ARIMA [3], mediante el uso de análisis caótico [30], y la predicción del tráfico de datos de Internet móvil utilizando modelos SARIMA (*Seasonal ARIMA*) para redes 2G GSM [59], redes 2G/3G [67] y redes 3G/4G LTE [46]. Dichos estudios, a pesar de no realizar predicciones sobre la calidad de servicio directamente, tienen gran relevancia para el estudio de Internet móvil, ya que el volumen del tráfico presente en la red de telefonía celular puede afectar directamente la calidad de servicio en un punto geográfico específico, dados los altos niveles de congestión de red que se pueden producir. Además, también se han desarrollado investigaciones con el fin de predecir la calidad de experiencia percibida por un usuario de Internet móvil a partir de mediciones pasivas de calidad de servicio [1, 15], lo cual también

⁵Por sus siglas en inglés: AutoRegressive Integrated Moving Average

⁶Por sus siglas en inglés: Generalized AutoRegressive Conditional Heteroscedastic

es un buen predecesor para el trabajo a realizar.

2.4. Origen de Datos

2.4.1. Datos Internos de la Red

Con respecto a los datos utilizados por los estudios de calidad de servicio, una gran parte de ellos realiza análisis haciendo uso solamente de medidas de calidad de servicio, que corresponden a aquellas descritas en la Sección 2.2. Estas medidas son tomadas por diversos dispositivos conectados a Internet, tales como computadores personales [6, 16, 37], *routers* [14], dispositivos móviles [12, 19, 40, 49, 64] o *hardware* desarrollado especialmente para este tipo de tareas [63]. Esta fuente de datos tiene la característica de informar acerca de parámetros internos a la red, recopilados por ciertos nodos que forman parte de ella.

2.4.2. Datos Externos

Otros estudios han tomado los datos técnicos descritos anteriormente y han llevado a cabo análisis para determinar la correlación existente entre estos datos técnicos y datos acerca de factores externos a las redes de Internet, buscando un mayor entendimiento de los patrones temporales presentes en la calidad de servicio de Internet, logrando así enriquecer los análisis de calidad llevados a cabo.

Entre estos estudios se encuentran las investigaciones acerca del impacto de factores climáticos en la calidad de servicio de Internet, como el efecto de la temperatura y humedad en la calidad de redes inalámbricas [41], principalmente en el indicador de intensidad de señal; y el efecto de lluvias, tormentas y huracanes en la estabilidad y disponibilidad de las conexiones de Internet [28, 56]. Además, se ha estudiado el efecto de lluvias y nevadas en el volumen de tráfico de redes móviles [3], lo cual es importante de tomar en cuenta ya que el volumen de tráfico en la red tiene incidencia en la calidad de servicio a recibir.

Además, recientes investigaciones han realizado los primeros pasos para determinar el efecto en el desempeño de la red de telefonía celular frente a agrupaciones masivas de personas en grandes eventos, utilizando para esto tanto el uso de mediciones activas como mediciones pasivas [23, 57].

El trabajo presentado, propone el análisis de datos técnicos (provenientes de la aplicación *Adkintun Mobile*) en conjunto con datos externos acerca de las condiciones climáticas presentes en el lugar donde las mediciones de calidad de servicio a utilizar han sido tomadas. Estos factores externos han sido considerados por el hecho de que representan, según el análisis realizado, el factor externo mayormente estudiado en investigaciones acerca de calidad de servicio de Internet, demostrando la importancia de su consideración. Una explicación más detallada de los datos utilizados se encontrará en el Capítulo 3.

2.5. Resumen

Dado el estudio del estado del arte realizado y presentado anteriormente, es posible categorizar las investigaciones llevadas a cabo en el área de estudio de la calidad de servicio de Internet según las siguientes distinciones:

1. Análisis de **Internet fijo** o **Internet móvil**
2. Aplicación de **mediciones activas** o **mediciones pasivas**
3. **Análisis estadísticos** o **análisis predictivos**
4. Uso de **datos internos de la red** o **datos externos**

A continuación, se encuentra una recapitulación multidimensional de los estudios analizados para el presente trabajo de tesis, en donde cada trabajo es clasificado según los cuatro puntos presentados. Por simplicidad, este resumen se encuentra dividido en dos tablas, una representando los estudios acerca de Internet fijo y otra para los estudios acerca de Internet móvil.

Tabla 2.1: Estudios de calidad de servicio para Internet fijo.

	Mediciones Activas	Mediciones Pasivas
Análisis Estadísticos	[6, 24, 42, 44] [28, 56]	[13, 14, 32, 35, 37, 43]
Análisis Predictivos	[5, 66, 68]	[16]

Datos Internos de la Red
Datos Internos de la Red + Datos Externos

Tabla 2.2: Estudios de calidad de servicio para Internet móvil.

	Mediciones Activas	Mediciones Pasivas
Análisis Estadísticos	[11, 12, 19, 39, 49, 53, 54, 60, 64] [23, 31, 57]	[8, 40, 49, 60, 64] [23, 31, 41, 57]
Análisis Predictivos	[63]	

Datos Internos de la Red
Datos Internos de la Red + Datos Externos

De esta manera, resulta claro observar que los análisis de calidad de servicio de Internet fijo y móvil son en su mayoría del tipo estadístico, por sobre los análisis predictivos, los cuales utilizan como base tanto mediciones activas como pasivas.

También, es importante notar que para análisis estadísticos de calidad de servicio la cantidad de estudios que considera factores externos a la red es considerablemente menor a los estudios que no consideran factores externos. A su vez, para análisis predictivos, no se halló ningún estudio que considerara la inclusión de factores externos a la calidad de servicio de Internet.

Por lo tanto, el trabajo presentado, busca hacer un aporte en el área de análisis predictivos de calidad de servicio de Internet móvil utilizando mediciones pasivas y considerando factores externos, lo cual no ha sido aún estudiado. Así, se busca comparar técnicas estadísticas y de aprendizaje de máquinas, con el fin de desarrollar modelos de predicción de la calidad de servicio de Internet móvil, específicamente de la intensidad de señal, tomando como base mediciones pasivas realizadas constantemente por dispositivos móviles. Además, se propone la consideración de factores externos que pudiesen tener un efecto en la calidad de servicio de Internet móvil para una área geográfica específica, como información acerca de las condiciones climáticas locales.

Capítulo 3

Descripción de Datos

3.1. Adkintun Mobile

Para este trabajo se utilizaron los datos pertenecientes a la aplicación chilena para dispositivos Android llamada *Adkintun Mobile*. Esta aplicación funciona como un monitor pasivo de la calidad de los servicios entregados por las compañías de telefonía celular, recolectando constantemente datos acerca del estado del dispositivo, para luego ser enviados a una base de datos central. *Adkintun Mobile* ejecuta diferentes procesos (llamados monitores), cada uno encargado de recolectar un tipo específico de información, los cuales son explicados a continuación.

3.1.1. Telephony Monitor

Este monitor accede a información acerca de la señal recibida por el dispositivo por parte de las antenas de telefonía celular, mediante la extensión de la clase `PhoneStateListener`¹, sobrescribiendo los métodos `onSignalStrengthsChanged`, gatillado cada vez que hay un cambio en la intensidad de señal recibida, y `onCellLocationChanged`, gatillado cada vez que el dispositivo se conecta a una nueva antena. El primer método recibe un objeto `SignalStrength` con información acerca del nivel de potencia de señal recibida, mientras que el segundo método recibe un objeto `CellLocation` con información de la tecnología de red que se está recibiendo por parte de la antena (3G UMTS, 3G+ HSDPA, 4G LTE, entre otras) y con los identificadores de la antena *Cell ID* (CID) y *Location Area Code* (LAC), los cuales generan una llave única para cada estación base de transmisión dentro de la red de telefonía móvil, correspondiente a una antena de telefonía celular, ya que el identificador CID es único dentro de cada LAC.

A través de este método, los dispositivos generan continuamente tuplas de datos con el formato mostrado en la Tabla 3.1, las cuales son enviadas al servidor de *Adkintun Mobile* y guardadas en una base de datos en la tabla `gsm_events`.

¹<https://developer.android.com/reference/android/telephony/PhoneStateListener>

Campo	Significado	Ejemplo
device_id	Código IMEI identificador del dispositivo	356938035643808*
sim_serial_number	Identificador único de tarjeta SIM	8956070200012908633*
date	Marca de tiempo en formato de Tiempo Unix o Tiempo POSIX	1526489185[s]
mcc	Código MCC ² identificador del país	730
mnc	Código MNC ³ identificador del operador de telefonía móvil	2
gsm_cid	Código CID identificador de la estación base de transmisión	83766
gsm_lac	Código LAC de área de ubicación	3853
network_type	Tecnología de red utilizada	4G_LTE
signal_strength	Intensidad de señal registrada	-70[dBm]

* Valores ficticios

Tabla 3.1: Registros de intensidad de señal recolectados por *Adkintun Mobile*.

A su vez, el servidor de *Adkintun Mobile* realiza la geolocalización de cada una de las antenas, mediante el uso de la herramienta *OpenCellID*, la cual cuenta con una API de libre acceso para hacer consultas sobre información de antenas de la red móvil, y así poder asociar los identificadores CID y LAC con una posición geográfica, información almacenada en la tabla `antennas`, cuyo formato se presenta en la Tabla 3.2.

Campo	Significado	Ejemplo
gsm_cid	Código CID identificador de la estación base de transmisión	83766
gsm_lac	Código LAC de área de ubicación	3853
lat	Latitud geográfica	-33,453475
lon	Longitud geográfica	-70,660727

Tabla 3.2: Registros de antenas generados en el servidor de *Adkintun Mobile*.

3.1.2. Connectivity Monitor

Este monitor se encarga de informar acerca de cambios en la conectividad a Internet del dispositivo, atendiendo la notificación `CONNECTIVITY_ACTION` entregada por la clase

²Por sus siglas en inglés: *Mobile Country Code*

³Por sus siglas en inglés: *Mobile Network Code*

`ConnectivityManager`⁴. Esta notificación es gatillada cuando ocurre un cambio en la conectividad a la red, indicando que la conexión ha sido establecida o perdida. Luego de recibir este aviso, el monitor obtiene un objeto de la clase `NetworkInfo` con información acerca del estado actual de la red, pudiendo conocer el tipo de conexión utilizada (WiFi o Móvil), si se cuenta o no con acceso a Internet, si se están utilizando servicios de *roaming* e información detallada del estado en que se encuentra la conexión a la red (en proceso de autenticación, conectando, escaneando red, conexión fallida, entre otras). Con esta información se generan continuamente tuplas de datos con el formato mostrado en la Tabla 3.3, las cuales son enviadas al servidor de *Adkintun Mobile* y guardadas en una base de datos en la tabla `connectivity_events`.

Campo	Significado	Ejemplo
<code>device_id</code>	Código IMEI identificador del dispositivo	356938035643808*
<code>sim_serial_number</code>	Identificador único de tarjeta SIM	8956070200012908633*
<code>date</code>	Marca de tiempo en formato de Tiempo Unix o Tiempo POSIX	1526489185[s]
<code>roaming</code>	Uso de servicios de <i>roaming</i>	<i>false</i>
<code>available</code>	Disponibilidad de acceso a Internet	<i>true</i>
<code>connected</code>	Conectado a alguna red	<i>false</i>
<code>detailed_state</code>	Estado detallado de la conexión	AUTHENTICATING
<code>connectivity_type</code>	Tipo de acceso a Internet	WiFi

* Valores ficticios

Tabla 3.3: Registros de estado de la conectividad de red recolectados *Adkintun Mobile*.

3.1.3. Traffic Monitor

Este monitor setea una alarma cada diez segundos para analizar el flujo de datos que ha experimentado el dispositivo móvil en el último intervalo de tiempo. Para esto, monitorea cada una de las aplicaciones instaladas, utilizando la clase `TrafficStats`⁵, que provee estadísticas relacionadas con el tráfico de red generado por las aplicaciones. Estas estadísticas incluyen información acerca del total de bytes transmitidos y recibidos, acerca del total de paquetes transmitidos y recibidos y acerca de la interfaz por la cual se registró el flujo de datos. Ya que estas estadísticas entregan información del total del tráfico generado desde la última vez que el dispositivo fue encendido, el monitor guarda continuamente información respecto de los últimos valores obtenidos por cada aplicación, con el fin de en cada iteración registrar solo la información correspondiente al flujo de tráfico de los últimos diez segundos, lo cual se calcula fácilmente restando los datos obtenidos con los últimos valores guardados. Con esta información se generan tuplas de datos con el formato mostrado en la Tabla 3.4, las cuales son enviadas al servidor de *Adkintun Mobile* y guardadas en una base de datos en la tabla `traffic_events`.

⁴<https://developer.android.com/reference/android/net/ConnectivityManager>

⁵<https://developer.android.com/reference/android/net/TrafficStats>

Campo	Significado	Ejemplo
device_id	Código IMEI identificador del dispositivo	356938035643808*
sim_serial_number	Identificador único de tarjeta SIM	8956070200012908633*
date	Marca de tiempo en formato de Tiempo Unix o Tiempo POSIX	1526489185[s]
package_name	Identificador único de aplicaciones Android	cl.niclabs.adkintunmobile
rx_bytes	Cantidad de bytes recibidos	250
tx_bytes	Cantidad de bytes transmitidos	250
rx_packets	Cantidad de paquetes recibidos	10
tx_packets	Cantidad de paquetes transmitidos	10
network_type	Tecnología de red utilizada	4G_LTE

* Valores ficticios

Tabla 3.4: Registros de tráfico de datos por aplicación recolectados por *Adkintun Mobile*.

3.2. Privacidad de Datos

Es importante notar que los datos recopilados sin duda alguna generan un conflicto de privacidad, ya que la información acerca de las antenas a las cuales los dispositivos se han conectado, puede ser usada para geolocalizar fácilmente el recorrido de un usuario de la aplicación, razón por la cual se han desarrollado métodos para el contexto de mediciones en masa que permitan obtener información agregada de los usuarios sin revelar información sensible de usuarios particulares [22].

De todas maneras, al instalar la aplicación *Adkintun Mobile*, los usuarios toman conocimiento de la política de privacidad establecida⁶, en la cual se describen los datos recolectados por la aplicación y se manifiesta que los datos recopilados son utilizados por NIC Chile Research Labs y que solo pueden ser utilizados con fines científicos, el cual es precisamente el contexto del presente trabajo de tesis.

3.3. Procesamiento de Datos

Los datos provenientes de la base de datos de *Adkintun Mobile* utilizados en este trabajo, tienen ubicación en la ciudad de Santiago de Chile, provenientes de aproximadamente 300 dispositivos móviles en dicha área.

Dado que una de las grandes motivaciones de este trabajo es probar la factibilidad de predecir valores futuros de calidad de servicio de Internet móvil en un área específica, se

⁶<http://adkintunmobile.cl/#privacy>

decidió realizar los experimentos sobre un área acotada de la ciudad de Santiago. El área escogida corresponde a la parte central de la comuna de Santiago, ubicada en el corazón de la ciudad de Santiago y que cubre una extensión de 15 km^2 aproximadamente. Esta elección fue realizada principalmente por el hecho de que esta comuna es la que concentra mayor cantidad de datos de usuarios, y además, corresponde a un área de la ciudad fuertemente caracterizada como área de oficina, por lo que se generan patrones de movilidad de usuarios con marcadas componentes periódicas, regidas según los horarios laborales, lo cual es importante al momento de analizar datos de calidad de servicio, puesto que el uso de la red puede influir en el rendimiento de esta.

Los datos de la comuna de Santiago utilizados, corresponden a registros de la tabla `gsm_events` presentada en la Tabla 3.1, obteniendo cerca de 20000 mediciones de intensidad de señal diarias, referentes a antenas de la red de telefonía celular ubicadas en el área seleccionada.

El problema de predecir valores futuros de intensidad de señal en la región mencionada, es abordado desde el punto de vista de predicción de series temporales. Por lo tanto, una tarea primordial contempla agregar temporalmente los datos provenientes de los usuarios para crear una única serie de tiempo, en donde cada punto de la serie corresponde a un valor representativo de la intensidad de señal recibida en el área en ese momento. Así, los registros son agregados en intervalos de dos horas, reportando valores correspondientes a la mediana de intensidad de señal en ese período. La elección de intervalos de dos horas permite obtener una curva de intensidad de señal agregada con una forma mejor definida que utilizando intervalos de tiempo menor, ya que así no se generan puntos faltantes en la curva, logrando una mayor completitud de la curva. Lo anterior resulta también beneficioso para la tarea de predecir valores futuros de intensidad de señal, ya que los modelos de predicción entregan mejores resultados al ser entrenados sin datos faltantes.

Ya que en cada intervalo de dos horas un usuario puede haber estado conectado a la red móvil por medio de múltiples antenas dentro del área de medición, y puede haber recibido distinta intensidad de señal a lo largo de una conexión con una antena, los valores de mediana son calculados ponderando cada valor de intensidad de señal por el tiempo (en segundos) en que el dispositivo estuvo recibiendo dicho nivel de intensidad de señal. De esta forma, un dispositivo con mayor cantidad de mediciones no tiene más peso en el valor final agregado, sino que todos tienen un peso proporcional al tiempo que permanecieron dentro del área de mediciones en cada intervalo.

Con el propósito de validar el uso de modelos de predicción de la calidad de servicio de Internet móvil utilizando series temporales, se realizó un análisis preliminar de la serie de tiempo de intensidad de señal, creada según lo descrito anteriormente. Los resultados obtenidos muestran que al agrupar las mediciones por ubicación geográfica (basado en la latitud y longitud de las antenas involucradas en cada medición) es posible observar patrones en la intensidad de señal recibida en función del tiempo. A modo de ejemplo, en la Figura 3.1 se muestra la intensidad de señal registrada en la comuna de Santiago durante una semana de lunes a domingo. Los datos del gráfico corresponden a valores agregados de 266637 mediciones tomadas por 109 dispositivos móviles, obteniendo la intensidad de señal recibida por las antenas presentes en la comuna. Es posible observar una periodicidad de 24 horas en los

datos obtenidos, mayormente identificable hasta el día viernes. Los cambios visibles para los días sábado y domingo, indican la posibilidad de la existencia también de una periodicidad de 7 días.

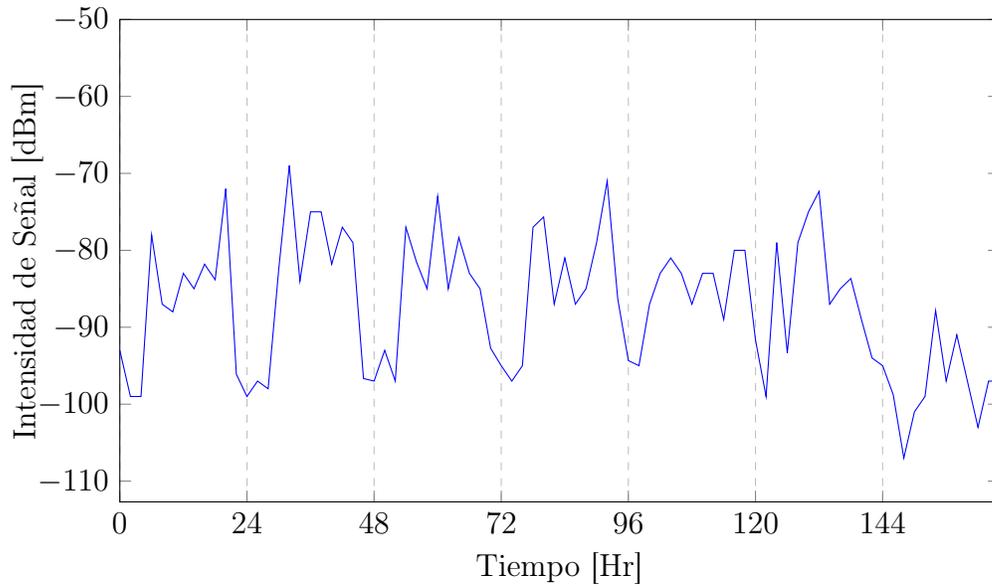


Figura 3.1: Intensidad de señal recibida en la comuna de Santiago durante una semana.

La dependencia del tiempo observada en las mediciones de intensidad de señal, promueven la factibilidad de la implementación de modelos de predicción basados en series temporales para la calidad de servicio de Internet móvil.

Capítulo 4

Modelos de Predicción

Con el fin de seguir los métodos utilizados en trabajos relacionados, se propone el uso de los modelos estadísticos de predicción ARIMA y SARIMA, así como el uso de redes neuronales comúnmente utilizadas en el análisis de series temporales. Estos métodos son descritos a continuación:

4.1. ARIMA

Los modelos ARIMA, también llamados modelos de Box-Jenkins, son modelos estadísticos que incluyen términos autorregresivos, términos de media móvil y operaciones diferenciales. Los modelos autorregresivos (AR) definen valores futuros de una serie de tiempo estacionaria (estable a través del tiempo) como una combinación lineal de los valores pasados, mientras que los modelos de media móvil (MA) definen valores futuros de una serie de tiempo estacionaria como un promedio ponderado de los valores pasados. Los modelos ARIMA unen ambos métodos junto con la incorporación de una sucesión de operaciones diferenciales (I), las cuales reemplazan cada valor de la serie por la diferencia entre dicho valor y el valor anterior, llevadas a cabo con el fin de convertir la serie de tiempo original en estacionaria.

Un modelo ARIMA(p, d, q), donde p es el orden de la componente autorregresiva, d es la cantidad de diferencias involucradas y q es el orden de la componente de media móvil, puede ser escrito como

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d y_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) e_t, \quad (4.1)$$

donde $(1 - \phi_1 B - \dots - \phi_p B^p)$ es la parte autorregresiva de orden p AR(p), $(1 - B)^d$ representa las d diferencias y $(1 + \theta_1 B + \dots + \theta_q B^q)$ es la parte de media móvil de orden q MA(q).

4.2. Seasonal ARIMA

Los modelos Seasonal ARIMA se encuentran formados por la inclusión de términos estacionales adicionales a los modelos ARIMA mencionados anteriormente, con el propósito de abordar explícitamente la existencia de patrones periódicos en las series. Son escritos como $ARIMA(p, d, q)(P, D, Q)_m$, donde (p, d, q) son los parámetros de la parte no estacional del modelo, (P, D, Q) son los parámetros de la parte estacional del modelo y m es la cantidad de puntos de la serie en cada estación.

La parte estacional del modelo, consiste en términos similares a los de la parte no estacional, pero que utilizan un operador de retardo del período estacional (B^m en vez de B), es decir, para un valor y_t , su valor anterior es considerado como y_{t-m} y no y_{t-1} . Así, la fórmula general de los modelos Seasonal ARIMA, es similar a la fórmula de los modelos ARIMA mostrada en la ecuación 4.1, pero en donde las componentes no estacionales se encuentran multiplicadas por las componentes estacionales de la siguiente manera:

- Parte autorregresiva no estacional es multiplicada por parte autorregresiva estacional $(1 - \Phi_1 B^m - \dots - \Phi_p B^{mp})$.
- Parte diferencial no estacional es multiplicada por parte diferencial estacional $(1 - B^m)^d$.
- Parte de media móvil no estacional es multiplicada por parte de media móvil estacional $(1 + \theta_1 B^m + \dots + \theta_q B^{mq})$.

4.3. Redes Neuronales

Los modelos de redes neuronales artificiales, basados en las redes neuronales naturales del sistema nervioso humano, consisten en la interconexión de grandes cantidades de nodos, los cuales definen individualmente pesos y operaciones para transmitir datos a través de sus conexiones y en consecuencia, a través de la red completa.

El almacenamiento y la actualización de los datos percibidos por los nodos, permite el entrenamiento y el aprendizaje de la red a partir de los datos de entrada, con el fin de poder realizar predicciones lógicas acerca de información futura.

Diferentes configuraciones respecto a las conexiones entre nodos, generan diferentes arquitecturas de redes neuronales, algunas de las cuales son frecuentemente utilizadas en el análisis de series temporales.

4.3.1. LSTM

Las redes LSTM (Long Short-Term Memory) corresponden a un tipo de red neuronal recurrente, es decir, una red cuyas conexiones contienen ciclos, pudiendo retener información con el fin de mantenerla a través de las etapas de la red. La principal característica de una

capa de red LSTM es que puede retener información a largo plazo (*long term*) y puede aprender cuándo tomar o no en cuenta esta información.

La implementación de un bloque de red LSTM consiste de tres compuertas: compuerta de entrada (*input gate*), compuerta de salida (*output gate*) y compuerta de olvido (*forget gate*), las cuales se relacionan mediante las siguientes ecuaciones:

$$c_t = i_t \circ \tanh(W_c x_t + V_c y_{t-1} + b_c) + f_t \circ c_{t-1} \quad (4.2)$$

$$y_t = o_t \circ \tanh(c_t), \quad (4.3)$$

donde i_t , f_t , o_t son las funciones de activación de cada compuerta respectivamente:

$$g_t = \text{sigmoid}(W_g x_t + V_g y_{t-1} + b_g), \quad (4.4)$$

donde g es la compuerta correspondiente, W y V son matrices de pesos, b es un vector de sesgo, x_t e y_t son los vectores de entrada y salida de cada etapa t , y \circ corresponde al producto de Hadamard entre dos matrices.

4.3.2. CNN-LSTM

Las redes CNN (Convolutional Neural Network) difieren de una red neuronal normal en la utilización de máscaras (o *kernels*) para aplicar una operación convolucional sobre los datos de entrada, con el fin de transformarlos y obtener información específica en que enfocarse.

La ecuación de una capa convolucional para datos de entrada con dos dimensiones se describe a continuación:

$$X_{ij} * K_{ij} = \sum_{a=0}^{m-1} \sum_{b=0}^{n-1} w_{ab} X_{(i+a)(j+b)} \cdot K_{ab} + b \quad (4.5)$$

donde X es la entrada y K es la máscara de tamaño $m \times n$. w y b corresponden al peso y sesgo respectivamente.

Las redes de tipo CNN-LSTM resultan tras la combinación de las dos arquitecturas de redes neuronales mencionadas, y consisten en una capa convolucional antes de una capa LSTM.

4.3.3. Naïve Persistent Model

Un modelo persistente se define como un modelo de predicción en donde los últimos datos observados corresponden a los próximos valores de predicción, y puede ser escrito como:

$$y'_t = y_{t-m}, \quad (4.6)$$

donde los datos presentan un período m .

Con respecto a los experimentos de predicción a realizar, se considera el modelo *Naïve* semanal, que repite la última semana de datos como predicción para el próximo período. Este modelo ingenuo es utilizado para evaluar el desempeño de los otros modelos de predicción descritos y ha sido escogido por el hecho de que la calidad de los servicios de telefonía celular es afectada por factores externos con fuertes componentes periódicas, como el clima [41, 56] o el movimiento de masas [23, 57] según las actividades humanas, por lo que repetir valores de días anteriores como predicción resulta un modelo aceptable para colocar como línea de base.

4.4. Evaluación de Modelos

Para comparar los resultados obtenidos por los diferentes modelos de predicción propuestos, se calcula en cada caso un porcentaje de precisión de la cantidad de puntos predichos correctamente con respecto al total de puntos predichos. Una predicción es considerada correcta si la diferencia entre el valor real y el valor predicho es inferior a 3 dBm. Este intervalo de confianza propuesto, fue escogido tras realizar el siguiente análisis.

Se tomaron 30 días de datos de intensidad de señal provenientes de la comuna de Santiago Centro y se calculó la mediana de la intensidad de señal para cada intervalo de dos horas, según lo descrito en la Sección 3.3. Además, para cada período de dos horas, se calcularon los intervalos de confianza que agruparan al 90% de los datos en ese punto a partir de la mediana. Este análisis se realizó de tres formas distintas: considerando todos los registros de intensidad de señal, considerando solo los registros de redes 3G UMTS y considerando solo los registros de redes 4G LTE (las dos tecnologías con mayor número de registros en el set de datos), obteniendo los siguientes resultados:

- Para intensidad de señal general (considerando todos los registros) se determinó que el intervalo de confianza se ubica, en promedio, en 2.50 dBm bajo la mediana y en 2.86 dBm sobre la mediana de cada intervalo de dos horas.
- Para intensidad de señal 3G UMTS se determinó que el intervalo de confianza se ubica, en promedio, en 3.60 dBm bajo la mediana y en 3.90 dBm sobre la mediana de cada intervalo de dos horas.
- Para intensidad de señal 4G LTE se determinó que el intervalo de confianza se ubica, en promedio, en 2.23 dBm bajo la mediana y en 2.97 dBm sobre la mediana de cada intervalo de dos horas.

Por lo tanto, se decidió estandarizar un intervalo de confianza de 3 dBm bajo y sobre la mediana, considerando los resultados obtenidos para los tres tipos de predicciones a realizar: intensidad de señal general, 3G UMTS y 4G LTE, descritos en profundidad en el Capítulo 5. Además, cabe destacar que la banda de confianza de 3 dBm escogida resulta inferior a los rangos utilizados comúnmente para caracterizar la intensidad de señal recibida, mostrados en la Tabla 4.1, en donde los niveles de intensidad de señal abarcan rangos de al menos 15 dBm.

Calidad	Rango en dBm
Excelente	$[-70, +\infty[$
Buena	$[-85, -70[$
Media	$[-100, -85[$
Mala	$] - \infty, -100[$

Tabla 4.1: Niveles generales de intensidad de señal recibida.

En adición, es necesario establecer un conjunto de métricas a considerar durante los procesos de configuración de modelos y evaluación de resultados obtenidos, con el fin de establecer un análisis con mayor completitud. Las métricas escogidas se dividen en dos subconjuntos, errores de predicción y distancias entre series de tiempo.

4.4.1. Errores de Predicción

En primer lugar, es necesario evaluar los resultados con respecto a la diferencia entre los valores reales y los predichos, para lo cual se consideran los siguientes indicadores:

1. **Raíz del Error Cuadrático Medio (RMSE):** Raíz cuadrada del promedio de la diferencia vertical al cuadrado existente entre cada valor real y su valor predicho.

$$RMSE = \sum_{i=1}^n \sqrt{\frac{(r_i - p_i)^2}{n}}$$

2. **Error Absoluto Medio (MAE):** Promedio de la distancia vertical existente entre cada valor real y su valor predicho.

$$MAE = \sum_{i=1}^n \frac{|r_i - p_i|}{n}$$

3. **Error Porcentual Absoluto Medio (MAPE):** Promedio de la distancia vertical existente entre cada valor real y su valor predicho, en términos porcentuales.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{r_i - p_i}{r_i} \right|$$

4.4.2. Distancia Entre Series de Tiempo

En adición de los errores de predicción es necesario tomar en cuenta otros factores al momento de evaluar los modelos de predicción, principalmente debido a que los errores de predicción mostrados anteriormente, no toman en cuenta características importantes al momento de comparar series temporales, como la similitud entre la forma de las curvas de datos reales y predichos o la distribución de los datos. Las medidas de distancia entre series de tiempo consideradas, utilizadas principalmente en procesos de *clustering* de series de tiempo, son presentadas a continuación.

1. **Distancia de Fréchet:** Es una medida de similitud de curvas que toma en consideración la ubicación y el orden de los puntos a lo largo de la curva. Formalmente es definida como sigue.
 - Sea S un espacio métrico con función de distancia d .
 - Una curva A en S es un mapeo continuo desde el intervalo unitario a S , es decir, $A : [0, 1] \rightarrow S$.
 - Una reparametrización α de $[0, 1]$ es una función sobreyectiva, continua y creciente $\alpha : [0, 1] \rightarrow [0, 1]$.
 - Sean A, B dos curvas dadas en S . La distancia de Fréchet entre A y B es definida como el ínfimo sobre todas las reparametrizaciones α y β de $[0, 1]$ del máximo sobre todo $t \in [0, 1]$ de la distancia en S entre $A(\alpha(t))$ y $B(\beta(t))$:

$$F(A, B) = \inf_{\alpha, \beta} \max_{t \in [0, 1]} \left\{ d(A(\alpha(t)), B(\beta(t))) \right\}$$

Intuitivamente, si las curvas representan las trayectorias de dos objetos unidos por una cuerda, la distancia de Fréchet es la longitud mínima de la cuerda necesaria para que ambos objetos pueden avanzar unidos.

2. **Distancia Basada en el Test de Razón de Verosimilitud Generalizada (SGLK):** Compara las dos curvas por medio del test de razón de verosimilitud generalizada, con el fin de chequear si las densidades espectrales de las series de tiempo pertenecen o no a la misma familia paramétrica [20]. Por lo tanto, mientras más similar sea la distribución de los datos a lo largo de ambas curvas, menor será su distancia SGLK.
3. **Distancia de Edición para Secuencias Reales (EDR):** Compara las dos curvas en términos de cuántas operaciones de edición (eliminar, insertar o reemplazar) son necesarias para transformar una curva en la otra. La distancia entre puntos se reduce a 0 o 1, en donde si la distancia entre dos puntos r_i y p_j es menor a un ε dado, entonces los puntos son considerados iguales.

Capítulo 5

Predicción de Intensidad de Señal

5.1. Experimentos Realizados

En este capítulo se presentan los resultados experimentales obtenidos al predecir valores de intensidad de señal recibida en un área determinada. Los datos utilizados pertenecen a dispositivos en un área reducida de la ciudad de Santiago, dentro de la comuna de Santiago Centro.

Todos los modelos de predicción mencionados en el Capítulo 4, toman como entrada la información respecto a ocho semanas previas de intensidad de señal agregada para el área central de la comuna de Santiago. La decisión de considerar ocho semanas, se tomó a partir del hecho de que, dado los datos utilizados, este es el mayor intervalo de tiempo en el cual se mantiene un número constante de usuarios activos en la zona de medición.

Las predicciones de intensidad de señal se realizan de tres formas distintas: considerando todos los registros de intensidad de señal sin discriminar según tecnología de red (en adelante, intensidad de señal general), considerando solamente los registros de redes 3G UMTS y considerando solamente los registros de redes 4G LTE, debido a que 3G UMTS y 4G LTE son las dos tecnologías de red más utilizadas en Santiago según los datos utilizados. Para cada caso, se reportan acerca de tres procesos de predicción independientes, llevados a cabo para predecir tres semanas consecutivas de datos.

Para cada uno de los modelos de predicción, se ha realizado una búsqueda exhaustiva con el fin de encontrar una configuración ideal de los parámetros, logrando obtener los mejores resultados posibles para las predicciones, según los indicadores descritos en la Sección 4.4. En cada caso, los resultados son comparados con los obtenidos por el modelo *Naïve* propuesto.

5.2. Intensidad de Señal General

Las Figuras 5.1, 5.2 y 5.3 muestran los resultados obtenidos al predecir tres semanas de datos de intensidad de señal general utilizando los cuatro modelos considerados. En la Tabla 5.1 se presentan las medidas de precisión calculadas en promedio para las tres semanas. Las medidas de precisión obtenidas para cada semana pueden ser consultadas en detalle en el Anexo A.

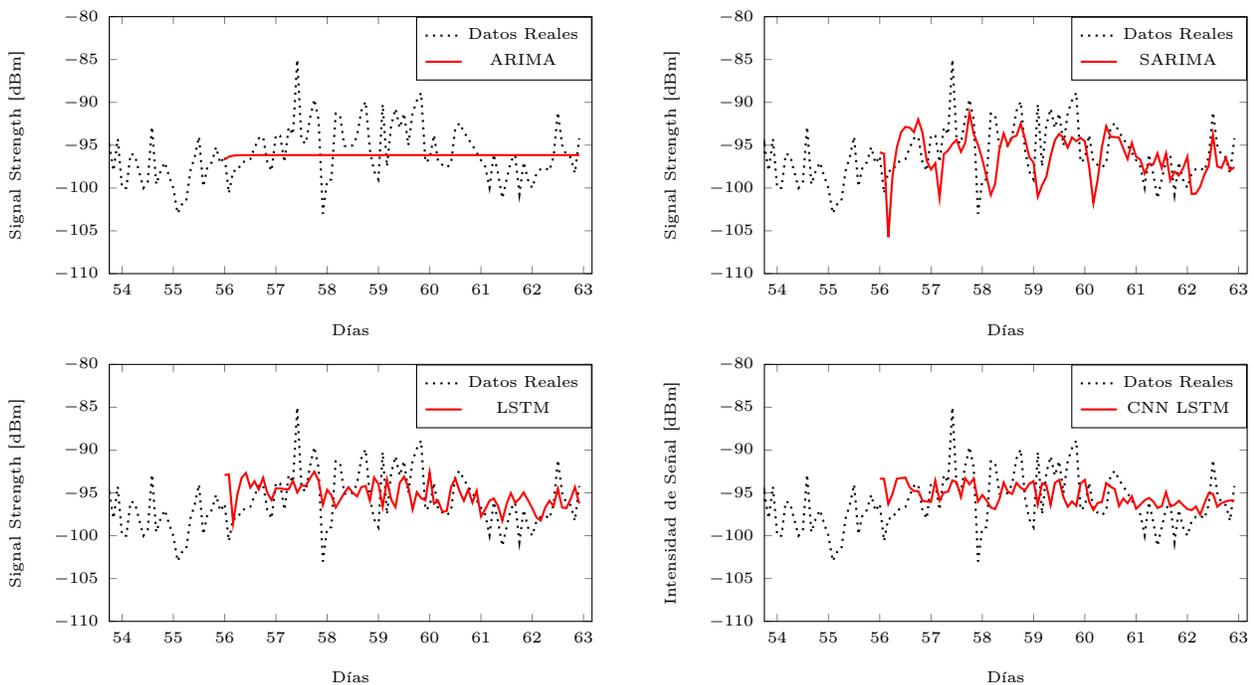


Figura 5.1: Predicción semana 1 de intensidad de señal general.

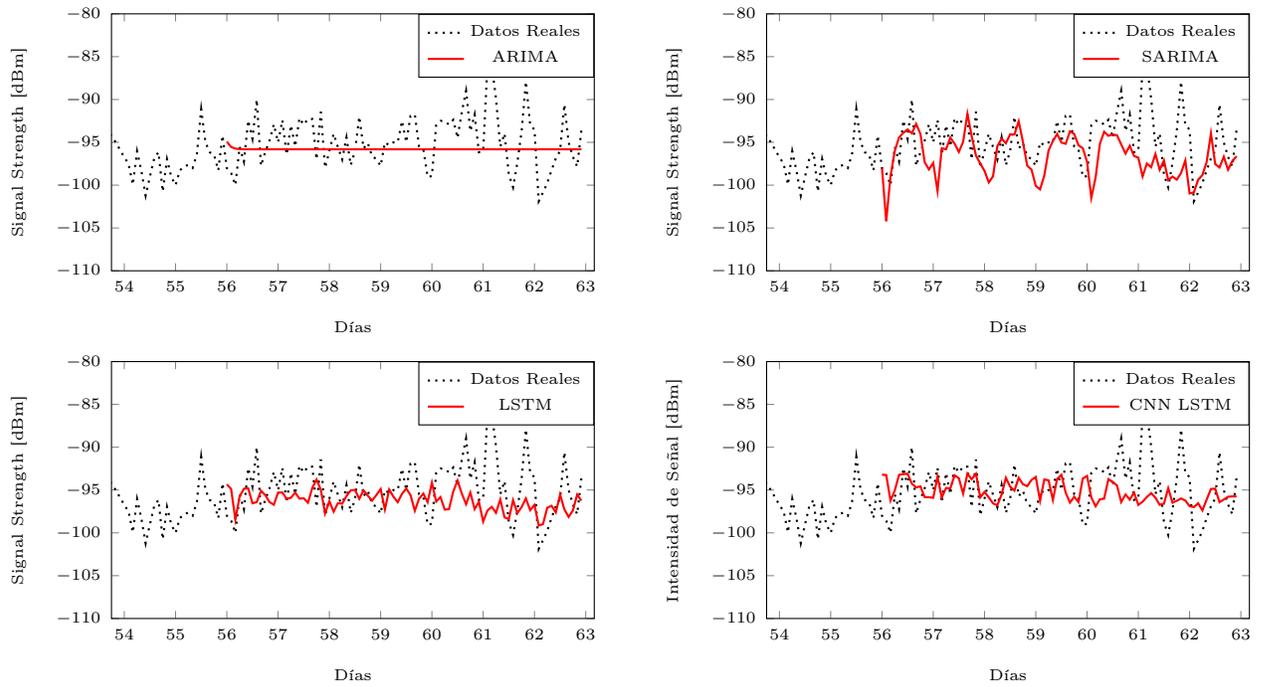


Figura 5.2: Predicción semana 2 de intensidad de señal general.

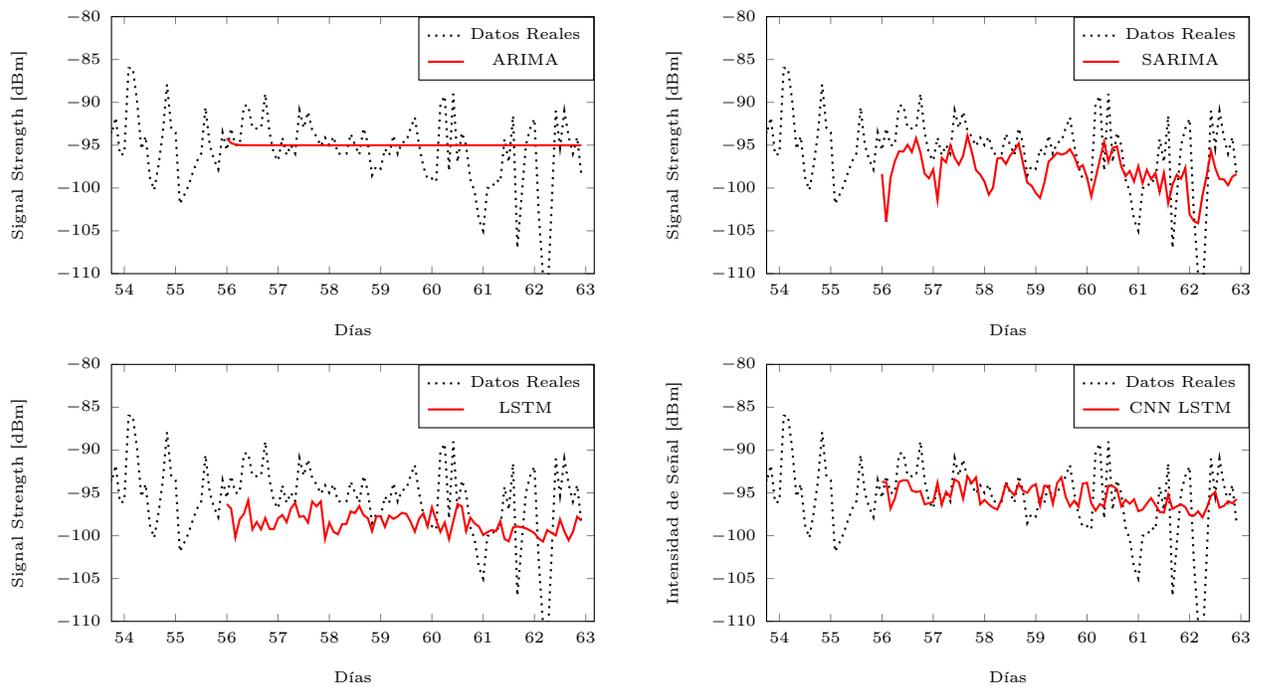


Figura 5.3: Predicción semana 3 de intensidad de señal general.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
Naïve	4.63	3.39	3.55	9.90	7.59	28.33	55.33
ARIMA	3.64	2.68	2.81	13.10	565.27	29.33	65.00
SARIMA	3.98	3.03	3.22	7.50	20.00	27.00	62.00
LSTM	3.88	3.04	3.23	9.37	37.15	30.00	61.00
CNN-LSTM	3.88	2.90	3.05	11.78	38.27	28.33	62.00

Tabla 5.1: Medidas de precisión para intensidad de señal general.

5.3. Intensidad de Señal 3G UMTS

Las Figuras 5.4, 5.5 y 5.6 muestran los resultados obtenidos al predecir tres semanas de datos de intensidad de señal en redes 3G UMTS utilizando los cuatro modelos considerados. En la Tabla 5.2 se presentan los medidas de precisión calculadas en promedio para las tres semanas. Las medidas de precisión obtenidas para cada semana pueden ser consultadas en detalle en el Anexo B.

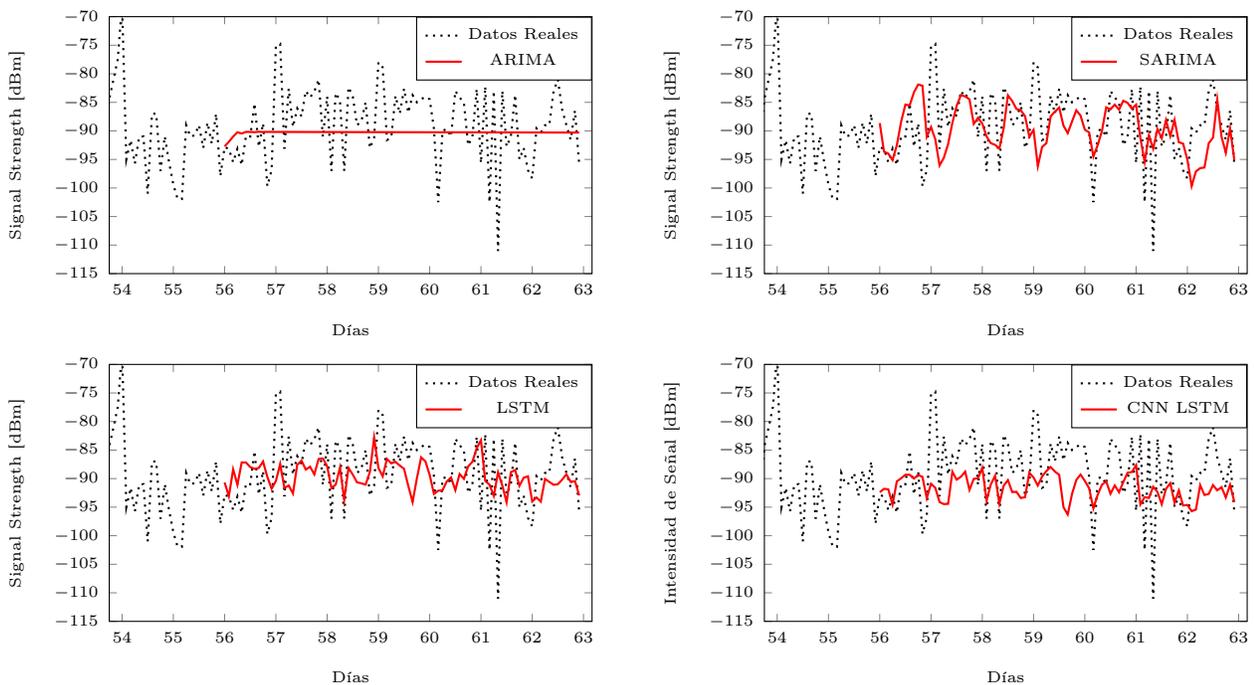


Figura 5.4: Predicción semana 1 de intensidad de señal 3G UMTS.

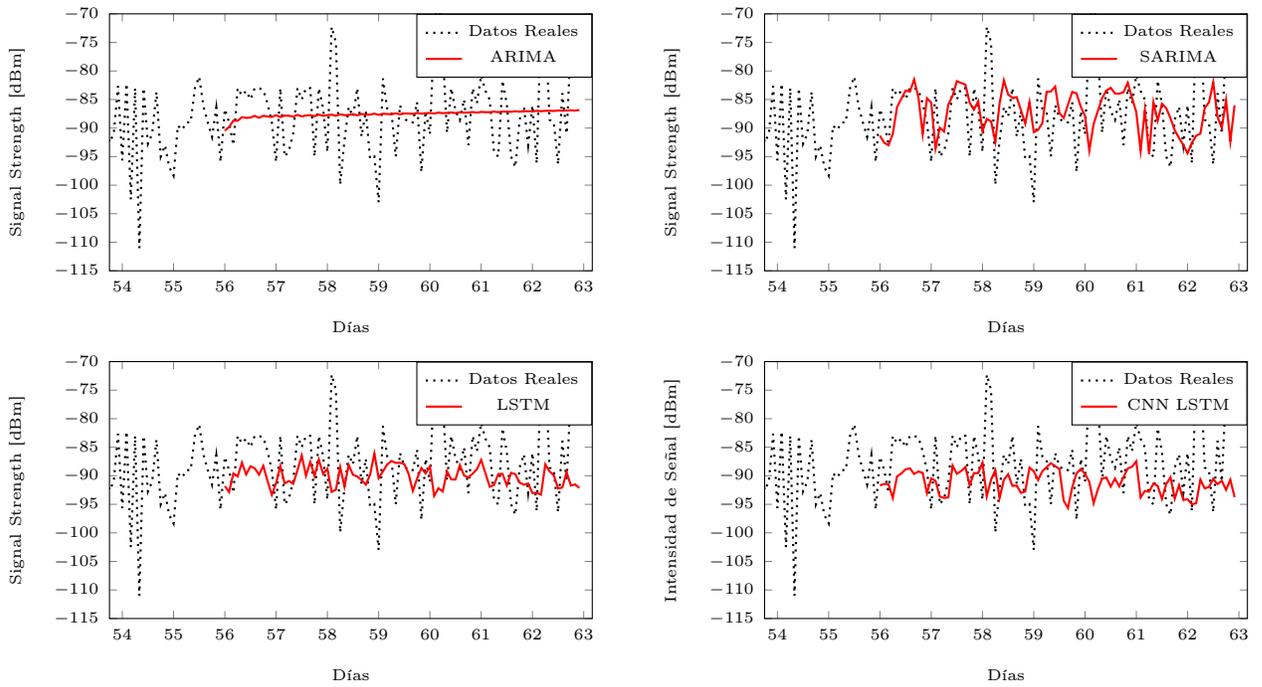


Figura 5.5: Predicción semana 2 de intensidad de señal 3G UMTS.

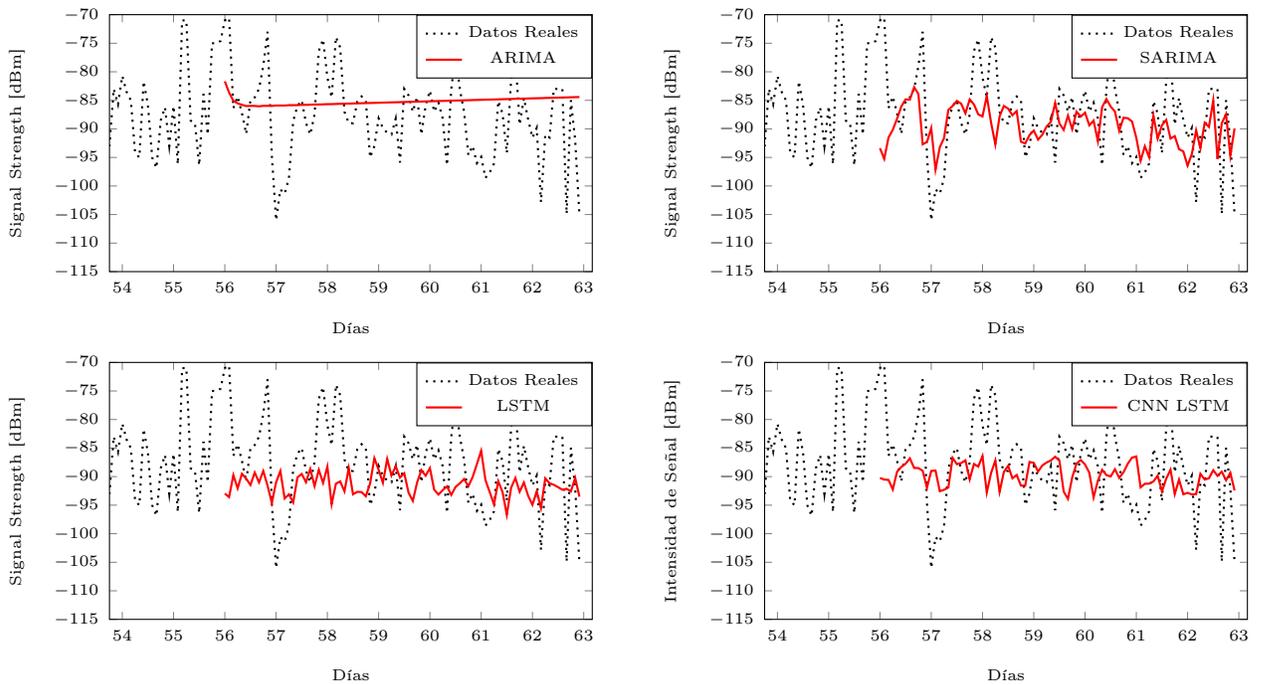


Figura 5.6: Predicción semana 3 de intensidad de señal 3G UMTS.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
Naïve	9.83	7.44	8.61	21.80	5.97	48.67	31.67
ARIMA	7.14	5.52	6.35	20.59	380.96	53.00	37.00
SARIMA	7.43	5.53	6.50	17.72	29.14	42.00	43.00
LSTM	7.67	6.05	7.17	20.01	43.74	52.67	33.00
CNN-LSTM	7.70	6.07	7.21	18.87	46.53	51.00	34.00

Tabla 5.2: Medidas de precisión para intensidad de señal 3G UMTS.

5.4. Intensidad de Señal 4G LTE

Las Figuras 5.7, 5.8 y 5.9 muestran los resultados obtenidos al predecir tres semanas de datos de intensidad de señal en redes 4G LTE utilizando los cuatro modelos considerados. En la Tabla 5.3 se presentan los medidas de precisión calculadas en promedio para las tres semanas. Las medidas de precisión obtenidas para cada semana pueden ser consultadas en detalle en el Anexo C.

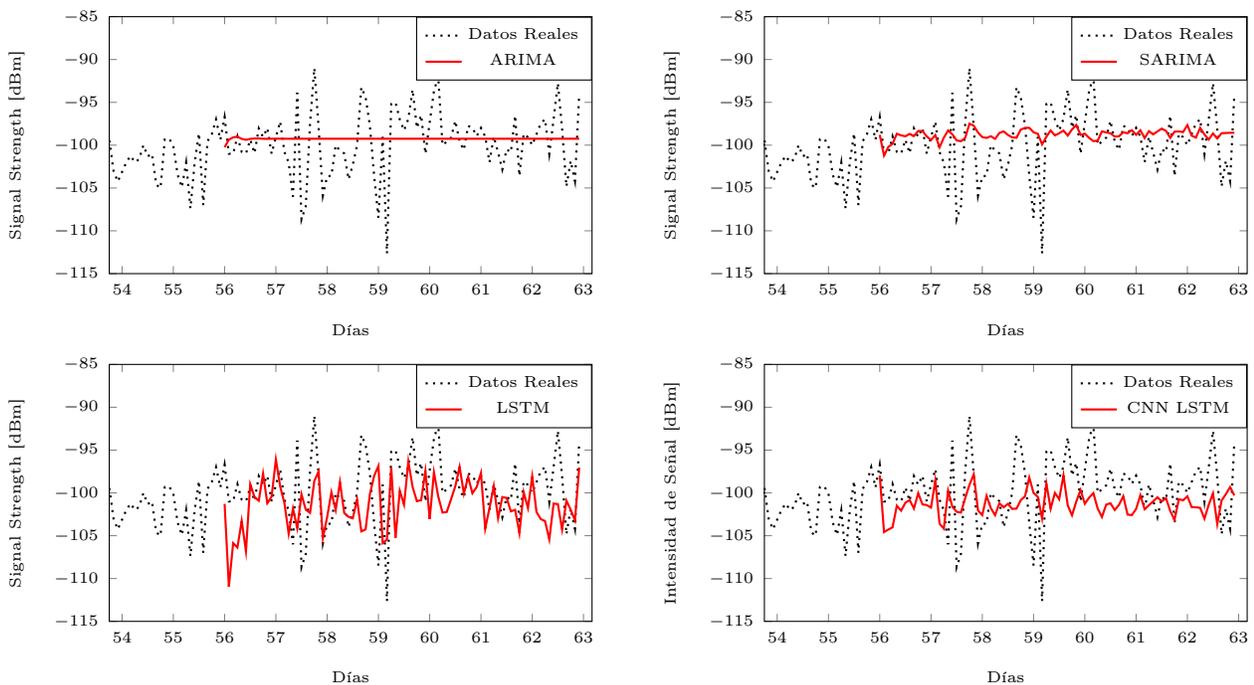


Figura 5.7: Predicción semana 1 de intensidad de señal 4G LTE.

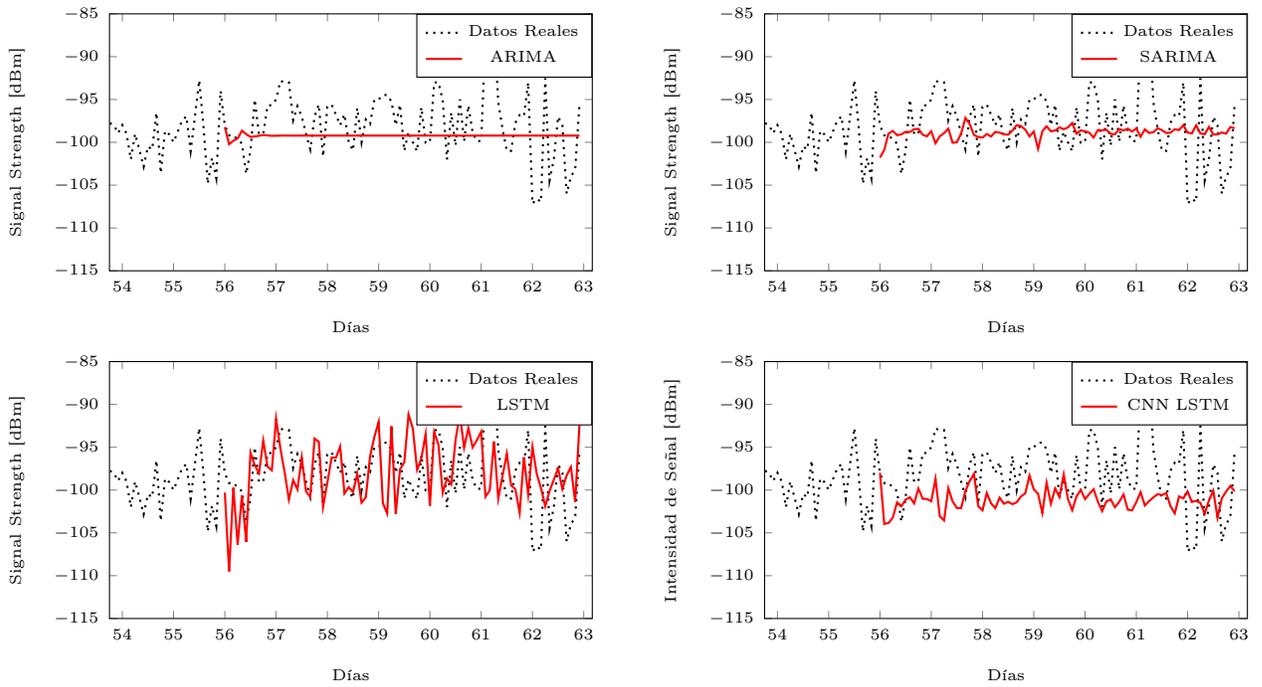


Figura 5.8: Predicción semana 2 de intensidad de señal 4G LTE.

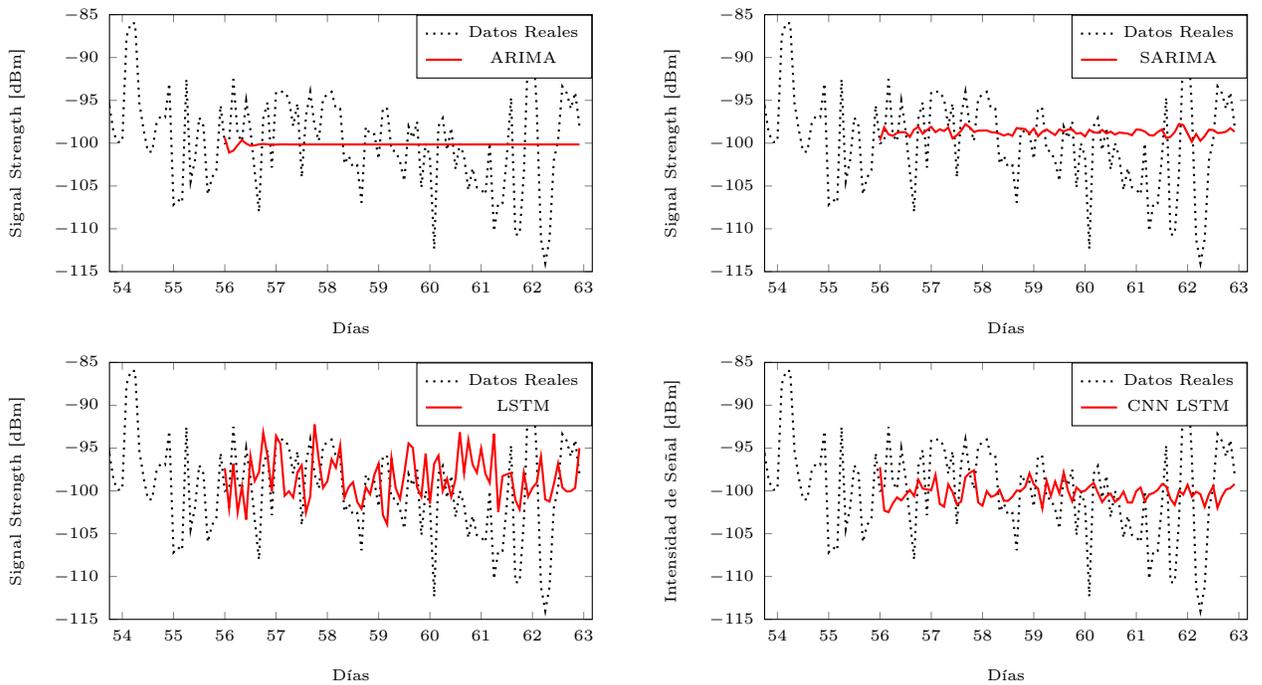


Figura 5.9: Predicción semana 3 de intensidad de señal 4G LTE.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
Naïve	6.19	4.59	4.63	10.32	4.89	38.67	45.33
ARIMA	4.47	3.44	3.48	13.37	356.37	37.33	54.00
SARIMA	4.52	3.45	3.46	12.69	116.88	35.33	57.00
LSTM	5.27	4.16	4.19	8.10	10.16	36.67	48.33
CNN-LSTM	4.92	3.94	4.02	11.05	47.09	41.33	48.00

Tabla 5.3: Medidas de precisión para intensidad de señal 4G LTE.

5.5. Discusión de Resultados

En términos generales, los resultados obtenidos muestran un desempeño positivo al predecir intensidad de señal, en donde los modelos de predicción obtienen, en la mayoría de los casos, mejores valores en las medidas de precisión y mayor porcentaje de precisión en las predicciones en comparación con el modelo *Naïve* de comparación.

Es importante destacar que en todos los casos, el modelo ARIMA no captura ningún patrón en los datos al realizar predicciones, en donde sus valores predichos son, en su mayor parte, una línea con valor constante en una ubicación cercana a los valores centrales de la curva de datos reales. El caso de las predicciones ARIMA, sin embargo, resulta muy importante, ya que a pesar de que claramente no es una buena predicción, obtiene buenos resultados para las medidas de errores de predicción (RMSE, MAE y MAPE), siendo en la mayoría de los casos inferiores aún a los otros modelos de predicción, que al menos gráficamente, presentan un comportamiento más cercano al esperado. Esto se debe al hecho de que la predicción realizada por ARIMA minimiza los errores promedios obtenidos al entregar una predicción constante con valores cercanos a la media de los datos, pudiendo en algunos casos incluso obtener un mayor porcentaje de precisión que los modelos basados en el uso de redes neuronales. Es por esto que es de gran importancia la utilización de otras medidas de precisión en adición a las ya mencionadas, ya que a pesar de presentar errores de predicción bajos, este modelo obtiene en general los valores más altos al considerar las distancias entre series temporales como EDR y de Fréchet. Lo anterior, es aún más notable al analizar la distancia SGLK. Según lo visto en la Sección 4.4, SGLK tiene gran importancia al tratarse de series de tiempo en escala logarítmica, ya que el modelo ARIMA obtiene valores mucho más altos al resto, lo cual es claro indicio de que la curva predicha se aleja bastante de la curva real a predecir.

Para los casos de predicción de intensidad de señal general y predicción en redes 3G UMTS, se observan comportamientos similares al analizar los gráficos, en donde el modelo SARIMA realiza predicciones marcando una clara periodicidad diaria y con una mayor variabilidad en cuanto a los valores predichos, abarcando un amplio intervalo de intensidad de señal y logrando una curva más ajustada a los datos reales. Por otra parte, ambos modelos basados en redes neuronales realizan predicciones con un intervalo más reducido que SARIMA, centrándose en valores más cercanos a la media (en especial en las curvas de intensidad de señal general) pero capturando también los patrones periódicos en las curvas de datos reales. En el caso de intensidad de señal general, los tres modelos son similares en sus valores de medi-

das de precisión y porcentaje de precisión, pero SARIMA obtiene menores valores para las distancias entre series temporales (de Fréchet, SGLK y EDR), confirmando el análisis visual de que la curva proporcionada por SARIMA es más ajustada a los datos reales. En el caso de intensidad de señal 3G UMTS, SARIMA obtiene consistentemente mejores resultados en todos los parámetros medidos en comparación con los modelos de redes neuronales, quienes presentan resultados muy similares entre ellos.

Para el caso de predicción en redes 4G LTE, SARIMA presenta un comportamiento diferente al anteriormente descrito, en donde su predicción entrega solamente valores cercanos a la media de los datos, exhibiendo un comportamiento muy similar al entregado por ARIMA, lo cual también se evidencia al observar el valor obtenido para la distancia SGLK, el cual es notablemente mayor que el obtenido por los modelos LSTM y CNN-LSTM. Con respecto a los modelos basados en redes neuronales, estos entregan mayor variabilidad en sus predicciones, en donde la curva otorgada por el modelo LSTM presenta una mayor similitud en forma a la curva real, lo que se refleja en menores valores de las distancias de Fréchet, SGLK y EDR. No obstante, ambas redes neuronales obtienen valores muy cercanos respecto al porcentaje de precisión calculado.

Cabe mencionar que el modelo *Naïve* obtiene resultados bajos respecto a las medidas de errores de precisión consideradas, lo cual evidencia la existencia de patrones periódicos en los datos, mencionados en la Sección 3.3. Sin embargo, el hecho de que los modelos de predicción utilizados obtengan un mejor desempeño, es muestra de la existencia de otros factores que influyen en la intensidad de señal recibida por los usuarios, y que la predicción de la intensidad de señal no es una tarea trivial que pueda ser abordada como una curva puramente periódica.

A modo de resumen es posible decir que los modelos de predicción logran predecir la intensidad de señal con un porcentaje de precisión promedio de 62.5% para intensidad de señal general, de 36.75% para redes 3G UMTS y de 51.83% para redes 4G LTE, por sobre los resultados obtenidos por el modelo *Naïve*, el cual obtiene un porcentaje de precisión de 53.33%, 31.67% y 45.33% en cada caso respectivamente.

Capítulo 6

Inclusión de Factores Externos

En adición a los datos provenientes de la aplicación *Adkintun Mobile*, fueron considerados datos externos, no tomados por los dispositivos móviles ni atinentes a calidad de Internet, con el motivo de analizar su efecto en la calidad de servicio recibida en un área geográfica específica.

6.1. Condiciones Climáticas

Siguiendo la literatura relacionada, se decidió utilizar información respecto a condiciones climáticas, ya que se ha probado su impacto en el uso y calidad de Internet, lo cual ha sido descrito en la Sección 2.4.2. Para esto, se consideraron los datos pertenecientes a la Dirección General de Aguas del Ministerio de Obras Públicas de Chile, quienes cuentan con una gran cantidad de estaciones meteorológicas a lo largo del país. Estas estaciones reportan cada una hora los valores medidos de temperatura ($^{\circ}C$), humedad (%) y precipitación (mm). Los datos mencionados son de libre acceso y se encuentran disponibles a través del sitio web de la Dirección General de Aguas [21]. Para los experimentos realizados, se consideró una de las estaciones meteorológicas ubicada en la comuna de Santiago, en el centro del área de mediciones de intensidad de señal mencionada en la Sección 3.3. En la Figura 6.1, se puede apreciar la ubicación de la estación meteorológica utilizada, rodeada por las estaciones base o antenas de telefonía móvil presentes en el área.

Para relacionar los datos de intensidad de señal agregados cada dos horas con la información de condiciones climáticas, resultó necesario representar los datos climáticos en intervalos de dos horas y no de una hora, como fueron originalmente obtenidos. Para esto, se tomó de a pares de registros climáticos consecutivos y se calculó el promedio entre ellos, obteniendo información agrupada cada dos horas acerca de los factores climáticos considerados.

Es importante mencionar que para todos los análisis realizados, se consideró que las condiciones climáticas medidas por la estación meteorológica son percibidas de igual manera en todo punto dentro del área de medición. Con relación a lo anterior, algunos estudios han analizado la precisión espacial de la información entregada por las estaciones meteorológicas

para ciertas ubicaciones geográficas [18, 34], estableciendo que a una distancia menor a 25 *km* de ellas, los valores reportados siguen siendo altamente representativos, obteniendo, por ejemplo, variaciones de temperatura del aire inferiores a 1°C. A pesar de que la precisión espacial depende en gran parte del área geográfica medida, el hecho de que para los datos utilizados en este trabajo, cada punto dentro del área de medición se encuentre a una distancia menor a 10 *km* de la estación meteorológica, es un indicio de que la variabilidad de las condiciones climáticas dentro de dicha área, debiese ser pequeña.

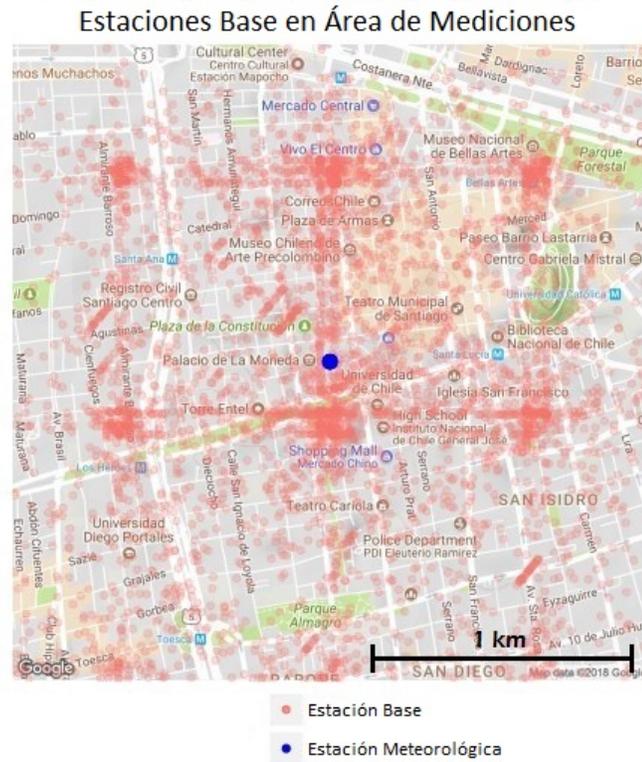


Figura 6.1: Ubicación geográfica de estaciones base y estación meteorológica en la comuna de Santiago.

Con el fin de validar la inclusión de los factores climáticos como regresores externos en los modelos de predicción, se realizaron análisis para determinar la existencia de una correlación entre los datos de intensidad de señal y los datos climáticos, sugerida por los trabajos relacionados estudiados. En la Tabla 6.1 se presentan los valores obtenidos al calcular el coeficiente de correlación de Pearson y el valor p , midiendo la relación lineal entre la intensidad de señal general recibida por los usuarios dentro del área de medición, y las condiciones climáticas. En cada caso, se obtiene un valor p menor que el nivel de significación convencionalmente establecido en 0.05, lo que permite rechazar la hipótesis nula (inexistencia de correlación). Los resultados indican la existencia de una relación lineal estadísticamente significativa entre la intensidad de señal y las condiciones climáticas, en donde la intensidad de señal aumenta junto con la temperatura, y cuando la humedad y las precipitaciones disminuyen. Dichas observaciones se encuentran en estrecha relación con los resultados obtenidos en [41] al realizar el mismo análisis.

Cabe destacar que según los resultados, la temperatura y la humedad serían los factores con mayor incidencia sobre la intensidad de señal recibida, seguidos por las precipitaciones,

las cuales tendrían una influencia menor, lo que puede deberse en cierta medida a que durante la mayor parte del tiempo, los valores de precipitaciones medidas son nulos.

Variable	Correlación de Pearson	Valor p
Temperatura	0.44	2.982e-09
Humedad	-0.42	1.534e-08
Precipitaciones	-0.17	0.02512

Tabla 6.1: Factores climáticos y su correlación de Pearson con la intensidad de señal recibida.

A modo de ejemplo, en la Figura 6.2 se muestra la relación observada entre intensidad de señal y humedad del aire, en donde los valores de porcentaje de humedad son agrupados en 9 intervalos y, para cada intervalo, se presenta el valor de la mediana de la intensidad de señal. Adicionalmente, los valores de intensidad de señal son agrupados para las tres tecnologías de red móvil más utilizadas por los usuarios de acuerdo al set de datos de *Adkintun Mobile* (UMTS, LTE y HSPAP), y sus medianas de intensidad de señal también son presentadas para cada intervalo de porcentaje de humedad. En el gráfico se observa una tendencia para todas las tecnologías de disminuir los valores de intensidad de señal cuando la humedad aumenta.

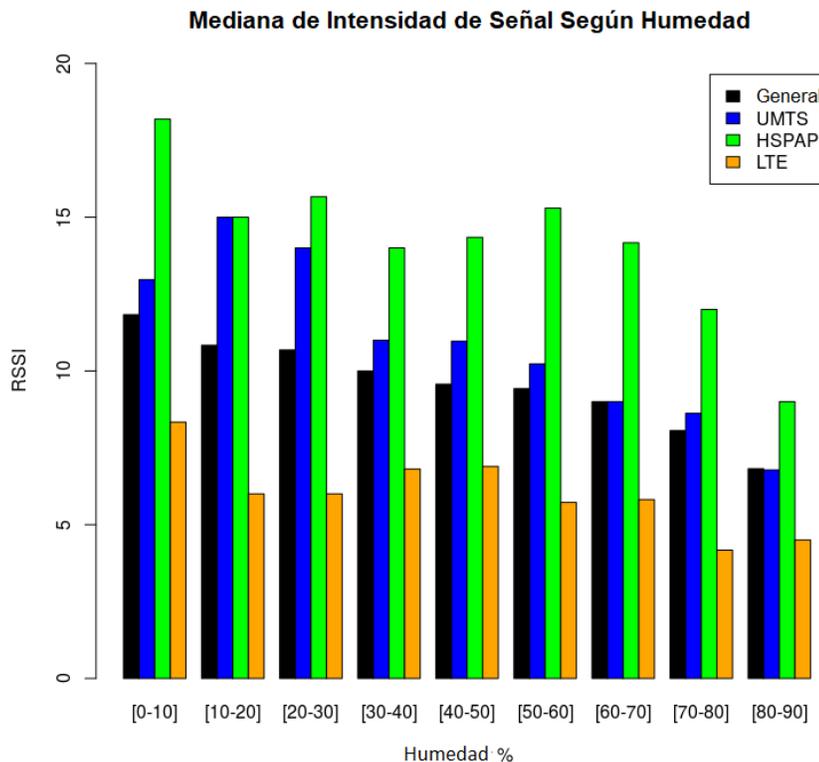


Figura 6.2: Relación entre intensidad de señal recibida y porcentaje de humedad en la comuna de Santiago.

Finalmente, en adición a los reportes de temperatura, humedad y precipitaciones obtenidos de la estación meteorológica, fueron calculados valores de cambio de temperatura y humedad (Δ), dado que esta información ya ha sido incluida en estudios anteriores acerca

del indicador de intensidad de señal recibida, obteniendo resultados satisfactorios [41]. Estos valores fueron calculados a partir de los datos originales, registrados en intervalos de una hora, y posteriormente se creó una serie de tiempo con intervalos de dos horas siguiendo el mismo proceso mencionado recientemente. Al igual que los datos de *Adkintun Mobile*, todos los valores de datos climáticos fueron considerados a partir de octubre de 2016.

6.2. Agrupaciones Masivas

Como se mencionó también en la Capítulo 2, investigaciones recientes han tomado los primeros pasos para determinar y cuantificar el impacto sobre la calidad y el desempeño de Internet móvil producto de la agrupación de personas durante eventos masivos. Para esto, se realizó un trabajo de recopilación de información acerca de múltiples eventos masivos realizados dentro del área de mediciones (comuna de Santiago Centro) entre los meses de octubre y diciembre del año 2016, considerando eventos deportivos, eventos musicales, manifestaciones públicas, entre otros. Con esta información, se creó una serie de tiempo con intervalos de dos horas, en donde cada dato corresponde a 1 o 0, funcionando como una variable indicatriz que muestra si en dicho período de tiempo ocurrió o no una agrupación masiva dentro del lugar de mediciones.

Al igual que con los datos climáticos, con el fin de validar la inclusión de la información acerca de agrupaciones masivas como factor externo en los modelos de predicción, se realizó un análisis para determinar la existencia de una correlación entre los datos de intensidad de señal y los datos de eventos masivos, sugerida por los trabajos relacionados estudiados. En la Tabla 6.2 se presentan los valores obtenidos al calcular el coeficiente de correlación de Pearson y el valor p , midiendo la relación lineal entre las series de tiempo para la intensidad de señal general recibida por los usuarios dentro del área de medición, y para las agrupaciones masivas. En este caso, se obtuvo una correlación baja en comparación a la obtenida para los factores climáticos, y se obtuvo un valor p mayor que el nivel de significación convencionalmente establecido en 0.05. Los resultados indican la inexistencia de una relación lineal entre la intensidad de señal y las agrupaciones masivas para el set de datos estudiado.

Variable	Correlación de Pearson	Valor p
Agrupaciones Masivas	0.09	0.2412

Tabla 6.2: Agrupaciones masivas y su correlación de Pearson con la intensidad de señal recibida.

Es importante mencionar que a pesar de que dentro de la literatura estudiada se ha mostrado el impacto de las agrupaciones masivas sobre la calidad de servicio de Internet móvil, no se ha podido concluir respecto a una correlación existente entre agrupaciones masivas e intensidad de señal, dejando como trabajo futuro la realización de análisis más profundos al respecto [23]. En el caso de los datos utilizados en este trabajo, tampoco se encontró información concluyente acerca de dicha correlación, lo cual también puede explicarse dado los siguientes puntos:

- Al igual que en el caso de las precipitaciones, que presentaron la menor correlación con la intensidad de señal dentro de los factores climáticos estudiados, las agrupaciones masivas corresponden a eventos muy esporádicos y de corta duración a lo largo del período considerado para las mediciones. Por lo tanto, como en la mayoría de los puntos de la serie de tiempo no hay eventos masivos, no se obtiene una correlación importante con la intensidad de señal recibida.
- Los estudios que analizan el impacto de la agrupación de masas en la calidad de la red móvil son realizados en localidades muy reducidas como por ejemplo, el área donde se realiza un festival de música. Sin embargo, el área de mediciones considerada en este trabajo corresponde, según lo descrito en la Sección 3.3, a una porción de la comuna de Santiago 15 km^2 aproximadamente, por lo cual, es razonable que un evento masivo llevado a cabo en un punto de esta área, no genere un impacto en la calidad de servicio de Internet móvil recibida en el área total.

Dado los resultados obtenidos, la información acerca de agrupaciones masivas de personas no fue incluida como factor externo para los modelos de predicción.

6.3. Experimentos Realizados

A los modelos de predicción para intensidad presentados en el Capítulo 4, se les entregó como parámetros de entrada adicionales las series de tiempo correspondientes a las primeras ocho semanas de datos correspondientes a los cinco factores climáticos considerados: temperatura, Δ temperatura, humedad, Δ humedad y precipitaciones.

Al igual que anteriormente, se realizaron predicciones semanales de intensidad de señal, diferenciando entre intensidad de señal general, intensidad de señal en redes 3G UMTS e intensidad de señal en redes 4G LTE. Cada uno de los modelos llevó a cabo predicciones para tres semanas de datos consecutivas según lo descrito en la Sección 5.1, y los resultados son presentados a continuación.

Intensidad de Señal General

Las Figuras 6.3, 6.4 y 6.5 muestran los resultados obtenidos al predecir las tres semanas de datos de intensidad de señal general utilizando los cuatro modelos considerados e incluyendo información externa de datos climáticos. En la Tabla 6.3 se presentan las medidas de precisión calculadas en promedio para las tres semanas. Las medidas de precisión obtenidas para cada semana pueden ser consultadas en detalle en el Anexo D.

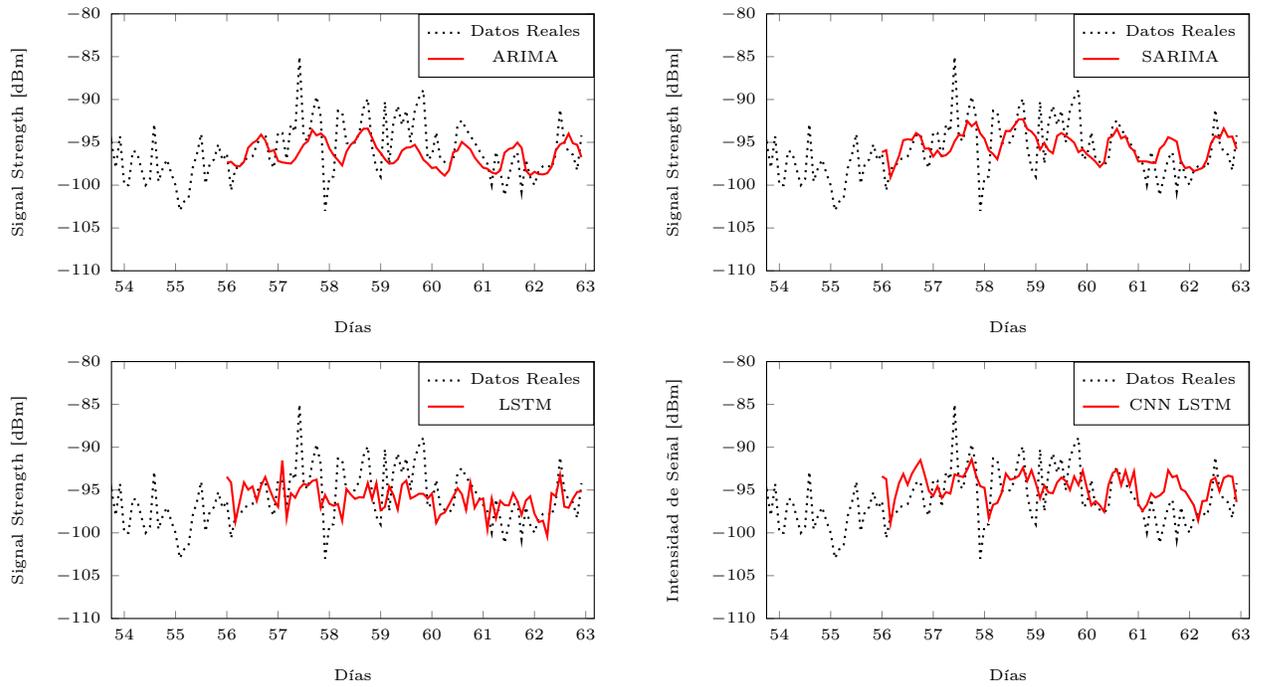


Figura 6.3: Predicción semana 1 de intensidad de señal general considerando condiciones climáticas.

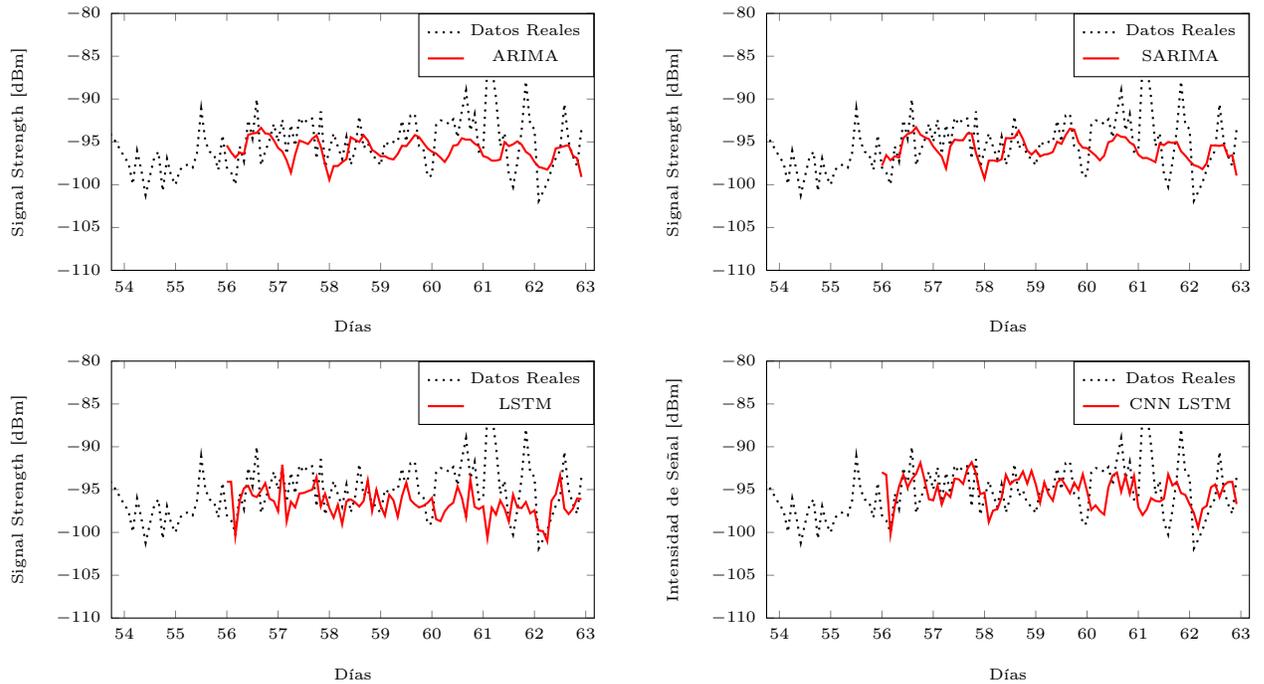


Figura 6.4: Predicción semana 2 de intensidad de señal general considerando condiciones climáticas.

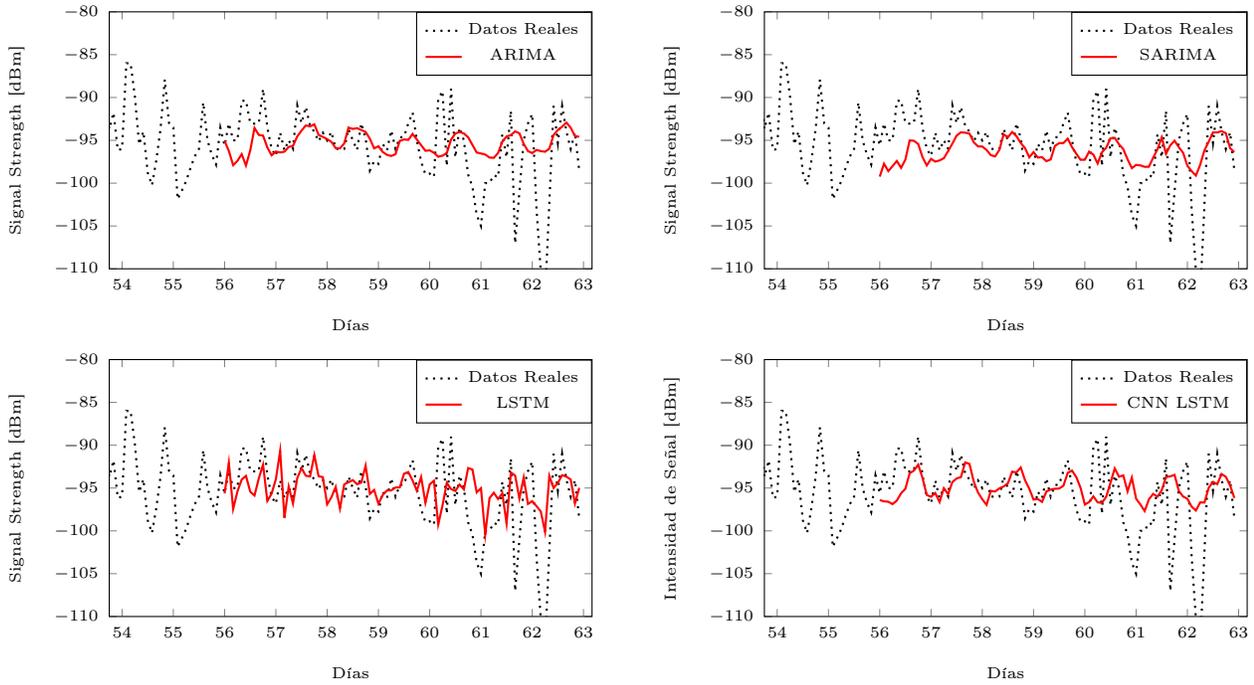


Figura 6.5: Predicción semana 3 de intensidad de señal general considerando condiciones climáticas.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
ARIMA	3.59	2.60	2.73	11.09	75.39	23.33	68.33
SARIMA	3.43	2.53	2.66	9.79	63.46	22.67	68.00
LSTM	3.75	2.81	2.96	9.03	21.29	24.67	66.33
CNN-LSTM	3.65	2.75	2.87	9.72	44.12	23.00	65.33

Tabla 6.3: Medidas de precisión para intensidad de señal general considerando condiciones climáticas.

Intensidad de Señal 3G UMTS

Las Figuras 6.6, 6.7 y 6.8 muestran los resultados obtenidos al predecir las tres semanas de datos de intensidad de señal en redes 3G UMTS utilizando los cuatro modelos considerados e incluyendo información externa de datos climáticos. En la Tabla 6.4 se presentan los medidas de precisión calculadas en promedio para las tres semanas. Las medidas de precisión obtenidas para cada semana pueden ser consultadas en detalle en el Anexo E.

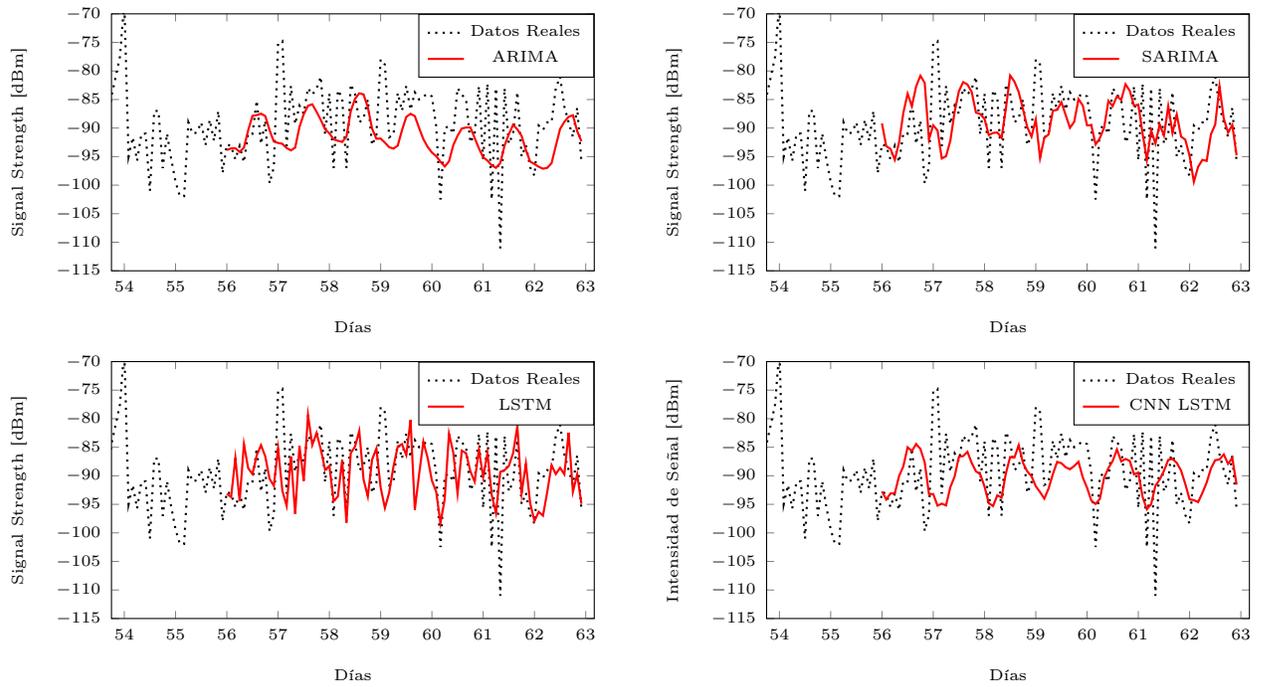


Figura 6.6: Predicción semana 1 de intensidad de señal 3G UMTS considerando condiciones climáticas.

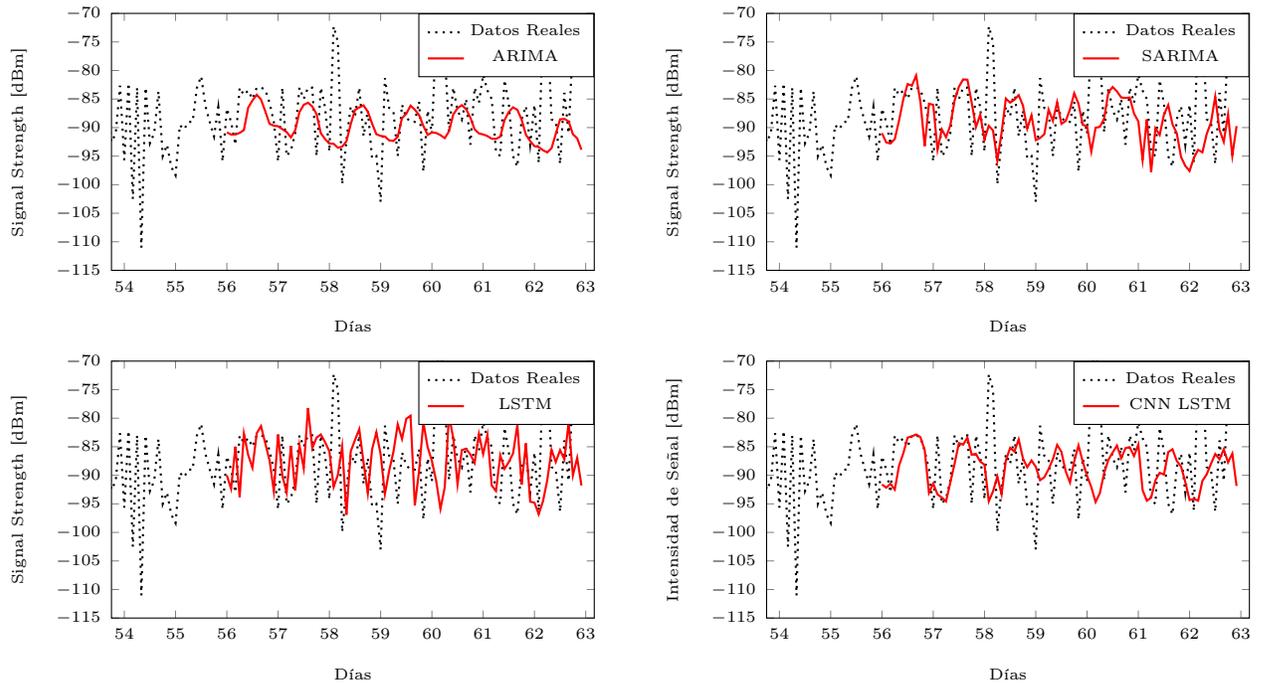


Figura 6.7: Predicción semana 2 de intensidad de señal 3G UMTS considerando condiciones climáticas.

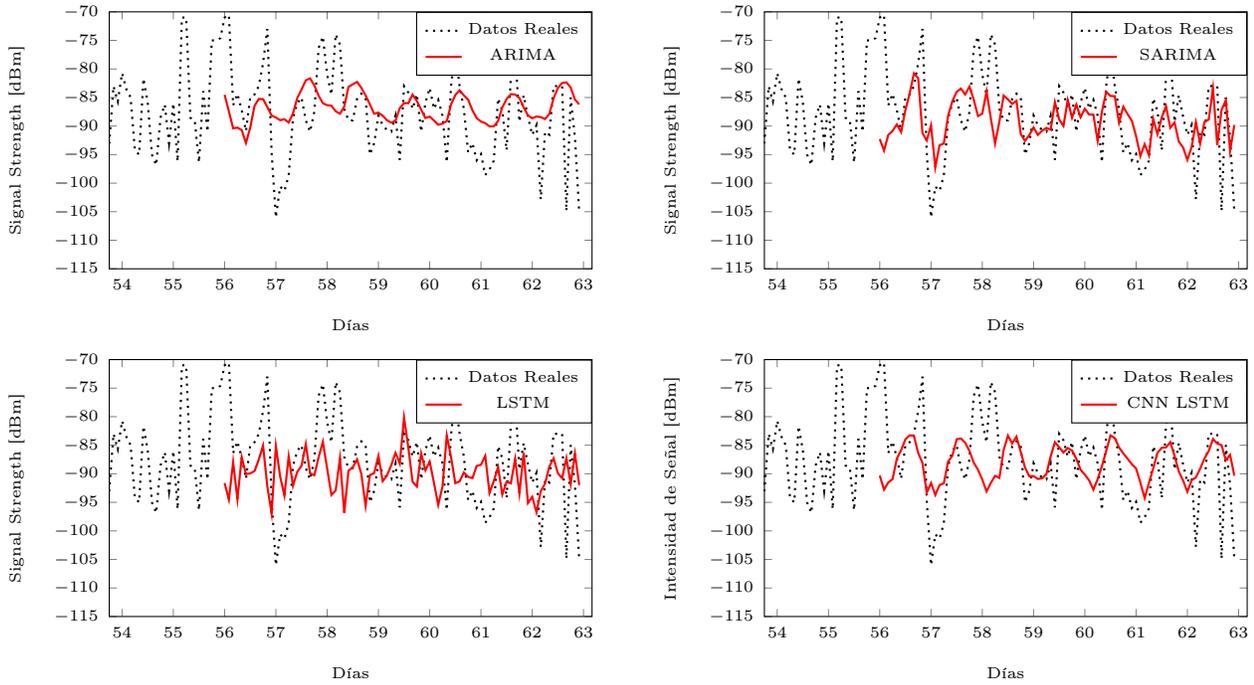


Figura 6.8: Predicción semana 3 de intensidad de señal 3G UMTS considerando condiciones climáticas.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
ARIMA	7.46	5.59	6.58	17.42	112.33	47.67	39.33
SARIMA	7.46	5.47	6.44	17.33	26.34	41.00	42.67
LSTM	8.05	6.15	7.20	17.02	10.62	44.67	38.00
CNN-LSTM	7.49	5.45	6.43	17.61	61.23	43.67	44.00

Tabla 6.4: Medidas de precisión para intensidad de señal 3G UMTS considerando condiciones climáticas.

Intensidad de Señal 4G LTE

Las Figuras 6.3, 6.4 y 6.5 muestran los resultados obtenidos al predecir las tres semanas de datos de intensidad de señal en redes 4G LTE utilizando los cuatro modelos considerados e incluyendo información externa de datos climáticos. En la Tabla 6.5 se presentan las medidas de precisión calculadas en promedio para las tres semanas. Las medidas de precisión obtenidas para cada semana pueden ser consultadas en detalle en el Anexo F.

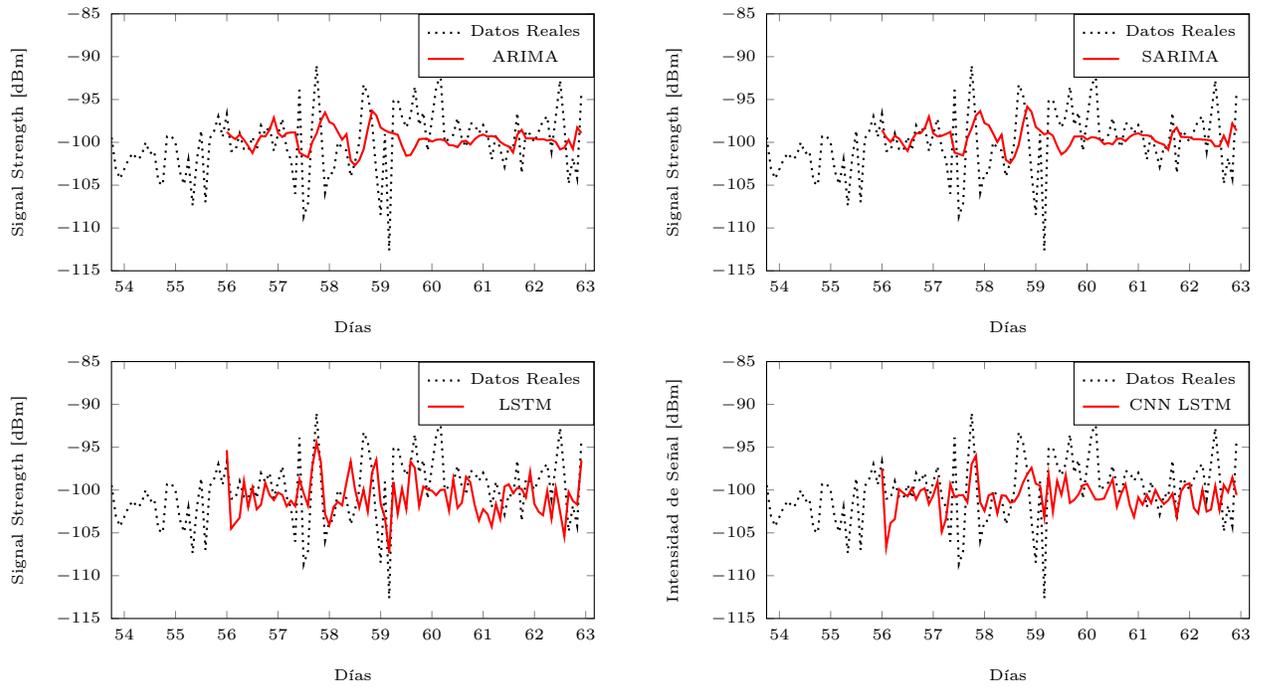


Figura 6.9: Predicción semana 1 de intensidad de señal 4G LTE considerando condiciones climáticas.

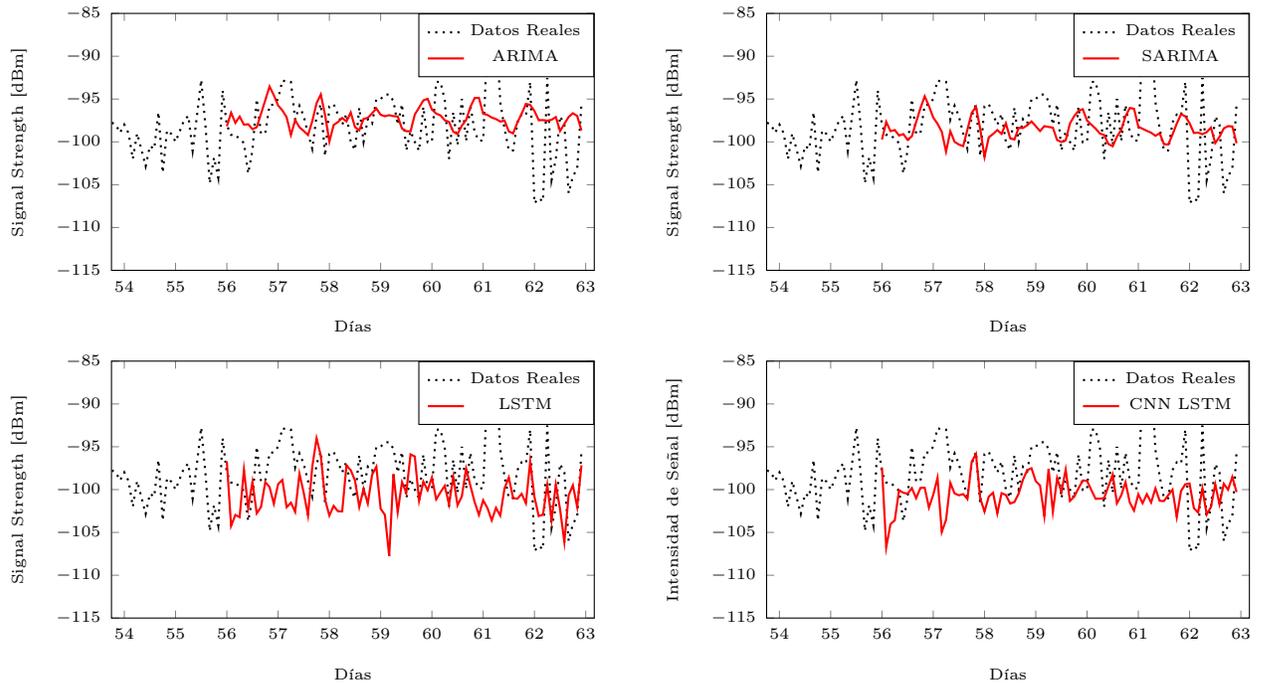


Figura 6.10: Predicción semana 2 de intensidad de señal 4G LTE considerando condiciones climáticas.

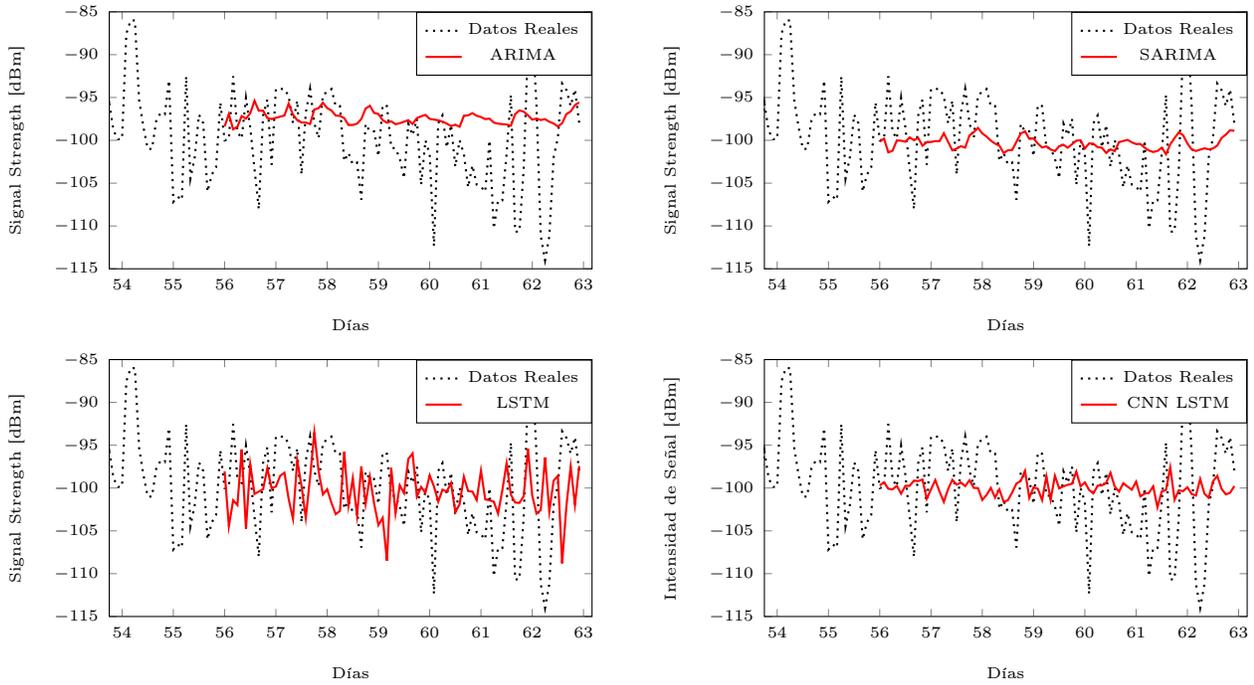


Figura 6.11: Predicción semana 3 de intensidad de señal 4G LTE considerando condiciones climáticas.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
ARIMA	4.75	3.58	3.57	11.50	89.96	32.33	57.67
SARIMA	4.50	3.42	3.46	11.05	88.14	34.33	55.00
LSTM	5.12	4.10	4.18	7.94	13.51	35.67	50.33
CNN-LSTM	4.86	3.76	3.83	10.44	38.86	39.00	50.00

Tabla 6.5: Medidas de precisión para intensidad de señal 4G LTE considerando condiciones climáticas.

6.4. Discusión de Resultados

En terminos generales, los resultados obtenidos muestran un buen desempeño al incluir información acerca de las condiciones climáticas en los modelos predictivos de intensidad de señal, en donde mayoritariamente, los modelos disminuyen sus errores de predicción y aumentan su porcentaje de precisión. Esto es principalmente notable en el modelo ARIMA, el cual anteriormente solo realizaba predicciones básicas, correspondientes en su mayor medida a una recta de valor constante cercana a la media de los datos reales.

Para el caso de predicción de intensidad de señal general, los modelos ARIMA y SARIMA presentan un comportamiento similar, con un marcado comportamiento periódico y con una curva de predicción suavizada, sin muchos cambios abruptos en sus datos. Dichos modelos

obtienen menores valores de errores de predicción (RMSE, MAE y MAPE) que los modelos basados en redes neuronales. Por otro lado, tanto LSTM como CNN-LSTM generan curvas de predicción con un comportamiento no suavizado, logrando detectar de mejor manera los cambios en la forma de la curva de datos reales, especialmente en el caso del modelo LSTM. Lo anterior, se puede evidenciar en los menores valores respecto a distancias de series temporales de Fréchet y SGLK que obtienen estos modelos.

El comportamiento anterior se repite en el caso de predicción de intensidad de señal en redes 4G LTE, en donde los modelos ARIMA y SARIMA presentan una curva de predicción suave, sin mucha variación ni cambios abruptos y obteniendo mejores valores de RMSE, MAE y MAPE en comparación a las redes neuronales. LSTM y CNN-LSTM presentan curvas de predicción más ajustadas a la curva original, pudiendo detectar una gran cantidad de cambios abruptos en los datos, pero fallando en otros casos, lo que genera que sus errores de predicción sean mayores que los modelos estadísticos. Sin embargo, su mayor similitud en la forma de la curva se refleja en menores valores respecto a las distancias entre series temporales de Fréchet, EDR y SGLK.

Respecto a la predicción de intensidad de señal 3G UMTS, el modelo ARIMA realiza una curva de predicción más suavizada y más centrada en valores medios que el resto de los modelos, lo cual se refleja en mayores valores obtenidos para las distancias EDR y SGLK. El modelo LSTM presenta un comportamiento más dinámico, capaz de predecir cambios abruptos y logrando una curva muy similar en forma para algunos períodos a la curva de datos reales, obteniendo valores bajos respecto a las distancias de Fréchet y SGLK. Sin embargo, su gran dinamismo provoca ciertos grandes errores, lo que genera la obtención de mayores valores para RMSE, MAE y MAPE. Finalmente, los modelos SARIMA y CNN-LSTM, presentan un carácter intermedio, con una curva más suave en comparación a LSTM, pero aún detectando algunos cambios abruptos y capturando de buena manera los patrones generales de la serie de tiempo real, lo cual es observable principalmente en SARIMA, que presenta una curva bastante ajustada a la real. Estos dos últimos modelos, son los que obtienen mejores resultados en relación al porcentaje de precisión de sus predicciones.

A modo de resumen es posible decir que los modelos de predicción logran predecir la intensidad de señal con un porcentaje de precisión promedio de 67.00 % para intensidad de señal general, de 41.00 % para redes 3G UMTS y de 53.25 % para redes 4G LTE, logrando un incremento de 4.50 %, 4.25 % y 1.42 % respectivamente al incorporar información acerca de las condiciones climáticas presentes en el área de mediciones.

Según los resultados obtenidos, se puede observar que la inclusión de condiciones climáticas en las predicciones para 4G LTE, genera una mejora menor en relación a las otras predicciones realizadas de intensidad de señal general y señal 3G UMTS. Una posible explicación para este fenómeno, es el hecho de que las diferentes condiciones climáticas afectan a la intensidad de señal recibida de manera distinta según la frecuencia de la onda de radio en cuestión [38]. En la Tabla 6.6, se encuentran las bandas de frecuencias utilizadas por las tres grandes compañías proveedoras de servicios móviles para entregar acceso a las redes 3G UMTS y 4G LTE [17]. De esta manera, se puede observar que los tres principales proveedores entregan servicios de 4G LTE por medio de dos bandas de frecuencias distintas, una de baja frecuencia (B28) y una de alta frecuencia (B7), las cuales operan siguiendo los estándares de LTE-A (LTE-*Advanced*),

usando la tecnología *Carrier Aggregation*, diseñada para aumentar la estabilidad, ancho de banda y velocidad de las conexiones LTE, permitiendo a los dispositivos móviles permanecer conectados a ambas bandas de frecuencia simultáneamente. Esta tecnología permitiría a su vez, que las conexiones LTE presenten mayor estabilidad frente a fenómenos naturales que tengan mayor incidencia en ciertos rangos de frecuencia del espectro electromagnético, ya que si una de las bandas se ve impactada por ciertas condiciones climáticas, los usuarios aún podrían continuar conectados normalmente a la red por medio de la otra banda de frecuencia.

Proveedor \ Banda	3G UMTS	4G LTE
Entel	B2 (1900 MHz)	B7 (2600 MHz) B28 (700 MHz)
Movistar	B2 (1900 MHz) B5 (850 MHz)	B7 (2600 MHz) B28 (700 MHz)
Claro	B2 (1900 MHz)	B7 (2600 MHz) B28 (700 MHz)

Tabla 6.6: Bandas de frecuencias para redes móviles utilizadas por los principales proveedores de Chile.

Conclusión

Como se mostró anteriormente, el continuo y rápido aumento en el uso de Internet móvil involucra grandes desafíos en términos de la constante evaluación del funcionamiento correcto de los servicios entregados. Es por esto que durante los últimos años una gran cantidad de estudios han sido desarrollados tanto en Chile como en el resto del mundo, buscando caracterizar la calidad de servicio entregado por las redes móviles, con el fin de entregar a proveedores y usuarios información útil acerca del estado de la red.

A lo largo del presente trabajo se abordó el problema de cómo anteponerse al estado futuro de la red móvil para un lugar geográfico determinado, planteando la necesidad de implementar modelos de predicción, con el propósito de conocer de antemano la calidad de servicio a recibir por los usuarios de servicios móviles en un futuro próximo.

En este trabajo se utilizaron datos acerca de la intensidad de señal recibida por usuarios de la aplicación chilena *Adkintun Mobile* dentro de una porción de la comuna de Santiago, con los cuales se confeccionaron líneas de tiempo con datos agregados que representasen la intensidad de señal recibida por los usuarios en dicha área a lo largo del tiempo. Posteriormente, se llevaron a cabo experimentos con el fin de predecir la intensidad de señal representada como una serie de tiempo, utilizando modelos de predicción basados en técnicas estadísticas y en el uso de redes neuronales.

Como conclusión del trabajo de tesis realizado, es posible mencionar que los modelos de predicción utilizados lograron predecir de forma favorable la intensidad de señal recibida, obteniendo un porcentaje de precisión promedio de 62.5 % para intensidad de señal general, de 36.75 % para redes 3G UMTS y de 51.83 % para redes 4G LTE, por sobre los resultados obtenidos por el modelo de persistencia empleado como modelo base, el cual obtuvo un porcentaje de precisión de 53.33 %, 31.67 % y 45.33 % en cada caso respectivamente.

Además, se hizo uso de información externa correspondiente a datos meteorológicos con el fin de mejorar las predicciones de intensidad de señal realizadas dentro de la comuna de Santiago. Esta mejora en la precisión de los modelos predictivos, se obtiene producto de la influencia de las condiciones climáticas sobre la calidad de servicio de Internet móvil, demostrada en trabajos de investigación anteriores, y demostrada también en el análisis de correlación previo a la inclusión de los datos climáticos en los modelos. La inclusión de esta información implica una mejora promedio de 4.50 % para las predicciones de intensidad de señal general, 4.25 % para las predicciones de intensidad de señal en redes 3G UMTS y 1.42 % para las predicciones de intensidad de señal en redes 4G LTE, considerando los cuatro modelos distintos de predicción utilizados.

Por lo tanto es posible dar respuesta a la hipótesis y preguntas de investigación planteadas al comienzo de este documento, demostrando la factibilidad de predecir calidad de servicio de Internet móvil para un área determinada, utilizando técnicas propias del análisis de series temporales utilizando modelos estadísticos y basados en el aprendizaje de máquinas, mostrando también la importancia de la consideración de factores externos que influyen en la calidad de servicio recibido por los usuarios, tales como las condiciones climáticas registradas, con las cuales es posible aumentar la precisión de los modelos predictivos empleados.

Con respecto a las medidas de error calculadas para evaluar los resultados obtenidos por los distintos modelos de predicción utilizados, es importante mencionar que a pesar de que la mayoría de trabajos que realizan labores de predicción en contextos similares utilizan solamente errores de predicción tradicionales como RMSE, MAE y MAPE, estos deben ser acompañados del uso de otras métricas de precisión al trabajar con series temporales, principalmente dado que dichos errores de predicción no toman en cuenta la esencia de las series temporales, es decir, el hecho de que se encuentran ordenadas lógicamente, en un orden establecido por el tiempo.

Trabajo Futuro

Mapas de Cobertura

Una forma recurrente de presentar la calidad de servicio entregado por las empresas de telefonía móvil a sus usuarios, es mediante la creación de los llamados mapas de cobertura, en donde se muestra de forma visual la intensidad de señal media para un área geográfica determinada. Un ejemplo de esto se muestra en la Figura 6.12, en donde los sectores con mayor intensidad de señal medida son representados con color verde, mientras que los sectores con menor intensidad de señal medida se representan con color rojo.



Figura 6.12: Promedio de intensidad de señal para un ISP en Chile.

A pesar de su común uso, se ha determinado que los mapas de cobertura no representan de buena forma la intensidad de señal ni la calidad de servicio que realmente se percibe en un lugar específico, ya que a pesar de que la visualización establezca una intensidad baja para un lugar, en ese mismo sector también se podrían experimentar momentos en los cuales la intensidad de señal es alta [60]. El trabajo de tesis presentado también sustenta el argumento del mal uso de los mapas de cobertura, ya que en un área geográfica pequeña la intensidad de señal puede variar significativamente a lo largo de un día, por lo que resulta muy difícil describir en un solo valor la calidad de la señal recibida en un lugar.

Por lo tanto, como trabajo futuro se propone la confección de herramientas visuales que consideren la variabilidad en el tiempo de la intensidad de señal así como las técnicas de predicción de dicho indicador que han sido presentadas en este trabajo, con el fin de confeccionar mapas de cobertura de red que representen la intensidad de señal recibida de forma dinámica en el tiempo, para dar información realmente útil a los usuarios de la red móvil.

Inclusión de Actividades Humanas

Es importante mencionar que el área geográfica de donde se obtuvieron los datos utilizados en los experimentos corresponde a un sector con una acentuada actividad humana propia de un área de trabajo o área de oficina, según lo mencionado en la Sección 3.3. Esto podría indicar que los mismos modelos de predicción (algoritmos y parámetros) utilizados en este trabajo no necesariamente tendrán el mismo buen comportamiento al predecir indicadores de calidad de servicio móvil en otras áreas de la ciudad. Esto se debe en parte a que las actividades humanas actúan como otro factor externo con efectos sobre la calidad de servicio recibida por los dispositivos, principalmente dado los niveles de congestión de la red en distintos momentos del día, relacionados con el movimiento de masas entre distintas áreas de actividades, como por ejemplo, desde áreas residenciales a áreas de oficina y viceversa, lo cual ya ha sido profundamente estudiado anteriormente [67].

Dado lo anterior, es importante tomar en cuenta a qué tipo de área corresponde el lugar donde se realiza la predicción de calidad de servicio, ya que las distintas actividades humanas pueden impactar de forma diferente la calidad de servicio que reciben los usuarios. Como primer aporte para este fin, y siguiendo lo realizado por estudios anteriores [67], se dividió a la ciudad de Santiago en celdas de 15 km^2 aproximadamente, y para cada una de ellas se calculó la serie de tiempo que indica la intensidad de señal recibida en el área durante una semana, siguiendo el mismo proceso descrito en la Sección 3.3. Posteriormente, se realizó un proceso de *clustering* de las series temporales utilizando un algoritmo de *k-means* para series de tiempo [33]. En la Figura 6.13 se presentan los resultados al ejecutar el algoritmo *k-means* con $k = 3$ y con la función de distancia *Dynamic Time Wrapping* para series temporales. En esta figura (en donde el área utilizada para los experimentos de este trabajo se encuentra marcada con un recuadro rojo), es posible interpretar de forma preliminar y teniendo un conocimiento aproximado de las actividades humanas predominantes en los distintos sectores de la ciudad, que las celdas se agrupan según las actividades humanas realizadas en ellas, en donde las celdas verdes tienden a encontrarse en sectores de áreas laborales, mientras que las celdas azules tienden a ubicarse en sectores de áreas residenciales, lo cual sirve como base para plantear la hipótesis de que áreas en donde se desarrollan actividades humanas

Clustering de Intensidad de Señal

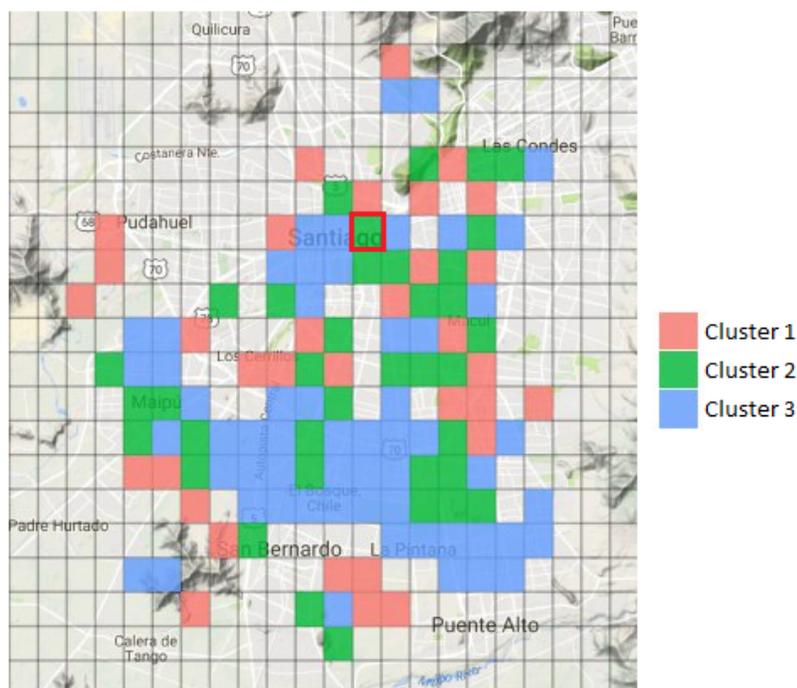


Figura 6.13: Áreas en Santiago, Chile agrupadas según sus patrones de intensidad de señal.

similares, presenten patrones similares en la calidad de servicio de Internet móvil recibido. Es importante reiterar que el análisis realizado tiene carácter preliminar y superficial, por lo que es necesario llevar a cabo un trabajo de investigación que analice en profundidad el efecto de las diferentes actividades humanas sobre la calidad de servicio de Internet móvil, pudiendo utilizar esta información en una generalización de los modelos predictivos empleados en el presente trabajo.

Predicción de Otros Indicadores de Calidad de Servicio

Como se mencionó con anterioridad, existen estudios que sugieren una fuerte correlación entre la intensidad de señal recibida por usuarios de las redes móviles y otros indicadores de calidad de servicio de Internet móvil como latencia o *throughput*, por lo que resulta promisorio la utilización de métodos de predicción, similares a los empleados en este trabajo, para predecir otros indicadores de calidad.

Uno de los posibles indicadores corresponde a la disponibilidad o estabilidad de la conexión (o qué tan probable es sufrir una desconexión de Internet), ya que el set de datos proveniente de la aplicación *Adkintun Mobile* presentado en la Sección 3.1, contiene información acerca de las desconexiones de la red móvil sufridas por los usuarios día a día. Esta información, obtenida a través de mediciones pasivas de la red, puede ser unida con la información acerca de las antenas utilizadas por los dispositivos móviles, con el fin de mapear cada evento

de desconexión con un área geográfica según la localización de la antena empleada en ese momento, y de esta manera representar la estabilidad de la conexión a Internet como una serie de tiempo a predecir, según los procedimientos seguidos en la Sección 3.3. Además, dada la correlación existente entre las desconexiones de Internet y las condiciones climáticas [56], también se propone el uso de esta información externa para incluir en las predicciones, siguiendo pasos similares a los descritos en el presente trabajo.

Otras Fuentes de Datos

Como se mencionó en la Sección 3.1, los datos de calidad de servicio utilizados provienen de la aplicación Android *Adkintun Mobile*, tomando en consideración solamente un área reducida de la ciudad de Santiago de Chile con una alta concentración de usuarios de la aplicación, con el fin de obtener series de tiempo para la intensidad de señal con mayor grado de representatividad y sin puntos faltantes.

Los buenos resultados obtenidos al utilizar los datos mencionados anteriormente, son un gran incentivo para realizar predicción de calidad de servicio de Internet móvil, por medio de la predicción de intensidad de señal u otros indicadores de calidad, en otras localidades geográficas. Lo anterior resulta factible dado el existente interés en la realización de estudios de investigación que, al igual que lo hecho por *Adkintun Mobile*, buscan caracterizar la infraestructura y la calidad presente en la red de telefonía móvil para una ubicación geográfica particular, en donde se llevan a cabo las mediciones. Entre estas fuentes de datos, cabe destacar la generada por el proyecto MONROE¹ en Europa [4], en donde equipos especializados y portadores de interfaces móviles han sido colocados en puntos fijos y móviles, como buses que día a día recorren algunas de las principales ciudades de Europa, tomando múltiples mediciones de la red móvil como mediciones pasivas de tráfico de datos, intensidad de señal recibida, inferencia de la topología de la red, entre muchas otras. Además, al utilizar tanto puntos de mediciones fijos como otros en constante movimiento, el área de mediciones resulta amplia en cada ciudad participante, pudiendo realizar los mismos experimentos propuestos en este trabajo en una mayor cantidad de áreas de forma simultánea. Los datos provenientes de este proyecto, han sido liberados para fines de investigación y usados en múltiples artículos científicos en el último tiempo².

Consideraciones Para Escenario Real

Los experimentos presentados en este trabajo fueron realizados en un ambiente acomodado, utilizando datos históricos provenientes de una única localidad geográfica. Sin embargo, para la implementación de los métodos descritos en un escenario real, sería necesario llevar a cabo la predicción de calidad de servicio para múltiples áreas a la vez, por lo que resulta importante tomar en consideración el rendimiento de los algoritmos utilizados en cuanto a su tiempo de ejecución, con el fin de realizar todos los procesos de predicción requeridos en

¹Por sus siglas en inglés: Measuring Mobile Broadband Networks in Europe

²<https://www.monroe-project.eu/resources/papers/>

el menor tiempo posible. Por lo tanto, es necesario llevar a cabo estudios que determinen la existencia de un *trade-off* entre el tiempo de ejecución y la precisión de los modelos de predicción, con el fin de seleccionar, de acuerdo a los recursos y requerimientos establecidos, el modelo más conveniente.

Además, en este trabajo se utilizaron valores históricos de las condiciones climáticas presentes en el área de mediciones, los cuales no estarían disponibles en un escenario real, en donde se deberían incluir en los modelos valores predichos acerca de las condiciones climáticas. La predicción meteorológica mediante modelos numéricos ha sido abordada desde el año 1922 y ha ido en constante evolución y mejora, por lo que hoy en día es posible contar con predicciones con alto grado de precisión [7]. De todas maneras, las predicciones son más acertadas cuando los valores predichos se encuentran en un rango de tiempo más cercano, por lo que en una aplicación real, resultaría conveniente realizar también predicciones de calidad de servicio de Internet móvil diarias, en adición a las predicciones semanales presentadas, las cuales se beneficiarían al considerar valores climáticos predichos con respecto al día siguiente, los cuales son en promedio más precisos que los valores predichos semanales.

Contribuciones del Trabajo de Tesis

Durante el desarrollo del presente trabajo, se realizaron las siguientes publicaciones científicas, en donde se presentó parte de la investigación realizada en la presente tesis de magíster. Las publicaciones fueron evaluadas y aceptadas en procesos de evaluación por pares (*peer review*), por parte de expertos en áreas relacionadas a redes y redes móviles.

1. **Madariaga, D.**, Panza, M. and Bustos-Jiménez, J. *I'm Only Unhappy When It Rains: Forecasting Mobile QoS With Weather Conditions*. In Network Traffic Measurement and Analysis Conference (TMA), 2018. Vienna, Austria. (to appear).

Principal Contribución: Factibilidad de predecir calidad de servicio de Internet móvil basándose en el uso de mediciones pasivas. También, se muestra la mejora de las predicciones realizadas al incorporar factores externos acerca de condiciones climáticas locales.

2. **Madariaga, D.**, Panza, M. and Bustos-Jiménez, J. *DNS Traffic Forecasting Using Deep Neural Networks*. In International Conference on Machine Learning for Networking (MLN), 2018. Paris, France. (to appear).

Principal Contribución: Uso de los modelos de predicción basados en redes neuronales presentados en este trabajo, aplicados de forma similar para la predicción de tráfico de datos DNS, con el fin de proporcionar una herramienta simple y eficaz para detectar anomalías en servidores DNS.

3. **Madariaga, D.** *A Proposal for a Mobile Internet QoS Forecasting Method Based on Passive Measurements*. In III Spring School on Networks (SSN), 2017. Pucón, Chile.

CEUR-WS. pages 24-26.

Principal Contribución: Validación del proceso de predicción de calidad de servicio de Internet móvil utilizando técnicas propias del análisis de series temporales, considerando la periodicidad presente en las mediciones de calidad obtenidas.

Bibliografía

- [1] Vaneet Aggarwal y col. “Prometheus: Toward quality-of-experience estimation for mobile apps from passive network measurements”. En: *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications*. ACM. 2014, pág. 18.
- [2] Inc. Akamai Technologies. “akamai’s [state of the internet] Q1 2017 report”. En: 10.1 (2017).
- [3] Yoshikazu Akinaga y col. “A proposal for a mobile communication traffic forecasting method using time-series analysis for multi-variate data”. En: *Global Telecommunications Conference, 2005. GLOBECOM’05. IEEE*. Vol. 2. IEEE. 2005, 6-pp.
- [4] Özgü Alay y col. “Measuring and assessing mobile broadband networks with MONROE”. En: *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2016 IEEE 17th International Symposium on A*. IEEE. 2016, págs. 1-3.
- [5] Ayman Amin, Alan Colman y Lars Grunske. “An approach to forecasting QoS attributes of web services based on ARIMA and GARCH models”. En: *2012 IEEE 19th International Conference on Web Services*. IEEE. 2012, págs. 74-81.
- [6] Demetres Antoniadis y col. “Available bandwidth measurement as simple as running wget”. En: *Proc. of Passive and Active Measurement Conference (PAM 2006)*. Citeseer. 2006, págs. 61-70.
- [7] Peter Bauer, Alan Thorpe y Gilbert Brunet. “The quiet revolution of numerical weather prediction”. En: *Nature* 525.7567 (2015), pág. 47.
- [8] Peter Benko, Gabor Malicsko y Andras Veres. “A large-scale, passive analysis of end-to-end TCP performance over GPRS”. En: *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*. Vol. 3. IEEE. 2004, págs. 1882-1892.
- [9] BrightEdge. *Mobile-First: 57% of Traffic is Now Mobile*. 2017. URL: <https://www.brightedge.com/resources/research-reports/mobile-first-57-traffic-now-mobile> (visitado 10-02-2018).
- [10] K Brown. “Internet society global internet report 2015”. En: *Virginia, United States: Internet society* (2016).
- [11] Edy Budiman y Ummul Hairah. “End-to-End QoS Tool Development and Performance Analysis for Network Mobile”. En: *International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET)* 3.2 (2017), págs. 128-135.
- [12] Edy Budiman y Oki Wicaksono. “Measuring quality of service for mobile internet services”. En: *Science in Information Technology (ICSITech), 2016 2nd International Conference on*. IEEE. 2016, págs. 300-305.

- [13] Javier Bustos-Jiménez y Camila Fuenzalida. “All packets are equal, but some are more equal than others”. En: *Proceedings of the Latin America Networking Conference on LANC 2014*. ACM. 2014, pág. 5.
- [14] Javier Bustos-Jimenez y col. “Adkintun: SLA monitoring of ISP broadband offerings”. En: *Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on*. IEEE. 2013, págs. 1445-1449.
- [15] Pedro Casas y col. “Predicting QoE in cellular networks using machine learning and in-smartphone measurements”. En: *Quality of Multimedia Experience (QoMEX), 2017 Ninth International Conference on*. IEEE. 2017, págs. 1-6.
- [16] John G Cleary y H Stele Martin. “Estimating bandwidth from passive measurement traces”. En: *In Passive and Active Measurement Workshop (PAM-2001*. Citeseer. 2001.
- [17] Halberd Bastion Consulting. *Mobile Networks*. 2018. URL: <https://halberdbastion.com/intelligence/mobile-networks> (visitado 10-02-2018).
- [18] Andrea J Coop y col. “Spatial accuracy of climate networks: a case study in Nebraska”. En: *Journal of Applied Meteorology and Climatology* 53.8 (2014), págs. 1932-1942.
- [19] IIT Delhi CUTS. *Mobile Internet Services in India: Quality of Service*. 2016. URL: http://www.iitd.ac.in/research/1615%5C_QoS%5C_Report%5C_CUTS%5C_IIT.pdf (visitado 10-02-2018).
- [20] Sonia Pértega Díaz y José A Vilar. “Comparing several parametric and nonparametric approaches to time series clustering: a simulation study”. En: *Journal of classification* 27.3 (2010), págs. 333-362.
- [21] Ministerio de Obras Públicas de Chile Dirección General de Aguas. *Selección de Estaciones y Parámetros*. 2018. URL: <http://www.dga.cl/Paginas/estaciones.aspx> (visitado 10-02-2018).
- [22] Giselle Font y col. “Location privacy for a quality of access to mobile Internet monitoring system”. En: *Communications and Network Security (CNS), 2014 IEEE Conference on*. IEEE. 2014, págs. 518-519.
- [23] Alexander Frömmgen y col. “Crowdsourcing measurements of mobile network performance and mobility during a large scale event”. En: *International Conference on Passive and Active Network Measurement*. Springer. 2016, págs. 70-82.
- [24] Fotis Georgatos y col. “Providing active measurements as a regular service for ISPs”. En: *PAM*. 2001.
- [25] Marta C Gonzalez, Cesar A Hidalgo y Albert-Laszlo Barabasi. “Understanding individual human mobility patterns”. En: *nature* 453.7196 (2008), pág. 779.
- [26] Miniwatts Marketing Group. *Penetración de Internet marca alza del 25 % y llega a los 16,7 millones de accesos*. 2017. URL: <http://www.subtel.gob.cl/penetracion-de-internet-marca-alza-del-25-y-llega-a-los-167-millones-de-accesos> (visitado 10-02-2018).
- [27] Maheen Hasib y John A Schormans. “Limitations of passive & active measurement methods in packet networks”. En: *London Communications Symposium (LCS), London, UK*. Vol. 38. 2003.
- [28] John Heidemann, Lin Quan y Yuri Pradkin. *A preliminary analysis of network outages during hurricane sandy*. University of Southern California, Information Sciences Institute, 2012.
- [29] Gabriel Hourton y col. “Crowd-measuring: Assessing the quality of mobile internet from end-terminals”. En: *Network Games, Control and Optimization (NetGCooP), 2012 6th International Conference on*. IEEE. 2012, págs. 145-148.

- [30] Xianhua Hu y Jie Wu. “Traffic forecasting based on chaos analysis in GSM communication network”. En: *Computational Intelligence and Security Workshops, 2007. CISW 2007. International Conference on*. IEEE. 2007, págs. 829-833.
- [31] Junxian Huang y col. “Anatomizing application performance differences on smartphones”. En: *Proceedings of the 8th international conference on Mobile systems, applications, and services*. ACM. 2010, págs. 165-178.
- [32] Lisheng Huang, Mingtian Zhou y Wenyong Wang. “A Passive Mode QoS Measurer for ISP”. En: *World Congress on Software Engineering*. IEEE. 2009, págs. 308-313.
- [33] Xiaohui Huang y col. “Time series k-means: A new k-means type smooth subspace clustering for time series data”. En: *Information Sciences* 367 (2016), págs. 1-13.
- [34] KG Hubbard. “Spatial variability of daily weather variables in the high plains of the USA”. En: *Agricultural and Forest Meteorology* 68.1-2 (1994), págs. 29-41.
- [35] Pollere Inc. *Transport Level Passive Ping Network Observer*. 2017. URL: <http://www.pollere.net/pping.html> (visitado 10-02-2018).
- [36] V Jacobson, R Braden y D Borman. “RFC 1323: TCP extensions for high performance, May 1992”. En: *Obsoletes RFC1072, RFC1185 [12, 13]. Status: PROPOSED STANDARD* ().
- [37] Hao Jiang y Constantinos Dovrolis. “Passive estimation of TCP round-trip times”. En: *ACM SIGCOMM Computer Communication Review* 32.3 (2002), págs. 75-88.
- [38] Donnie Jones. *Navy Electricity and Electronics Training Series: Introduction to Wave Propagation Transmission Lines and Antennas. Module 10*. Naval Education, Training Professional Development y Technology Center, 1998.
- [39] Marko Jurvansuu y col. “HSDPA performance in live networks”. En: *Communications, 2007. ICC'07. IEEE International Conference on*. IEEE. 2007, págs. 467-471.
- [40] Felipe Lalanne y col. “Adkintun Mobile: Towards using personal and device context in assessing mobile QoS”. En: *Wireless Communications and Mobile Computing Conference (IWCMC), 2015 International*. IEEE. 2015, págs. 49-54.
- [41] Jari Luomala e Ismo Hakala. “Effects of temperature and humidity on radio signal strength in outdoor wireless sensor networks”. En: *Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on*. IEEE. 2015, págs. 1247-1255.
- [42] Ratul Mahajan y col. “Uncovering Performance Differences Among Backbone ISPs with Netdiff.” En: *NSDI*. 2008, págs. 205-218.
- [43] HS MARTIN. “Analysis of Internet Delay Times”. En: *Proc. The 1st Passive and Active Measurement Workshop*. 2000, págs. 141-148.
- [44] Warren Matthews y Les Cottrell. “The PingER project: active Internet performance monitoring for the HENP community”. En: *IEEE Communications Magazine* 38.5 (2000), págs. 130-136.
- [45] Aref Meddeb. “Internet QoS: Pieces of the puzzle”. En: *IEEE Communications Magazine* 48.1 (2010).
- [46] Dandan Miao, Xiaowei Qin y Weidong Wang. “The periodic data traffic modeling based on multiplicative seasonal ARIMA model”. En: *Wireless Communications and Signal Processing (WCSP), 2014 Sixth International Conference on*. IEEE. 2014, págs. 1-5.
- [47] Venkat Mohan, YR Janardhan Reddy y K Kalpana. “Active and passive network measurements: a survey”. En: *International Journal of Computer Science and Information Technologies* 2.4 (2011), págs. 1372-1385.
- [48] Al Morton. *Active and passive metrics and methods (with hybrid types in-between)*. Inf. téc. 2016.

- [49] Ashkan Nikravesh y col. “Mobile network performance from user devices: A longitudinal, multidimensional analysis”. En: *International Conference on Passive and Active Network Measurement*. Springer. 2014, págs. 12-22.
- [50] Eduardo Mucelli Rezende Oliveira y col. “On the regularity of human mobility”. En: *Pervasive and Mobile Computing* 33 (2016), págs. 73-90.
- [51] OpenSignal. *OpenSignal Reports*. 2018. URL: <https://opensignal.com/reports/> (visitado 18-07-2018).
- [52] Vern Paxson. “Bro: a system for detecting network intruders in real-time”. En: *Computer networks* 31.23-24 (1999), págs. 2435-2463.
- [53] Jarmo Prokkola y col. “3G/HSPA performance in live networks from the end user perspective”. En: *Communications, 2009. ICC'09. IEEE International Conference on*. IEEE. 2009, págs. 1-6.
- [54] Feng Qian y col. “Characterizing radio resource allocation for 3G networks”. En: *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. ACM. 2010, págs. 137-150.
- [55] Martin Roesch y col. “Snort: Lightweight intrusion detection for networks.” En: *Lisa*. Vol. 99. 1. 1999, págs. 229-238.
- [56] Aaron Schulman y Neil Spring. “Pingin’ in the rain”. En: *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. ACM. 2011, págs. 19-28.
- [57] Muhammad Zubair Shafiq y col. “A first look at cellular network performance during crowded events”. En: *ACM SIGMETRICS Performance Evaluation Review*. Vol. 41. 1. ACM. 2013, págs. 17-28.
- [58] Nikhil Shetty, Galina Schwartz y Jean Walrand. “Internet QoS and regulations”. En: *IEEE/ACM Transactions on Networking (TON)* 18.6 (2010), págs. 1725-1737.
- [59] Yantai Shu y col. “Wireless traffic modeling and prediction using seasonal ARIMA models”. En: *Communications, 2003. ICC'03. IEEE International Conference on*. Vol. 3. IEEE. 2003, págs. 1675-1679.
- [60] Sebastian Sonntag, Lennart Schulte y Jukka Manner. “Mobile network measurements- It’s not all about signal strength”. En: *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*. IEEE. 2013, págs. 4624-4629.
- [61] Zhili Sun y col. “Internet QoS and traffic modelling”. En: *IEE Proceedings-Software* 151.5 (2004), págs. 248-255.
- [62] Subsecretaría de Telecomunicaciones. *Internet Usage Statistics, The Internet Big Picture. World Internet Users and 2018 Population Stats*. 2018. URL: <https://www.internetworldstats.com/stats.htm> (visitado 18-07-2018).
- [63] Pedro Torres y col. “Data analytics for forecasting cell congestion on LTE networks”. En: *Network Traffic Measurement and Analysis Conference (TMA), 2017*. IEEE. 2017, págs. 1-6.
- [64] Narseo Vallina-Rodriguez y col. “RILAnalyzer: a comprehensive 3G monitor on your phone”. En: *Proceedings of the 2013 conference on Internet measurement conference*. ACM. 2013, págs. 257-264.
- [65] Huandong Wang y col. “Understanding mobile traffic patterns of large scale cellular towers in urban environment”. En: *Proceedings of the 2015 Internet Measurement Conference*. ACM. 2015, págs. 225-238.
- [66] Qing Peng Wang y col. “The control oriented QoS: Analysis and prediction”. En: *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*. Vol. 2. IEEE. 2001, págs. 1897-1902.

- [67] Fengli Xu y col. “Big data driven mobile traffic understanding and forecasting: A time series approach”. En: *IEEE transactions on services computing* 9.5 (2016), págs. 796-805.
- [68] Liangzhao Zeng y col. “Event-driven quality of service prediction”. En: *International Conference on Service-Oriented Computing*. Springer. 2008, págs. 147-161.

Anexos

Anexo A

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
Naïve	5.10	3.71	3.90	13.14	267.41	30.00	45.00
ARIMA	3.27	2.55	2.71	11.16	585.25	26.00	58.00
SARIMA	3.49	2.65	2.81	6.24	18.51	22.00	59.00
LSTM	3.18	2.43	2.57	7.54	30.64	24.00	56.00
CNN-LSTM	3.64	2.81	3.01	9.95	34.70	24.00	57.00

Medidas de precisión para semana 1 de intensidad de señal general.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
Naïve	4.12	3.17	3.40	5.10	197.28	25.00	46.00
ARIMA	3.26	2.52	2.70	9.81	534.00	30.00	53.00
SARIMA	3.96	2.93	3.15	7.05	15.07	24.00	56.00
LSTM	3.56	2.72	2.93	7.92	30.86	24.00	52.00
CNN-LSTM	3.73	2.89	3.12	8.83	30.08	31.00	48.00

Medidas de precisión para semana 2 de intensidad de señal general.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
Naïve	4.68	3.29	3.36	11.46	239.93	30.00	50.00
ARIMA	4.38	2.97	3.02	18.32	576.56	32.00	52.00
SARIMA	4.49	3.52	3.70	9.22	26.41	35.00	45.00
LSTM	4.90	3.97	4.18	12.65	49.93	42.00	36.00
CNN-LSTM	4.26	2.99	3.04	16.57	50.03	30.00	49.00

Medidas de precisión para semana 3 de intensidad de señal general.

Anexo B

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
Naïve	8.64	6.49	7.37	13.53	428.03	48.00	27.00
ARIMA	6.41	5.09	5.83	19.39	385.93	52.00	32.00
SARIMA	6.61	5.09	5.81	13.96	38.14	43.00	33.00
LSTM	6.38	5.16	5.87	16.87	41.68	46.00	23.00
CNN-LSTM	6.75	5.39	6.25	14.87	49.78	49.00	32.00

Medidas de precisión para semana 1 de intensidad de señal 3G UMTS.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
Naïve	10.43	7.88	9.39	21.33	495.95	48.00	23.00
ARIMA	7.08	5.51	6.58	21.66	371.33	57.00	27.00
SARIMA	8.00	5.90	7.09	16.35	25.59	39.00	29.00
LSTM	8.29	6.22	7.62	20.74	58.96	55.00	22.00
CNN-LSTM	8.75	6.65	8.16	22.05	50.08	54.00	24.00

Medidas de precisión para semana 2 de intensidad de señal 3G UMTS.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
Naïve	10.41	7.95	9.07	30.55	542.33	50.00	26.00
ARIMA	7.92	5.97	6.65	20.73	385.61	50.00	34.00
SARIMA	7.68	5.60	6.59	22.85	23.69	44.00	34.00
LSTM	8.33	6.77	8.02	22.43	30.59	57.00	19.00
CNN-LSTM	7.61	6.17	7.21	19.69	39.72	50.00	24.00

Medidas de precisión para semana 3 de intensidad de señal 3G UMTS.

Anexo C

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
Naïve	5.49	4.28	4.32	13.17	317.38	39.00	37.00
ARIMA	3.79	2.80	2.79	12.97	428.99	27.00	57.00
SARIMA	3.78	2.75	2.74	12.02	109.21	27.00	57.00
LSTM	4.62	3.57	3.61	7.34	5.97	33.00	46.00
CNN-LSTM	4.09	3.33	3.37	8.78	36.60	35.00	43.00

Medidas de precisión para semana 1 de intensidad de señal 4G LTE.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
Naïve	5.58	4.16	4.33	8.06	297.46	37.00	41.00
ARIMA	4.23	3.26	3.39	13.21	286.48	39.00	45.00
SARIMA	4.28	3.34	3.47	11.89	92.36	34.00	44.00
LSTM	5.01	3.93	4.04	6.22	4.96	34.00	39.00
CNN-LSTM	5.22	4.16	4.36	12.34	40.16	47.00	31.00

Medidas de precisión para semana 2 de intensidad de señal 4G LTE.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
Naïve	7.50	5.33	5.24	9.73	341.79	40.00	39.00
ARIMA	5.39	4.27	4.26	13.93	353.65	46.00	38.00
SARIMA	5.50	4.25	4.17	14.17	149.06	45.00	34.00
LSTM	6.19	4.98	4.92	10.75	19.55	43.00	30.00
CNN-LSTM	5.44	4.33	4.32	12.03	64.50	42.00	36.00

Medidas de precisión para semana 3 de intensidad de señal 4G LTE.

Anexo D

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
ARIMA	3.19	2.32	2.48	8.55	73.71	21.00	60.00
SARIMA	2.99	2.18	2.30	7.47	60.07	18.00	62.00
LSTM	3.18	2.48	2.63	6.56	22.76	21.00	61.00
CNN-LSTM	3.20	2.59	2.71	6.43	42.81	23.00	57.00

Medidas de precisión para semana 1 de intensidad de señal general considerando condiciones climáticas.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
ARIMA	3.35	2.60	2.79	8.40	63.86	25.00	54.00
SARIMA	3.26	2.48	2.66	7.68	52.70	23.00	57.00
LSTM	3.89	2.98	3.21	7.56	17.12	28.00	51.00
CNN-LSTM	3.56	2.79	2.97	7.05	24.53	22.00	54.00

Medidas de precisión para semana 2 de intensidad de señal general considerando condiciones climáticas.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
ARIMA	4.23	2.86	2.92	16.33	88.60	24.00	59.00
SARIMA	4.04	2.94	3.04	14.22	77.62	27.00	54.00
LSTM	4.18	2.98	3.04	12.98	23.98	25.00	56.00
CNN-LSTM	4.19	2.87	2.93	15.69	65.03	24.00	57.00

Medidas de precisión para semana 3 de intensidad de señal general considerando condiciones climáticas.

Anexo E

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
ARIMA	6.75	5.30	6.17	14.00	114.84	46.00	30.00
SARIMA	6.52	5.00	5.69	13.81	31.17	41.00	29.00
LSTM	6.45	4.89	5.59	12.80	8.48	43.00	34.00
CNN-LSTM	6.64	5.13	5.91	15.16	78.39	45.00	28.00

Medidas de precisión para semana 1 de intensidad de señal 3G UMTS considerando condiciones climáticas.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
ARIMA	8.32	5.94	7.31	19.30	136.80	51.00	31.00
SARIMA	8.38	5.99	7.31	16.43	27.03	40.00	38.00
LSTM	9.36	7.01	8.36	17.19	6.94	43.00	28.00
CNN-LSTM	8.43	5.76	7.04	17.91	43.65	41.00	39.00

Medidas de precisión para semana 2 de intensidad de señal 3G UMTS considerando condiciones climáticas.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
ARIMA	7.30	5.52	6.27	18.96	85.36	46.00	36.00
SARIMA	7.49	5.41	6.34	21.74	20.83	42.00	35.00
LSTM	8.34	6.57	7.66	21.08	16.44	48.00	23.00
CNN-LSTM	7.41	5.46	6.34	19.77	61.65	45.00	34.00

Medidas de precisión para semana 3 de intensidad de señal 3G UMTS considerando condiciones climáticas.

Anexo F

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
ARIMA	4.07	2.98	2.98	10.01	74.32	26.00	54.00
SARIMA	4.00	2.90	2.90	10.25	73.46	24.00	56.00
LSTM	3.86	3.14	3.17	5.87	10.76	30.00	45.00
CNN-LSTM	3.99	3.15	3.18	8.52	23.45	30.00	48.00

Medidas de precisión para semana 1 de intensidad de señal 4G LTE considerando condiciones climáticas.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
ARIMA	4.15	3.19	3.26	8.87	60.81	29.00	50.00
SARIMA	4.25	3.26	3.38	10.05	52.41	33.00	44.00
LSTM	5.42	4.20	4.41	9.05	8.88	39.00	35.00
CNN-LSTM	5.09	3.86	4.05	10.67	21.74	43.00	37.00

Medidas de precisión para semana 2 de intensidad de señal 4G LTE considerando condiciones climáticas.

	RMSE	MAE	MAPE	Fréchet	SGLK	EDR	% Precisión
ARIMA	6.03	4.58	4.45	15.62	134.75	42.00	40.00
SARIMA	5.25	4.09	4.09	12.85	138.55	46.00	37.00
LSTM	6.08	4.96	4.96	8.91	20.88	38.00	29.00
CNN-LSTM	5.49	4.28	4.26	12.13	71.39	44.00	37.00

Medidas de precisión para semana 3 de intensidad de señal 4G LTE considerando condiciones climáticas.