



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

DESARROLLO DE UN MODELO PARA PREDECIR INTERRUPCIONES EN LA RED DE UNA EMPRESA DE TELECOMUNICACIONES

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL

BENJAMÍN IGNACIO O'RYAN BLÁZQUEZ

PROFESOR GUÍA:
NICOLÁS CISTERNAS GONZÁLEZ

MIEMBROS DE LA COMISIÓN:
GASTÓN SUAREZ CROTHERS
FELIPE VILDOSO CASTILLO

SANTIAGO DE CHILE
2024

RESUMEN DE LA MEMORIA PARA OPTAR AL TÍTULO DE: Ingeniero Civil Industrial

ESTUDIANTE: Benjamín Ignacio O’Ryan Blázquez

FECHA: 2024

PROFESOR GUÍA: Nicolás Cisternas González

DESARROLLO DE UN MODELO PARA PREDECIR INTERRUPCIONES EN LA RED DE UNA EMPRESA DE TELECOMUNICACIONES

El presente proyecto de titulación se desarrolla en el contexto de la empresa Entel, una destacada compañía en el sector de las telecomunicaciones, líder en servicios de telefonía móvil con un 34% de participación de mercado. La relevancia de mantener la posición de liderazgo en el mercado y mantener a los clientes satisfechos, lleva a la empresa a siempre tratar de mantener sus servicios activos. Para lograr esto, es fundamental minimizar el impacto de las fallas que pueda tener la red de telecomunicaciones, específicamente la red Core, ya que esta es la base para mantener todos los servicios funcionales. El objetivo central de este proyecto radica en la creación de un modelo de aprendizaje automático capaz de prever posibles fallas en la red de la Región Metropolitana, permitiendo así que la empresa pueda tomar acciones de mantenimiento preventivas y evitar problemas en los servicios que podrían disminuir la satisfacción y experiencia del cliente.

Para lograr el desarrollo de este modelo se empleó la metodología CRISP-DM. Esta metodología consta en entender la problemática desde la mirada del negocio para dar soluciones adecuadas, entender que datos se tiene disponibles, el procesamiento de estos datos, la creación y evaluación del modelo. El objetivo de predicción se definió como la posibilidad de que existan fallas importantes dentro de un nodo en las próximas 2 horas. Para lograr esto, se eligió el algoritmo XGBoost ya que presentó buenos resultados.

Los resultados del modelo fueron favorables, logrando prever con un 73% de precisión el estado de fallas importantes en los nodos de la red Core con dos horas de antelación, superando el umbral de éxito definido en 60% según los estándares de la empresa. Este logro representa un beneficio significativo para Entel al brindar la capacidad de anticiparse a posibles fallos, programar mantenimiento técnico preventivo y elevar la eficiencia operativa y la confiabilidad de la red. Es esencial destacar que, al prevenir problemas en las redes, se garantiza una mayor confiabilidad en la prestación de servicios, generando un impacto positivo directo en la satisfacción de los clientes.

En caso de implementar este modelo, se estimaría una reducción sustancial de 182 millones de pesos chilenos en los riesgos financieros asociados a una hora de falla en la red. Además, existe la posibilidad de lograr una disminución significativa de un 73% en la fuga de clientes motivada por problemas en la red. Este enfoque proactivo no solo resguarda los recursos financieros, sino que también fortalece la retención de clientes y la reputación de Entel en el mercado.

Todo por el futuro.

Agradecimientos

Tras un extenso trayecto dedicado a la elaboración de este informe, deseo expresar mi agradecimiento en primer lugar a los profesores de la comisión, quienes, con su cercanía y valioso aporte de conocimientos, fueron guías fundamentales en el desarrollo de este proyecto. En segundo término, quiero reconocer el respaldo constante de mi núcleo familiar a lo largo de los seis años de carrera. Destaco el apoyo invaluable de mi querida hermana, quien colaboró activamente en la redacción de esta sección. Asimismo, agradezco a mi madre por tomarse el tiempo de leer esta memoria y proporcionarme comentarios constructivos que contribuyeron significativamente a perfeccionar el trabajo realizado.

Finalmente, extendiendo mi gratitud a mi querida novia, quien ha sido un pilar fundamental a lo largo de toda mi carrera, brindándome un apoyo inquebrantable en mi desarrollo académico, personal y profesional.

Quiero expresar mi agradecimiento a la Universidad de Chile, en especial a la FCFM, así como a los académicos que contribuyeron a la formación de mis conocimientos. Su orientación me permitió superar pruebas difíciles, desarrollar habilidades para resolver problemas y cultivar la resiliencia frente a situaciones complejas.

Tabla de Contenido

1. Antecedentes generales	1
1.1. Análisis del Mercado	1
1.2. Características de la Empresa	3
1.3. Estructura Organizacional	5
1.4. Red de Telecomunicaciones.....	6
1.5. Fallas en la Red de Telecomunicaciones	8
1.6. Oportunidad de Mejora.....	9
1.7. Antecedentes de los nodos.....	10
2. Justificación del Proyecto	11
2.1. Rol del Estudiante	12
3. Objetivos	12
3.1. Objetivo General.....	12
3.2. Objetivos Específicos	13
3.3. Alcances	13
4. Marco Conceptual.....	14
4.1. Data Science	14
4.2. Aprendizaje de Maquinas	14
4.2.1. Tipos de Aprendizaje de Maquinas.....	14
4.2.2. Aprendizaje Supervisado.....	14
4.2.3. Extreme Gradient Boosting (<i>XGBOOST</i>).....	15
4.3. Métricas de Evaluación	16
5. Metodología.....	17
6. Desarrollo.....	19
6.1. Comprensión del Negocio	19
6.2. Entendimiento y Análisis de los Datos	20
6.2.1. Pre-Procesamiento de los Datos.....	20
6.2.2. Entendimiento de las Variables.....	21

7.2.3. Análisis descriptivo de los datos	22
6.3. Preparación de los Datos.....	26
6.3.1. Preparación Indicadores de Rendimiento y Conexiones	26
6.3.2. Preparación datos de Alarmas	27
6.3.3. Transformaciones Finales.....	27
6.4. Modelado.....	28
6.4.1. Bases de entrenamiento y de prueba.....	28
6.4.2. Construcción del Modelo	29
6.4.3. Evaluación del Modelo	29
7. Resultados.....	30
7.1. Limitaciones y Espacios de Mejora.....	33
7.2. Futuras Líneas de Investigación.....	33
8. Conclusiones.....	34
Bibliografía	36
Anexo	38
Anexo A.....	38
Anexo B	39
Anexo C.....	40
Anexo D.....	41

Índice de Tablas

Tabla 1: Características Bases de Datos	21
Tabla 2: Estadísticas Descriptivas Alarmas en Nodos	24
Tabla 3: Estadísticas Descriptivas Indicadores Rendimiento.....	24
Tabla 4: Estadísticas Descriptivas Conexiones Entre Nodos	25
Tabla 5: Resultados del Modelo.....	29
Tabla 6: Resultados del Modelo por Clases	30

Índice de Ilustraciones

Ilustración 1: Cantidad de Servicios de Telecomunicaciones Contratados Chile... 1	1
Ilustración 2: Participación de Mercado en Internet Móvil Chile, 2023	2
Ilustración 3: Participación de Mercado en Telefonía Móvil Chile, 2023	3
Ilustración 4: Organigrama Entel	5
Ilustración 5: Diagrama Funcionamiento de Red	7
Ilustración 6: Red Core y Servicios a Empresas	7
Ilustración 7: Falla de Enlace en la Red Core.....	8
Ilustración 8: Diagrama Metodología CRISP-DM	18
Ilustración 9: Cantidad de Alarmas por Hora del Dia	22
Ilustración 10: Mediciones de Rendimiento por Hora del Dia	23
Ilustración 11: Cruce Bases Indicadores de Rendimiento con Conexiones.....	26
Ilustración 12: Ejemplo Shifteo de Variable.....	28
Ilustración 14: Variables Importantes del Modelo.....	32

1. Antecedentes generales

1.1. Análisis del Mercado

El mercado de las telecomunicaciones en Chile ha presentado altas cifras de crecimiento en los últimos 20 años; donde se distinguen importantes aumentos en los servicios de internet y telefonía contratados por las personas, lo que genera mejoras en la calidad de vida de estas.

En el 2002 las personas promediaban 0,5 servicios de telecomunicaciones contratados, es decir, solo la mitad de la población tenía acceso a telefonía, internet, etc. Actualmente en 2023, el promedio alcanzó los 3 servicios por cada habitante, demostrando que hoy en día hay más necesidad para contratar alguno de estos y que las empresas proveedoras tienen una mayor habilidad para generar la oferta suficiente y satisfacer las necesidades de los clientes ((SUBTEL), 2023). A continuación, un gráfico que detalla la cantidad de los principales servicios contratados al primer trimestre del 2023.

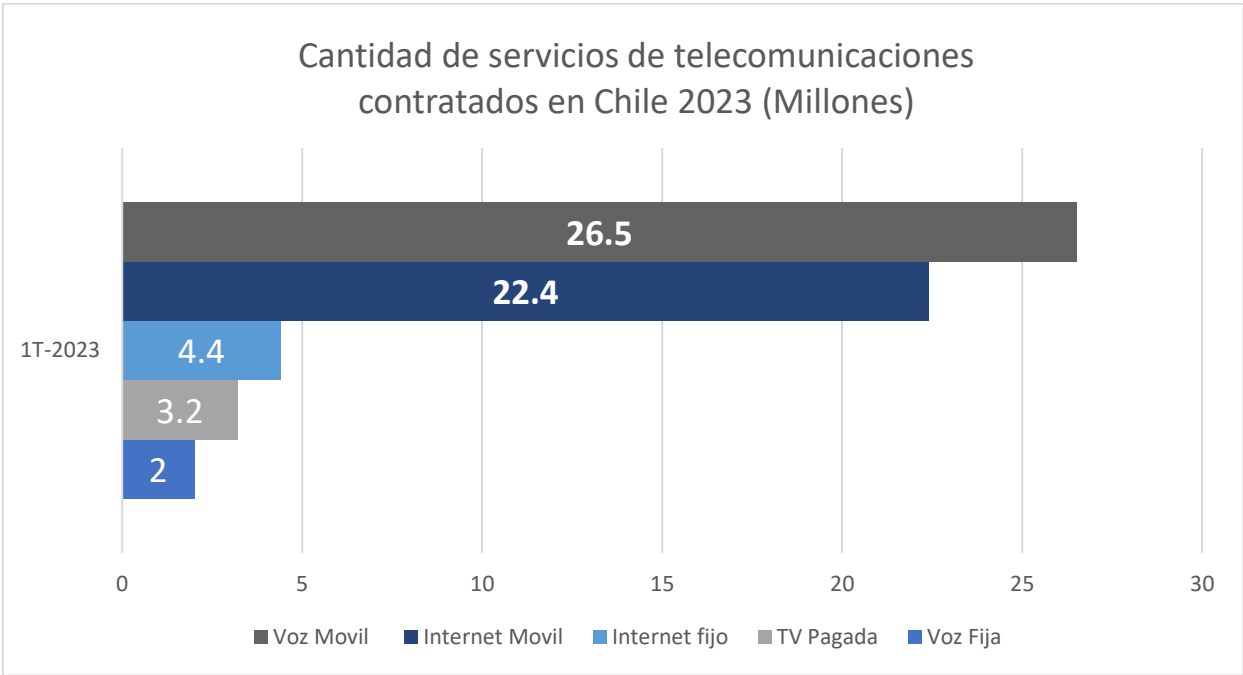


Ilustración 1: Cantidad de Servicios de Telecomunicaciones Contratados Chile

Se puede notar que existe una importante cantidad respecto a servicios de internet móvil, el cual llegó a ser cercano al de telefonía móvil y logrando entre ellos un total de 48,9 millones de contrataciones lo que representa aproximadamente un 83% del total de

servicios contratados en Chile. Esto indica, que existe una clara tendencia respecto a los servicios relacionados a dispositivos móviles, los cuales son relevantes para las utilidades de las empresas de telecomunicaciones.

El tráfico de internet móvil en Chile aumenta aproximadamente un 31,3% anualmente, reflejando la creciente demanda. Dentro de este crecimiento es importante notar que la conexión 5G implementada a inicios del 2022, ha mostrado un gran crecimiento interanual de 642% alcanzando un total aproximado de 2 millones de suscripciones móviles ((SUBTEL), 2023). Estas cifras demuestran la rápida adopción de nuevas tecnologías que buscan potenciar los segmentos móviles en busca de una mayor satisfacción del cliente y que reafirman la importancia de estos servicios tanto para las empresas como para la población.

El mercado de telecomunicaciones en Chile se caracteriza por su alta competitividad, con seis grandes empresas dominando el panorama: Entel, Movistar, WOM, Claro, VTR y Virgin. A diferencia del promedio mundial de 2 a 3 proveedores, esta competencia ha llevado a beneficios tangibles para los consumidores, como una buena relación entre velocidad/precio de 1GB de datos por 5,5 dólares convirtiendo al país en el mejor de Latinoamérica en estos aspectos (Larocca, 2019). A continuación, se muestran 2 gráficos que representan los líderes de la industria en el segmento móvil:

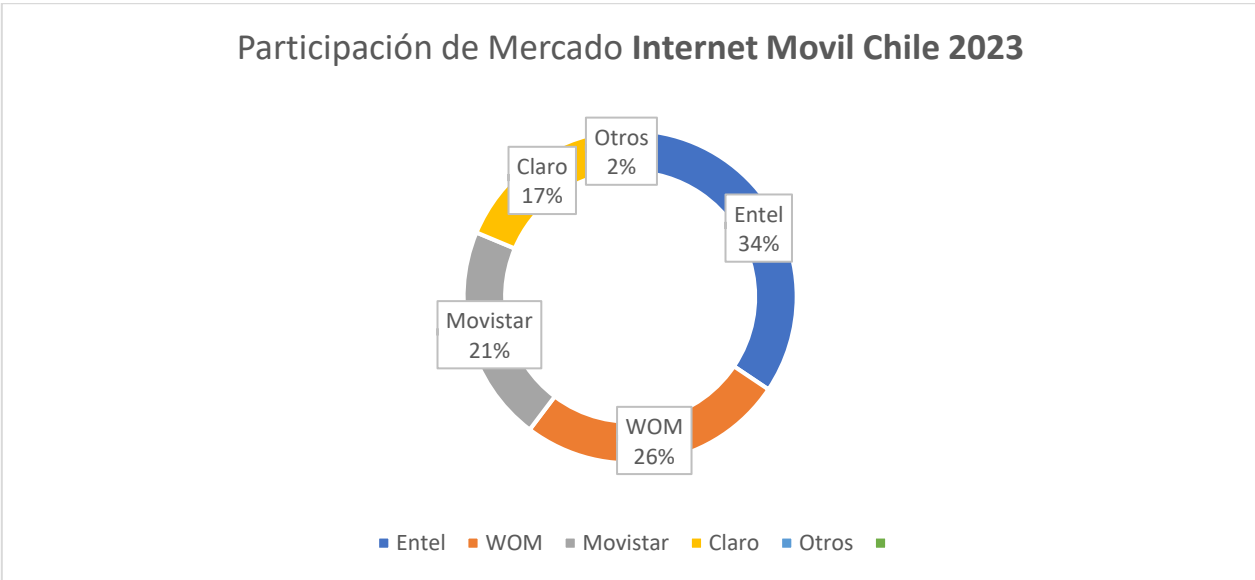


Ilustración 2: Participación de Mercado en Internet Móvil Chile, 2023

Participación de Mercado Telefonía Movil Chile 2023

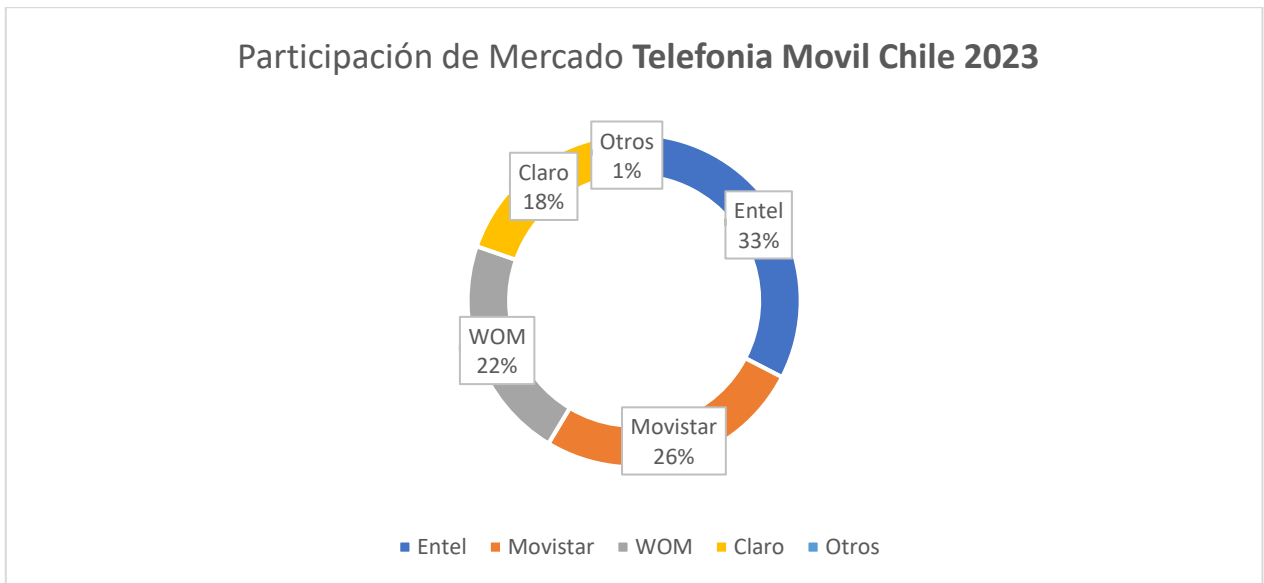


Ilustración 3: Participación de Mercado en Telefonía Móvil Chile, 2023

De estos gráficos se puede notar claramente que Entel lidera el mercado en los segmentos móviles con alrededor de un 33,5% de participación de mercado considerando telefonía e internet móvil.

1.2. Características de la Empresa

Como bien se mencionó en la sección anterior, el presente análisis se realizará dentro de la empresa Entel. Esta empresa es líder dentro del sector de las telecomunicaciones por su alta participación de mercado; además, ha sido galardonada como la 15va empresa más responsable en ámbitos ASG (ambiental, social, económico) del ranking Merco 2022, líder en 10 categorías en estudio sobre calidad de servicios destacando en la categoría general de redes 5G y consistencia de la red según el ranking de OPEN SIGNAL e incontables reconocimientos que ponen a la empresa en una muy buena posición (Entel, 2023).

La Empresa nacional de telecomunicaciones (ENTEL) fue fundada en 1964 por el Ministerio de Hacienda luego del terremoto de 1960 que dejó en evidencia la necesidad de contar con un sistema que asegurase la continuidad de las telecomunicaciones nacionales e internacionales. Entel pasó por una etapa de crecimiento e internacionalización para luego ser privatizada en 1992, luego de eso la empresa adoptó tecnologías de red móvil como el 3.5G, banda ancha, adquirió data centers, adoptó tecnologías 4G y se abrió la marca Entel Perú durante el 2014, expandiendo las operaciones al país vecino. En los últimos 10 años la empresa se destaca por seguir adoptando tecnologías de red, obtener constantes reconocimientos y ser de los pioneros en implementar la tecnología 5G (Entel, 2022).

Actualmente Entel es una sociedad anónima abierta (S.A) siendo presidida por Juan Hurtado Vicuña y su CEO actual es Antonio Buchi Buc. Cuenta con operaciones en Chile y Peru, empleando alrededor de 12.000 personas, ofreciendo productos de telefonía fija y móvil, servicios de internet, banda ancha móvil, transporte satelital de señales, call center, data center y servicios de TI. Sus ingresos a finales del año 2022 alcanzaron \$3.137.710 millones, un 27,5% superior respecto del año anterior, en tanto la utilidad anual pasó de \$74.975 millones en 2021, a \$456.787 millones en el año 2022. Actualmente cuentan con aproximadamente 20 millones de clientes entre Chile y Perú y el mayor hito que se logró en el año 2022 fue alcanzar una presencia de red 5G en 270 comunas a lo largo del país, alcanzando una participación de 44,3% en conexiones superando ampliamente a la competencia (Entel, 2022).

La misión de Entel Chile se define como “Hacer que todos vivamos mejor conectados contribuyendo responsablemente a transformar nuestra sociedad.” y su propósito como “Ofrecer el mejor servicio a nuestros clientes para transformar responsablemente nuestra sociedad a través de una experiencia distintiva.” Por otro lado, los principales objetivos estratégicos de la empresa son los siguientes (Entel, 2023) (Entel, 2022):

- Innovación y adaptación: Explorar nuevas oportunidades como realidad virtual, inteligencia artificial, tomar negocios digitales con Entel Ocean, entre otros.
- Organización y cultura ágil: Transformar la compañía en un entorno más digital y buscando ser una empresa 'Data Driven'¹.
- Experiencia distintiva: Se busca tener un rol significativo en la sociedad con la sostenibilidad en el centro, mantener el liderazgo en telefonía móvil y crecer en Perú.
- Infraestructura moderna y robusta: Buscar la constante modernización de la infraestructura de redes usando tecnología para posicionar a Entel como líder en 5G

¹ Ser una empresa Data Driven significa tomar decisiones estratégicas basadas en el análisis e interpretación de los datos almacenados en entornos digitales.

1.3. Estructura Organizacional

Para contextualizar el área en donde el estudiante se desenvuelve y donde se realizará el trabajo de título; se adjunta un diagrama que explicita la estructura organizacional de la empresa:

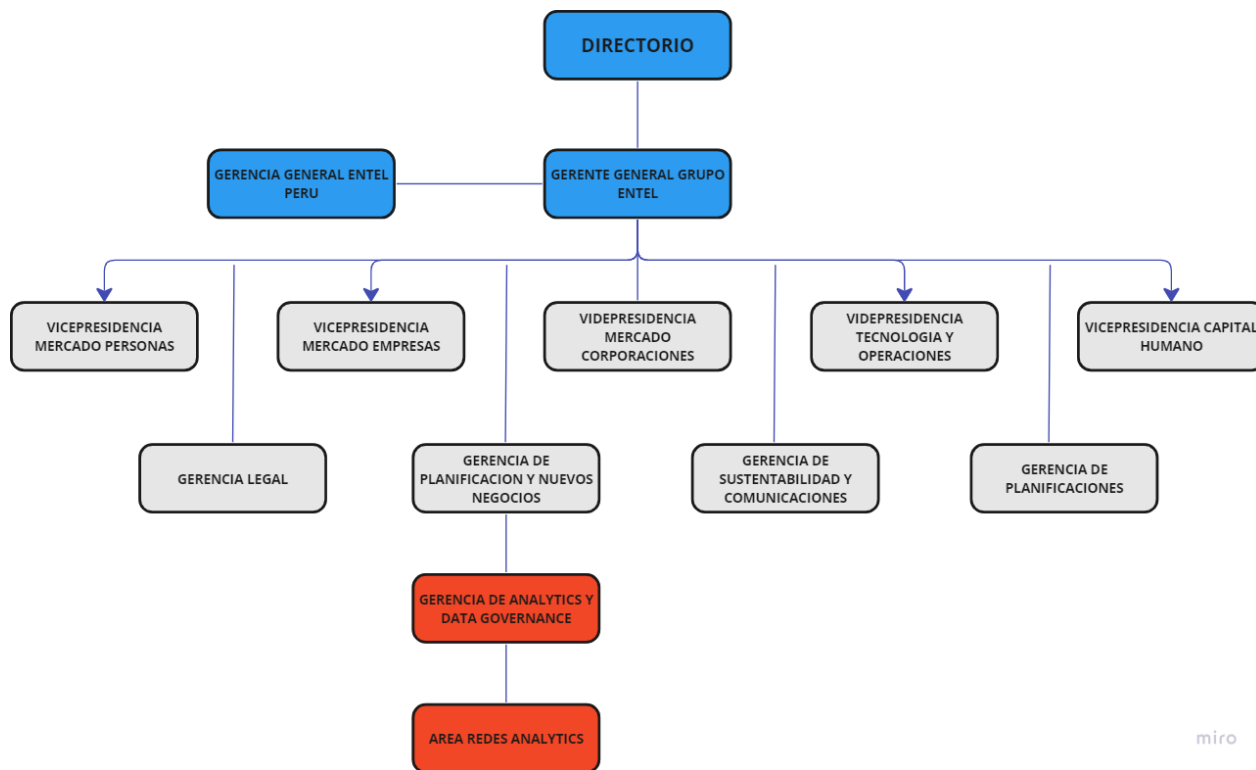


Ilustración 4: Organigrama Entel

En el previo diagrama se pueden ver las diferencias áreas y gerencias dentro de la organización, en esta ocasión, el estudiante se encuentra realizando su memoria en la gerencia de Planificación y Nuevos Negocios. Dentro de ella, se encuentra la gerencia de Analytics & Data Governance la cual cuenta con 47 trabajadores y dentro de esta última gerencia, específicamente en el área de Redes Analytics, en donde el estudiante realiza su trabajo de título.

Dada la creciente popularidad de las herramientas de analítica avanzada y el objetivo de la empresa de basar sus decisiones en los datos, se creó el área de Redes Analytics la cual tiene como objetivo atender a las necesidades del negocio usando técnicas de Data Science con la misión de realizar proyectos sobre las redes de la empresa y poder beneficiar a las operaciones internas.

1.4. Red de Telecomunicaciones

Para la empresa es de vital importancia que los servicios móviles estén activos y existan las oportunidades tecnológicas para mejorar su desempeño y adoptar nuevas tecnologías para poder mantenerse como líderes en el mercado. La base de todos los servicios de comunicaciones desemboca en la infraestructura tecnológica que se encarga de a la creación de la red, la cual sustenta todos los servicios que la empresa ofrece a sus clientes.

Una red de telecomunicaciones se puede diferenciar en dos grandes segmentos, la red de radio y la red Core. La red de radio corresponde al segmento que se encarga de captar las señales de los dispositivos (computadores, celulares, etc.) y mediante las torres de telefonía, llevar a información a su destino.

La red Core, también conocida como "red central", es el núcleo de una red de comunicación y garantiza la entrega eficiente y confiable del tráfico de datos entre las torres de telefonía (denominadas POP) principales de la red. Esta red proporciona servicios de conectividad y enrutamiento entre diferentes partes de la red, controlando el flujo de tráfico entre estas áreas. Conecta regiones o países y dirige el tráfico hacia redes externas, como internet y servicios en la nube.

La red Core está diseñada para ser rápida, confiable, escalable y segura, y se considera la parte más crítica de una red de comunicación. Está compuesta por hardware, dispositivos y software de red, como conmutadores y enrutadores. El funcionamiento completo de la red se puede ejemplificar en 4 pasos:

1. **Dispositivos de Origen y Destino:** En una red de telecomunicaciones, tienes dispositivos que envían información (como teléfonos, computadores) y dispositivos que reciben esa información en algún otro lugar.
2. **Transmisión de Datos:** La información se convierte en señales eléctricas, ondas de radio u otras formas que viajan hasta las antenas (denominadas POPs), donde se transmiten por medios físicos como fibra óptica.
3. **Enrutamiento y Conmutación:** En el camino hacia su destino, los datos a menudo necesitan ser dirigidos a través de nodos de red, como routers y switches, que toman decisiones sobre cómo enviar los datos al próximo nodo en el camino correcto hacia su destino. Es en este paso donde la red Core se hace presente.
4. **Llegada a su Destino:** Finalmente, los datos llegan al destino, es decir, llegan a la antena más cercana al usuario de destino, en donde se transforman en información más legible para ser recibida en el dispositivo de destino (teléfonos, computadores).

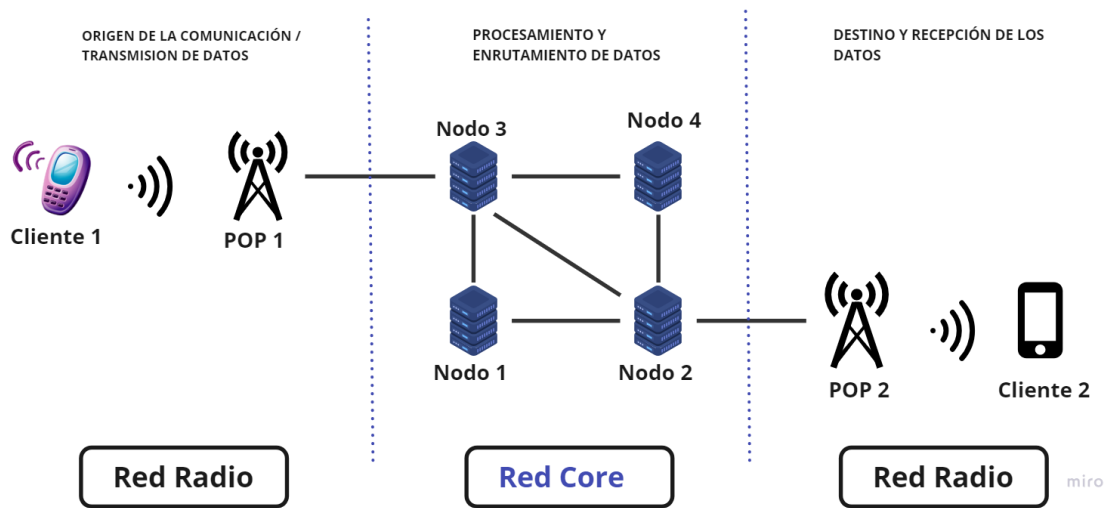


Ilustración 5: Diagrama Funcionamiento de Red

En la *ilustración 5* se ejemplifica el funcionamiento de la comunicación entre dos clientes. Es así como se observa el rol de la red de radio, la cual es la encargada de la recepción y transmisión de los datos, así como también de generar la recepción en el dispositivo de destino. La red Core, en tanto, corresponde al conjunto de nodos que se encuentran en medio, encargándose del procesamiento y enrutamiento de la información, buscando el camino más eficiente.

Por otro lado, la red Core no sólo está presente al momento de ofrecer servicios a personas naturales, sino que también a clientes institucionales, como empresas, bancos, aseguradoras, entre otros.

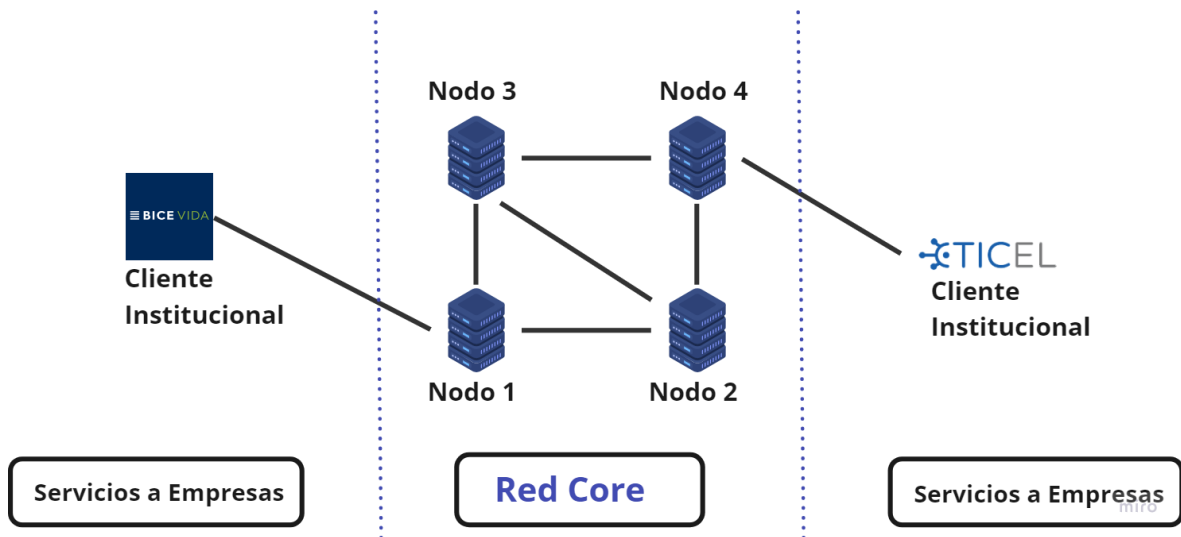


Ilustración 6: Red Core y Servicios a Empresas

En la *ilustración 6* se muestra la relación entre Entel y sus empresas de clientes. Para ofrecer servicios a grandes compañías, Entel genera conexiones directas entre la red Core y las empresas (evitando la conexión a la red de radio) ofreciéndoles servicios de fibra óptica e internet o servicio móvil. Este mecanismo implica mayor eficiencia y velocidad. A su vez, cubre los requisitos de transmisión de datos, de los clientes institucionales.

Es importante notar que el uso de la red Core está presente en toda la oferta, puesto que es la encargada del transporte de información. Con ello, constituye un pilar fundamental para la continuidad del negocio y por consecuencia, los ingresos de la empresa. Es por esto, que Entel tiende a hacer todo lo posible para evitar cualquier tipo de falla que pueda ocurrir en la red Core, ya que puede afectar tanto a las personas como a empresas que sustentan el negocio.

1.5. Fallas en la Red de Telecomunicaciones

Cuando se produce un corte de enlace o fallas entre nodos de la red Core, esto impacta directamente en la red general, ya que la información se ve obligada a seguir rutas alternativas que podrían llegar a saturarse, resultando en una disminución de velocidad e intermitencia del servicio. Estos errores pueden ocurrir por: fallas en los equipos tecnológicos, problemas de conexión entre equipos, error humano, bajas en el voltaje y diversos factores externos.

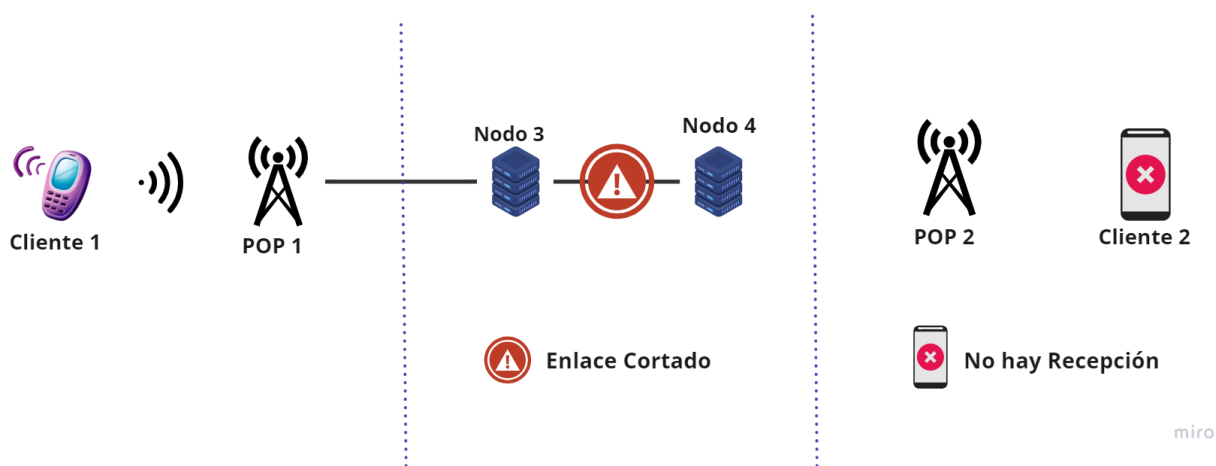


Ilustración 7: Falla de Enlace en la Red Core

La desconexión completa de un nodo de la red Core, es decir, cuando todos los enlaces fallan, impide la transmisión de información a la red de radio lo que inhabilita una posible comunicación entre dos clientes o servicios de internet. Como consecuencia, todas las

torres de telefonía de origen quedan inhabilitados o experimentan interrupciones, afectando así áreas completas de la región en la que están ubicados los POP (En este caso, el área inhabilitada corresponde a todas las antenas que se conectan al nodo 3).

La inhabilitación completa de una zona tiene un impacto directo en todos los usuarios de servicios de Entel que se encuentran en esos puntos geográficos. En este escenario, los servicios o productos contratados dejan de funcionar correctamente, generando una significativa afectación para los usuarios. Para el caso de los clientes institucionales, también se presentan interrupciones o cortes de servicio debido a que las rutas de transporte de información estarían cortadas. Esto podría implicar que todo el servicio de internet de una empresa falle por completo y afectar sus operaciones.

Cuando una empresa de telecomunicaciones enfrenta cortes, interrupciones o lentitud en la conexión de sus servicios, se traduce en una disminución del prestigio. Cualquier falla en la calidad del servicio, afecta las expectativas que los agentes consumidores tienen hacia la compañía.

1.6. Oportunidad de Mejora

Como se destacó en secciones anteriores, Entel persigue la meta de transformarse en una empresa denominada "Data Driven", lo que implica fundamentar sus decisiones estratégicas en la información almacenada de sus procesos. La finalidad es tomar decisiones más fundamentadas y prescindir de la especulación. Este objetivo estratégico surge ante el constante incremento de datos en el entorno moderno, los cuales pueden ser aprovechados mediante diversas herramientas tecnológicas para comprender a profundidad a los clientes, optimizar procesos internos y gestionar recursos de manera eficiente. De esta manera, las empresas que optan por este enfoque estratégico logran mejorar su eficiencia operativa, obteniendo así una ventaja competitiva y una mayor confianza en las decisiones tomadas (Consulting, s.f.).

Alcanzar este objetivo estratégico implica: mejorar la infraestructura tecnológica, implementar el almacenamiento de datos en la nube, fomentar un cambio cultural en el personal, garantizar la seguridad de los datos y emplear herramientas de analítica avanzada, entre otros. En algunos casos, podría necesitarse una transformación digital completa en las organizaciones, con inversiones cercanas a 27.5 millones de dólares en grandes empresas (Amazon, 2022) (Solutions, 2023).

Dado que la empresa ha optado por seguir este objetivo estratégico debido a sus beneficios y otro de sus objetivos estratégicos es buscar una infraestructura moderna de redes usando tecnología, resulta lógico aplicar metodologías propias de una empresa "Data Driven". Específicamente, en el ámbito de las redes, utilizar la analítica avanzada

permite tomar acciones y decisiones estratégicas más fundamentadas. El propósito es mejorar la calidad y el funcionamiento continuo de los servicios ofrecidos.

Al ser tan importante la red Core puesto que es la base de la red completa, tiene que contar con mayores medidas de seguridad, equipos más capacitados, que sean menos propensos a fallas y una mayor prioridad en términos de servicio técnico. Si bien la red Core cuenta con medidas adecuadas para su funcionamiento continuo, es inevitable encontrarse con problemas imprevistos, principalmente fallas de las conexiones entre los nodos, que pueden cortar rutas de transmisión de información y afectar enormemente a los servicios que los clientes contrataron.

Es por esto, que la oportunidad de mejora que se plantea es utilizar la información y datos que generan los nodos de la red Core y usar herramientas de aprendizaje automático para crear un modelo el cual sea capaz de predecir con cierto grado de probabilidad las fallas o interrupciones generadas en la red, para que de esta forma el personal técnico de mantenimiento pueda adelantarse a los problemas y solucionarlos anticipadamente, mejorando la eficiencia operativa.

Para la elaboración de esta oportunidad de mejora, se proponen instrumentos de aprendizaje automático, pues son capaces de manejar grandes volúmenes de datos. Además, el estudiante trabajará con la gerencia de Core Transporte (cliente interno), la cual es la encargada de gestionar y tomar las decisiones estratégicas, respecto a la red Core.

1.7. Antecedentes de los nodos

Tal como se señaló anteriormente, es necesario utilizar la información y datos que producen los nodos de la red Core para poder aprovechar la oportunidad de mejora y desarrollar el modelo predictivo. Es importante detallar que información se utilizará para asegurar un correcto entendimiento del trabajo. Esta información corresponde a 3 elementos principales:

- **Alarmas en los nodos:** Las alarmas en los nodos corresponden a una serie de advertencias que emiten los equipos tecnológicos dentro de cada nodo cuando algo no está funcionando de manera adecuada o a ocurrido un error. Estas advertencias se recopilan y se almacenan en bases de datos a las cuales se pueden acceder para entender que fue lo que sucedió. Los errores que se generan en los equipos pueden ocasionar perdidas de conexiones entre dos nodos de la red Core, es decir, pueden llegar a generar una interrupción del enlace o un corte total
- **Indicadores de rendimiento:** Los equipos dentro de cada nodo de la red Core también generan información acerca de su desempeño, es decir, indicadores que miden el rendimiento actual. Estos buscan medir el estado de las conexiones entre nodos, indicando la diferencia de potencias en decibelios entre el origen y la

conexión de destino. De esta forma, si un indicador de rendimiento marca una potencia más bajo de los rangos normales, indicaría que existe una pérdida de enlace parcial o incluso una pérdida de enlace completa cuando la potencia del enlace disminuye drásticamente entre los nodos afectados. Esta información proviene de una base de datos distinta a la de las alarmas en la red, a la cual se puede acceder a través de un servidor especializado.

- **Conexiones de los nodos:** Se tiene una base de datos que contiene las conexiones entre los nodos, es decir, el nombre del nodo de origen y el nombre de los nodos a los cuales se está conectado. Además, se tiene los valores de referencia de cada conexión, es decir, el valor en decibelios que representa el estado normal de la conexión sin que existan fallas o pérdidas de enlace.

2. Justificación del Proyecto

El proyecto tiene como objetivo principal la creación de un modelo predictivo que utilice la información de los nodos para anticipar posibles fallos en segmentos o conexiones de la red en un futuro cercano. Este enfoque permitirá que la gerencia de Core Transporte tome medidas preventivas para evitar cortes o problemas significativos en la red. Además, el modelo puede servir como un indicador valioso para tomar decisiones estratégicas relacionadas con la arquitectura, tecnología y equipos de esta sección de la red total.

La importancia de este proyecto radica en las consecuencias económicas de las fallas en la red. Se estima que cada hora de fallo en la red representa una pérdida de ingresos de \$250 millones, lo que equivale al 8% de los ingresos totales anuales, según fuentes internas. Esta cifra incluye multas que Entel debe pagar a sus clientes institucionales, costos operativos de mantenimiento, pérdida de clientes potenciales y fuga de clientes debido a la insatisfacción del servicio contratado.

Además de las repercusiones financieras, las fallas en la red pueden generar un impacto negativo en la reputación de la empresa. Esto incluye posibles denuncias de clientes insatisfechos en redes sociales y medios de comunicación masiva, lo que afecta la imagen de la compañía. Por otro lado, se sabe mediante encuestas que alrededor de un 30% de la fuga de clientes son debido a la insatisfacción con la calidad de señal o red.

Es así como la creación de un modelo predictivo proporcionaría a Entel la capacidad de prever y abordar proactivamente problemas en la red, mitigando así los costos asociados con las fallas. Además, al evitar pérdidas de clientes y la consiguiente mala reputación, la empresa puede ofrecer un servicio más sólido y mejorar su posición en el mercado. Este proyecto se justifica no solo por la optimización de la eficiencia operativa, sino también

por la protección de los ingresos, retención de clientes y la preservación de la reputación de la empresa.

2.1. Rol del Estudiante

A modo general, el estudiante desempeñara el rol de Data Scientist dentro de la empresa, específicamente trabajando con la información y datos de la infraestructura de redes.

El rol del Data Scientist es el de desarrollar un modelo predictivo que identifique los segmentos de la red core que podrían verse afectados por una falla en un enlace. Este es un rol clave en el proyecto, ya que el modelo predictivo será utilizado para tomar decisiones sobre cómo mejorar la fiabilidad, resiliencia de la red y anticiparse a posibles fallas futuras. El proceso de trabajo incluye la selección de los datos, análisis descriptivos, la selección de los algoritmos de aprendizaje automático la evaluación del rendimiento del modelo y la comunicación y presentación de resultados.

Para desempeñar este rol con éxito, el Data Scientist deberá usar su conocimiento de técnicas para el análisis de datos, incluyendo el aprendizaje de máquinas para crear modelos predictivos, la estadística avanzada y la visualización de datos. Además, deberá tener buenas habilidades de comunicación, tanto para comunicarte con los miembros del equipo como con la contra parte del proyecto.

3. Objetivos

Con el fin de poder abordar la oportunidad de mejora tratada anteriormente y lograr desarrollar el proyecto, se determinaron el objetivo general y los objetivos específicos.

3.1. Objetivo General

Crear un modelo de aprendizaje automático que permita la predicción de fallas de los enlaces entre nodos, de la red de Entel en la Región Metropolitana, para que la empresa pueda detectar anticipadamente las fallas y resolverlas sin afectar la experiencia del cliente.

3.2. Objetivos Específicos

- Procesar los datos de las alarmas producidas por los nodos de la red y la información de desempeño de los equipos durante el mes de octubre, para poder acotar las variables que se usarán.
- Construir un conjunto de datos usando la información de las alarmas generadas en los equipos, sus indicadores de desempeño y las conexiones de los nodos, para poder determinar mediante un análisis que variables añadir al modelo.
- Definir y crear un modelo de aprendizaje de máquinas que pueda predecir anticipadamente las fallas entre nodos de la red, usando las variables definidas previamente.
- Analizar la capacidad de predicción del modelo usando métricas de precisión para determinar si es eficaz en predecir las interrupciones en la red, comparando con los estándares de la empresa.

3.3. Alcances

La elaboración del modelo como el manejo de la información y los análisis respectivos quedaran dentro de la gerencia previamente mencionada y con la contraparte. Esto sucede porque se maneja información confidencial respecto a las redes de la empresa (ubicación de los nodos, fallas, etc.). Dentro de los principales alcances que se definen para el proyecto se encuentra lo siguiente:

- Todos los procesos de obtención de la información y datos lo realizarán la gerencia de Core y Transporte. Esto no formará parte del trabajo del estudiante, dado que este último solo se encargará de realizar los requerimientos de información que sea necesaria.
- Los datos utilizados tienen una temporalidad acotada debido a la extensa cantidad de registros de alarmas que se generan en la red, es por esto, que se utilizaran muestras de 1 mes como máximo, debido a que el procesamiento de la información puede verse entorpecido si se usan registros mayores.
- En primera instancia el trabajo se acotará a la red Core de la Región Metropolitana para luego analizar la posibilidad de extrapolarlo a todo país. Esto debido a que la red Core y la región son prioridad para la empresa.
- No se llevará el modelo a producción (que funcione con datos en tiempo real), esto porque requiere una validación y herramientas que forman parte de otra área de la gerencia y además requiere una mayor cantidad de tiempo invertido que no se adecua al contrato laboral del estudiante.

- Respecto a los indicadores de éxito del proyecto, estos corresponden a la evaluación del modelo predictivo, usando métricas de precisión y a la validación de la empresa. La definición del éxito del proyecto está sujeto a los estándares de la empresa, es decir, que el modelo predictivo tenga un acierto del 60% o más. Si se logra superar este umbral se considera que el modelo es apto para ser implementado.

4. Marco Conceptual

4.1. Data Science

Es una disciplina multidisciplinaria que se centra en la generación de conocimiento y toma de decisiones a partir de datos. Abarca diversas técnicas de matemáticas, estadística, ciencias de la computación, ingeniería y ciencias sociales. Una práctica común de esta disciplina es usar modelos de aprendizaje de máquinas con los datos recopilados para la generación del conocimiento.

4.2. Aprendizaje de Maquinas

Aprendizaje de máquinas o machine learning es una de las ramas de la inteligencia artificial, cuyo concepto consiste en la capacidad de máquinas o computadores para resolver problemas de forma inteligente, pues está basado en aprendizaje de patrones en grandes volúmenes de datos.

4.2.1. Tipos de Aprendizaje de Maquinas

El aprendizaje de máquinas se divide en tres grandes grupos, estos son el aprendizaje supervisado, no supervisado y reforzado, la principal diferencia es la salida esperada de los algoritmos según el grupo al que pertenezca. Para efectos de este proyecto se usará el aprendizaje supervisado.

4.2.2. Aprendizaje Supervisado

Grupo de algoritmos que requieren conocer tanto la entrada como salida de los datos, es decir que necesitan saber la respuesta esperada o etiqueta para aprender y así generalizar la salida para las futuras entradas que ingresen al modelo (Kotsiantis, 2007).

Existen dos grupos de algoritmos supervisados, el primer grupo son las técnicas de clasificación, cuya principal característica es que la respuesta es una variable categórica, un ejemplo de clasificación es la detección de spam en correos electrónicos, pues la salida esperada son solamente dos opciones: clasificar un correo como normal o spam.

El segundo grupo son los algoritmos de regresión, se diferencia de los clasificadores porque en este caso la etiqueta es un valor numérico y por lo tanto está en un rango continuo. un ejemplo de regresión es predecir el valor de un automóvil dado su kilometraje.

Para el desarrollo de este informe se explicita la construcción de un algoritmo de clasificación llamado XGBOOST.

4.2.3. Extreme Gradient Boosting (XGBOOST)

Extreme Gradient Boosting o XGBoost (XGBoost, 2022), es un software abierto (open source) que contiene un modelo de aprendizaje de máquinas para problemas de regresión y clasificación. Este modelo es una versión avanzada del algoritmo de Gradient Boosting Trees, el cual es un método de machine learning basado en el entrenamiento de varios árboles de decisión cuyas salidas son sumadas para obtener una salida final. Los árboles de decisión son estructuras de datos que se utilizan para tomar decisiones basadas en ciertas condiciones. Estos árboles dividen el conjunto de datos en subconjuntos más pequeños basándose en las características de entrada (Chen, 2016).

A grandes rasgos, el proceso de entrenamiento se realiza en varias iteraciones. En cada iteración se adiciona un único árbol de decisión al modelo el cual es entrenado para que pueda predecir los errores cometidos por el modelo en la iteración pasada. De esta forma, la salida del modelo será igual a la predicción en la iteración anterior más la predicción del nuevo árbol de decisión, ponderado por una tasa de aprendizaje para así reducir la influencia de cada árbol a la salida de todo el modelo y evitar el sobreajuste.

Este algoritmo presenta varias optimizaciones a momento de implementarse, lo que le da mucha ventaja de rendimiento:

- Función perdida: La función de pérdida se optimiza durante el proceso de entrenamiento para encontrar los mejores parámetros del modelo.
- División Optima: utiliza algoritmos para encontrar las mejores divisiones de cada nodo del árbol de decisión.
- Tasa de Aprendizaje: Controla la contribución de cada árbol al modelo final. Un aprendizaje más bajo significa una contribución más pequeña de cada árbol, lo que puede ayudar a evitar el sobreajuste.

- Optimización de Parámetros: Existen diversos parámetros del modelo que pueden ser ajustados con el objetivo de mejorar la medición.

4.3. Métricas de Evaluación

En modelos de clasificación se busca etiquetar con una clase a las distintas entradas, es por esto por lo que resulta natural evaluar el comportamiento de un clasificador según qué tan bien o mal acierta al predecir cada clase. Un modelo de clasificación tiene tecnicismos que indican como fue catalogada la clase (Vujović, 2021) :

Verdadero positivo (VP): Predicción correcta que ocurre cuando una muestra de clase A fue etiquetada como clase A.

Verdadero negativo (VN): Predicción correcta que ocurre cuando una muestra distinta a la clase A fue correctamente catalogada con su respectiva clase real.

Falso positivo (FP): Predicción incorrecta que ocurre cuando una muestra de clase distinta a clase A fue catalogada como clase A.

Falso negativo (FN): Predicción incorrecta que ocurre cuando una muestra de clase de clase A fue catalogada como una de las clases distintas a la A.

Para determinar el desempeño de un algoritmo de clasificación existen 4 métricas principales:

- **Accuracy:** Tasa de aciertos considerando tanto los verdaderos positivos como los negativos sobre el total de las muestras.

$$Accuracy = \frac{VP + VN}{VP + FP + VN + FN}$$

- **Precisión:** Tasa de aciertos en las muestras etiquetadas como positivas para una respectiva clase, es decir cuantas de las muestras que fueron catalogadas como positivas realmente lo son, cuantifica la presencia de falsos positivos en el clasificador.

$$Precisión = \frac{VP}{VP + FN}$$

- **Recall:** Tasa de aciertos sobre todas las muestras de una clase real, cuantifica la presencia de falsos negativos en el clasificador.

$$Recall = \frac{VP}{VP + FP}$$

- **F1-Score:** Métrica compuesta por la precisión y el recall, es considerada para casos donde existe desbalanceo entre las distintas clases.

$$F1\ Score = \frac{2 \times Precisión \times Recall}{Precisión + Recall}$$

5. Metodología

Para la elaboración de este proyecto se usará la metodología CRISP-DM, la cual se usa normalmente en proyectos de Data Mining que tiene mucha similitud con los proyectos de Data Science. (IIC, 2021) (IBM, 2021). Usar esta metodología trae consigo múltiples beneficios, de los cuales los que más destacan son el enfoque iterativo, el cual es crucial para proyectos de Data Science ya que siempre se puede iterar para mejorar los modelos y obtener mejores resultados. Por otro lado, CRISP-DM se enfoca fuertemente en entender y comprender las necesidades de la empresa antes de realizar y planificar el proyecto, esto asegura que los resultados obtenidos tengan alineación con los objetivos estratégicos y sea útil para tomar decisiones.

Si bien existen beneficios, también existen desventajas de esta metodología, tal como lo explica (J. S. Saltz, 2021), esta metodología es débil en términos de priorización de tareas coordinación y comunicación entre equipos. En la *ilustración 8* se ve un diagrama global de la metodología.

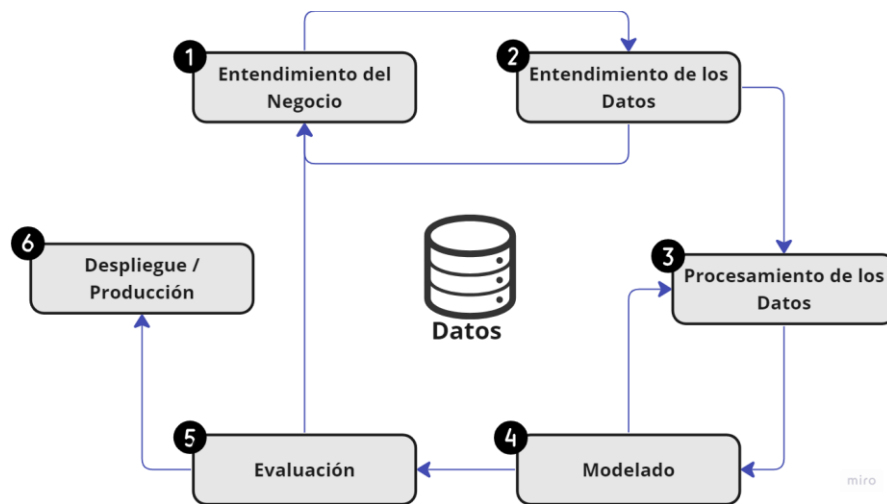


Ilustración 8: Diagrama Metodología CRISP-DM

Entendimiento del Negocio

En esta fase, se define el problema del negocio que se está tratando de resolver y se establecen los objetivos del proyecto. Esto puede incluir definir preguntas específicas que se desean responder, identificar los criterios de éxito y entender cómo el proyecto beneficiará al negocio.

Entendimiento y Análisis de los Datos

Se recopilan los datos necesarios para el proyecto y se realiza una exploración inicial para comprender la naturaleza de los datos. Esto implica recopilar datos brutos, evaluar su calidad, realizar estadísticas descriptivas y visualizaciones para comprender las características de los datos. Es aquí en donde se utilizan lenguajes de programación como Python y SQL para el manejo y análisis de los datos.

Procesamiento y preparación de los Datos

En esta etapa, los datos se limpian, transforman y preparan para su uso en el modelado. Esto puede incluir llenar valores faltantes, codificar variables categóricas, normalizar o estandarizar datos numéricos, unir bases de datos y seleccionar las variables más relevantes a incluir en el futuro modelo. Aquí se utiliza principalmente Python para realizar las preparaciones, limpieza y manejo de los datos.

Modelado

Se seleccionan y aplican técnicas de aprendizaje automático (modelado). Se entrenan varios modelos y se ajustan sus parámetros para mejorar el rendimiento. Esta fase implica iteración, donde se prueban diferentes algoritmos y configuraciones para encontrar el mejor modelo para el problema. La selección del modelo depende de que es lo que se busca predecir o clasificar, basado en esto, existen diferentes herramientas tecnológicas

que permiten hacer múltiples comparaciones de desempeño al mismo tiempo. Para la creación de este modelo, se utilizó una librería de Python llamada Scikit-Learn.

Evaluación

Los modelos entrenados se evalúan en datos no vistos para medir su rendimiento. Se utilizan métricas de evaluación adecuadas para el tipo de problema que se está abordando. Si el modelo no cumple con los requisitos, el proceso puede regresar a la etapa de modelado para hacer ajustes y mejorar el rendimiento. Para este paso también se usa la librería de Python mencionada anteriormente.

Despliegue

En esta fase, el modelo seleccionado se implementa en el entorno de producción. Esto implica integrar el modelo en los sistemas existentes para que pueda tomar decisiones en tiempo real. Esta fase de la metodología no se aplicará, debido a que esta fuera de los alcances definidos para el proyecto.

Ciclo Iterativo

Los proyectos que tienen como objetivo crear un modelo, por naturaleza son iterativos, ya que siempre pueden aparecer nuevas variables importantes que añadir o también mejorar constantemente su desempeño. Si bien esta fase es la última, es importante notar que la iteración es aplicada a lo largo de todo el proyecto con el objetivo de obtener el mejor resultado posible.

6. Desarrollo

6.1. Comprensión del Negocio

En esta fase inicial del proyecto, es esencial profundizar en la comprensión del problema del negocio que se busca resolver: la detección anticipada de posibles fallas en la red de telecomunicaciones de la empresa Entel. La infraestructura de red de Entel desempeña un papel crítico al ser el sustento de todos los servicios ofrecidos a sus clientes, ya sean particulares o empresas.

La importancia de abordar eficazmente este problema radica en los riesgos financieros y de reputación asociados con las fallas en la red. En términos económicos, las pérdidas potenciales derivadas de una hora de inactividad pueden ascender a cifras significativas, alcanzando hasta 250 mil millones de pesos. Estas cifras no solo representan un riesgo financiero directo, sino que también subrayan la urgencia de implementar un sistema eficiente de detección temprana para poder tomar medidas de prevención anticipadas.

Además de las implicaciones económicas, una falla en la red puede tener un impacto directo en la percepción de la empresa por parte de sus clientes. La confiabilidad y la disponibilidad de los servicios de telecomunicaciones son fundamentales en un mercado altamente competitivo. La reputación de Entel y la fuga de clientes está intrínsecamente ligada a su capacidad para mantener operativa su red de manera consistente y confiable.

En este contexto, la creación de un modelo de aprendizaje supervisado se presenta como una solución estratégica. Este modelo no solo buscará prever fallas potenciales, sino que también permitirá implementar medidas proactivas para minimizar el tiempo de inactividad de la red y las consecuencias de las fallas en esta. Al lograr esto, el proyecto no solo cumple con su objetivo técnico, sino que contribuye directamente a la viabilidad y la sostenibilidad del negocio. El éxito del proyecto radica en los estándares de la empresa respecto a los modelos predictivos, esto significa que el modelo pueda predecir anticipadamente alrededor de un 60% de las posibles fallas e interrupciones en la red Core.

6.2. Entendimiento y Análisis de los Datos

6.2.1. Pre-Procesamiento de los Datos

Como bien se mencionó en el apartado 2.7 se tienen 3 fuentes de datos: las alarmas en los nodos, los indicadores de rendimiento y las conexiones de los nodos. Los datos de rendimiento y las conexiones se almacenan en un servidor específico al cual se puede acceder generando una solicitud especial. Los datos de alarmas están almacenados en AWS (la nube de la empresa) y se puedan acceder de forma fácil mediante una herramienta de la nube llamada S3 (herramienta que sirve para almacenar datos).

Estos datos se encontraban almacenados en un formato "log"², es decir, toda la información estaba condensada haciendo que fuese muy difícil trabajar directamente con ella. Para resolver este problema se tuvo que realizar un pre-procesamiento de estos "logs" el cual fue bastante complejo debido a la poca estandarización que presentaban los datos. Para realizar el procesamiento y abordar su complejidad se utilizaron múltiples programas y plataformas informáticas, como SQL, Python, AWS y VS Code principalmente. En el Anexo A se puede ver una imagen de los datos en su formato de origen.

Es importante notar que parte del preprocesamiento es limpiar la base de datos, es decir, quitar valores duplicados, quitar filas con nulos, filas con valores atípicos que no tienen relevancia para el proyecto y definir en una primera instancia las variables mas

² Los logs o registros en el sector IT hacen referencia a los archivos de texto en los que se incluyen de forma cronológica los acontecimientos como cambios, actualizaciones y demás que han ocurrido dentro de un sistema informático.

importantes y representativas de cada base de datos. Esto último, se hizo gracias a un estudio preliminar de la información.

Luego del preprocesamiento y una ardua limpieza de la información se crearon 3 bases de datos con las siguientes características:

Base de Datos	Filas	Variables Relevantes (Columnas)
Alarmas	1,1 millones	10
Indicadores de Rendimiento	19 millones	9
Conexión entre nodos	25	7

Tabla 1: Características Bases de Datos

6.2.2. Entendimiento de las Variables

Gracias al conocimiento experto de los trabajadores de la gerencia de Analytics y reuniones con la gerencia de Operaciones y Mantenimiento se validó que la información recopilada es suficiente para poder crear una primera iteración del modelo. Sin embargo, esto no descarta futuros ajustes y mejoras continuas con el fin de obtener mejores métricas de evaluación y por ende un mejor desempeño.

Después de haber llevado a cabo el proceso de preprocesamiento de las bases de datos, resultó fundamental comprender la naturaleza y el significado de las variables existentes. Este entendimiento era esencial para determinar si la información recopilada era adecuada para llevar a cabo el proyecto. En el Anexo B, C y D, se encuentran un diccionario detallado que describe cada variable en cada una de las bases de datos.

Por otra parte, es fundamental valorar la calidad de los datos adquiridos. En este sentido, se llevaron a cabo revisiones de los datos con el objetivo de garantizar su precisión, su actualización adecuada y constante, su fácil accesibilidad e integridad. Con el propósito de realizar estas revisiones, el estudiante participó en reuniones con el equipo de Big Data para contrastar la información. Además, llevó a cabo análisis de manera independiente. Como resultado de estas evaluaciones, se llegó a la conclusión de que los datos exhiben una buena calidad y son aptos para su utilización en el proyecto.

7.2.3. Análisis descriptivo de los datos

Luego de generar conocimiento de que significa cada variable, es importante realizar análisis descriptivos para entender en profundidad el comportamiento de la información, como, por ejemplo, la temporalidad, frecuencia, cantidad de opciones posibles e importancia.

El análisis descriptivo es una herramienta fundamental para comprender la información de cualquier conjunto de datos. Una vez que se ha definido el significado de cada variable, el análisis descriptivo permite identificar las siguientes características:

- **Temporalidad:** ¿Cuándo se recopilaron los datos? ¿Existe algún patrón temporal en los datos?
- **Categorías:** ¿Cuántos opciones o categorías diferentes hay para cada variable?
- **Tipo de dato:** ¿Las variables son categóricas o numéricas?
- **Importancia:** ¿Qué variables son más relevantes para el análisis?

Para analizar la temporalidad de la información, se realizaron gráficos de barras los cuales muestran la cantidad de registros por hora promedio a lo largo de un mes, de las bases de datos “Alarmas” e “Indicadores de Rendimiento”. Respecto a la base de datos “Conexiones entre nodos”, esta es información estática, es decir, no cambia durante el tiempo y como consecuencia no se puede analizar su temporalidad³.

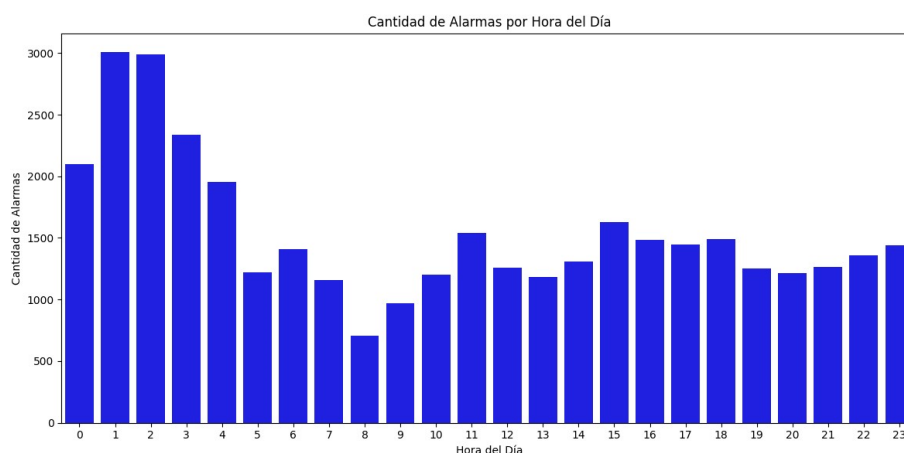


Ilustración 9: Cantidad de Alarmas por Hora del Día

³ A pesar de que las conexiones entre los nodos podrían experimentar variaciones debido a cambios en la arquitectura tecnológica, estas alteraciones son tan infrecuentes que, para los propósitos del proyecto, se asume que no habrá cambios.

En la *ilustración 9* se puede ver la cantidad de alarmas que ocurren en un día considerando todos los nodos de la Región Metropolitana, esto demuestra que existe un patrón de aparición más alto durante horas de la madrugada, esto se debe a una menor supervisión técnica durante ese rango horario, debido a la baja cantidad de personal que trabaja durante esas horas

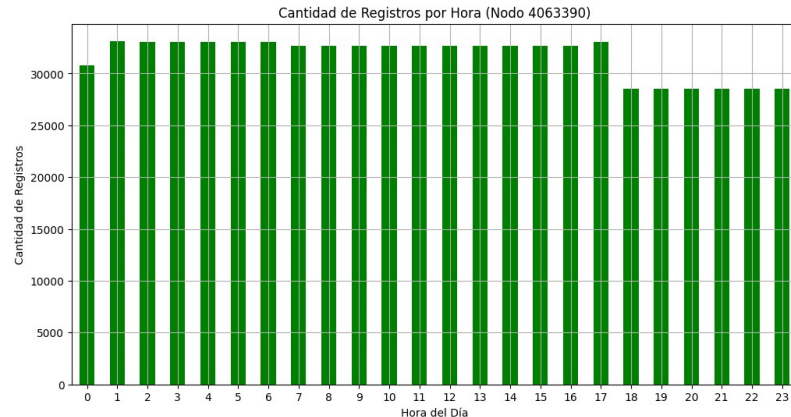


Ilustración 10: Mediciones de Rendimiento por Hora del Día

La *ilustración 10* muestra la cantidad de mediciones de rendimiento que se hacen sobre un nodo en particular a lo largo del día. Se puede notar que presenta una tendencia constante y luego de analizarlo en profundidad se descubrió que están mediciones se realizan cada 15 minutos exactos.

Luego de realizar este análisis de temporalidad, se determinó dos aspectos importantes. Primero, la creación de este proyecto y el modelo en particular, pueden ayudar significativamente a reducir la cantidad de alarmas que suceden en la madrugada dado que otorgaría un mayor control y posibilidad de anticipar los errores. Segundo, las bases de datos tienen temporalidades diferentes, es decir, dentro de un nodo suceden indicadores de rendimiento cada 15 minutos, pero las alarmas pueden suceder en cualquier momento, esto puede dificultar una futura unión de ambas bases de datos al momento de añadir la información al modelo.

Para analizar la cantidad de categorías u opciones posibles de las variables, se usaron medidas de tendencia central y un conteo simple. En las tablas 2, 3 y 4 se pueden ver las características de las variables⁴ de las tres bases de datos.

⁴ Para efectos de la creación de las tablas, no se incluyeron variables de fechas y números identificadores de nodos ni alarmas, debido a que estas variables son identificadoras y cambian dependiendo de la muestra de datos.

Variable (Alarmas)	Tipo	Cantidad de Opciones	Promedio
Tipo_error	Categórica	226	-
Clasificacion_severidad	Categórica	4	-
Estado_resolucion	Categórica	4	-
Tipo_alarma	Categórica	4	-
Duración_alarma	numérica	-	6 segundos

Tabla 2: Estadísticas Descriptivas Alarmas en Nodos

Variable (Desempeño)	Tipo	Cantidad de Opciones	Promedio
Nombre_evento	Categórica	18	-
Valor_evento	numérica		20 (dB)

Tabla 3: Estadísticas Descriptivas Indicadores Rendimiento

Variable (Conexiones)	Tipo	Cantidad de Opciones	Promedio
Origen	Categórica	24	-
Destino	Categórica	24	-
Referencia	Numérica	-	23 (dB)

Tabla 4: Estadísticas Descriptivas Conexiones Entre Nodos

De estas tablas se puede ver que variables existen en las bases de datos. Las variables expuestas son las que se consideraron relevantes para una primera iteración del modelo.

Para validar la importancia de la información se realizaron múltiples reuniones en las cuales se validó que información es importante analizar para tener un conocimiento más amplio del proyecto y sus objetivos. La información validada y recolectada es la siguiente:

- Si bien existen múltiples *Tipo_error*, existen un total de 27 errores importantes que pueden afectar las conexiones entre los nodos, estos errores se clasifican como críticos. Se determinó que son errores importantes porque de una u otra forma tienen relación con las interrupciones o cortes de enlace entre nodos y por ende afectan el funcionamiento de la red. Además, estos errores tienen relación entre sí, es decir, una alarma o error puede llegar a desencadenar otras alarmas distintas.
- La variable de "Referencia" en la base de datos "Conexiones entre nodos" indica el valor de potencia en (dB) normal que debiese existir entre nodos (origen y destino). Esta variable sirve como punto de comparación.
- Respecto a los indicadores de rendimiento, se puede notar un total de 18 eventos, es decir, cada 15 minutos se miden un total de 18 indicadores. Luego de entender en profundidad el significado de cada indicador, se determinó que los indicadores *LSOOPMIN* (Potencia de Salida) y *LSIOPMIN* (Potencia de Entrada) son de importancia, puesto que estos indican la potencia de entrada y salida mínima que se registró en el nodo destino, es decir, si estos indicadores están muy por debajo en comparación a la variable "Referencia" indicaría que la potencia no es suficiente para alimentar el enlace y por ende no estaría funcionando de manera óptima ni transmitiendo los datos de forma correcta. En casos más extremos, cuando la potencia de entrada es muy menor al estado normal, se considera que el enlace se cortó y por ende el traspaso de información por esa ruta queda inhabilitado.
- Las alarmas, ocurren generalmente cuando hay fluctuaciones en la potencia de los enlaces.

6.3. Preparación de los Datos

6.3.1. Preparación Indicadores de Rendimiento y Conexiones

En la sección anterior se detalla parte del análisis y entendimiento de los datos que se utilizarán para el proyecto. Luego de esto, es necesario procesar la información a modo de obtener solo una base de datos, integrando toda la información recolectada.

El primer paso en procesar los indicadores de rendimiento, como bien se mencionó antes es necesario considerar solo los indicadores LSIOPMIN y LSOOPMIN. Para realizar esto, se filtró la base tomando solo los registros para estos indicadores obteniendo un total de 5 millones de registros. Para calcular el valor correcto de la medición de potencia de los enlaces es necesario restar ambos indicadores, puesto que uno representa la potencia de entrada y el otro a potencia de salida. Teóricamente, la resta de ambos indicadores en un estado normal debiese ser igual a la variable de "Referencia". Tomando esto en cuenta, se creó una variable llamada "Diferencia de Potencia", la cual se origina de la siguiente operación:

$$\text{Diferencia de Potencia (dB)} = \text{Referencia} - (\text{LSOOPMIN} - \text{LSIOPMIN})$$

Esta variable indica si existió una variación de potencia alejada de su valor teórico normal, entre dos nodos. Luego de realizar estas operaciones se unió la base resultante con la base de conexiones entre los nodos para obtener el "destino" de cada conexión y agregarle la variable previamente creada. En la siguiente figura se puede ver una muestra de la base de datos resultante:

ONEID	Destino	Origen	diferencia de potencia	EndTime
6592	4063235	DBC-WDM-CM01	MPU-WDM-CM01	1.7 2023-10-11 02:15:00
75979	4063377	HCH-WDM-CM01	IDP-WDM-CM01	0.1 2023-10-11 02:15:00
77640	4063378	IDP-WDM-CM01	HCH-WDM-CM01	2.4 2023-10-11 02:15:00
84236	4063378	IDP-WDM-CM01	QLC-WDM-CM01	5.0 2023-10-11 02:15:00
84237	4063378	IDP-WDM-CM01	QLC-WDM-CM01	5.0 2023-10-11 02:15:00

Ilustración 11: Cruce Bases Indicadores de Rendimiento con Conexiones

Luego de tener esta base de datos creada, fue necesario agrupar la información por cada nodo en intervalos de 2 horas para cada origen y destino del enlace. Al momento de agrupar la información se consideró que la variable de "Diferencia de Potencia" sea el valor máximo registrado en ese intervalo, de esta forma se puede saber con mayor precisión si durante esas 2 horas ocurrieron diferencias significativas que pudieron ocasionar alarmas.

Este último supuesto es basado en que previamente se mencionó que las alarmas pueden ser originadas por fluctuaciones de potencia en los enlaces.

6.3.2. Preparación datos de Alarmas

La integración de la base de datos de alarmas fue bastante más compleja debido a que las alarmas ocurren de forma aleatoria y los indicadores de rendimiento cada 15 minutos. Luego de analizar esta problemática más en profundidad y sumado a que de la información levantada se determinó que lo importante son las secuencias de errores consecutivos que pueden desembocar en una pérdida de enlace.

Para poder juntar las alarmas con la base de datos creada en la sección anterior se decidió agrupar todas las alarmas ocurridas por cada nodo en intervalos de 2 horas, es decir, se creó una base de datos en la cual se crean variables en relación con lo que sucedió dentro de este intervalo. Para realizar esto, se crearon variables que se encargan de contar las características de las alarmas ocurridas (Las características corresponden por ejemplo a la severidad, tipo de alarma, estado de alarma, etc). Por ejemplo, esto corresponde a contar cuantas alarmas fueron críticas, cuantas afectaron al servicio móvil, cuantas alarmas fueron identificadas, etc. Todo esto, para cada intervalo de 2 horas.

Luego de hacer estas agrupaciones, se juntó esta base de datos de alarmas con la base de indicadores y conexiones, para tener toda la información integrada en intervalos de 2 horas.

6.3.3. Transformaciones Finales

Luego de tener toda la información integrada en una sola base de datos, fue necesario crear la variable que buscamos predecir. En este caso, se definió que el objetivo de predicción es un estado de alarmas críticas en las próximas 2 horas. Esta variable objetivo se llamó "Falla" y para su construcción se aplicó la siguiente regla:

$$Falla \begin{cases} \text{Importante,} & \text{Si ocurrieron alarmas criticas en las siguientes 2 horas} \\ \text{Normal,} & \text{No ocurrieron alarmas criticas en las siguientes 2 horas} \end{cases}$$

Esta variable objetivo es lo que el modelo buscaría predecir, es decir, predecir un estado en el cual un nodo pueda presentar alarmas críticas. Para que el modelo pueda aprender de lo que sucedió en el pasado y lograr predecir con cierta probabilidad lo que ocurrirá en el futuro, es importante aplicar una técnica llamada "Shifteo". Esta técnica consiste en que a cada intervalo de 2 horas se le asigna la variable "Falla" que ocurrió en las siguientes 2 horas. A continuación, una figura que ejemplifica esto:

Hora	Diff. potencia	Falla
2 pm	5 dB	Normal (t)
4 pm	7.5 dB	Importante
6 pm	1 dB	Normal



Hora	Diff. potencia	Falla
2 pm	5 dB	Importante (t + 1)
4 pm	7.5 dB	Normal
6 pm	1 dB	Normal

Ilustración 12: Ejemplo Shifteo de Variable

Esta ilustración ejemplifica la operación realizada. Al registro de las 2pm se le asigna la variable “Falla” que ocurrió a las 4pm, de esta forma el modelo puede aprender que lo que ocurrió a las 2pm género alarmas críticas a las 4pm. Esta técnica se utiliza ya que estamos tratando con datos temporales.

Al crear la variable “Falla”, se crearon dos clases diferentes, “Importante” y “Normal”. Estas clases están desbalanceadas, ya que la cantidad de registros que se marcaron que ocurrió una falla importante son 230, en cambio, los registros de un estado normal son 2700. Esto puede afectar el rendimiento de modelo, es por esto que hay que aplicar un balanceo de las clases. Esto último, se logra agregando una ponderación a la clase con menos registros dentro del modelo, de esta forma, se balancea la muestra y el modelo entiende que la clase “Importante” tiene un mayor peso dentro de los resultados.

6.4. Modelado

Antes de comenzar a crear un modelo es fundamental entender que es lo que se quiere predecir, es decir, la variable dependiente (Y) y las variables independientes o atributos (X) que ayudaran a realizar las predicciones.

En este caso Y corresponde a la variable “Falla”, la cual indica si ocurrieron alarmas criticas/importantes o si no ocurrieron alarmas importantes. Como se quiere predecir entre 2 posibles categorías o clases, se tienen que utilizar modelos de clasificación.

6.4.1. Bases de entrenamiento y de prueba

Para poder utilizar modelos de aprendizaje de máquinas, es necesario dividir la base de datos en 2 partes, la base de entrenamiento y la base de prueba. La base de entrenamiento

se utilizará para entrenar el modelo. Por otro lado, la base de prueba se utilizará para probar el modelo y determinar su desempeño real.

Como se trata de datos temporales a lo largo del mes de octubre, la división se hará de la siguiente manera:

- Base de Entrenamiento: Todo lo que ocurrió previo al 25 de octubre
- Base de Prueba: Todo lo que ocurrió después del 25 de octubre

6.4.2. Construcción del Modelo

Para la ejecución de este proyecto, se optó por implementar el modelo XGBoost (Extreme Gradient Boosting) debido a su destacado rendimiento y eficiencia. XGBoost es reconocido por su capacidad para manejar grandes conjuntos de datos, lo que resulta esencial dada la magnitud de la información a analizar en este caso. Su algoritmo de boosting y la manera en que gestiona los árboles de decisión lo hacen especialmente efectivo para prever patrones complejos y mejorar la precisión de las predicciones.

La elección de XGBoost se respalda no solo en su eficacia general en tareas de aprendizaje automático, sino también en su historial de éxito específico en proyectos internos de la empresa. Además, ha demostrado ser una opción confiable en competiciones a nivel global, ganando reconocimiento en el ámbito de la ciencia de datos. La flexibilidad de XGBoost para ajustarse a diferentes tipos de datos, tiene capacidad de manejar variables predictoras tanto numéricas como categóricas y es eficiente con clases desbalanceadas (como en este caso), son aspectos adicionales que respaldan su elección para abordar las complejidades y necesidades particulares de este proyecto.

6.4.3. Evaluación del Modelo

Para evaluar el modelo creado es necesario hacer estimaciones en la base de datos de prueba. Como el modelo no ha sido entrenado con la información de esta base de datos, es una muy buena forma de obtener un desempeño más real de sus métricas. Luego de realizar este procedimiento se lograron los siguientes resultados:

Clase	Precisión	Recall	F1 Score	Cantidad de Clases
Falla Importante	73%	57%	0,64	230
Estado Normal	96%	98%	0,97	2762

Tabla 5: Resultados del Modelo

Estos resultados se discutirán en la siguiente sección.

7. Resultados

En este proyecto, el propósito es anticipar posibles fallas o interrupciones en la red de telecomunicaciones de la empresa Entel. Con el fin de alcanzar esta meta, se diseñó un modelo de aprendizaje automático con la finalidad de prever un estado de alerta en los nodos que conforman la red Core. El modelo demostró resultados de predicción de un 73% lo que supera las condiciones de éxito establecidas de lograr un 60%, por lo tanto, se considera un modelo efectivo que puede contribuir a abordar la problemática mencionada anteriormente.

En cuanto a los principales resultados, evaluados mediante las métricas comúnmente empleadas en modelos de clasificación, se obtuvo lo siguiente:

Clase	Accuracy	Precisión	Recall	F1-Score
Falla Importante (clase 0)	90%	73%	57%	0,64
Estado Normal (clase 1)	90%	96%	98%	0,97

Tabla 6: Resultados del Modelo por Clases

Es relevante destacar que los resultados más significativos corresponden a la predicción de la clase "Falla Importante". Como se mencionó anteriormente, esta clase indica si un nodo va a experimentar alarmas críticas en las próximas 2 horas, y es precisamente lo que se busca anticipar en función de la problemática de la empresa. La interpretación de estos resultados se expone de la siguiente manera:

1. **Accuracy:** La métrica de "Accuracy" se refiere a la precisión de un modelo en clasificar correctamente los datos. En este caso, se menciona que el modelo tiene un 90% de precisión, lo que significa que acierta en la clasificación el 90% de las veces. Sin embargo, esta cifra puede ser engañosa, especialmente cuando hay un desbalance en la cantidad de datos de cada clase. Por ejemplo: un conjunto de datos donde el 90% de las instancias pertenecen a la Clase A y solo el 10% a la Clase B. Un modelo que prediga siempre la Clase A podría alcanzar un 90% de Accuracy

solo por adivinar la clase más común. Esto no hace que el modelo sea efectivo, especialmente si estamos más interesados en la Clase B.

Para evitar este sesgo, es importante observar otras métricas y no guiarse tanto por estos resultados de Accuracy.

2. **Precisión:** Es crucial, porque indica que cuando el modelo predice una posible falla, es probable que esté acertando. En este escenario, se logra una precisión del 73% en relación con las fallas importantes, lo cual es positivo basado en que para que un modelo se considere útil se tiene que superar el umbral del 60% como se explicó en los alcances del proyecto. Sin embargo, es esencial contrastar este resultado con otras métricas, ya que, en este contexto, los falsos positivos, es decir, predecir fallas que no ocurrieron, pueden conllevar a un gasto innecesario de tiempo y recursos por parte de Entel. Mas adelante se discutirá el impacto de los falsos positivos en la empresa.
3. **Recall:** Un recall del 57% indica que el modelo es capaz de capturar el 57% de todas las instancias positivas, es decir, las futuras fallas importantes. En otras palabras, hay un 43% de instancias positivas que el modelo no está identificando (predice que no ocurrirán fallas importantes, pero en realidad sí suceden). En resumen, esta métrica revela la proporción de fallas importantes que el modelo logró clasificar en comparación con el total de fallas importantes que realmente ocurrieron.
4. **F1-Score:** Esta métrica es particularmente útil dada la desigualdad en la distribución de las clases y sirve para contrastar la precisión y el recall del modelo. Principalmente se utiliza para comparar modelos; en este caso, si un modelo futuro obtiene un F1-Score más alto, indica que se minimizarían los falsos positivos y falsos negativos, generando predicciones más precisas.

La evaluación del modelo reveló resultados prometedores, destacando una precisión del 73% en la predicción de fallas importantes. A pesar de la desigualdad en la distribución de clases, el modelo demostró una capacidad sólida para distinguir entre estados normales y situaciones críticas, como se evidenció en el F1-Score de 0,64 para fallas importantes.

Al momento de analizar los resultados y ver la congruencia de estos, es importante ver cuáles fueron las variables que el modelo considero más importantes. En la siguiente figura se muestran las 6 variables más importantes:

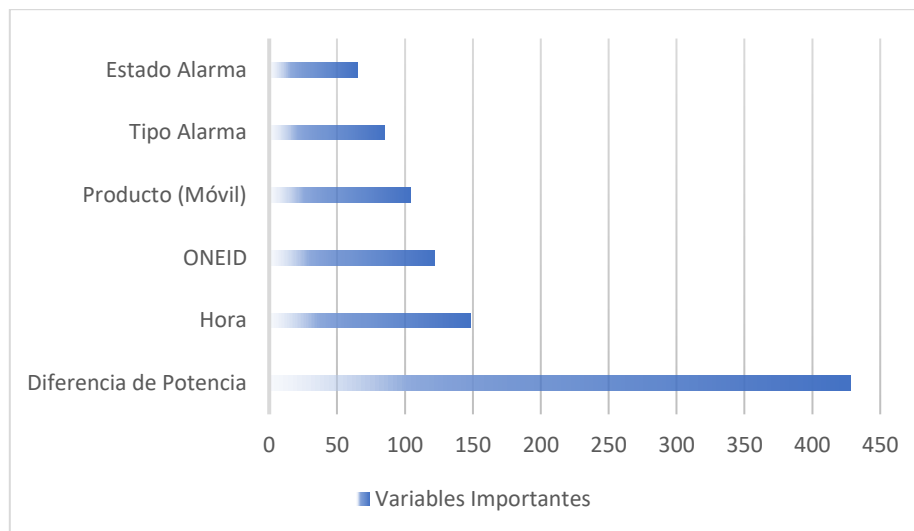


Ilustración 13: Variables Importantes del Modelo

La *ilustración 14* muestra cuales fueron las variables más importantes para los resultados del modelo. Este grafico demuestra cuanto pesa cada variable al resultado, se puede notar que la variable “Diferencia de potencia” juega un rol crucial en los resultados dado que aporta bastante. La variable de “Hora” también es relevante porque estamos tratando con datos temporales. Por último, analizar cuáles fueron las variables más relevantes es una muy buena forma de ver que el modelo creado tenga sentido, en este caso, que el modelo considere estas variables como importante demuestra consistencia con la información levantada y los procesos de investigación.

En general, el modelo ha arrojado resultados positivos, cumpliendo con el objetivo de predicción. Para la gerencia de Operaciones y Mantenimiento, la implementación de este modelo implica estar alerta 2 horas antes de que ocurra una falla en un nodo, permitiendo coordinar a un equipo técnico que pueda realizar mantenimiento anticipado en el lugar físico. Asimismo, este modelo puede aplicarse para organizar y optimizar rutas de mantenimiento a lo largo de la red, buscando una sinergia entre la reducción de fallas y las rondas de mantenimiento técnico.

En determinadas instancias, el modelo podría generar falsos positivos, dando lugar a la movilización de técnicos y la asignación de recursos para abordar una falla inexistente. Tras dialogar al respecto con la gerencia de Operaciones y Mantenimiento, se ha establecido que el costo asociado a un falso positivo es de aproximadamente \$900.000 pesos chilenos. No obstante, esta cifra no representa un inconveniente significativo para la empresa, ya que es considerablemente inferior al impacto que podría tener una hora de interrupción en la red esto último evaluado en \$250 millones de pesos chilenos. En este sentido, se destaca que el cliente interno considera que la verdadera prioridad radica en mejorar la situación actual, evidenciando y validando así la mejora del 73%

7.1. Limitaciones y Espacios de Mejora

Este enfoque se encuentra centrado en la Región Metropolitana, donde el traslado de un extremo a otro requiere aproximadamente 2 horas. No obstante, en regiones más alejadas del país, como el norte, los equipos de mantenimiento necesitarían más de 6 horas para llegar a su destino o a nodos ubicados en lugares más remotos. Esta situación presenta una oportunidad de mejora notable, al intentar extrapolar el modelo a diferentes regiones del país, considerando las distancias de desplazamiento de los equipos técnicos.

Para alcanzar este objetivo, es fundamental contemplar una ventana de predicción más amplia, es decir, permitir que el modelo pueda prever situaciones hasta con 6 horas de antelación, asegurando siempre métricas de evaluación satisfactorias. Lograr esto implica la incorporación de nuevas fuentes de datos, la comprensión de las redes presentes en otras regiones del país, la adaptación de los parámetros del modelo, la exploración de otros tipos de modelos, o incluso la inclusión de más datos históricos. Estas medidas permitirán obtener resultados óptimos y generar un valor significativo para las operaciones en las zonas norte y sur del país.

7.2. Futuras Líneas de Investigación

En futuras etapas de desarrollo del modelo, podría resultar provechoso incorporar datos adicionales y mejorar la precisión respecto al tipo específico de falla y los recursos necesarios para el mantenimiento en un nodo. Al hacerlo, al despachar a los equipos de mantenimiento, podrían determinar con exactitud las herramientas y repuestos requeridos. Asimismo, se sugiere llevar a cabo una investigación más detallada sobre las discrepancias de potencia entre los enlaces de los nodos (actualmente asumidas como eventos aleatorios) para obtener una comprensión completa de la problemática. Si se identifica alguna relación importante con otra variable, esto podría fortalecer el modelo.

En el proceso de implementar este modelo en la producción, se podría considerar evaluar la integración con un sistema de alertas móviles o algún software interno que notifique a los usuarios en tiempo real ante la detección de una posible falla importante, mejorando así el tiempo de respuesta de los equipos de mantenimiento.

Por último, es esencial destacar que, dentro de las áreas operativas de cualquier empresa, existen diversas oportunidades para aprovechar la tecnología con el objetivo de optimizar los procesos. Esta sinergia genera un valor sustancial en las empresas, como una toma de decisiones más fundamentada, reducción de costos, mayor eficiencia y una gestión mejorada. Se propone, para futuras iteraciones o proyectos vinculados a datos y análisis avanzados, explorar el ecosistema de operaciones, ya que constituye una excelente manera de generar un valor significativo dentro de las empresas.

8. Conclusiones

En conclusión, este proyecto de investigación se enfocó en desarrollar e implementar un modelo de aprendizaje automático destinado a prever fallas críticas en la red de telecomunicaciones de Entel. A través de un análisis exhaustivo de datos de alarmas, indicadores de rendimiento y conexiones entre nodos, se identificaron patrones y relaciones cruciales que sirvieron de base para la creación de un modelo predictivo.

Este enfoque no solo proporciona una anticipación efectiva de las fallas, sino que también permite una planificación proactiva de mantenimiento. La capacidad de alertar hasta dos horas antes brinda a los equipos técnicos la oportunidad de intervenir antes de interrupciones significativas en los nodos, mejorando así la eficiencia operativa y la calidad del servicio.

Es fundamental destacar que, a pesar de los resultados alentadores, el modelo presenta ciertas limitaciones. La disparidad en la distribución de clases y la necesidad constante de ajustes ante cambios en el entorno son consideraciones importantes durante la implementación práctica. Para abordar el desbalance, se aplican técnicas estadísticas que asignan ponderaciones a la clase que indica fallas importantes. De esta manera, se logran métricas ligeramente mejoradas y se resuelve este desafío específico.

Para determinar los ajustes continuos necesarios en el modelo, resulta crucial recopilar información que contraste el rendimiento teórico del modelo con su implementación real. Este enfoque permite identificar diversas áreas de mejora, como la adición de variables, la optimización de parámetros del modelo, la exploración de otros modelos, entre otros. Estas medidas buscan potenciar aún más el valor que el modelo puede generar para la empresa.

Aunque la ejecución del modelo fue apropiada, es crucial resaltar que las fases iniciales de exploración resultaron notablemente lentas. Esta demora se atribuye a la falta de orientación hacia el cliente interno en la gerencia de Analytics, aspecto crucial para comprender las verdaderas necesidades y problemáticas. A menudo, se enfatiza la adopción de tecnologías vanguardistas y la aplicación de algoritmos avanzados, pero se descuida las definiciones precisas de las problemáticas, la evaluación del verdadero valor que los proyectos pueden agregar y el entendimiento en profundidad de las otras áreas de negocio.

A pesar de resultados positivos en el proyecto y el desarrollo del modelo, es fundamental reconocer limitaciones y áreas de mejora. La falta de implementación impide medir el impacto real, limitando la comparación con resultados teóricos. Esta comparación es vital para identificar áreas de mejora y planificar iteraciones futuras del modelo.

Una implementación exitosa de este modelo podría resultar en una reducción del 73% en las fallas de la red, siempre y cuando los equipos técnicos sean capaces de responder adecuadamente a las señales proporcionadas por el modelo. Basándonos en este rendimiento, se estima que los riesgos financieros disminuirán en aproximadamente un 73%, pasando de 250 millones por hora de falla a un total de 67,5 millones. Adicionalmente, la capacidad de reducir la incidencia de fallas significativas que impactan los servicios de la empresa tiene un efecto indirecto en la satisfacción del cliente. Este resultado positivo puede mejorar la retención de clientes, considerando que aproximadamente el 30% de los clientes que abandonan lo hacen debido a problemas con la red. Además, este logro podría incluso tener un efecto positivo en la atracción de nuevos clientes debido a una mejora en la reputación de la empresa otorgada por la capacidad de mantener la red y los servicios en constante funcionamiento. En última instancia, el valor radica en la eficiencia operativa que el modelo proporciona, posicionando a la empresa para mantenerse competitiva en un entorno dinámico y demandante.

Bibliografía

- (SUBTEL), S. d. (2023). *Sector telecomunicaciones primer trimestre 2023*. Obtenido de https://www.subtel.gob.cl/wp-content/uploads/2023/06/PPT_Series_MARZO_2023_V0.pdf
- Amazon. (2022). *AWS*. Obtenido de <https://aws.amazon.com/es/what-is/digital-transformation/#:~:text=La%20transformaci%C3%B3n%20digital%20es%20el,ofrece%20valor%20a%20los%20clientes>.
- Chauhan, K. (2022). "Automated Machine Learning: The New Wave of Machine Learning,". *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*.
- Chen, T. &. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785-794.
- Consulting, K. (s.f.). *kaitsconsulting*. Obtenido de <https://www.kaitsconsulting.com/ventajas-de-ser-una-organizacion-data-driven/>
- Entel. (2022). *Memoria Integrada Entel 2022*. Obtenido de https://entel.modyocdn.com/uploads/a01e1ba5-7821-4e6f-ac82-4c4521107583/original/230419_Entel_Memoria_2023_Libro_interactiva.pdf
- Entel. (2023). *Información Corporativa*. Obtenido de <https://informacioncorporativa.entel.cl/nuestra-compa%C3%B1ia>
- Equifax-Diario Financiero. (2023). *Las claves de empresas Data Driven*. Obtenido de <https://equifax.diariofinanciero.cl/data-driven-las-claves-que-explican-por-que-este-enfoque-es-mas-relevante-que-nunca-en-el-mundo-corporativo/>
- IBM. (2021). *CRISP DM Help Overview*. Obtenido de <https://www.ibm.com/docs/zh/spss-modeler/18.0.0?topic=dm-crisp-help-overview>
- IIC. (2021). *La metodología CRISP-DM en ciencia de datos*. Obtenido de <https://www.iic.uam.es/innovacion/metodologia-crisp-dm-ciencia-de-datos/>
- J. S. Saltz. (2021). CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps,". *IEEE International Conference on Big Data*.
- Kenley, R., & Seppänen, O. (2010). *Location-Based Management for Construction: Planning, Scheduling and Control*. Abingdon, Inglaterra: Spon Press.

- Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*,. 3-24.
- Larocca, N. (2019). *Chile presenta la mejor relación precio/velocidad en datos móviles de la región*. Obtenido de <https://www.telesemana.com/blog/2019/03/26/chile-presenta-la-mejor-relacion-precio-velocidad-en-datos-moviles-de-la-region/>
- Solutions, N. (Agosto de 2023). Obtenido de <https://www.netsolutions.com/insights/cost-of-digital-transformation/#:~:text=A%20recent%20study%20from%20International,on%20complete%20digital%20transformation%20projects>.
- Statista. (2017). *PIB por sector económico Chile 2023*. Obtenido de <https://es.statista.com/estadisticas/1285944/participacion-de-las-actividades-economicas-en-el-pib-de-chile/>
- SUBTEL. (2014). Obtenido de Capítulo Telecomunicaciones: https://www.subtel.gob.cl/images/stories/apoyo_articulos/notas_prensa/cta_publica_2010_2014/capitulo_telecomunicaciones_06012014.pdf
- Vujović, Ž. (2021). Classification model evaluation metrics. *International Journal of Advanced Computer Science and Applications*. 599-606.
- XGBoost. (2022). Obtenido de Kaggle: <https://www.kaggle.com/code/dansbecker/xgboost/noteb>

ANEXOS

Anexo A

	<code>_raw</code>
0	<code>",2023/8/17 09:59:35 GMT-04:00,322207324,,0</code>
1	<code>346358821,"ONE=4063255,NE=3145769,FR=0,S=8,CP=...</code>
2	<code>346358813,"NE=167772968,S=17,CP=-1,PP=16 1 ,F...</code>
3	<code>Alarm Parameter(hex) 0xff 0xff 0xff 0xff" ...</code>
4	<code>346358703,"NE=3146038,S=11,CP=-1,PP=0 65535 ,...</code>
...	...
1831804	<code>Alarm Parameter(hex) 0x01 0x00 0x01 0xff 0xff" ...</code>
1831805	<code>343487017,"NE=3150462,S=6,CP=-1,PP=1 1 ,FB=1 ...</code>
1831806	<code>343486955,"NE=3155888,S=3,CP=-1,PP=16 1 ,FB=9...</code>
1831807	<code>343486870,"NE=167775157,S=60,CP=-1,PP=16 1 ,F...</code>
1831808	<code>343486889,"NE=167775930,S=6,CP=-1,PP=16 1 ,FB...</code>
1831809	rows × 1 columns

Anexo A: Formato origen de los datos extraídos del servidor

Anexo B

Variable (BD Alarmas)	Explicación
ONEID	Numero identificador del sito o nodo
Nombre_nodo	Nombre técnico del nodo (sitio físico) en donde se encuentra el subnodo
Nombre_subnodo	Nombre del subnodo en el cual ocurrió la alarma
Alarm_name	Nombre del error que origino la alarma
type	Que tipo de alarma es
Producto_name	A que tipo de servicio afecto la alarma
severidad	Clasificación de severidad de alarma
status	Estado de la alarma o error (resuelta o no)
Fecha_inicio	Fecha exacta cuando se inició la alarma
Fecha_fin	Fecha cuando se resolvió la alarma

Anexo B: Diccionario Variables Base de Datos Alarmas

Anexo C

Variable (BD Indicadores)	Explicación
ONEID	Numero identificador del sitio o nodo
NEID	Numero identificador del nodo
Nombre_nodo	Nombre del nodo
Nombre_subnodo	Nombre del sub-nodo en el cual se hizo la medición de desempeño
Numero_shelf	Sección del subnodo en la cual esta ubicado el equipo al cual se le midió desempeño
Numero_slot	Puerto del Shelf en la cual está ubicado el equipo al cual se le midió desempeño
Nombre_Evento	Nombre de los indicadores de desempeño que se midieron
Valor_Evento	Valor en (dB)(Potencia) de las mediciones de desempeño
Fecha_finalizacion_evento	Fecha exacta cuando se midió el desempeño del equipo

Anexo C: Diccionario base de datos Indicadores de Desempeño

Anexo D

Variable (Conexiones entre nodos)	Explicación
ONEID	Numero identificador del sitio o nodo
NEID	Numero identificador del nodo
origen	Nombre del nodo de origen de la conexión
destino	Nombre del nodo de destino de la conexión
SheldID	Numero de Shelf de la conexión
BoardID	Puerto del Shelf en el cual sale la conexión
Referencia	Valor en (dB) que indica el estado normal de la conexión (Potencia adecuada)

Anexo D: Diccionario de variables base de datos Conexión entre Nodos