



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

DESARROLLO DE UNA METODOLOGÍA PARA
PREDECIR LA SATISFACCIÓN DE CLIENTES
POSTPAGO DE UNA COMPAÑÍA DE
TELECOMUNICACIONES USANDO VARIABLES
DE RED MÓVIL MEDIANTE LA
IMPLEMENTACIÓN DE LA METAHEURÍSTICA
HARMONY SEARCH

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL
INDUSTRIAL

GONZALO HERNÁN SERRADILLA WERNER

PROFESOR GUÍA:
BLAS DUARTE ALLEUY

INTEGRANTES DE LA COMISIÓN:
ROCÍO RUIZ MORENO
JOSÉ NALDA REYES

SANTIAGO DE CHILE
2024

**DESARROLLO DE UNA METODOLOGÍA PARA PREDECIR LA SATISFACCIÓN
DE CLIENTES POSTPAGO DE UNA COMPAÑÍA DE TELECOMUNICACIONES
USANDO VARIABLES DE RED MÓVIL MEDIANTE LA IMPLEMENTACIÓN DE
LA METAHEURÍSTICA HARMONY SEARCH**

El presente informe detalla el desarrollo de una metodología para predecir la satisfacción de clientes postpago de una compañía de telecomunicaciones, utilizando variables de red móvil y la implementación de la metaheurística Harmony Search. El proyecto, llevado a cabo en la Vicepresidencia Mercado Personas de Entel Chile, se centró en mejorar la precisión del modelo de predicción de satisfacción, el cual se basa en variables críticas como RSRQ y RSRP, que miden la calidad y potencia de la señal de red recibida por los clientes.

El desarrollo del proyecto se estructuró en varias etapas, comenzando con la implementación del algoritmo Harmony Search para optimizar los parámetros de temporalidad y umbral que maximicen el nivel de predicción de la satisfacción general, medido a través del Information Value (IV). Se recopiló y procesó la data de sesiones de red móvil y resultados de encuestas de satisfacción para utilizarlos como insumo en el algoritmo. Posteriormente, se aplicó la metodología a las variables RSRQ y RSRP, evaluando las mejoras obtenidas en el modelo de predicción actual al incluir estas variables procesadas.

Los resultados del proyecto mostraron mejoras significativas en los valores de IV de las armonías optimizadas, lo que refleja una mayor capacidad predictiva de las variables RSRQ y RSRP para determinar la satisfacción del cliente. La incorporación de estas variables en el modelo de predicción actual resultó en un aumento notable en las métricas ROC AUC y en las curvas Lift, destacándose una mejora en el primer quintil, donde el valor de Lift ascendió de 1.9 a 2.25, demostrando una mayor eficacia en la identificación de clientes detractores en comparación con el modelo base.

En la discusión se abordaron las limitaciones del estudio, tales como la representatividad de los datos y los posibles sesgos introducidos por la metodología. Además, se discutió la sostenibilidad y el mantenimiento del modelo a largo plazo, destacando la importancia de gestionar adecuadamente las actualizaciones del modelo y de considerar las barreras tecnológicas y operativas que podrían surgir en su implementación dentro de la organización.

En conclusión, la metodología desarrollada ha demostrado ser efectiva en la optimización de las variables de red móvil para predecir la satisfacción del cliente. A futuro, se recomienda validar los resultados en diferentes períodos y contextos, así como integrar las variables procesadas en el modelo predictivo de producción para asegurar su aplicabilidad y eficacia a largo plazo. Además, se sugiere explorar mejoras metodológicas, como considerar combinaciones de días no consecutivos, para identificar periodos críticos que afecten la satisfacción del cliente de manera más precisa, maximizando así el impacto comercial directo de la metodología, mejorando la retención de clientes y la eficiencia operativa de la compañía.

A mi madre que es incondicional, mi padre que me guía desde el cielo y mi hermano que me acompaña día tras día.

Agradecimientos

Llegar a esta etapa de mi vida es un momento emocionante y lleno de satisfacción. A pesar de las diversas dificultades que he enfrentado a lo largo de mi vida escolar y universitaria, estar ad portas de obtener mi título como ingeniero civil industrial me llena de orgullo y satisfacción personal. Sin embargo, sería injusto atribuirme este logro únicamente a mí mismo. Por ello, quiero aprovechar esta instancia para expresar mi más profundo agradecimiento a quienes han sido fundamentales en mi camino, tanto en lo personal como en lo académico.

En primer lugar, quiero agradecer a mi madre. A pesar de las dificultades y pérdidas que ha tenido que afrontar, ella es un ejemplo de sacrificio, determinación y amor incondicional. Es tan responsable como yo de que haya llegado a este momento en mi vida, y me faltarían palabras para expresar todo el agradecimiento que siento hacia ella.

En segundo lugar, quiero agradecer a mi padre. Aunque no está físicamente conmigo, su ejemplo me ha guiado a lo largo de este camino. Sin duda, espero que, dondequiera que esté, se sienta orgulloso de mis logros.

A mi hermano, mi pilar y compañero de vida, le agradezco profundamente por su constante apoyo. A Lalo, Pame y mi prima, que han estado a mi lado en las buenas y en las malas, siempre brindándome su cariño y apoyo, les estoy infinitamente agradecido.

A mis amigos de la vida, quienes me alegran constantemente y me hacen parte de un grupo donde la amistad verdadera y el amor fraternal existen, les agradezco por su constante presencia y por ser parte esencial de mi vida.

Quiero agradecer también a Blás Duarte, quien ha demostrado una disposición y generosidad invaluable durante el proceso de desarrollo de esta memoria, guiándome con su experiencia y sabiduría.

Finalmente, agradezco a todas las personas que, a lo largo de mi carrera, han aportado a mi crecimiento, tanto en lo académico como en lo laboral y lo personal. A todos aquellos que, desinteresadamente, han contribuido con un granito de arena a mi vida, les expreso mi más sincero agradecimiento por ser parte de este viaje.

Tabla de Contenido

| | |
|---|-----------|
| 1. Antecedentes generales: sector de telecomunicaciones y compañía Entel | 1 |
| 1.1. Servicios y estadísticas del sector de telecomunicaciones en Chile | 1 |
| 1.2. Entel | 1 |
| 1.2.1. Acerca de Entel | 1 |
| 1.2.2. Historia y principales hitos de la compañía | 2 |
| 1.2.3. Vicepresidencia Mercado Personas, Equipo de Experiencia y Satisfacción del Cliente | 2 |
| 2. Proyecto | 4 |
| 2.1. Descripción del problema u oportunidad | 4 |
| 2.2. Impacto indirecto a partir del desarrollo del proyecto | 5 |
| 2.3. Descripción y justificación del proyecto | 6 |
| 3. Objetivos | 7 |
| 3.1. Objetivo General | 7 |
| 3.2. Objetivos específicos | 7 |
| 4. Alcances | 8 |
| 4.1. Limitantes del proyecto | 8 |
| 4.2. Enfoque de solución escogido | 9 |
| 4.3. Metas e indicadores de éxito | 10 |
| 4.4. Objetivos y resultados fuera del alcance del proyecto | 10 |
| 5. Marco Conceptual | 12 |
| 5.1. Metodologías, algoritmos y métricas de predicción | 12 |
| 5.1.1. Metodología de desarrollo de proyectos Analytics Entel | 12 |
| 5.1.2. Proceso ETL | 12 |
| 5.1.3. Metaheurísticas de Optimización | 12 |
| 5.1.4. Metaheurística Harmony Search | 13 |
| 5.1.5. Métricas para evaluar capacidad predictiva de una variable | 14 |
| 5.2. Conceptos | 15 |
| 5.2.1. Variables de red móvil | 15 |
| 5.2.2. Encuestas de satisfacción | 15 |
| 5.2.3. Amazon Web Services (AWS) | 15 |
| 5.2.4. Amazon Elastic Compute Cloud (Amazon EC2) | 16 |
| 5.2.5. Amazon Simple Storage Service (Amazon S3) | 16 |
| 5.2.6. Amazon Athena | 16 |
| 5.2.7. Information Value | 16 |
| 5.2.8. Relación entre IV y WoE | 17 |
| 5.2.9. Modelo de predicción | 17 |

| | |
|---|-----------|
| 5.2.10. Información de sesiones de los clientes | 17 |
| 5.3. Disciplinas de la Ingeniería Civil Industrial relacionadas con el proyecto | 17 |
| 5.3.1. Data Science (Ciencia de datos) | 17 |
| 5.3.2. Big Data | 18 |
| 5.3.3. Data Analytics | 18 |
| 5.3.4. Machine Learning | 18 |
| 5.4. Lenguajes de programación | 18 |
| 5.4.1 Python y Visual Studio Code | 18 |
| 5.4.2. Structured Query Language | 19 |
| 6. Metodología | 20 |
| 6.1. Fase 1: Investigación | 20 |
| 6.1.1. Etapa de Ideación | 20 |
| 6.1.2. Etapa de Prueba de Concepto (POC) | 20 |
| 6.2. Fase 2: Desarrollo del proyecto | 21 |
| 7. Desarrollo | 22 |
| 7.1. Selección e ingesta de datos | 22 |
| 7.2. Análisis exploratorio de los datos | 22 |
| 7.3. Preprocesamiento | 22 |
| 7.4. Elaboración de metodología | 23 |
| 7.4.1. Versión inicial (primera iteración) | 23 |
| 7.4.2. Versión final (segunda iteración) | 24 |
| 7.5. Ejecución de metodología | 24 |
| 7.5.1. Ejecución versión inicial (primera iteración) | 24 |
| 7.5.2. Ejecución versión final (primera iteración) | 25 |
| 7.6. Incorporación de variables procesadas en modelo de predicción actual | 25 |
| 8. Resultados | 28 |
| 8.1. Aplicación metodología primera iteración | 28 |
| 8.2. Aplicación metodología segunda iteración (final) | 29 |
| 8.2.1. Variable RSRQ | 29 |
| 8.2.1.1. Tablón 1 | 29 |
| 8.2.1.2. Tablón 2 | 31 |
| 8.2.2. Aplicación metodología a variable RSRP | 32 |
| 8.2.2.1. Tablón 1 | 32 |
| 8.2.2.2. Tablón 2 | 33 |
| 8.3. Pruebas en modelo de predicción actual | 34 |
| 8.3.1. Binarización óptima | 34 |
| 8.3.2. Resultados de predicción modelos | 35 |
| 9. Discusión | 38 |
| 10. Conclusiones | 40 |

| | |
|--|-----------|
| 11. Bibliografía | 42 |
| 12. Anexos | 45 |
| Anexo A: Principales servicios del sector de telecomunicaciones, total de usuarios, penetración internet móvil y participación de mercado de tráfico móvil por operador. | 45 |
| Anexo B: Organigrama Grupo Entel 2023 | 48 |
| Anexo C: Modelo móvil satisfacción de clientes postpago Entel | 49 |
| Anexo D: Clientes PortOut Encuestados y Churn por Motivos de Red | 50 |
| Anexo E: Importancia de variables en modelo de predicción actual (julio 2024) de la satisfacción | 51 |
| Anexo F: Metodología desarrollo de proyecto | 52 |
| Anexo G: Flujo ejecución Algoritmo harmony Search | 53 |
| Anexo G: Opt Binning y métrica WoE | 54 |
| Anexo H: Comparación de la memoria de armonías antes y después de iterar en la ejecución de la metodología sobre las variables RSRQ y RSRP | 55 |

1. Antecedentes generales: sector de telecomunicaciones y compañía Entel

1.1. Servicios y estadísticas del sector de telecomunicaciones en Chile

Los principales servicios del sector de telecomunicaciones en Chile incluyen Internet fijo, Internet móvil (3G, 4G, 5G), telefonía fija, telefonía móvil y televisión de pago. La penetración de Internet móvil (una medida que indica el número de accesos móviles por cada 100 habitantes) alcanzó los 112,2 accesos por cada 100 habitantes en septiembre de 2023, con un total de 22,55 millones de accesos móviles. De estos, 3.144.038 conexiones correspondieron a la tecnología 5G, lo que subraya la creciente importancia de las redes móviles en la conectividad nacional (Subsecretaría de Telecomunicaciones de Chile, 2023).

El crecimiento constante en el total de servicios de telecomunicaciones ha sido impulsado por la adopción de tecnologías avanzadas como 4G y 5G. Estas tecnologías han jugado un papel crucial en la mejora de la infraestructura de telecomunicaciones del país, facilitando el acceso a servicios digitales avanzados con mayores velocidades y mejor cobertura para los usuarios.

En cuanto al mercado de Internet móvil, se concentra principalmente en cuatro actores: Claro, Wom, Movistar y Entel. Según el Informe Trimestral del Sector de Telecomunicaciones del tercer trimestre de 2023 (Subsecretaría de Telecomunicaciones de Chile, 2023), Wom supera a Movistar y Claro en participación de mercado, tanto en número de conexiones como en cantidad de tráfico, aunque Entel lidera en ambas categorías.

Para más detalles sobre los distintos servicios de la industria de telecomunicaciones, la cantidad de usuarios de cada servicio, el desglose de la penetración de Internet móvil durante 2023 y la participación de mercado de conexiones y tráfico móvil por operador, consulte el [Anexo A](#).

1.2. Entel

1.2.1. Acerca de Entel

Tal cómo se declara en la Memoria Integrada Entel 2023 (Grupo Entel, 2024), Entel es una compañía con 60 años de trayectoria, destacada como una de las mayores sociedades anónimas que cotizan en la Bolsa de Santiago y líder en el área de tecnología y telecomunicaciones. Sus operaciones se extienden a Chile y Perú, donde cuenta con más de 20,2 millones de abonados móviles y ha registrado ingresos anuales por CLP 2.573.142 millones en 2023.

A lo largo de estos casi 60 años, Entel ha evolucionado con el objetivo de no solo conectar a las personas, sino también acercar la tecnología y habilitar su uso para aprovechar las posibilidades que esta brinda en la vida cotidiana. La misión de Entel es fomentar la conectividad, habilitando su uso y acercando las infinitas posibilidades de la tecnología a la vida diaria de las personas.

La visión de Entel se centra en que la tecnología tiene el potencial de mejorar positivamente la vida diaria de las personas, favoreciendo el desarrollo socioeconómico, la productividad y la sostenibilidad de las empresas y sus colaboradores. La empresa se esfuerza por transformar y mejorar la sociedad en la que vivimos, contribuyendo de manera responsable y sostenible.

Entel ofrece servicios de conectividad móvil y fija, junto con una amplia gama de soluciones digitales e informáticas para personas y empresas de todos los tamaños. Su estrategia busca entregar experiencias simples, eficientes y sorprendentes, respaldadas por una red robusta de última generación y una sólida imagen de marca.

1.2.2. Historia y principales hitos de la compañía

Entel fue fundada poco después del terremoto que afectó a Chile en 1960, con el objetivo de dotar al país de un sistema de telecomunicaciones moderno, seguro y resiliente. En 1964, se creó la Empresa Nacional de Telecomunicaciones (Entel) para asegurar la continuidad de las comunicaciones y proporcionar conectividad de larga distancia internacional. Durante las décadas siguientes, Entel realizó importantes avances tecnológicos, como la digitalización de su red en 1987, el despliegue de fibra óptica en 1990 y el lanzamiento del servicio de telefonía móvil PCS en 1997.

En 2000, Entel se expandió a Perú bajo la marca Americatel y continuó innovando con el lanzamiento de la red 3G en 2006 y la obtención de espectro para redes LTE 4G en 2012. La adquisición de Nextel Perú en 2013 consolidó su presencia en el país. En 2021, Entel obtuvo espectro para desarrollar tecnología 5G, y en 2023 completó el despliegue de la red 5G en Chile, lanzando también Entel Digital. Esta evolución demuestra el compromiso de Entel con la innovación y el desarrollo tecnológico, consolidándose como líder en telecomunicaciones en la región. (Grupo Entel, 2024)

1.2.3. Vicepresidencia Mercado Personas, Equipo de Experiencia y Satisfacción del Cliente

Entel Chile está estructurada en diversas vicepresidencias (VP) que reportan directamente al gerente general del Grupo Entel, tal como se aprecia en el [Anexo B: Organigrama Grupo Entel 2023](#). Una de estas vicepresidencias es la de Mercado Personas, la cual se encarga de gestionar todo lo relacionado con los clientes individuales (B2C). Este sector incluye el área de Big Data y Advanced Analytics, con equipos dedicados a la experiencia del cliente, campañas y portabilidad.

La VP de Mercado Personas atiende a más de ocho millones de usuarios finales en Chile, ofreciendo servicios de telefonía móvil, internet, televisión y venta de equipos. Esta área es

fundamental para alinearse con la estrategia de Entel, centrada en mejorar la calidad y cobertura de la red, con inversiones significativas en tecnología 5G y el fortalecimiento de la red 4G. Además, la compañía destaca la satisfacción y experiencia del cliente como un habilitador estratégico dentro de su memoria anual integrada. Este enfoque se refleja en el proyecto Centro de Comando, el cual analiza indicadores de red y experiencia del cliente para mejorar la satisfacción.

En un mercado competitivo y desafiante, Entel ha mantenido una evolución positiva en clientes de telefonía móvil de mayor valor, profundizando su penetración en datos móviles y consolidando su liderazgo en ingresos. La experiencia del cliente es central en la propuesta de valor de la compañía. El programa de transformación Entel UP, lanzado en 2022, se enfoca en mejorar la satisfacción del cliente mediante el monitoreo de indicadores operacionales y de satisfacción, proporcionando experiencias simples, cercanas y resolutivas.

Es importante mencionar que la compañía realiza encuestas de satisfacción con periodicidad mensual que corresponden a estudios relacionales donde el objetivo es poder realizar un seguimiento acerca de la satisfacción general y del indicador de recomendación que tienen los clientes de la compañía a través de una muestra de clientes que se distribuye a lo largo y ancho del territorio nacional.

2. Proyecto

2.1. Descripción del problema u oportunidad

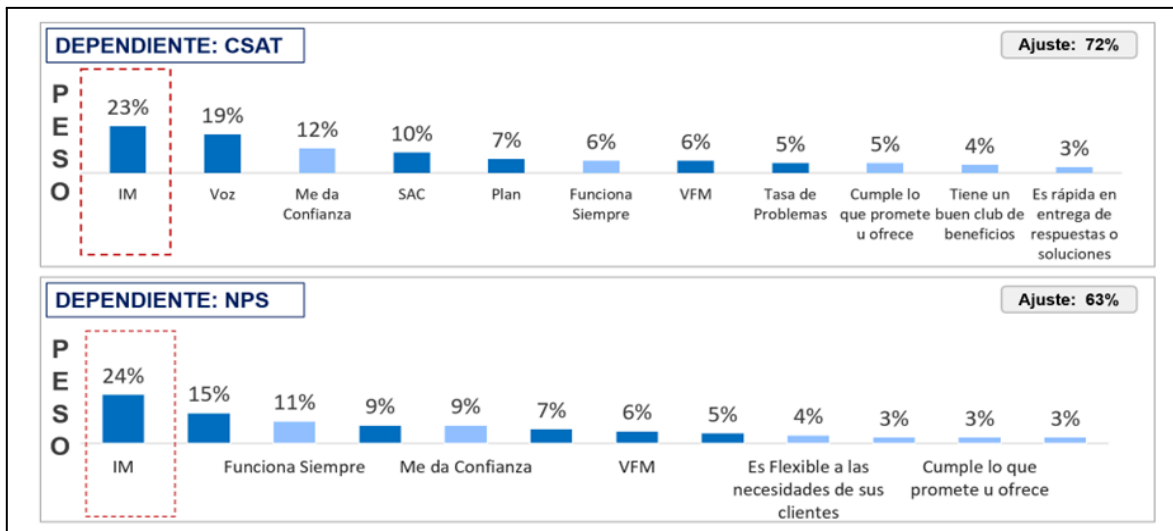
Tal como se mencionó en la sección de antecedentes generales, uno de los principales focos de interés de la Vicepresidencia Mercado Personas del Grupo Entel corresponde a mejorar la experiencia del cliente evaluada a través de la satisfacción que mantienen los clientes con la compañía. Esta métrica se monitorea mediante los resultados de las encuestas de satisfacción realizadas por la Empresa Ipsos a los clientes postpago de Entel. Las encuestas evalúan la satisfacción con la compañía, la calidad del servicio, los canales de atención, entre otras dimensiones.

A modo de dejar en evidencia la importancia de las variables de red móvil en la satisfacción que tienen los clientes postpago de Entel con la compañía, se muestran a continuación los resultados del Informe Trimestral Tracking NPS y Satisfacción Neta Clientes Móvil Entel, el cual considera los resultados del primer trimestre de 2024.

El Informe Trimestral Tracking NPS y Satisfacción Neta Clientes Móvil Entel reveló que el área (dimensión) que más peso e influencia tiene tanto en el nivel de recomendación (medido por el Net Promotor Score) como en la satisfacción general de la compañía (CSAT), corresponde al **Internet Móvil** (cobertura y señal del internet).

A continuación, se observa en el recuadro superior el análisis realizado acerca de los pesos que tienen las distintas dimensiones en las respuestas de satisfacción general respecto a la compañía y en la parte inferior el mismo análisis realizado sobre el nivel de recomendación. Éste, permite concluir que el área prioritaria para mejorar incluye la cobertura y señal del internet para mejorar la satisfacción del cliente.

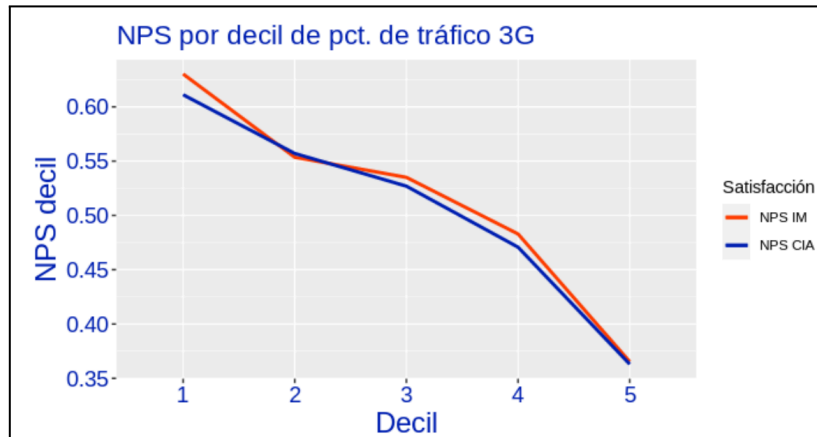
Figura 1: Análisis de las palancas de Entel. Clientes Suscripción. Resultados 1Q 2024



Fuente: Informe Trimestral Resultados Tracking movil NPS, Ipsos Chile

Adicionalmente, tal como se muestra en la Figura 2, se ha identificado que existe una correlación importante cercana a 1 entre la calidad de las conexiones móviles con el nivel de satisfacción de los clientes, por lo que a la fecha, dentro del equipo de Advanced Analytics de la compañía, se cuenta con un [modelo predictivo](#) que permite predecir futuros clientes detractores de la compañía utilizando variables de red móvil 4G que explican la satisfacción.

Figura 2: Correlación NPS y KPI's de red móvil



Fuente: Elaboración equipo Advanced Analytics Entel Chile

Considerando lo expuesto anteriormente, es que dentro del Equipo de Experiencia de la Vicepresidencia (VP) Mercado Personas surge la oportunidad de mejorar la predicción de la satisfacción de los clientes de la compañía, con el objetivo de comunicar al equipo de Redes Entel con mayor nivel de precisión las oportunidades de realizar mejoras estructurales en las zonas que los indicadores de las variables de red están perjudicando la satisfacción de los clientes respecto a la compañía.

2.2. Impacto indirecto a partir del desarrollo del proyecto

De acuerdo con información confidencial manejada por el equipo de Analytics de Entel, el **ARPU** (Average Revenue Per User) de los clientes postpago en junio de 2024 fue de \$22,577 pesos chilenos, lo que significa que, en promedio, cada cliente con suscripción genera ingresos mensuales de \$22,577 CLP.

Entre marzo y diciembre de 2023, una muestra de la población que se portó a la competencia reveló que 5,597 móviles realizaron portabilidad hacia otras compañías (PortOut). De estos, un 28% lo hizo debido a problemas relacionados con la señal y cobertura de la red móvil 4G, lo que representa aproximadamente 1,567 clientes insatisfechos (véase [Anexo D: Clientes PortOut Encuestados y Churn por Motivos de Red](#) para mayor detalle).

Aunque no se puede afirmar de manera directa que una metodología predictiva de satisfacción basada en variables de red móvil pueda reducir el **churn** (tasa de abandono), el desarrollo de

este proyecto podría mejorar la predicción del modelo actual utilizado en Analytics. Esto, a su vez, podría ayudar indirectamente a disminuir la fuga de clientes mediante la implementación de acciones correctivas y mejoras estratégicas basadas en los resultados mejorados en la predicción de detractores.

2.3. Descripción y justificación del proyecto

El proyecto consiste en desarrollar una metodología para encontrar el mayor nivel de predicción de la satisfacción general de los clientes postpago de la compañía, utilizando variables de red móvil 4G. Para esto, se desarrollará un **algoritmo metaheurístico estocástico** que busca soluciones óptimas para maximizar una función objetivo basada en el **Information Value (IV)**.

Las soluciones óptimas se determinan a partir de tres parámetros: dos que establecen un período de fechas (día inicial y día final) y un tercero que identifica un umbral de la variable de red móvil al que se aplicará la metodología.

Una vez que la metodología esté desarrollada, se probará su efectividad con las variables de red móvil RSRP (indicadora de la potencia de la red que reciben los clientes) y RSRQ (indicadora de la calidad de la red que reciben los clientes), que han mostrado una fuerte correlación con la satisfacción del cliente.

A partir de una reunión llevada a cabo con la contraparte, área de Redes Entel, se estableció que existen 2 variables de mayor interés que correlacionan fuertemente con la satisfacción de los clientes, las cuales son RSRP (indicadora de la potencia de la red que reciben los clientes) y RSRQ (indicadora de la calidad de la red que reciben los clientes). La importancia de las variables se determinó anteriormente con la ejecución del modelo de predicción actual y de ahí es que la contraparte hace sus recomendaciones, por consiguiente, las variables antes mencionadas se utilizan para probar la ejecución de la metodología una vez desarrollada ésta. Puede verse mayor detalle de la importancia de variables en [Anexo E: Importancia de variables en modelo de predicción actual \(julio 2024\) de satisfacción](#).

La metodología a desarrollar se basa en la implementación de una metaheurística, específicamente el **Harmony Search Algorithm**. Este algoritmo se utilizará para encontrar los parámetros óptimos que maximicen el **Information Value** en la predicción de la satisfacción utilizando variables de red. El objetivo del proyecto es desarrollar una metodología que permita optimizar el uso de variables de red para mejorar la predicción de clientes que podrían ser **detractores** (aquellos con una nota de satisfacción menor o igual a 4). Esto permitirá aprovechar plenamente la oportunidad existente en el área de Mercado Personas para mejorar la satisfacción del cliente.

El proyecto se centra en desarrollar una metodología que, utilizando un algoritmo metaheurístico, optimice la predicción de la satisfacción del cliente mediante variables de red móvil 4G. Las variables clave RSRP y RSRQ se usarán para probar la metodología, con el objetivo final de mejorar la predicción de clientes detractores y, en última instancia, la satisfacción general del cliente.

3. Objetivos

3.1. Objetivo General

El objetivo general del proyecto corresponde a desarrollar una metodología que permita encontrar el mayor nivel de predicción de la satisfacción de los clientes postpago de una compañía de telecomunicaciones mediante la implementación de la metaheurística Harmony Search utilizando las variables de red móvil 4G RSRQ y RSRP.

3.2. Objetivos específicos

Los objetivos específicos del proyecto son los siguientes:

- Implementar la metaheurística Harmony Search para obtener combinaciones óptimas de los parámetros de temporalidad y umbrales de las variables de red móvil que permiten obtener el mayor nivel de predicción de la satisfacción, evaluado mediante la métrica Information Value.
- Evaluar metodología desarrollada en distintos conjuntos de datos para asegurar robustez.
- Evaluar los niveles de predicción de la satisfacción medidos a través de la métrica Information Value aplicando la metodología en las variables RSRP y RSRQ.
- Realizar pruebas en el modelo actual de predicción de satisfacción de clientes incluyendo las variables a las cuales se les aplicó la metodología para evaluar la utilidad y efectividad de ésta.
- Concluir y validar la utilidad de incluir en el modelo actual (el usado por la compañía) las variables de red a las cuales se les aplicó la metodología, evaluando cómo varían las métricas de predicción del modelo.

4. Alcances

4.1. Limitantes del proyecto

El proyecto tiene como objetivo principal desarrollar una metodología para determinar el nivel predictivo de una variable independiente respecto a una variable objetivo, enfocándose en el área de Redes de la compañía. Este enfoque busca generar **insights** y **KPIs** que sirvan como indicadores de satisfacción, en lugar de analizar directamente la variable en cuestión sin ninguna modificación. Por ejemplo, se pretende recomendar al área de redes que preste atención a los clientes que tengan un alto porcentaje de sus sesiones bajo cierto umbral durante un periodo determinado, ya que se espera que estos clientes sean detractores en el siguiente mes.

Dentro de las limitaciones del proyecto, destaca la disponibilidad de datos para probar la metodología desarrollada. Por políticas internas de Entel, los datos costosos de almacenar en el servicio de almacenamiento en la nube (AWS S3) se eliminan después de seis meses. Debido a esto, se decidió probar la metodología utilizando tableros de datos que abarcan periodos de dos meses, con los cuales se realiza la predicción de la satisfacción. Además, durante febrero de 2024, se implementaron cambios en la encuesta de satisfacción (Ipsos), como la modificación de la estructura y orden de las preguntas, la inclusión de nuevas preguntas relacionadas con aspectos emocionales cliente-compañía, y la incorporación de la pregunta de recomendación (NPS) como un proxy de lealtad de los clientes.

Para evitar que estos cambios afecten la consistencia de los resultados, se utilizaron los siguientes tableros de datos para el desarrollo del proyecto:

- **Tablón 1:** Datos de noviembre y diciembre de 2023, entrenando con resultados de la encuesta Ipsos de enero de 2024.
- **Tablón 2:** Datos de febrero y marzo de 2024, entrenando con resultados de la encuesta Ipsos de abril de 2024.

Las limitaciones del proyecto incluyen la disponibilidad de datos y los cambios en las encuestas de satisfacción, lo que llevó a la utilización de tableros de datos específicos para mantener la consistencia en la prueba de la metodología.

Actix: Base de datos que contiene información de las distintas sesiones de conexión de cada cliente de la compañía. En ella, está disponible la información de las distintas variables de red móvil 4G, como la variable "RSRP" y "RSRQ".

4.2. Enfoque de solución escogido

Dentro del desarrollo del proyecto, se tomaron varias decisiones respecto a cómo proceder, y en este apartado se justifica el enfoque de solución escogido. El desarrollo de la metodología requería el uso de una heurística de optimización para probar distintos parámetros que pudieran encontrar cómo obtener el nivel de predicción óptimo a partir de las variables a utilizar, maximizando la función objetivo Information Value. Luego de haber investigado los distintos algoritmos que permiten iterar sobre distintas combinaciones de parámetros y que cumplen la función para la cual se utiliza una metaheurística, se decidió implementar la metaheurística de optimización Harmony Search, principalmente por su flexibilidad para incorporar o modificar los parámetros que forman parte de los conjuntos de soluciones a lo largo de las iteraciones que se deciden realizar.

Cabe destacar que, como se definió dentro de los objetivos del proyecto, se busca desarrollar una metodología mediante la implementación del algoritmo Harmony Search, independientemente de que pudiera haber algún otro algoritmo más eficiente, ya que no es el foco del proyecto. Para más detalle de la comparación de algoritmos, véase la sección de Investigación en la descripción de la metodología utilizada para el desarrollo del proyecto ([6.1. Fase 1: Investigación](#)).

Luego, se procedió a definir los parámetros a considerar en las soluciones que genera el algoritmo para una primera iteración, los cuales corresponden al día de inicio de los datos para la variable de red en cuestión, día de fin y umbral. Se decidió usar como variable predictora de la satisfacción la cantidad de veces que los clientes están bajo el umbral fijo definido por la mejor solución al acabar las iteraciones del Harmony Search.

Producto de hallazgos obtenidos al momento de interpretar los resultados, se decidió realizar una segunda iteración donde los parámetros que conforman las soluciones para evaluar la función objetivo son el día de inicio de los datos, día de fin y 'p', donde 'p' permite construir un umbral óptimo de la variable que se utilizará. Respecto a los primeros dos parámetros, existe interés en tener flexibilidad en la ventana de datos para comprobar teorías sobre cuánto tiempo hacia atrás recuerda el cliente cuando responde una encuesta de satisfacción, y el modelo actual toma las variables del mes completo, por lo que no se captura qué fechas son más representativas de la satisfacción. Respecto al parámetro que ayuda a determinar el umbral de la variable, en un comienzo se utilizaron valores fijos (primera iteración) definidos de negocio, y aunque no lograban mejoras en el modelo (correlación baja con nota de satisfacción), se decidió fijar umbrales personalizados según cada cliente para ratificar una hipótesis de que esto podría ser útil.

Cabe destacar que se hace uso de la métrica Information Value (IV) en la maximización de la función objetivo del algoritmo por su capacidad para evaluar la capacidad predictiva de una variable independiente respecto a una variable objetivo binaria, lo que es crucial para determinar la importancia de las variables de red a las cuales se les aplique la metodología en términos de obtener la mejor predicción de la satisfacción del cliente.

A continuación, una vez se finalizó con el desarrollo de la metodología para predecir la satisfacción, se decidió probar ésta utilizando dos variables de red, RSRQ y RSRP.

¿Por qué? Tal como se mencionó anteriormente en la sección [Descripción y justificación del proyecto](#), las variables RSRQ y RSRP correlacionan fuertemente con la satisfacción y por juicio

de expertos, son variables miradas por el mercado. Sumado a lo anterior, al observar la importancia de variables en el modelo de predicción actual que se utiliza en el equipo de Advanced Analytics Entel, puede observarse que ambas se encuentran posicionadas dentro de las variables más importantes para predecir la satisfacción. Para más detalle de la importancia de variables, véase el anexo [Anexo E: Importancia de variables](#).

Se seleccionó este enfoque de solución con base en consideraciones prácticas y técnicas, teniendo en cuenta factores como el tiempo disponible, los recursos disponibles y la necesidad de generar insights relevantes para la mejora continua del servicio. Es importante destacar que el alcance del proyecto no abarca el análisis exhaustivo de la variable en sí misma, sino que se centrará en el desarrollo de la metodología predictiva y la identificación de indicadores clave de satisfacción.

Finalmente, se considera el método de Logistic Regression para comprobar las variaciones en el rendimiento del modelo al incluir las variables procesadas por sus bajos costos computacionales, alta interpretabilidad y rápidos resultados. Lo anterior, puede sustentarse en que “Los modelos de ML creados mediante regresión logística ayudan a las organizaciones a obtener información procesable a partir de sus datos empresariales. Pueden usar esta información para el análisis predictivo a fin de reducir los costos operativos, aumentar la eficiencia y escalar más rápido” (Amazon Web Services, 2024).

4.3. Metas e indicadores de éxito

Dado que el objetivo general del proyecto tiene relación con completar la metodología de implementación del algoritmo Harmony Search para maximizar una función objetivo (IV) y no con mejorar la performance del modelo de predicción actual, es que la meta es obtener las soluciones óptimas propias de la ejecución del algoritmo que determinen la temporalidad óptima de los datos a considerar y el tercer parámetro que permite definir el umbral de la variable en estudio, sin embargo, indicadores de éxito que comprobarían que la metodología desarrollada es útil corresponde a obtener ganancias/aumentos en las métricas de predicción ROC AUC y Curva LIFT de los modelos una vez se incluyan las variables procesadas al modelo de predicción actual.

4.4. Objetivos y resultados fuera del alcance del proyecto

Parte de lo que queda fuera del alcance del proyecto desarrollado es considerar en el flujo normal del modelo (modelo ya en producción) la inclusión de las variables a las cuales se les aplique la metodología. Además, queda fuera del alcance del proyecto ejecutar y desarrollar la metodología de predicción con otros algoritmos distintos al Harmony Search. También se propone entrenar con la metodología con más periodos de datos para asegurar mayor robustez.

Otro objetivo que se tiene a futuro es desarrollar la metodología de manera análoga a como se hace en el proyecto en cuestión, pero utilizando como métrica de evaluación el estadístico K-S el cual puede utilizarse para comparar si los datos observados se ajustan o no a los valores predichos. “La K-S también puede utilizarse para comprobar si dos distribuciones de

probabilidad unidimensionales subyacentes difieren. Es una forma muy eficiente de determinar si dos muestras son significativamente diferentes entre sí ". (Singh, 2023)

Otro objetivo que se tiene y que está fuera del alcance es aplicar la metodología desarrollada en este proyecto a la variable de red móvil "Delay" que es de interés en el equipo de Advanced Analytics la cual no está productiva y disponible hasta la fecha en las bases de datos de Entel, sin embargo se está trabajando en ello. Esta variable tiene relación con la demora (delay en inglés) en que se cargan , descargan y/o procesan los datos móviles.

5. Marco Conceptual

5.1. Metodologías, algoritmos y métricas de predicción

5.1.1. Metodología de desarrollo de proyectos Analytics Entel

Para el desarrollo del proyecto, se utilizó una metodología propia del Equipo de Advanced Analytics de Entel, que consta de cinco etapas: descubrimiento, prueba de concepto, desarrollo del producto mínimo viable (MVP), fase de piloto, y automatización. Esta metodología está diseñada para definir roles y responsabilidades específicos para cada miembro del equipo (data scientist, machine learning engineer, data engineer), lo que facilita que cada contraparte se enfoque en su especialidad. Para más detalle sobre la metodología, consulte el [Anexo F: Metodología desarrollo de proyecto](#).

5.1.2. Proceso ETL

Extracción, transformación y carga (ETL) es el proceso consistente en combinar datos de diferentes orígenes un gran repositorio central llamado almacenamiento de datos. ETL utiliza un conjunto de reglas comerciales para limpiar y organizar datos en bruto y prepararlos para el almacenamiento, el análisis de datos y el machine learning (ML). Puede abordar necesidades de inteligencia empresarial específicas mediante análisis de datos (como la predicción del resultado de decisiones empresariales, la generación de informes y paneles, la reducción de la ineficacia operativa, entre otras). (Amazon Web Services, 2024.)

5.1.3. Metaheurísticas de Optimización

Las metaheurísticas son técnicas de optimización de búsqueda basadas en algoritmos heurísticos que se utilizan para resolver problemas complejos en los que el espacio de búsqueda es muy grande o desconocido. Las metaheurísticas son algoritmos de propósito general que no están diseñados para un problema específico, sino que se pueden adaptar a una amplia variedad de problemas de optimización en diferentes campos, como la inteligencia artificial, la ingeniería, la economía, la biología y muchos otros. Las metaheurísticas son algoritmos que se basan en la

exploración iterativa del espacio de soluciones, utilizando estrategias para moverse de forma inteligente por el espacio de búsqueda y para escapar de los óptimos locales. (Gamco, 2021)

En el proyecto en cuestión, se investigaron distintas metaheurísticas que permitieran iterar sobre distintas combinaciones de parámetros de temporalidad y umbral a partir de una variable de red, de manera de optimizar ésta y encontrar de qué manera podía predecirse mejor la satisfacción de clientes a partir de la variable. Algunas metaheurísticas que cumplen lo requerido son:

- **Algoritmos Genéticos (GA):** Basados en la evolución biológica, los GA utilizan operadores como la selección, el cruce y la mutación para explorar el espacio de soluciones y encontrar la mejor combinación de parámetros.
- **Recocido Simulado (SA):** Imita el proceso de enfriamiento de metales para encontrar una solución óptima global, evitando quedar atrapado en óptimos locales.
- **Búsqueda Tabú (TS):** Utiliza memoria a corto plazo para evitar ciclos y mejorar la exploración del espacio de soluciones, permitiendo escapar de óptimos locales y encontrar mejores soluciones a largo plazo (Geem, 2001).
- **Harmony Search (HS):** Este algoritmo está inspirado en el proceso de improvisación musical, donde los músicos ajustan sus notas para crear la mejor armonía posible. HS es flexible y permite ajustar parámetros de manera dinámica para optimizar la función objetivo (Geem, 2001; Geem, 2009).

5.1.4. Metaheurística Harmony Search

La Heurística de Búsqueda de Armonía (Harmony Search, HS) es un algoritmo de optimización inspirado en el proceso de improvisación musical. El algoritmo está diseñado para encontrar soluciones óptimas en diversos problemas, imitando la manera en que los músicos buscan la mejor armonía. Este algoritmo consta de varios componentes clave: Tamaño de la Memoria de Armonía (HMS), la Tasa de Consideración de Memoria de Armonía (HMCR), la Tasa de Ajuste de Pitch (PAR) y las Iteraciones del algoritmo.

La Memoria de Armonía (HMS) es una estructura que almacena un conjunto de soluciones viables, similar a cómo los músicos guardan diferentes combinaciones de notas que suenan bien juntas. Al inicio, esta memoria se llena con soluciones generadas aleatoriamente. Durante el proceso de optimización, nuevas soluciones son generadas y evaluadas; si una nueva solución es mejor que la peor solución en la memoria, reemplaza a esta última. De este modo, HM siempre contiene las mejores soluciones encontradas hasta el momento.

La Tasa de Consideración de Memoria de Armonía (HMCR) es un parámetro que determina la probabilidad de que un valor de una nueva solución sea seleccionado de la HM en lugar de ser generado aleatoriamente. Un valor alto de HMCR significa que es más probable que las nuevas soluciones se construyan a partir de las mejores soluciones previas, lo cual puede acelerar la convergencia hacia la solución óptima. Por ejemplo, un HMCR de 0.95 indica que hay un 95% de probabilidad de elegir los valores de la HM, y solo un 5% de probabilidad de generar nuevos valores aleatoriamente.

La Tasa de Ajuste de Pitch (PAR) introduce un mecanismo para ajustar ligeramente los valores seleccionados de la HM, similar a cómo los músicos ajustan la afinación de sus instrumentos para lograr una mejor armonía. Este parámetro ayuda al algoritmo a explorar soluciones cercanas a las ya conocidas, evitando que el proceso de búsqueda se estanque en óptimos locales. Un valor de PAR de 0.1, por ejemplo, implica que hay un 10% de probabilidad de ajustar un valor seleccionado hacia un valor vecino en cada iteración.

Finalmente, el número de iteraciones define cuántas veces el algoritmo repetirá el proceso de generación y evaluación de nuevas soluciones. Cada iteración representa una nueva "improvisación" en la búsqueda de la armonía óptima. El proceso continúa hasta que se cumpla un criterio de parada, como alcanzar un número máximo de iteraciones o lograr una mejora insignificante entre iteraciones consecutivas.

Estos cuatro componentes clave trabajan en conjunto para permitir que la Heurística de Búsqueda de Armonía explore eficientemente el espacio de soluciones y converja hacia la solución óptima en una amplia gama de problemas de optimización. (Geem & Kim, 2001)

Respecto a los valores de los hiperparámetros utilizados para construir y ejecutar la metodología de predicción, se realiza una búsqueda en rejilla ó "Grid Search" probando todas las combinaciones posibles entre los valores discretos definidos para cada hiperparámetro, los que se muestran a continuación.

HMS= [5, 10, 15, 20]
HMCR=[0.1, 0.3, 0.5, 0.7, 0.9]
PAR=[0.1, 0.3, 0.5, 0.7, 0.9]
ITER= [5.000, 10.000, 15.000, 20.000]

Finalmente, luego de ejecutar 400 veces el algoritmo ($4 \times 5 \times 5 \times 4 = 400$), los parámetros óptimos definidos para aplicar la metodología a alguna variable de red son los siguientes: HMS=20; HMCR=0.7; PAR=0.1; ITER=5000.

5.1.5. Métricas para evaluar capacidad predictiva de una variable

Para medir la capacidad predictiva de una variable independiente respecto a una variable objetivo binaria, se utilizan varias métricas como **Information Value (IV)**, **Gini Index**, **Test de Kolmogorov-Smirnov (KS)**, y **Entropía**. Estas métricas evalúan la relación y la importancia de la variable predictora en la predicción de resultados binarios (Gareth et al., 2013).

- **Information Value (IV)**: Mide la capacidad predictiva de una variable independiente respecto a una variable binaria.
- **Chi-Cuadrado (Chi-Square)**: Evalúa la independencia entre la variable predictora y la variable objetivo binaria.
- **Gini Index**: Mide la pureza de una división en la variable objetivo binaria.
- **Entropía**: Mide la impureza o desorden en las particiones de datos.

- **Test de Kolmogorov-Smirnov (KS):** Mide la diferencia máxima entre las distribuciones acumuladas de eventos y no eventos.

5.2. Conceptos

5.2.1. Variables de red móvil

En el contexto de las redes móviles, las variables relacionadas con la potencia de la señal y la calidad de la misma desempeñan un papel fundamental en la evaluación y optimización del rendimiento de la red para proporcionar una experiencia satisfactoria a los usuarios. Dos de estas variables clave son la Potencia de Señal Recibida de la Referencia (RSRP) y la Calidad de Señal Recibida de la Referencia (RSRQ). La importancia de RSRP y RSRQ radica en su capacidad para proporcionar información crucial sobre la intensidad y calidad de la señal recibida por los dispositivos móviles desde las estaciones base de la red. La RSRP se enfoca en la medición de la potencia promedio de la señal recibida, mientras que la RSRQ evalúa la calidad de la señal considerando la presencia de interferencia y ruido en la red. (International Journal of Wireless & Mobile Networks, 2015)

5.2.2. Encuestas de satisfacción

Las encuestas de satisfacción del cliente son instrumentos de medición de tipo relacionales que se utilizan para determinar la satisfacción de los clientes con respecto a los productos, servicios o la impresión general de una empresa. Además, se hace un seguimiento de los niveles de recomendación que mantiene la muestra de clientes con la compañía. Esto permite identificar las necesidades del cliente, focos prioritarios de mejoras y cualquier problema con el producto o servicio. (Qualtrix.xm, 2024)

5.2.3. Amazon Web Services (AWS)

Amazon Web Services (AWS) como una plataforma líder en servicios de computación en la nube ofrecida por Amazon.com. AWS proporciona una amplia gama de servicios en la nube, como almacenamiento, bases de datos, análisis, aprendizaje automático, redes y más, diseñados para ayudar a las organizaciones a desarrollar y escalar sus aplicaciones de forma eficiente. Estos servicios en la nube son altamente flexibles, escalables y seguros, lo que permite a las empresas gestionar sus recursos de TI de manera eficaz y optimizar sus operaciones. Con la infraestructura confiable y flexible de AWS, las empresas pueden acceder a recursos informáticos a través de Internet de manera rentable y eficiente. (AWS, 2023)

5.2.4. Amazon Elastic Compute Cloud (Amazon EC2)

EC2 es un servicio de computación en la nube que ofrece instancias virtuales escalables y personalizables para diversas necesidades de computación.

5.2.5. Amazon Simple Storage Service (Amazon S3)

S3 es un servicio de almacenamiento de objetos, con un enfoque en escalabilidad, disponibilidad, seguridad y rendimiento. Permite almacenar y proteger grandes cantidades de datos para diversos casos de uso y cumplir con requisitos de seguridad y cumplimiento.

5.2.6. Amazon Athena

Athena es un servicio de análisis interactivo y sin servidor que soporta formatos abiertos de archivos y tablas. Permite analizar grandes cantidades de datos en Amazon S3 y otros orígenes de datos, con una interfaz simple compatible con SQL y Python. (AWS, 2023)

5.2.7. Information Value

El Valor de la Información (IV) es una técnica esencial en el análisis de modelos predictivos, inicialmente desarrollada para la industria de scoring crediticio. Esta herramienta evalúa y selecciona variables en modelos de riesgo crediticio, como la probabilidad de incumplimiento, y también se aplica en proyectos de análisis de marketing, como modelos de abandono de clientes y respuesta a campañas. El IV mide la capacidad predictiva de una variable independiente respecto a una variable dependiente, calculado mediante la fórmula que relaciona las proporciones de no-eventos y eventos con el Weight of Evidence (WOE). (Guide & Bhalla, 2015)

Figura 3: Interpretación valores IV.

| Information Value | Variable Predictiveness |
|-------------------|-----------------------------|
| Less than 0.02 | Not useful for prediction |
| 0.02 to 0.1 | Weak predictive Power |
| 0.1 to 0.3 | Medium predictive Power |
| 0.3 to 0.5 | Strong predictive Power |
| >0.5 | Suspicious Predictive Power |

Fuente: Siddiqi 2006

5.2.8. Relación entre IV y WoE

El Information Value es una de las técnicas de selección de importancia de variables más usadas en modelos predictivos. Ayuda a rankear variables en base a su importancia. El IV se calcula usando la siguiente fórmula:

$$\text{Information Value} = \Sigma \text{WoE} * (\% \text{eventos} - \% \text{no eventos})$$

(Guide & Bhalla, 2016)

5.2.9. Modelo de predicción

Los modelos de análisis predictivo están diseñados para evaluar datos históricos, descubrir patrones, observar tendencias y utilizar esa información para predecir tendencias futuras. Los modelos de análisis predictivo más populares son los de clasificación, agrupación y series temporales. (IBM, 2024.)

5.2.10. Información de sesiones de los clientes

Las empresas de telecomunicaciones poseen torres con antenas distribuidas a lo largo del territorio donde proveen cobertura, en este caso, cobertura de internet móvil. Cada vez que un cliente navega o hace uso de internet, la antena más cercana registra diversa información acerca de la sesión en la cual el cliente estuvo conectado a la red. Esta información incluye múltiples variables de red, como el RSRQ (Received Signal Received Quality) y el RSRP (Received Signal Received Power), que son indicadores de la calidad y potencia de la red, respectivamente.

5.3. Disciplinas de la Ingeniería Civil Industrial relacionadas con el proyecto

5.3.1. Data Science (Ciencia de datos)

La ciencia de datos es el estudio de datos con el fin de extraer información significativa para empresas. Es un enfoque multidisciplinario que combina principios y prácticas del campo de las matemáticas, la estadística, la inteligencia artificial y la ingeniería de computación para analizar grandes cantidades de datos. (AWS, 2023).

En el contexto del proyecto, **Data Science** se emplea para el desarrollo de modelos predictivos que permitan anticipar la satisfacción del cliente a partir de variables de red. Este enfoque incluye la aplicación de algoritmos avanzados, como el metaheurístico Harmony Search, y el uso de métricas como el **Information Value** para evaluar el rendimiento del modelo.

5.3.2. Big Data

El término "big data" se refiere a los datos que son tan grandes, rápidos o complejos que es difícil o imposible procesarlos con los métodos tradicionales. (SAS, 2024). En el caso del presente proyecto, los datos provienen de millones de sesiones de usuarios de redes móviles y requieren de plataformas robustas como **AWS (Amazon Web Services)** para su almacenamiento y procesamiento. La capacidad de gestionar grandes volúmenes de datos en tiempo real es crucial para poder analizar patrones de comportamiento y mejorar la predicción de satisfacción de los clientes.

5.3.3. Data Analytics

Data analytics es el proceso de recopilar y analizar muchos datos del cliente para extraer conclusiones sobre patrones de comportamientos, intereses personales y tendencias de compra. El data analytics agiliza la toma de decisión, mejora los procesos empresariales, mejora el engagement del usuario, reduce los costes e impulsa el crecimiento y la rentabilidad. (Cognizant, 2024)

5.3.4. Machine Learning

El machine learning (ML) es un subconjunto de la inteligencia artificial (IA). Se enfoca en enseñar a las computadoras para que aprendan de los datos y mejoren con la experiencia -en lugar de ser explícitamente programadas para hacerlo. (SAP, 2024)

En este proyecto, ML es fundamental para la creación de modelos predictivos capaces de identificar patrones en las variables de red móvil que afectan la satisfacción del cliente. El uso de algoritmos como el Harmony Search permite ajustar automáticamente los parámetros del modelo para maximizar su precisión.

5.4. Lenguajes de programación

5.4.1 Python y Visual Studio Code

Python es un lenguaje de programación ampliamente utilizado en las aplicaciones web, el desarrollo de software, la ciencia de datos y el machine learning (ML). Este lenguaje permite procesar grandes volúmenes de datos y cuenta con diversas librerías para el manejo de dato, procesamiento, visualización y desarrollo de modelos computacionales diversos.

Para el caso en cuestión, el desarrollo del modelo de predicción de satisfacción se desarrolla mediante la utilización del lenguaje Python usando como entorno de desarrollo integrado (IDE) Visual studio code. Visual Studio Code es un editor de código fuente ligero pero eficaz que se ejecuta en el escritorio y está disponible para Windows, macOS y Linux. Incluye compatibilidad integrada con JavaScript, TypeScript y Node.js, y cuenta con un amplio ecosistema de extensiones para otros lenguajes y entorno de ejecución (como C++, C#, Java, Python, Go, .NET). (Microsoft, 2024.)

5.4.2. Structured Query Language

El Lenguaje de Consulta Estructurada (SQL) es un lenguaje gestor para el manejo de la información en las bases de datos relacionales. Este tipo de lenguaje de programación permite comunicarse con la base de datos y realizar operaciones de acceso y manipulación de la información almacenada. Permite operaciones como selección, inserción, actualización y eliminación de datos además de operaciones administrativas sobre las bases de datos. Por tanto, se trata de un lenguaje completo de bases de datos que va más allá de la recuperación de la información. (Universidad Europea, 2021)

Para el proyecto en cuestión, la data necesaria para el desarrollo del modelo se obtiene mediante la realización de consultas en lenguaje SQL, las cuales se ejecutan mediante el software ya mencionado Amazon Athena. Las consultas realizadas obtienen la información desde el sistema de almacenamiento Amazon S3.

6. Metodología

El desarrollo de una metodología dentro del área de Advanced Analytics en la compañía se ha estructurado en varias fases y etapas, siguiendo un enfoque **End-to-End Automático**. Este enfoque abarca desde la investigación inicial hasta la automatización completa del modelo predictivo. En este capítulo, se detallan las distintas fases, etapas y las actividades específicas que conforman la metodología, con especial atención a las tareas propias del Data Scientist.

6.1. Fase 1: Investigación

La fase de investigación es fundamental para sentar las bases del proyecto. Su objetivo principal es comprender a fondo el proceso de negocio, identificar problemas relevantes, y analizar los datos disponibles para construir una base sólida sobre la cual se desarrollará el proyecto.

6.1.1. Etapa de Ideación

En la **etapa de ideación**, el primer paso es definir claramente el problema de negocio que se busca resolver. Es crucial que esta definición sea precisa y alineada con los objetivos estratégicos de la compañía. Durante esta fase, se realiza un brainstorming para generar y evaluar posibles soluciones. Este proceso debe ser estructurado, asegurando que todas las ideas se consideren y se evalúen en función de su viabilidad y alineación con los objetivos del proyecto.

Es vital que en esta etapa se seleccionen los datos correctos y que se establezca un flujo claro y lógico de cómo se manejarán estos datos a lo largo del proyecto. Aquí es donde el **canvas analítico** juega un papel crucial, proporcionando una visión estructurada del problema, las posibles soluciones, y cómo los datos fluirán a través del proceso. Este canvas debe ser revisado y validado para garantizar que todas las áreas críticas del problema se han cubierto antes de avanzar.

6.1.2. Etapa de Prueba de Concepto (POC)

La **prueba de concepto (POC)** es donde se valida la hipótesis inicial. En esta etapa, se construye un modelo simplificado utilizando los datos disponibles. Es crucial que este modelo sea lo suficientemente robusto para identificar patrones y señales relevantes, pero también lo suficientemente flexible para adaptarse a nuevos hallazgos.

Durante la POC, se deben realizar análisis exhaustivos para verificar la calidad de los datos y su relevancia para el problema. Es fundamental no solo identificar patrones, sino también entender las limitaciones de los datos y cómo pueden afectar la predicción de la satisfacción del cliente. Dependiendo de los resultados de la POC, se toma una decisión informada sobre si avanzar o regresar a la etapa de ideación para ajustes adicionales.

6.2. Fase 2: Desarrollo del proyecto

La fase de desarrollo del proyecto se enfoca en construir y validar una solución robusta. Esta fase incluye la creación de un Producto Mínimo Viable (MVP), la implementación de un piloto y la implementación en producción.

En la etapa del Producto Mínimo Viable (MVP), se desarrolla una aplicación analítica funcional. Esto implica el desarrollo de la solución inicial que pueda ser utilizada para análisis predictivo, la definición de pruebas de aseguramiento de calidad y pruebas unitarias, la revisión y optimización del código para asegurar su calidad y eficiencia, y la utilización de datos precisos y relevantes. Además, se documenta todo el proyecto y se valida el modelo utilizando datos históricos para asegurar su precisión. La entrega del MVP se realiza en Docker, proporcionando un entorno de despliegue preparado para el equipo de automatización.

La siguiente etapa es la implementación del piloto, donde se prueba la solución en un entorno real. Esto implica desplegar la solución en un entorno de prueba controlado y evaluar el rendimiento del modelo en condiciones reales para confirmar su eficacia. Si los resultados del piloto son satisfactorios, se avanza a la siguiente etapa.

En la etapa final de la fase de desarrollo del proyecto, se realiza la implementación en producción. El objetivo aquí es trasladar la aplicación analítica a servidores productivos y configurar la ejecución automática del modelo con verificaciones periódicas para asegurar su funcionamiento continuo y correcto. El resultado de esta fase es un modelo predictivo en producción, con ejecución periódica y monitoreo constante para mantener su precisión y relevancia.

Para el proyecto en cuestión, se destaca que la etapa de producción del proyecto queda fuera del alcance y de los plazos del proceso.

Para más detalle de la metodología, consultar el [Anexo F](#).

7. Desarrollo

Una vez identificado el problema y oportunidad descrita anteriormente, el desarrollo del proyecto consiste en una serie de pasos desde la extracción de la data a utilizar hasta la obtención de los resultados del algoritmo Harmony Search adaptado al caso de estudio con el fin obtener el mayor nivel de predicción posible a partir de ciertas variables de red y posteriormente, la realización de las pruebas en el modelo actual incluyendo las variables procesadas RSRQ y RSRP.

7.1. Selección e ingesta de datos

En primer lugar, se desarrolla una Query en lenguaje SQL para extraer a través del servicio AWS Athena, una base de datos de la compañía Entel que contiene información de las sesiones de la totalidad de los clientes de la compañía. A su vez, se realizó un cruce de la información de las sesiones con la información de las encuestas de satisfacción, de manera de considerar sólo las sesiones de los clientes para los cuales se tienen resultados de encuestas de satisfacción. El resultado de la query realizada se guarda en el servicio de almacenamiento S3 para tenerla disponible durante todo el desarrollo.

Respecto a la data con información de las sesiones (incluyen las variable de red móvil) de los clientes, se consideraron los meses de noviembre y diciembre 2023 más febrero y marzo de 2024 mientras que de la data de los resultados de las encuestas que incluye la variable de satisfacción a predecir se consideraron los meses de enero y abril 2024.

7.2. Análisis exploratorio de los datos

En esta sección se hace un análisis exploratorio de los datos de una manera no exhaustiva si no que con el fin de observar las variables de los data frames a utilizar.

7.3. Preprocesamiento

Luego, a partir de la información almacenada en S3 se procesa la data Ipsos para crear la columna dummy 'detractores', la cual toma el valor de 1 para los cliente con satisfacción menor o igual 4 y 0 para los clientes con satisfacción 5, 6 o 7.

7.4. Elaboración de metodología

7.4.1. Versión inicial (primera iteración)

La implementación del algoritmo comienza con la generación de armonías aleatorias, donde cada armonía se representa como una tupla de parámetros (notas) conformada por (día_inicio, día_fin, umbral). El día de inicio indica el primer día de datos disponibles para el algoritmo, el día de fin señala el último día de datos considerados, y el parámetro umbral indica un valor de umbral fijo para la variable en cuestión.

En lugar de considerar la variable de red original (como RSRQ o RSRP), el algoritmo selecciona la cantidad de sesiones en las que un cliente determinado (identificado por su móvil) está bajo el umbral fijado por el algoritmo dentro del periodo de datos determinado por las primeras dos notas de las armonías. De esta manera, la variable predictora se convierte en "*suma_sesiones_bajo_umbral_fijo*".

Se decidió utilizar *suma_sesiones_bajo_umbral_fijo* como variable predictora de la satisfacción en lugar de la variable original de red porque captura de manera más precisa las experiencias negativas repetidas de los clientes. Esta variable transformada permite identificar eventos críticos en los que la calidad del servicio no cumple con los niveles esperados (bajo umbral), lo que tiene un impacto directo en la percepción del cliente. Este enfoque proporciona una herramienta más interpretable y accionable para la toma de decisiones estratégicas dentro de la organización, ya que permite observar de manera simple la cantidad de sesiones que un cliente tiene dentro de un intervalo de valores.

Una vez generada una armonía, se calcula su Information Value usando como función objetivo o de predicción la etiqueta "Detractor", la cual representa a la satisfacción, ya que como se mencionó anteriormente, Detractor toma el valor de uno si la satisfacción general de un cliente es menor o igual a 4 y cero en el caso contrario. La armonía generada junto al valor de Information Value de ésta se guarda en la memoria de armonías (HM), la cual almacena cierto número de soluciones.

Cuando la memoria (HMS) está llena, el algoritmo comienza a iterar, generando nuevas soluciones (armonías) en cada iteración. Para generar una nueva armonía, se empieza con la primera nota de la armonía (día de inicio). Con cierta probabilidad, determinada por el parámetro HMCR, las notas se buscan dentro de la memoria que contiene armonías previamente guardadas o se genera una nueva nota. Además, también con cierta probabilidad, dependiente del parámetro PAR, a cada nota nueva buscada en la memoria se le aplica un ajuste, es decir, se le suma o resta un valor delta para obtener soluciones distintas a las ya guardadas en memoria.

Después de generar una nueva armonía, se compara su valor de IV con el de la peor armonía guardada en la memoria hasta el momento. Si el IV de la nueva solución es mejor, se reemplaza la peor armonía por la recién generada. El algoritmo continúa iterando hasta completar el número total de iteraciones definido como hiperparámetro.

Los hiperparámetros que se definen para ejecutar el algoritmo son HMS, HMCR, PAR y iteraciones, los cuales están definidos y descritos en la sección [5.1.4. Metaheurística Harmony Search](#).

Por último, se observó la distribución de la cantidad de sesiones de cada cliente considerado en los datos para verificar que si tenía un respaldo lógico él considerar la variable “*suma_sesiones_bajo_umbral_fijo*” como predictora de la satisfacción.

7.4.2. Versión final (segunda iteración)

La segunda iteración y versión final del desarrollo de la metodología considera ciertos cambios respecto a lo definido anteriormente. De manera análoga a la anterior, se implementa el algoritmo para la generación de armonías aleatorias, donde cada armonía ahora se representa como una tupla de parámetros (notas) conformadas como (dia_inicio, dia_fin, p). Los primeros dos parámetros se mantienen constantes respecto a la versión anterior; sin embargo, el parámetro de umbral fijo se reemplazó por la nota p, la cual corresponde a un parámetro (número entero) que permite construir un umbral personalizado (varía según cada cliente) de la variable en estudio, determinando la cantidad de desviaciones estándar que se consideran para generar el umbral, tal como se muestra en la fórmula a continuación:

$$Umbral_ (variable) = mean(variable) + p * std(variable)$$

Cabe destacar que en lugar de considerar directamente la variable en observación, el algoritmo selecciona umbrales personalizados según la data de las sesiones de cada cliente y para dicha variable, en esta segunda iteración, evalúa el porcentaje del total de las sesiones que un cliente (identificado por su móvil) estuvo bajo su propio umbral. Lo anterior, dentro del periodo de datos determinado por las primeras dos notas de las armonías. De esta manera, la variable predictora pasa a ser “porcentaje_sesiones_bajo_umbral_variable”.

El algoritmo continúa análogamente al caso anterior hasta alcanzar el criterio de parada definido como hiperparámetro (número de iteraciones). Los hiperparámetros que se definen para ejecutar el algoritmo son HMS, HMCR, PAR y iteraciones, los cuales están definidos y descritos en la sección “Marco Conceptual / Harmony Search Algorithm”.

El flujo de la ejecución del algoritmo puede consultarse en [Anexo G: Flujo ejecución Algoritmo harmony Search](#).

7.5. Ejecución de metodología

7.5.1. Ejecución versión inicial (primera iteración)

Para la ejecución de la metodología en una primera instancia, se construyó un tablón de datos de entrenamiento que consideró información de las sesiones de los clientes de la

compañía para los meses de noviembre y diciembre 2023 y la variable target “detractor” del mes de enero 2024. A continuación, se ejecuta el algoritmo Harmony Search aplicado a la variable RSRQ y, por último, los resultados del algoritmo en el periodo de datos descrito se aplican a los resultados de la encuesta Ipsos de febrero para testear los resultados fuera de un set de entrenamiento.

7.5.2. Ejecución versión final (primera iteración)

Para la ejecución de la metodología considerando la versión final (iteración 2), se construyeron 2 tablonas para obtener los valores de predicción óptimos a partir de los resultados del algoritmo, los cuales se describen a continuación:

- Tablón 1: Construido con las sesiones de los meses noviembre y diciembre 2023 y variable target “detractor” del mes de Enero 2024
- Tablón 2: Construido con las sesiones de los meses Febrero y Marzo 2024 y variable target “detractor” del mes de Abril 2024

Luego, se ejecuta para ambos periodos el algoritmo Harmony Search utilizando primero la variable RSRQ y luego la variable RSRP. Una vez obtenidas las armonías óptimas, se aplican a un periodo distinto (los resultados del algoritmo en tablón 1 se aplican a resultados de la encuesta ipsos de febrero para testear los resultados y los resultados del algoritmo en tablón 2 se aplican a resultados de la encuesta ipsos de mayo para testear los resultados).

7.6. Incorporación de variables procesadas en modelo de predicción actual

Una vez obtenidos los resultados de la ejecución del algoritmo Harmony Search, se procede a realizar pruebas con estos en el modelo de predicción actual, donde se entrena el modelo para predecir la satisfacción del mes de abril 2024. El objetivo de estas pruebas es evaluar el impacto de las variables optimizadas en el rendimiento del modelo, mediante la inclusión de éstas en el proceso de predicción de la satisfacción. Para esto, se realiza lo siguiente:

- 1) Carga de datos de las sesiones (tabla Actix del data lake S3) y posterior filtrado de datos actix según las soluciones óptimas (armonías) obtenidas a partir de la ejecución de la metodología en las variables RSRQ y RSRP.
- 2) Carga de la data que contiene el score de detractores definido en el [Anexo C: Modelo móvil satisfacción de clientes postpago Entel](#).
- 3) Carga de data con variable de satisfacción del mes de abril 2024.
- 4) Cruce de la distinta data para generar el data frame que se incluirá en la ejecución del modelo.

Antes de proceder con la prueba del modelo incorporando las variables transformadas a partir del desarrollo de la metodología harmony Search, se aplica el proceso de binarización óptima para mejorar la predicción de las variables a utilizarse en un modelo de regresión logística. Para más detalles del proceso de binarización óptima usando la métrica WoE consulte el [Anexo G: Opt Binning y métrica WoE](#).

Figura 4: Vista del data frame para evaluar la predicción de los distintos modelos

| | WoE_porcentaje_bajo_umbral_rsrp | WoE_porcentaje_bajo_umbral_rsrq | score_detractor | detractor |
|----|---------------------------------|---------------------------------|-----------------|-----------|
| 0 | 1.976331 | 0.395278 | 0.421768 | 0.0 |
| 1 | -1.114712 | 0.395278 | 0.417667 | 0.0 |
| 2 | 0.097041 | -0.251547 | 0.446156 | 0.0 |
| 3 | 0.097041 | 0.260283 | 0.461821 | 0.0 |
| 4 | 0.097041 | 1.669745 | 0.422869 | 0.0 |
| 5 | -1.114712 | -0.251547 | 0.465734 | 0.0 |
| 6 | 0.097041 | -0.251547 | 0.420470 | 0.0 |
| 7 | 0.097041 | -0.251547 | 0.429819 | 0.0 |
| 8 | 1.976331 | 0.260283 | 0.429400 | 0.0 |
| 9 | 0.097041 | -0.251547 | 0.429046 | 0.0 |
| 10 | 0.097041 | -0.251547 | 0.437963 | 0.0 |
| 11 | 0.378066 | -0.251547 | 0.446867 | 0.0 |
| 12 | 0.097041 | 0.395278 | 0.425383 | 0.0 |
| 13 | 0.097041 | 0.395278 | 0.420586 | 0.0 |
| 14 | 0.378066 | -0.251547 | 0.426165 | 0.0 |
| 15 | 0.097041 | -0.251547 | 0.465821 | 0.0 |
| 16 | 0.097041 | 0.395278 | 0.428413 | 1.0 |
| 17 | 0.097041 | 1.669745 | 0.414540 | 0.0 |
| 18 | 0.097041 | -0.251547 | 0.453600 | 0.0 |
| 19 | -0.348983 | 0.260283 | 0.429401 | 0.0 |

Fuente: Elaboración propia

Luego, se definen los siguientes 5 modelos con el fin se observar incrementalmente la utilidad de incluir las variables optimizadas en el modelo predictivo actualmente utilizado .La variable a predecir (y= detractor) se define de la siguiente manera:

$$p(y) = \frac{e^{f(x)}}{1 + e^{f(x)}}$$

- Modelo 1: $f(x) = \beta_0 + \beta_1 * score_detractor$

Este modelo utiliza únicamente el score de detractores actual, sin incluir ninguna de las variables optimizadas. Actúa como la línea base para comparar el rendimiento de los modelos subsiguientes.

- Modelo 2:

$$f(x) = \beta_0 + \beta_1 * score_detractor + \beta_2 * WoE_porcentaje_umbral_rsrq$$

En este modelo, se incorpora la variable optimizada **WoE_porcentaje_umbral_rsrq** junto con el score de detractores. Este modelo se diseñó para evaluar cómo la inclusión de la variable optimizada **RSRQ** mejora la capacidad predictiva del modelo en comparación con el Modelo 1.

- Modelo 3:

$$f(x) = \beta_0 + \beta_1 * score_detractor + \beta_2 * WoE_porcentaje_umbral_rsrp$$

Similar al Modelo 2, pero en lugar de **RSRQ**, se incluye la variable optimizada **WoE_porcentaje_umbral_rsrp**. Este modelo permite analizar el impacto de la variable **RSRP** en la predicción de la satisfacción, comparándolo tanto con el Modelo 1 como con el Modelo 2.

- Modelo 4:

$$f(x) = \beta_0 + \beta_1 * WoE_porcentaje_umbral_rsrq + \beta_2 * WoE_porcentaje_umbral_rsrp$$

Este modelo incluye ambas variables optimizadas, **WoE_porcentaje-umbral_rsrq** y **WoE_porcentaje_umbral_rsrp**, pero sin el score de detractores. El objetivo de este modelo es evaluar si la combinación de ambas variables optimizadas por sí solas puede predecir la satisfacción de los clientes, sin necesidad del score tradicional.

- Modelo 5:

$$f(x) = \beta_0 + \beta_1 * score_detractor + \beta_2 * WoE_porcentaje_umbral_rsrq + \beta_3 * WoE_porcentaje_umbral_rsrp$$

Finalmente, este modelo incorpora tanto el score de detractores como las dos variables optimizadas, **WoE_porcentaje_umbral_rsrq** y **WoE_porcentaje_umbral_rsrp**. La intención es observar si la combinación de todas las variables disponibles (score más variables optimizadas) proporciona el mejor rendimiento predictivo en comparación con los modelos anteriores.

Teniendo los modelos definidos, para cada uno de estos se aplicó una regresión logística con validación cruzada (5-fold cross-validation). Los modelos se entrenaron utilizando las combinaciones de variables definidas y se evaluaron en términos de la métrica ROC AUC para medir su desempeño en la predicción de detractores.

Adicionalmente, se generaron curvas ROC y curvas Lift para comparar la capacidad predictiva y la eficiencia de cada modelo. La curva ROC permite visualizar la tasa de verdaderos positivos frente a la tasa de falsos positivos, mientras que la curva Lift muestra la capacidad del modelo para identificar los casos positivos en comparación con una selección aleatoria.

8. Resultados

8.1. Aplicación metodología primera iteración

Tabla 1: Resultados ejecución RSRQ periodo sesiones Noviembre y Diciembre 2023

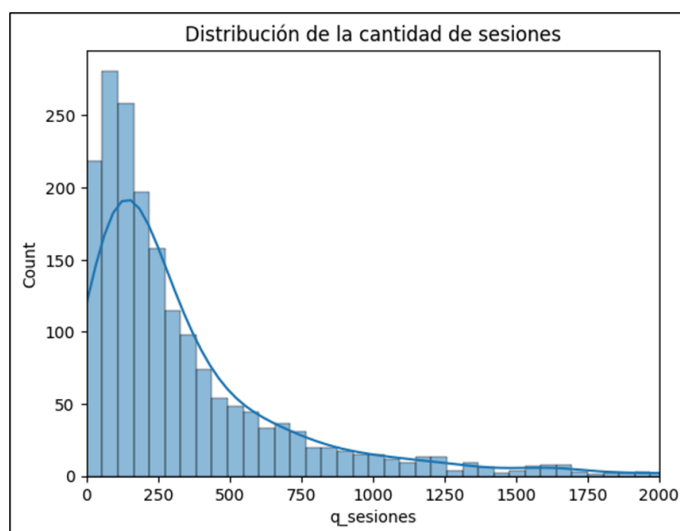
| | |
|--------------------------------------|---------------------------------------|
| Armonía Óptima | (‘30-11-2023’, ‘05-12-2023’, ‘-12.4’) |
| IV Armonía Óptima | 0.11 |
| Notas buscadas en la memoria | 10.340 |
| Reemplazos de armonías en la memoria | 107 |

Fuente: Elaboración propia

Como resultado de aplicar la metodología en una primera versión, se obtiene que la variable predictora óptima de la satisfacción, es decir, “*suma_bajo_umbral_fijo*” entre los días 30 de noviembre de 2023 y 05 de diciembre de 2023, donde el umbral fijo corresponde a -12.4 en la variable RSRQ, tiene un nivel de predicción de la satisfacción de enero 2024 medio bajo.

A continuación, se muestra el resultado de la distribución de la cantidad de sesiones por cliente:

Figura 5: distribución de cantidad de sesiones por cliente



Fuente: Elaboración propia

A partir del gráfico anterior, se observa que la cantidad de sesiones es variable entre los clientes considerados dentro de la muestra.

8.2. Aplicación metodología segunda iteración (final)

8.2.1. Variable RSRQ

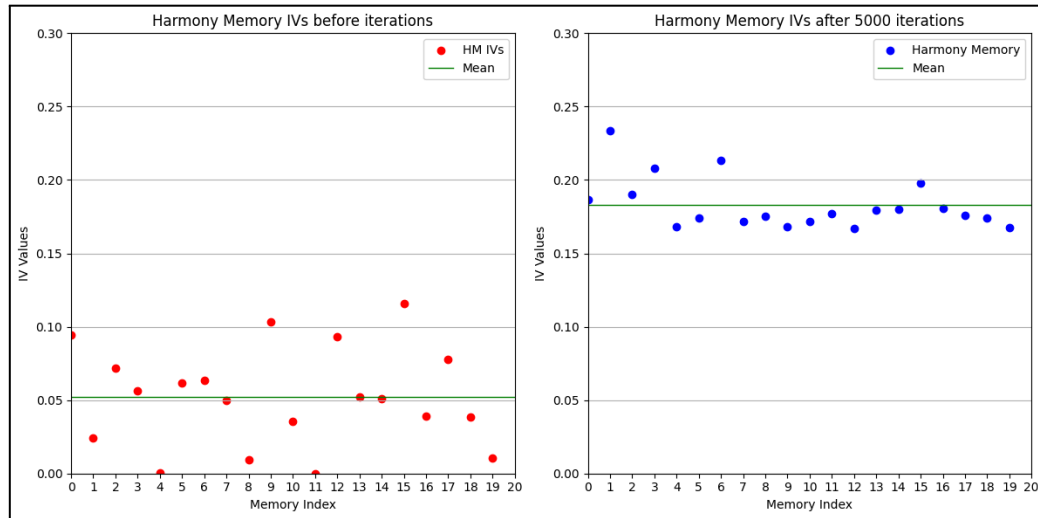
8.2.1.1. Tablón 1

Tabla 2: Resultados ejecución RSRQ tablón 1

| | |
|--------------------------------------|----------------------------------|
| Armonía Óptima | ('2023-12-06', '2023-12-12', -1) |
| IV Armonía Óptima | 0.233 |
| Notas buscadas en la memoria | 10.466 |
| Reemplazos de armonías en la memoria | 110 |

Fuente: Elaboración propia

Figura 6: Comparación IV's de las armonías en memoria antes y después de la ejecución



Fuente: Elaboración propia

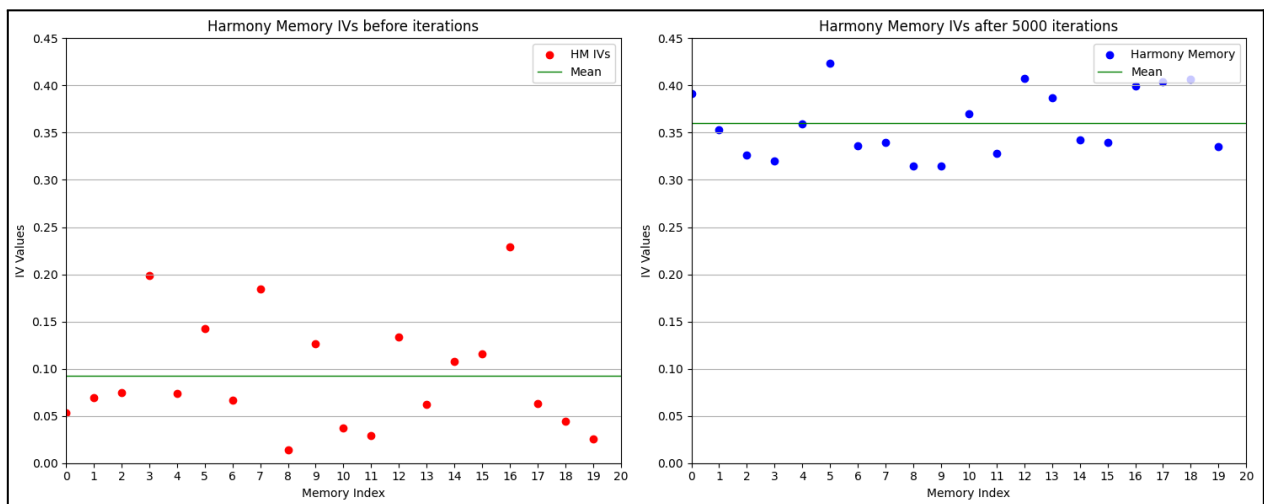
8.2.1.2. Tablón 2

Tabla 3: Resultados ejecución RSRQ tablón 2

| | |
|--------------------------------------|-----------------------------------|
| Armonía Óptima | ('2024-02-29', '2024-03-06', 0.5) |
| IV Armonía Óptima | 0.42 |
| Notas buscadas en la memoria | 10.747 |
| Reemplazos de armonías en la memoria | 111 |

Fuente: Elaboración propia

Figura 7: Comparación IV's de las armonías en memoria antes y después de la ejecución



Fuente: Elaboración propia

En el Tablón 1, los resultados de la ejecución del algoritmo Harmony Search para la variable RSRQ muestran una mejora significativa en los valores de Information Value (IV) de las armonías en memoria. La comparación de los IVs antes y después de la ejecución refleja un aumento en la capacidad predictiva de las armonías generadas. Este aumento sugiere que las armonías optimizadas tienen una mayor capacidad para predecir correctamente la satisfacción del cliente en el periodo de enero 2024.

En el Tablón 2, los resultados también indican una mejora en los IVs de las armonías después de la ejecución del algoritmo. La comparación de los IVs antes y después de la ejecución muestra un incremento notable, lo que refuerza la efectividad del algoritmo en la optimización de las armonías para predecir la satisfacción del cliente en el periodo de abril 2024.

8.2.2. Aplicación metodología a variable RSRP

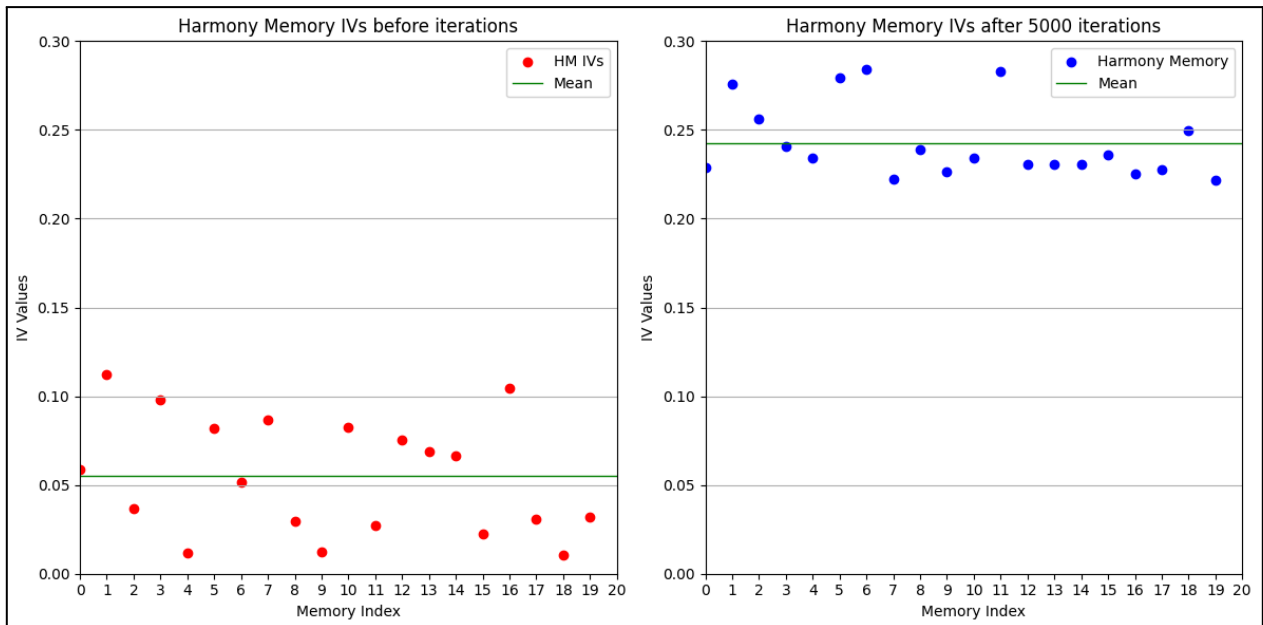
8.2.2.1. Tablón 1

Tabla 4: Resultados ejecución RSRP tablón 1

| | |
|--------------------------------------|----------------------------------|
| Armonía Óptima | (‘2023-11-31’, ‘2023-12-10’, -2) |
| IV Armonía Óptima | 0.28 |
| Notas buscadas en la memoria | 10.464 |
| Reemplazos de armonías en la memoria | 107 |

Fuente: Elaboración propia

Figura 8: Comparación IV's de las armonías en memoria antes y después de la ejecución



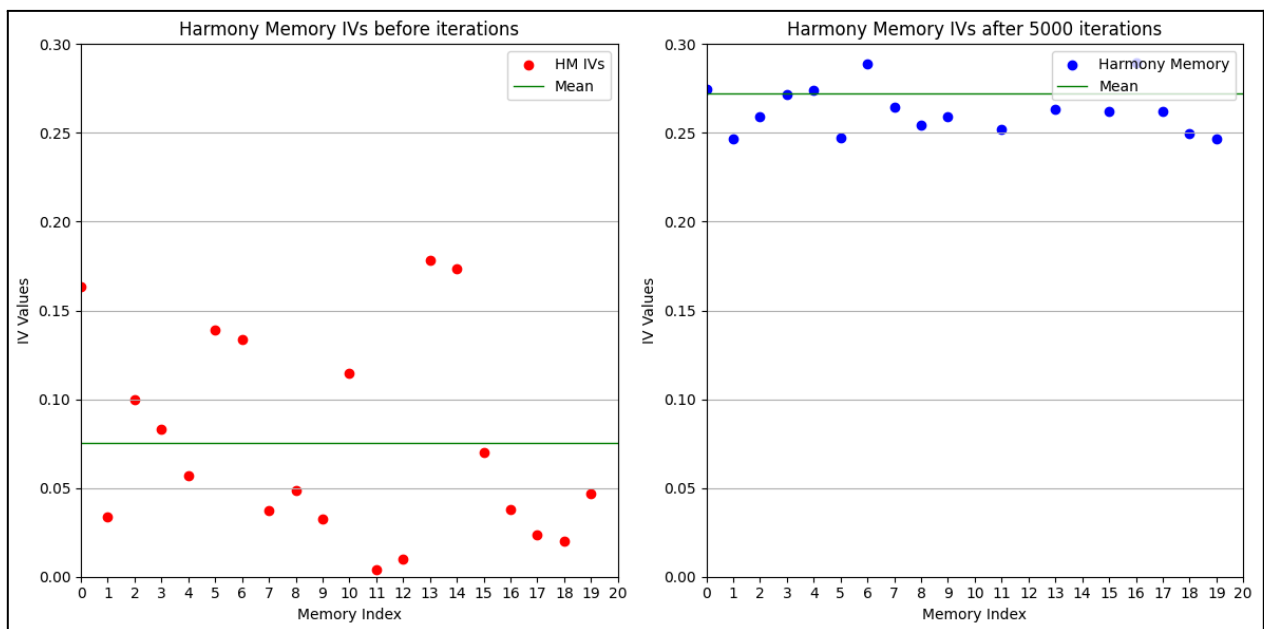
Fuente: Elaboración propia

8.2.2.2. Tablón 2

Tabla 5: Resultados ejecución RSRP tablón 2

| | |
|--------------------------------------|-----------------------------------|
| Armonía Óptima | (‘2024-02-28’, ‘2024-03-17’, 1.5) |
| IV Armonía Óptima | 0.35 |
| Notas buscadas en la memoria | 10.537 |
| Reemplazos de armonías en la memoria | 116 |

Figura 9: Comparación IV's de las armonías en memoria antes y después de la ejecución



Fuente: Elaboración propia

Para la variable RSRP en el Tablón 1, los resultados de la ejecución del algoritmo Harmony Search también muestran una mejora en los IVs de las armonías en memoria. La comparación de los IVs antes y después de la ejecución refleja un aumento en la capacidad predictiva de las armonías generadas, similar a los resultados obtenidos para la variable RSRQ.

En el Tablón 2, los resultados para la variable RSRP muestran una mejora significativa en los IVs después de la ejecución del algoritmo. La comparación de los IVs antes y después de la ejecución demuestra que las armonías optimizadas tienen una mayor capacidad para predecir correctamente la satisfacción del cliente en el periodo de abril 2024, confirmando la efectividad del algoritmo en la optimización de las armonías.

8.3. Pruebas en modelo de predicción actual

8.3.1. Binarización óptima

Figura 10: Resultados proceso Opt Binning variables porcentaje_bajo_umbral RSRP y RSRQ

| Binning table de variable 'porcentaje_bajo_umbral_rsrp' | | | | |
|---|-----------|------------|----------|--------|
| Bin | Count (%) | Event rate | WoE | IV |
| (-inf, 34.81) | 6,3% | 1,1% | 1,97633 | 0,1138 |
| [34.81, 43.32) | 11,9% | 5,3% | 0,37807 | 0,0145 |
| [43.32, 60.44) | 61,8% | 6,9% | 0,09704 | 0,0056 |
| [60.44, 63.94) | 7,7% | 19,8% | -1,11471 | 0,1526 |
| [63.94, 67.85) | 6,7% | 10,3% | -0,34898 | 0,0095 |
| [67.85, inf) | 5,6% | 6,2% | 0,20899 | 0,0023 |

| Binning table de variable 'porcentaje_bajo_umbral_rsrq' | | | | |
|---|-----------|------------|--------|-------|
| Bin | Count (%) | Event rate | WoE | IV |
| (-inf, 61.48) | 9,2% | 1,5% | 1,670 | 0,133 |
| [61.48, 65.35) | 21,5% | 5,2% | 0,395 | 0,028 |
| [65.35, 66.46) | 8,3% | 5,9% | 0,260 | 0,005 |
| [66.46, inf) | 61,0% | 9,4% | -0,252 | 0,043 |

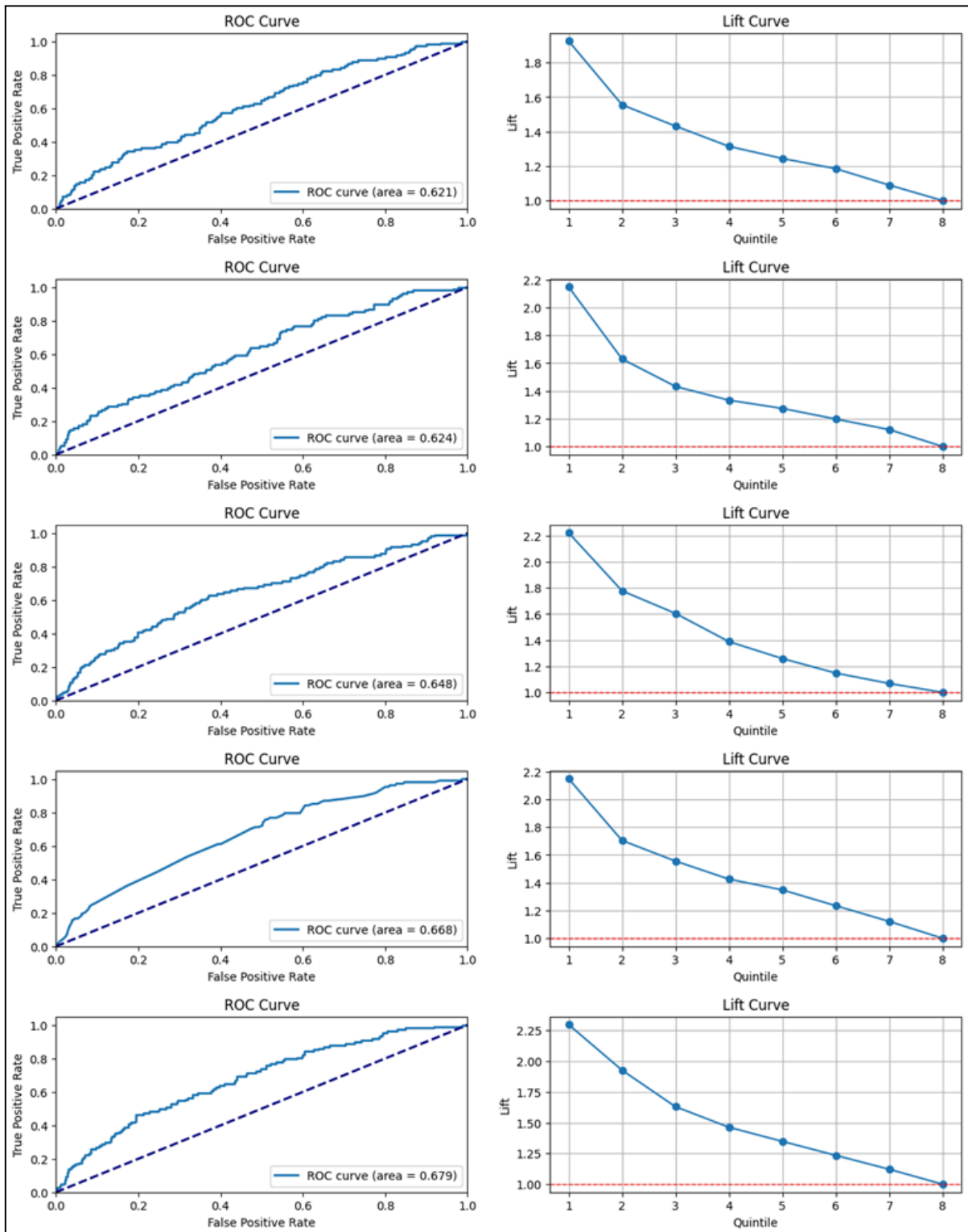
Fuente: Elaboración propia

La variable porcentaje_bajo_umbral_rsrp mide el porcentaje de sesiones en las que la potencia de la señal RSRP de un cliente estuvo por debajo de su umbral. El análisis de binning muestra que los clientes con un porcentaje bajo de sesiones bajo el umbral (rango (-inf, 34.81)) tienen una probabilidad significativamente menor de ser detractores, como lo indica un WoE positivo alto de 1.9763 y una tasa de detractores de 1.1%. A medida que el porcentaje de sesiones bajo el umbral aumenta, el WoE se vuelve negativo (especialmente en el rango [60.44, 63.94)), indicando una mayor probabilidad de detracción.

La variable porcentaje_bajo_umbral_rsrq mide el porcentaje de sesiones en las que la calidad de la señal RSRQ de un cliente estuvo por debajo de su umbral. Los resultados del binning indican que los clientes con un bajo porcentaje de sesiones bajo el umbral (rango (-inf, 61.48)) también tienen una menor probabilidad de ser detractores, reflejado en un WoE positivo alto de 1.670. Sin embargo, conforme aumenta el porcentaje de sesiones bajo el umbral, el WoE disminuye y eventualmente se vuelve negativo en el rango [66.46, inf), lo que indica una mayor probabilidad de insatisfacción.

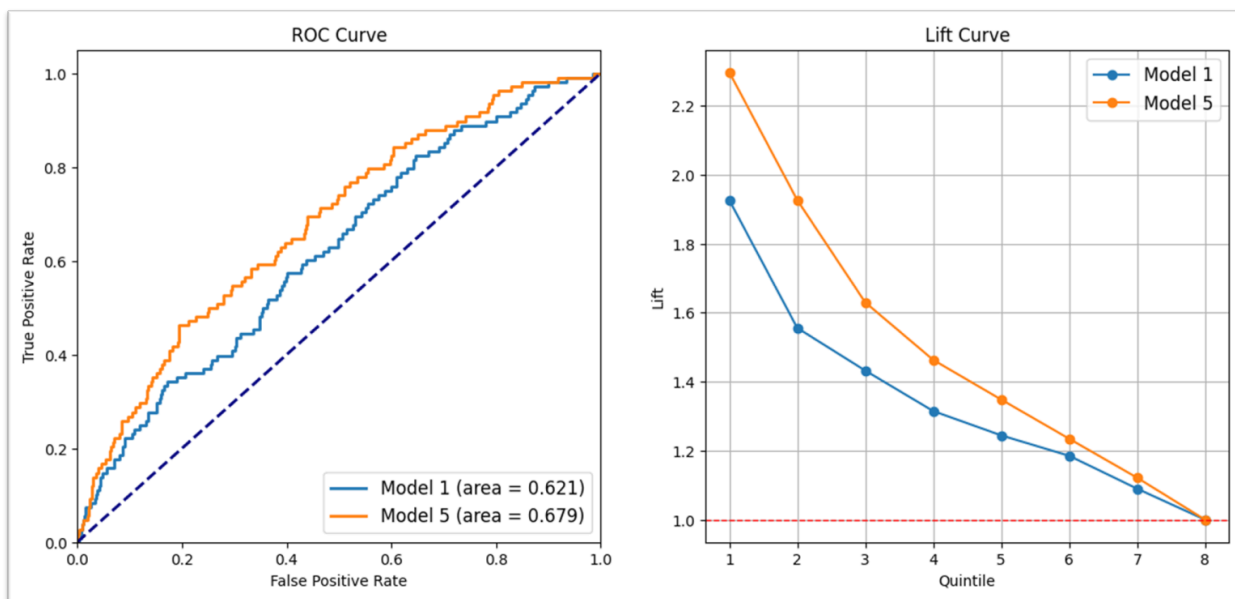
8.3.2. Resultados de predicción modelos

Figura 11: Curvas ROC y curvas Lift de los distintos modelos definidos



Fuente: Elaboración propia

Figura 12: Comparación entre rendimiento del modelo de predicción actual y modelo que incorpora variables RSRQ y RSRQ optimizadas por metodología desarrollada



Fuente: Elaboración propia

En esta sección se presentan los resultados obtenidos tras el entrenamiento y la evaluación de los cinco modelos definidos para la predicción de la satisfacción de los clientes. Los modelos se evaluaron mediante validación cruzada utilizando la métrica ROC AUC, y se generaron curvas ROC y Lift para cada modelo. A continuación se detallan los resultados obtenidos.

El **Modelo 1**, que solo incluye la variable `score_detractor`, obtuvo una ROC AUC promedio de 0.627. Este modelo mostró una precisión en la clasificación de los detractores de 0.10 y un recall de 0.52. La curva ROC generada para este modelo mostró un área bajo la curva (AUC) de 0.621, mientras que la curva Lift presentó un lift inicial de 1.93 en el primer quintil. Estos resultados indican una capacidad moderada del modelo para identificar correctamente a los detractores.

El **Modelo 2**, que incorpora las variables `score_detractor` y `WoE_porcentaje_bajo_umbral_rsrq`, mejoró ligeramente la ROC AUC promedio a 0.658. Sin embargo, la precisión para los detractores disminuyó a 0.09, aunque el recall aumentó significativamente a 0.77. La curva ROC mostró un AUC de 0.624, y la curva Lift mostró un lift inicial de 2.15 en el primer quintil. Esto sugiere que, aunque el modelo mejoró en la identificación de detractores, la precisión de las predicciones se vio comprometida.

El **Modelo 3**, que añade `WoE_porcentaje_bajo_umbral_rsrp` a `score_detractor`, alcanzó una ROC AUC promedio de 0.667. La precisión para los detractores fue de 0.15, con un recall de 0.30. La curva ROC presentó un AUC de 0.648, y la curva Lift mostró un lift inicial de 2.22 en el primer quintil. Este modelo demostró una mejora tanto en la precisión como en la capacidad de identificar detractores, en comparación con los modelos anteriores.

El **Modelo 4**, que incluye las variables transformadas `WoE_porcentaje_bajo_umbral_rsrq` y `WoE_porcentaje_bajo_umbral_rsrp`, obtuvo una ROC AUC promedio de 0.670. La precisión para los detractores fue de 0.11, y el recall alcanzó 0.76. La curva ROC mostró un AUC de 0.668, mientras que la curva Lift presentó un lift inicial de 2.15 en el primer quintil. Estos resultados indican que las variables transformadas por sí solas tienen un impacto significativo en la capacidad del modelo para identificar detractores.

Finalmente, el **Modelo 5**, que combina `score_detractor` con ambas variables transformadas (`WoE_porcentaje_bajo_umbral_rsrq` y `WoE_porcentaje_bajo_umbral_rsrp`), obtuvo la mejor ROC AUC promedio de 0.695. La precisión para los detractores fue de 0.11, con un recall de 0.74. La curva ROC generada para este modelo mostró un AUC de 0.679, y la curva Lift presentó un lift inicial de 2.30 en el primer quintil. Este modelo demostró ser el más robusto en términos de precisión y capacidad de identificación de detractores.

9. Discusión

En el desarrollo de este proyecto se tomaron diversas decisiones metodológicas y técnicas que, aunque fueron justificadas en su momento, merecen una reflexión crítica para identificar posibles mejoras y lecciones aprendidas. Al cuestionar estas decisiones, se pueden resaltar tanto las fortalezas como las áreas de oportunidad que podrían haber sido manejadas de manera diferente, de cara a mejorar los resultados y la aplicabilidad del modelo desarrollado.

Una de las decisiones más relevantes fue el uso de solo dos tabloneros de datos, cubriendo periodos de dos meses cada uno, para entrenar la metodología. Esta decisión, tomada conjuntamente por el equipo de Advanced Analytics de Entel y el estudiante, se basó en las limitaciones impuestas por la política interna de la compañía sobre la retención de datos. Sin embargo, al analizar más a fondo, surge la preocupación de si esta ventana de tiempo limitada es realmente representativa del comportamiento general de los clientes. Al restringir el periodo de entrenamiento a solo cuatro meses, se corre el riesgo de introducir sesgos estacionales, afectando la capacidad del modelo para generalizar los resultados a lo largo de todo el año. Esta limitación podría haber sido mitigada solicitando excepciones en la política de retención de datos para utilizar un rango temporal más amplio, o incorporando técnicas de simulación para estimar datos adicionales y evaluar la estabilidad del modelo bajo diferentes condiciones.

Otra decisión clave fue la elección del algoritmo Harmony Search para la optimización de los parámetros de las variables de red. Aunque esta metaheurística es flexible y adecuada para el tipo de problemas abordados en este proyecto, la falta de una comparación exhaustiva con otros algoritmos de optimización, como los Algoritmos Genéticos o el Recocido Simulado, limita nuestra capacidad para afirmar con certeza que Harmony Search fue la mejor opción disponible. En retrospectiva, habría sido valioso realizar un análisis comparativo entre diferentes algoritmos, no solo para identificar la mejor metodología, sino también para proporcionar un respaldo más sólido a la elección final.

Los resultados del proyecto también se ven influidos por la representatividad de los datos utilizados. La metodología desarrollada se basó en datos de sesiones móviles de clientes y encuestas de satisfacción, limitados a un periodo de cuatro meses. Esto introduce un posible sesgo, ya que no captura variaciones temporales más amplias ni circunstancias excepcionales que podrían afectar la satisfacción del cliente. Además, no se evaluó el impacto de posibles fallas en la red de Entel, lo que podría comprometer la fiabilidad del modelo en condiciones fuera de lo normal. La falta de mecanismos para ajustar automáticamente los parámetros en respuesta a eventos excepcionales podría resultar en predicciones incorrectas, afectando negativamente la toma de decisiones basada en el modelo.

La implementación y el mantenimiento del modelo desarrollado también presentan desafíos prácticos que deben considerarse. Estos incluyen la gestión de versiones del modelo, la frecuencia de las actualizaciones necesarias y cómo manejar cambios en los datos o en el entorno operativo que podrían requerir ajustes en la metodología. La sostenibilidad del modelo en un entorno operativo real depende en gran medida de la capacidad de la organización para integrarlo en su flujo de trabajo de manera eficiente, asegurando la calidad y disponibilidad

continúa de los datos y capacitando al personal para realizar ajustes y actualizaciones según sea necesario.

Finalmente, es importante reconocer que la metodología de Harmony Search, aunque efectiva, puede introducir sesgos dependiendo de la configuración de los parámetros y la naturaleza de las soluciones exploradas. La falta de un análisis de sensibilidad que evalúe cómo pequeños cambios en los parámetros afectan los resultados limita la comprensión completa de la robustez del modelo. Es posible que las soluciones encontradas no sean las globalmente óptimas, sino solo localmente óptimas dentro del espacio de búsqueda definido. En futuras investigaciones, sería valioso explorar la robustez de los resultados bajo diferentes configuraciones de parámetros y comparar estos con otras metodologías de optimización.

En resumen, si se volviera a comenzar este proyecto, varias decisiones clave se reconsiderarían para mejorar la robustez y aplicabilidad de los resultados. La ampliación del periodo de entrenamiento, la exploración de otras metaheurísticas y un enfoque más profundo en la sostenibilidad operativa del modelo son aspectos que podrían haber fortalecido aún más el impacto del proyecto. No obstante, los resultados obtenidos proporcionan una base sólida para futuras mejoras en la predicción de la satisfacción del cliente en la compañía, con un enfoque claro en la optimización de las variables de red móvil.

10. Conclusiones

El desarrollo y la aplicación de la metodología basada en el algoritmo Harmony Search han permitido identificar parámetros óptimos para maximizar la capacidad predictiva de las variables RSRQ y RSRP en la satisfacción del cliente postpago en una compañía de telecomunicaciones. Los resultados obtenidos muestran que la implementación de esta metaheurística ha sido exitosa en mejorar los valores de Information Value (IV), estableciendo una base sólida para la predicción de clientes detractores.

Específicamente, la metodología permitió seleccionar combinaciones óptimas de parámetros de temporalidad y umbrales, alcanzando un alto nivel de predicción de la satisfacción, evaluado a través de la métrica IV. La robustez de esta metodología se confirmó al aplicarla a distintos conjuntos de datos, lo que resultó en mejoras significativas en los niveles de predicción al utilizar las variables RSRP y RSRQ.

El análisis de las curvas Lift mostró una mejora notable, especialmente en el primer quintil, donde el valor subió de 1.9 a 2.25. Este resultado indica que la metodología desarrollada tiene un potencial considerable para mejorar la predicción de la satisfacción del cliente, permitiendo identificar más eficazmente a los detractores. Esto sugiere que, en futuras campañas con restricciones presupuestarias, se podría focalizar en el primer quintil de la muestra de datos, aumentando significativamente la probabilidad de identificar a los detractores y, por ende, posibilitando acciones más precisas y eficientes.

Además, es importante considerar el impacto comercial directo que esta metodología podría tener para la empresa. De acuerdo con la información del ARPU de clientes postpago, mejorar la predicción de detractores podría tener un efecto indirecto en la retención de clientes. Por ejemplo, entre marzo y diciembre de 2023, el 28% de los clientes que realizaron portabilidad lo hicieron debido a problemas con la señal y cobertura de la red móvil 4G, lo que representa aproximadamente 1,567 clientes insatisfechos. Aunque no se puede afirmar de manera directa que esta metodología reduzca el churn, la mejora en la predicción podría contribuir a una disminución de la fuga de clientes mediante la implementación de acciones correctivas basadas en los resultados obtenidos.

Sin embargo, es crucial realizar pruebas adicionales en el modelo de predicción actual para evaluar la utilidad y efectividad de las variables optimizadas incluidas en la metodología. La implementación exitosa de esta metodología dentro de la organización dependerá de la capacidad para integrarla en los procesos operativos existentes, lo cual incluye la gestión de actualizaciones, la formación del personal, y la superación de barreras tecnológicas y operativas.

En términos del cumplimiento del objetivo general, se ha desarrollado una metodología efectiva para encontrar el mayor nivel de predicción de la satisfacción de los clientes postpago utilizando las variables de red móvil 4G RSRQ y RSRP. La metodología basada en Harmony Search ha demostrado ser una herramienta robusta y flexible para optimizar la capacidad predictiva de estas variables, lo cual puede contribuir significativamente a la mejora de la satisfacción del cliente en la compañía de telecomunicaciones.

Finalmente, se propone una mejora futura en la metodología respecto a la determinación de los parámetros óptimos. Actualmente, el periodo de datos óptimo considera todos los días comprendidos entre un día de inicio y un día de fin. Sin embargo, sería beneficioso explorar combinaciones de días no necesariamente consecutivos, para identificar periodos específicos que tengan un mayor impacto en la satisfacción del cliente. Esta mejora permitiría una visión más detallada y precisa de los periodos críticos, optimizando aún más la capacidad predictiva de las variables de red móvil.

11. Bibliografía

Referencias

- Amazon Web Services. (2024.). *¿Qué es ETL? - Explicación de extracción, transformación y carga (ETL)*. AWS. Retrieved May 17, 2024, from <https://aws.amazon.com/es/what-is/etl/>
- Amazon Web Services. (2024). *¿Qué es la regresión logística? - Explicación del modelo de regresión logística*. AWS. Retrieved June 28, 2024, from <https://aws.amazon.com/es/what-is/logistic-regression/>
- Analytics Vidhya. (2021, July 20). *Understand Weight of Evidence and Information Value!* Analytics Vidhya. Retrieved June 28, 2024, from <https://www.analyticsvidhya.com/blog/2021/06/understand-weight-of-evidence-and-information-value/>
- AWS. (2023). *Consultas de datos al instante | Análisis de datos SQL | Amazon Athena*. AWS. Retrieved June 28, 2024, from <https://aws.amazon.com/es/athena/>
- AWS. (2023). *¿Qué es AWS? - Computación en la nube con Amazon Web Services*. AWS. Retrieved June 28, 2024, from <https://aws.amazon.com/es/what-is-aws/>
- AWS. (2023). *¿Qué es la ciencia de datos? - Explicación de la ciencia de datos*. AWS. Retrieved August 29, 2024, from <https://aws.amazon.com/es/what-is/data-science/>
- Cognizant. (2024). *¿Qué es el data analytics?* Cognizant. Retrieved August 29, 2024, from <https://www.cognizant.com/es/es/glossary/data-analytics>
- Gamco. (2021). *Qué es Metaheurísticas Concepto y definición. Glosario*. Gamco. Retrieved August 29, 2024, from <https://gamco.es/glosario/metaheuristicas/>
- Gareth, J., Witten, D., Hastie, T., & Tibshirani, R. (2013, Junio 24). *1 An Introduction to Statistical Learning*. Squarespace. Retrieved August 29, 2024, from <https://static1.squarespace.com/static/5ff2adbe3fe4fe33db902812/t/6009dd9fa7bc363aa822d2c7/1611259312432/ISLR+Seventh+Printing.pdf>

- Geem, Z. W., & Kim, J. H. (2001, Febrero). *A New Heuristic Optimization Algorithm: Harmony Search*. ResearchGate. Retrieved May 17, 2024, from https://www.researchgate.net/publication/220164617_A_New_Heuristic_Optimization_Algorithm_Harmony_Search
- Grupo Entel. (2024, Abril). *Memoria Integrada 2023*. Entel.modocdn. Retrieved May 17, 2024, from https://entel.modyocdn.com/uploads/5e637cce-97a7-49a1-9ca2-dce0b54dadfb/original/Memoria_Entel_2023.pdf
- Guide, S., & Bhalla, D. (2015, Marzo). *Weight of Evidence (WOE) and Information Value (IV) Explained*. ListenData. Retrieved May 17, 2024, from <https://www.listendata.com/2015/03/weight-of-evidence-woe-and-information.html>
- IBM. (2024.). *¿Qué es el análisis predictivo?* IBM. Retrieved May 17, 2024, from <https://www.ibm.com/es-es/topics/predictive-analytics>
- International Journal of Wireless & Mobile Networks. (2015, Noviembre 04). *SINR, RSRP, RSSI AND RSRQ MEASUREMENTS IN LONG TERM EVOLUTION NETWORKS*. SINR, RSRP, RSSI AND RSRQ MEASUREMENTS IN LONG TERM EVOLUTION NETWORKS. Retrieved May 17, 2024, from <https://www.airccse.org/journal/jwmn/7415ijwmn09.pdf>
- Kulkarni, K. (2021, July 20). *Understand Weight of Evidence and Information Value!* Analytics Vidhya. Retrieved July 22, 2024, from <https://www.analyticsvidhya.com/blog/2021/06/understand-weight-of-evidence-and-information-value/>
- Microsoft. (2023.). *Visual Studio Code*. Visual Studio: IDE y Editor de código para desarrolladores de software y Teams. Retrieved May 17, 2024, from <https://visualstudio.microsoft.com/es/>
- Qualtrix.xm. (2024). *Encuesta de satisfacción del cliente: qué es y cómo se hace* | Qualtrics. Qualtrics. Retrieved June 28, 2024, from

<https://www.qualtrics.com/es/gestion-de-la-experiencia/cliente/encuesta-de-satisfaccion-del-cliente/>

SAP. (2024). *¿Qué es machine learning? | Definición, tipos y ejemplos*. SAP. Retrieved August 29, 2024, from

<https://www.sap.com/latinamerica/products/artificial-intelligence/what-is-machine-learning.html>

SAS. (2024). *Big Data: Qué es y por qué importa*. SAS Institute. Retrieved August 29, 2024, from https://www.sas.com/es_mx/insights/big-data/what-is-big-data.html

Siddiqi. (2006). *Siddiqi, N. (2006). Credit Risk Scorecards, Developing and Implementing Intelligent Credit Scoring*. Hoboken, NJ John Wiley & Sons, Inc. - References. Scientific Research Publishing. Retrieved July 22, 2024, from

<https://www.scirp.org/reference/referencespapers?referenceid=2834218>

Singh, N. (2023, September 25). *Métricas De Evaluación De Modelos En El Aprendizaje Automático*. DataSource.ai. Retrieved June 28, 2024, from

<https://www.datasource.ai/es/data-science-articles/metricas-de-evaluacion-de-modelos-en-el-aprendizaje-automatico>

Subsecretaría de Telecomunicaciones de Chile. (2023). *Informe Trimestral del Sector Telecomunicaciones - Tercer Trimestre 2023*.

https://www.subtel.gob.cl/wp-content/uploads/2023/12/PPT_Series_SEPTIEMBRE_2023_V0.pdf

Universidad Europea. (2021, December 30). *¿Qué es SQL y para qué sirve? | Blog UE*.

Universidad Europea. Retrieved May 17, 2024, from

<https://universidadeuropea.com/blog/lenguaje-programacion-sql/>

12. Anexos

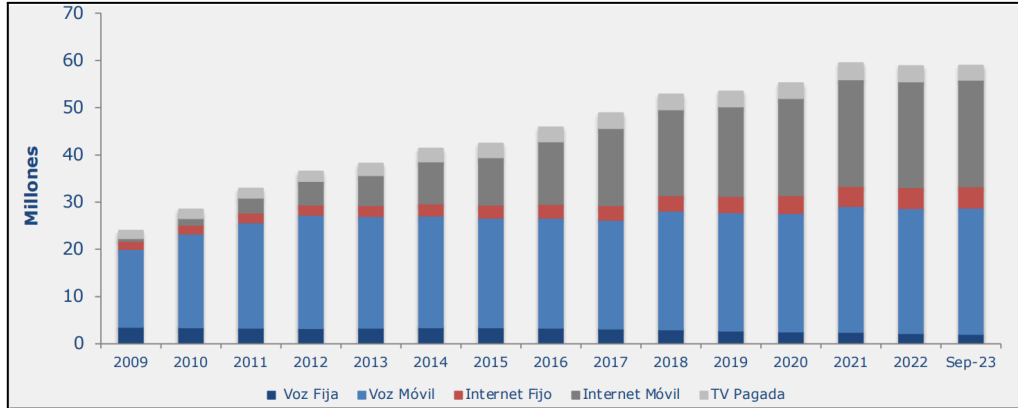
Anexo A: Principales servicios del sector de telecomunicaciones, total de usuarios, penetración internet móvil y participación de mercado de tráfico móvil por operador.

Figura 13: Resumen Principales Servicios

| Servicio | Unidad | 3T-2022 | 3T-2023 | Variación interanual |
|---------------------------|---------------------|----------|----------|----------------------|
| Internet fija | Miles | 4.481 | 4.507 | 0,6% |
| Internet móvil (3G+4G+5G) | Miles | 22.429 | 22.554 | 0,6% |
| 4G | Miles | 19.674 | 18.395 | -6,5% |
| 5G | Miles | 1.587 | 3.144 | 98,2% |
| Internet móvil (tráfico) | Miles de Tera Bytes | 3.586,5 | 4.449,7 | 24,1% |
| Internet fijo (tráfico) | Miles de Tera Bytes | 18.222,7 | 22.238,8 | 22,0% |
| Telefonía fija | Miles | 2.319 | 2.021 | -12,9% |
| Telefonía móvil | Miles | 26.434 | 26.754 | 1,2% |
| TV Pago | Miles | 3.342 | 3.121 | -6,6% |

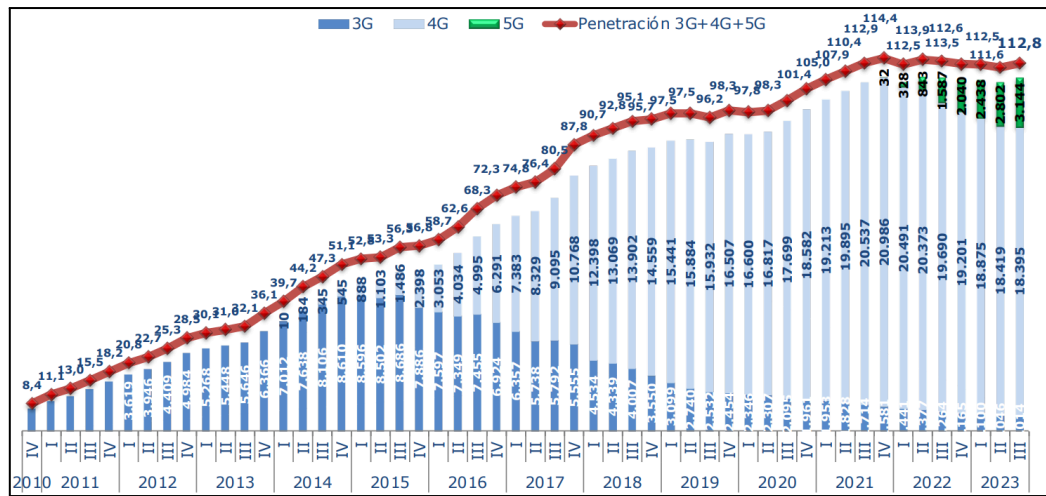
Fuente: Informe Trimestral Subtel, Septiembre 2023

Figura 14: Total de servicios a septiembre de 2023



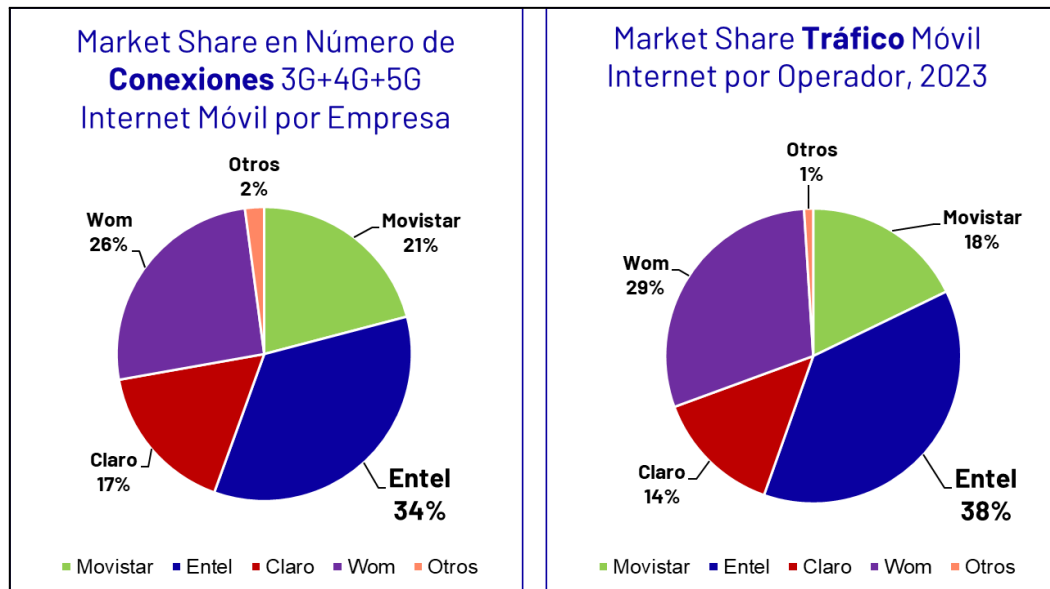
Fuente: Informe Trimestral Subtel, Septiembre 2023

Figura 15: Conexiones y penetración cada 100 habitantes



Fuente: Informe Trimestral Subtel, Septiembre 2023

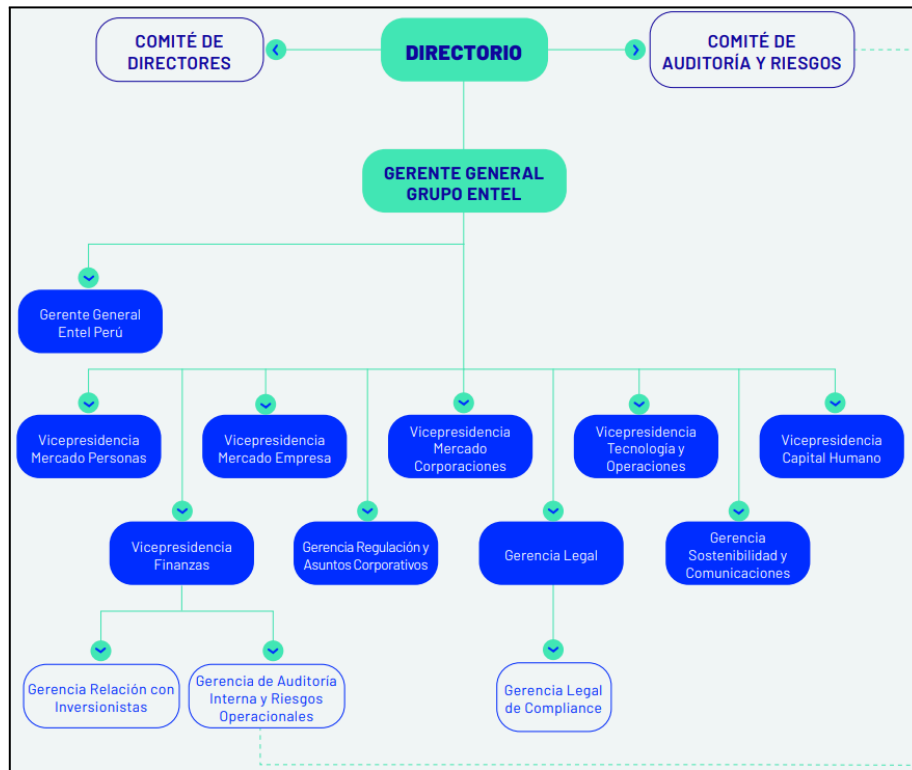
Figura 16: Participación de mercado número de conexiones móviles y tráfico móvil por operador



Fuente: Informe Trimestral Subtel, Septiembre 2023

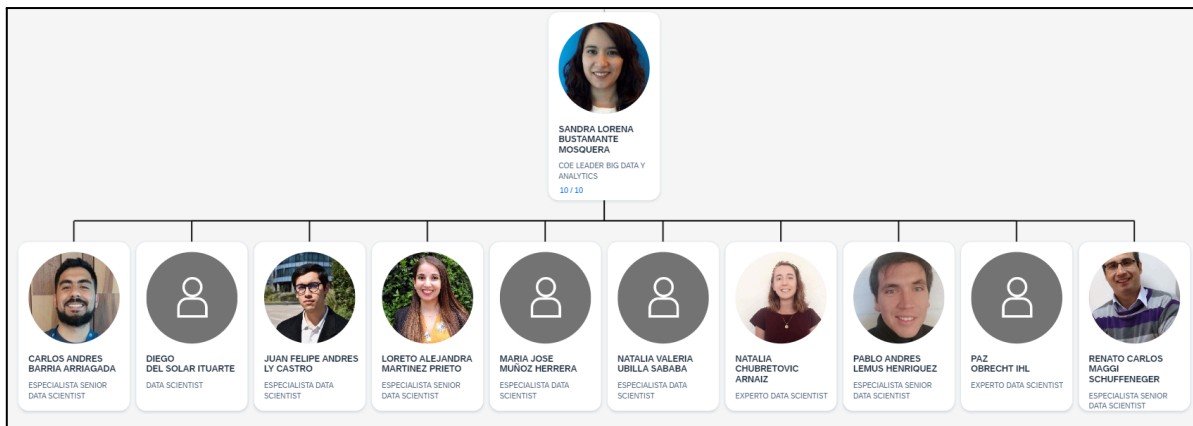
Anexo B: Organigrama Grupo Entel 2023

Figura 17: Organigrama Grupo Entel



Fuente: Memoria Anual Entel 2023

Figura 18: Organigrama área Big Data & Analytics B2C



Fuente: Información Corporativa Entel Chile

Anexo C: Modelo móvil satisfacción de clientes postpago Entel

Descripción del modelo: El modelo de predicción de detractores desarrollado por Entel es una herramienta analítica diseñada para identificar y predecir cuáles clientes tienen una alta probabilidad de convertirse en detractores. Este modelo se basa en un enfoque de machine learning, utilizando un modelo de clasificación específico, el *BalancedRandomForestClassifier*. El modelo genera un score de detractor, que es una probabilidad asignada a cada cliente, indicando la posibilidad de que se convierta en un detractor en el próximo mes. Este score se expresa como un valor entre 0 y 1, donde 0 indica ninguna probabilidad y 1 indica una alta probabilidad de ser detractor.

Objetivo del modelo: El objetivo principal del modelo es asignar un score de probabilidad a cada cliente de la base completa de datos de Entel, indicando la probabilidad de que un cliente sea un detractor en el próximo mes. Esto permite a la compañía anticipar posibles pérdidas de clientes y tomar medidas proactivas para mejorar la satisfacción del cliente y reducir la tasa de abandono.

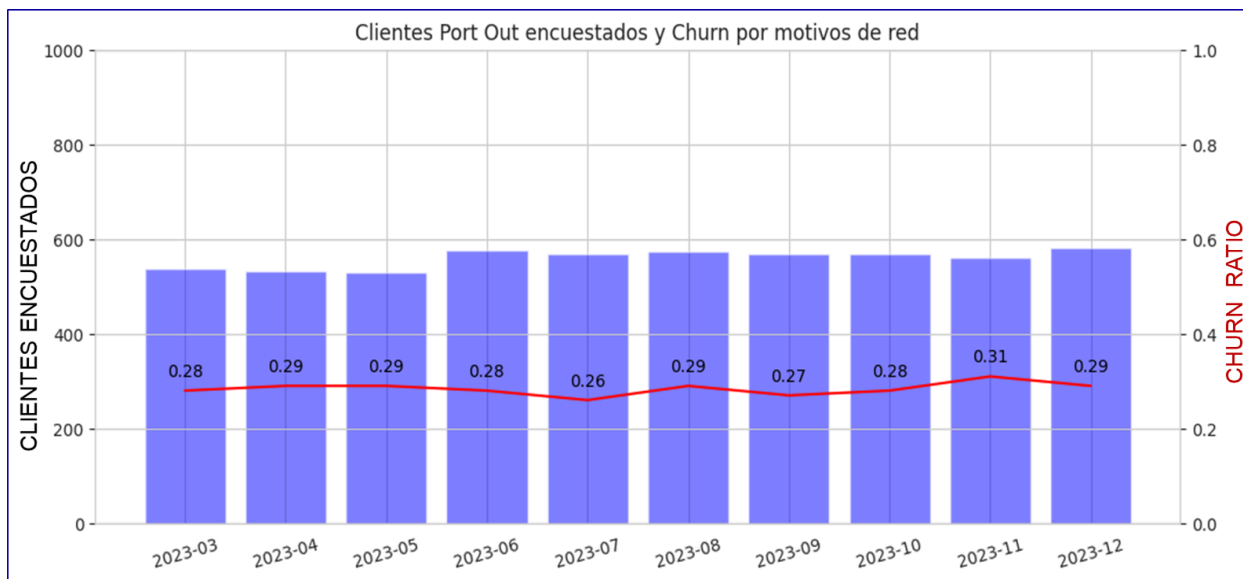
Los datos que se utilizan para el desarrollo del modelo son los siguiente:

| Fuente de Datos | Descripción |
|-----------------------|--|
| Calidad de Red | Accesibilidad, disponibilidad y recursos de celda. |
| Uso de datos | Tasa de uso por tecnología, tiempo estimado de navegación. |
| Movilidad | Selección de celdas más traficadas, score de cobertura, tasa de tráfico en zonas de menor cobertura. |
| Uso de Voz | Tráfico de llamadas, compañías favoritas de llamadas entrantes y salientes, cortes de llamadas. |
| Información Comercial | Antigüedad, cantidad de líneas, valor del plan, cantidad de reclamos e interacciones, recargas en prepago. |

El modelo se entrena utilizando un conjunto de datos históricos, incluyendo aproximadamente 8 meses de datos y 3000 clientes encuestados por mes. La variable objetivo es determinar si un cliente es un detractor o no. Para esto, se utilizó el algoritmo *BalancedRandomForestClassifier*, que permitió alcanzar una AUC (Área Bajo la Curva ROC) de 0.66 en el set de entrenamiento y 0.65 en el set de testeo, indicando un desempeño razonable del modelo en la discriminación entre clientes detractores y no detractores.

Anexo D: Clientes PortOut Encuestados y Churn por Motivos de Red

Figura 19: Clientes PortOut y Churn por red



Fuente: Análisis PortOut por Red, Analytics Entel Chile 2023.

Anexo E: Importancia de variables en modelo de predicción actual (julio 2024) de la satisfacción

Figura 20: Importancia de variables en Modelo Score Satisfacción

| Top 15 variables más importantes | | Valor promedio de score por decil | | | | | | | | | | |
|--|-------------|-----------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|-------------|
| Variables | Importancia | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Tecnología |
| Percentil 5 RSRP | | -114,99 | -115,80 | -116,18 | -116,86 | -117,84 | -118,56 | -119,14 | -119,70 | -120,41 | -121,81 | 4G |
| Percentil 95 RSRQ | | -84,22 | -83,57 | -83,55 | -84,18 | -84,81 | -85,70 | -86,24 | -88,17 | -88,17 | -91,82 | 4G |
| Tasa celdas 4G-5G vs 3G día | | 0,11 | 0,11 | 0,12 | 0,14 | 0,17 | 0,19 | 0,23 | 0,25 | 0,28 | 0,32 | 3G, 4G y 5G |
| Tasa celdas 4G y 5G noche | | 0,92 | 0,91 | 0,91 | 0,89 | 0,86 | 0,82 | 0,81 | 0,80 | 0,80 | 0,76 | 4G y 5G |
| Porcentaje sesiones bajo RSRQ -16 | | 41,44 | 59,86 | 72,63 | 87,66 | 104,85 | 120,38 | 139,96 | 167,68 | 214,18 | 366,02 | 4G |
| Porcentaje sesiones bajo RSRP -116 | | 8,34 | 8,52 | 8,78 | 9,83 | 11,61 | 13,54 | 15,53 | 17,58 | 20,61 | 28,90 | 4G |
| Tasa celdas 4G y 5G día | | 0,91 | 0,90 | 0,90 | 0,88 | 0,85 | 0,82 | 0,81 | 0,81 | 0,80 | 0,78 | 4G y 5G |
| Tasa celdas 4G-5G vs 3G noche | | 0,08 | 0,09 | 0,11 | 0,12 | 0,15 | 0,17 | 0,22 | 0,25 | 0,28 | 0,37 | 3G, 4G y 5G |
| Tasa celdas 4G-5G vs 3G | | 0,12 | 0,12 | 0,13 | 0,14 | 0,18 | 0,21 | 0,26 | 0,27 | 0,31 | 0,33 | 3G, 4G y 5G |
| Porcentaje tiempo tráfico en celdas 4G | | 0,97 | 0,97 | 0,97 | 0,95 | 0,92 | 0,89 | 0,86 | 0,84 | 0,82 | 0,75 | 4G |
| Handovers 3G-> 2G | | 1,70 | 2,20 | 2,51 | 3,51 | 4,75 | 6,13 | 7,62 | 9,73 | 16,19 | 52,07 | 4G |
| std_porcentaje_celda_mas_traficada | | 0,36 | 0,35 | 0,34 | 0,33 | 0,33 | 0,32 | 0,32 | 0,31 | 0,30 | 0,28 | 3G y 4G |
| sum_porcentaje_celda_mas_traficada | | 23,24 | 27,02 | 28,87 | 30,00 | 30,66 | 30,66 | 31,68 | 32,64 | 33,97 | 36,09 | 3G y 4G |
| Handovers 4G-> 2G | | 195,50 | 308,41 | 387,02 | 473,80 | 582,39 | 682,56 | 831,24 | 1061,71 | 1311,19 | 1676,80 | 3G y 4G |
| Promedio RSRP | | -101,91 | -102,17 | -102,38 | -103,15 | -104,19 | -105,07 | -106,67 | -107,77 | -106,67 | -110,29 | 4G |

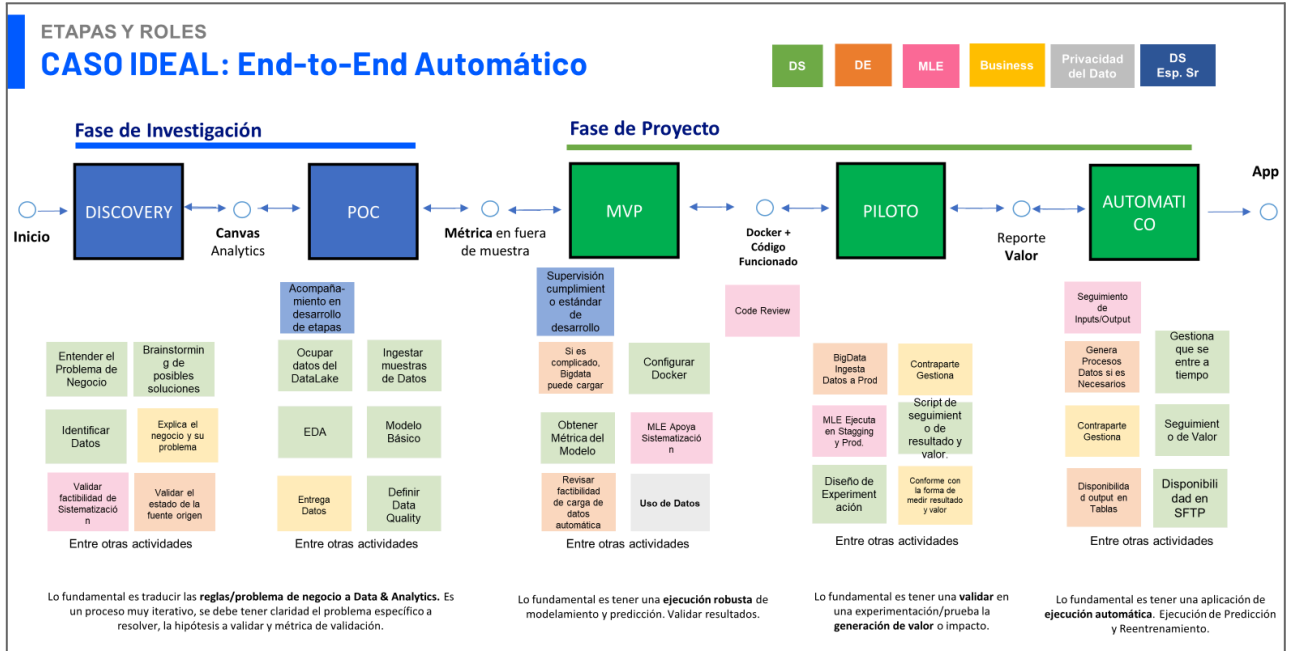
Menor prob. de ser detractor

Mayor prob. de ser detractor

Fuente: Modelo score satisfacción 2024, Advanced Analytics Entel

Anexo F: Metodología desarrollo de proyecto

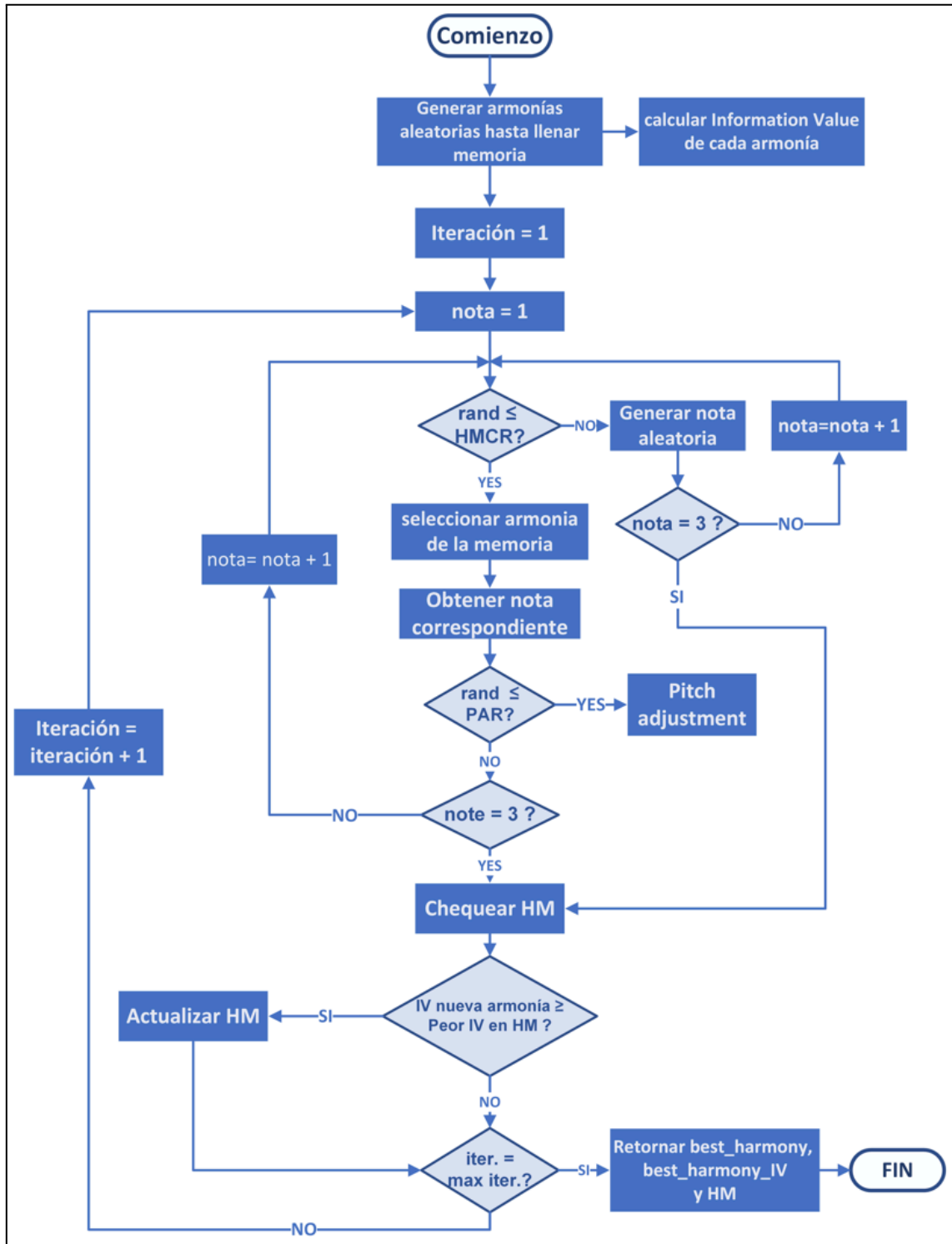
Figura 21: Metodología Advanced Analytics Entel



Fuente: Elaboración propia de Grupo entel, área AA.

Anexo G: Flujo ejecución Algoritmo harmony Search

Figura 22: Flujo funcionamiento implementación Harmony Search



Fuente: Elaboración propia

Anexo G: Opt Binning y métrica WoE

Como indica IBM, SPSS Statistics La binarización óptima es la discretización óptima de una variable en intervalos dados un objetivo numérico discreto o continuo. OptBinning es una biblioteca escrita en Python que implementa una formulación rigurosa y flexible de programación matemática para resolver el problema de la binarización óptima para un objetivo de tipo binario, continuo o multiclase.

Según (Analytics Vidhya, 2021), la técnica WOE permite realizar una transformación que logra aumentar el rendimiento de los Modelos de regresión Logística evitando el tema de outliers distribuciones sesgadas.

Para este proceso, se aplica la binarización de una variable a través de la siguiente fórmula:

$$WoE = \ln \frac{\% \text{ eventos}}{\% \text{ no eventos}}$$

Si una característica continua no tiene una relación lineal con el logaritmo de las probabilidades, la característica puede ser dividida en grupos y una nueva característica creada reemplazando cada bin por su valor de WoE puede ser utilizada en lugar de la característica original. Por lo tanto, el WoE es un buen método de transformación de variables para regresión logística.

Anexo H: Comparación de la memoria de armonías antes y después de iterar en la ejecución de la metodología sobre las variables RSRQ y RSRP

Figura 23: Comparación mejor y peor armonía de la memoria, ejecución RSRQ tablón ipsos enero 2024

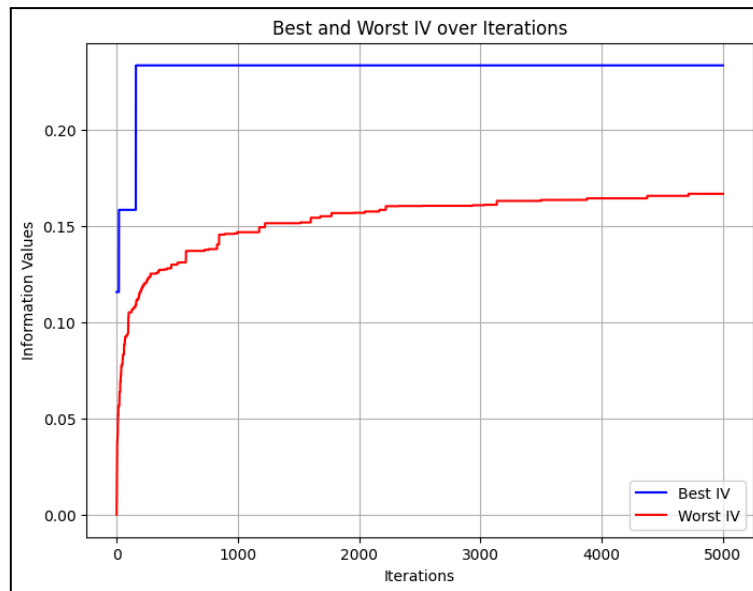


Figura 24: Comparación mejor y peor armonía de la memoria, ejecución RSRQ tablón ipsos Abril 2024

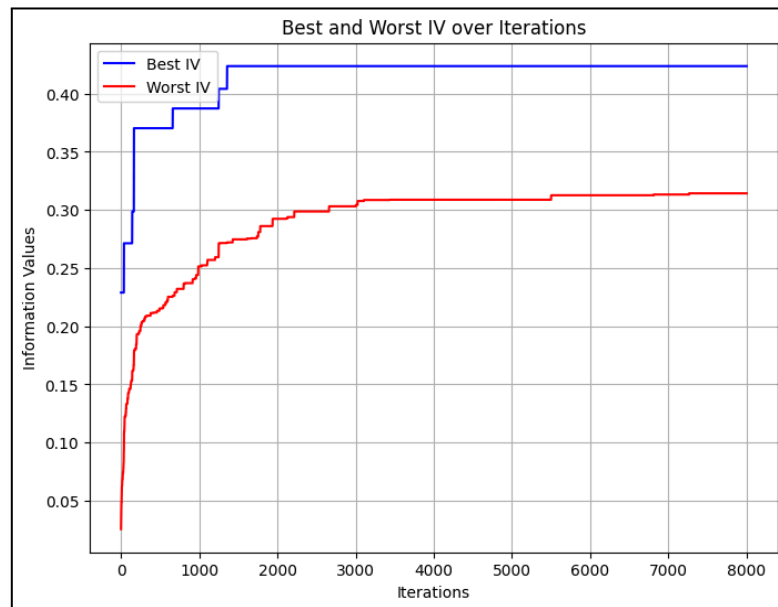


Figura 25: Comparación mejor y peor armonía de la memoria, ejecución RSRP tablón ipsos enero 2024

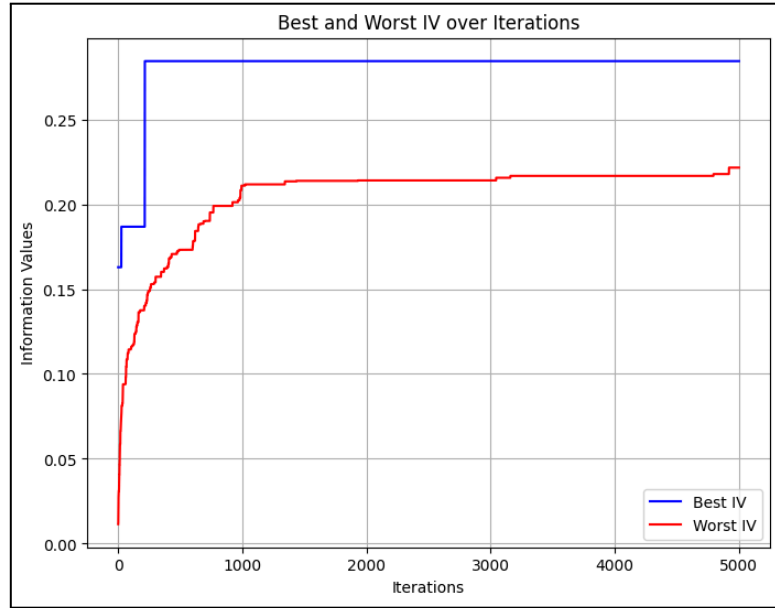
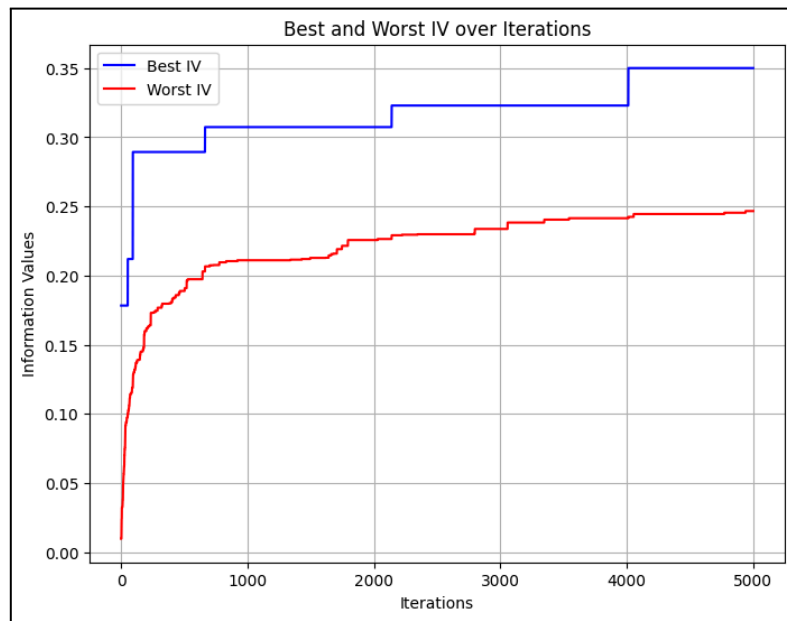


Figura 26: Comparación mejor y peor armonía de la memoria, ejecución RSRP tablón ipsos Abril 2024



Fuente figuras 23-26 : Elaboración Propia