



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

MODELO PREDICTIVO DE PROBABILIDADES DE CAMBIO EN  
EMPAQUETAMIENTO, A TRAVÉS DE APRENDIZAJE SUPERVISADO EN REDES  
NEURONALES ARTIFICIALES, COMO HERRAMIENTA PARA CAMPAÑAS DE  
CROSS-SELLING, PARA LA EMPRESA VTR COMUNICACIONES SPA.

MEMORIA PARA OPTAR AL TÍTULO DE  
INGENIERO CIVIL INDUSTRIAL

ANDRÉS PATRICIO EEDE MARTÍNEZ

PROFESOR GUÍA:  
IGNACIO CALISTO LEIVA

MIEMBROS DE LA COMISIÓN:  
EDGARDO SANTIBÁÑEZ VIANI  
MARCEL GOIC FIGUEROA

SANTIAGO DE CHILE  
2019

RESUMEN DE LA MEMORIA PARA OPTAR  
AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL  
POR: ANDRÉS PATRICIO EEDE MARTÍNEZ  
FECHA: 2019  
PROF. GUÍA: IGNACIO CALISTO LEIVA

MODELO PREDICTIVO DE PROBABILIDADES DE CAMBIO EN  
EMPAQUETAMIENTO, A TRAVÉS DE APRENDIZAJE SUPERVISADO EN REDES  
NEURONALES ARTIFICIALES, COMO HERRAMIENTA PARA CAMPAÑAS DE  
CROSS-SELLING, PARA LA EMPRESA VTR COMUNICACIONES SPA.

El presente informe detalla el análisis, construcción y evaluación de un modelo predictivo de cambios en empaquetamiento de clientes, a través de aprendizaje supervisado en redes neuronales artificiales. El trabajo es realizado para la empresa VTR Comunicaciones SpA y se adscribe a la gerencia de producto fijo. Se presentan primero los antecedentes del mercado relevante de telecomunicaciones en Chile, el contexto en que se sitúa la organización y los datos que justifican un análisis y modelamiento como el realizado.

El producto final corresponde a un modelo predictivo cuyo resultado es una probabilidad específica de cambio en empaquetamiento para cada cliente. La medición del logro de este objetivo se basa en métricas de precisión y recall, además de métricas de comparación de ajuste respecto a otros modelos y respecto al impacto al negocio. Se define un algoritmo de selección de datos y variables, un estudio de modelos de clasificación y su desempeño; y se estima el efecto de un modelo predictivo en términos cuantitativos generales y aquellos relevantes al negocio.

Ahora bien, la selección del modelo de red neuronal artificial es producto de un estudio de los principales modelos de clasificación en aprendizaje automático en cuanto a literatura asociada, estudios empíricos, benchmark relevantes y evaluación competitiva sobre datos reales referentes al trabajo de título. A su vez, la construcción integra un algoritmo de obtención de datos, tratamiento de desbalanceo, selección de variables, utilización de funciones sigmoide y rectificadoras en las capas de la red y una optimización basada en ajuste de hiper-parámetros.

Finalmente, el modelo logra un mínimo de 20 y un máximo de 84 puntos porcentuales en cuanto a razón de verdaderos positivos respecto a falsos positivos sobre un modelo aleatorio. Los perfiles de precisión acumulados muestran un mínimo de 14.71 y un máximo de 41.39 puntos porcentuales de observaciones positivas al 50 % sobre un modelo aleatorio. Las probabilidades de cambio de empaquetamiento resultantes generan la oportunidad de discriminar en cuanto a descuentos discretizados entregados y superar a cualquier modelo aleatorio o univariado como los actualmente utilizados en la empresa.

*A mi madre, Eliana Eede.*

# Agradecimientos

Me gustaría agradecer a todos quienes fueron parte de este proceso.

A mi madre, mi sustento principal de aprendizaje.

A Tamara, por su cariño, paciencia y compañía.

A mis amigos.

# Tabla de Contenido

<b>Introducción</b>	<b>1</b>
<b>1. Antecedentes</b>	<b>3</b>
1.1. Telecomunicaciones en Chile . . . . .	4
1.1.1. Participación de Mercado por Servicio Principal . . . . .	5
1.1.2. Empaquetamiento de Servicios en el Mercado . . . . .	8
1.2. Contexto VTR en Chile . . . . .	12
1.2.1. Datos estadísticos generales . . . . .	12
1.2.2. Resultados Liberty Global Latino América al primer trimestre 2019 . . . . .	14
1.2.3. Comportamiento de productos y clientes abonados . . . . .	15
1.2.4. ARPC por empaquetamiento . . . . .	18
1.3. Tasas de conversión (Cross-selling) . . . . .	21
1.4. Justificación de un modelo predictivo . . . . .	23
1.4.1. Hipótesis de causa y alternativas de solución . . . . .	24
1.4.2. Estimación del valor agregado del modelo . . . . .	25
<b>2. Objetivos y Alcance</b>	<b>27</b>
2.1. Objetivo General . . . . .	27
2.2. Objetivos Específicos . . . . .	27
2.3. Alcances . . . . .	28
2.4. Resultados esperados . . . . .	28
<b>3. Marco Teórico</b>	<b>30</b>
3.1. Estado del arte . . . . .	30
3.1.1. Modelos tradicionales en telecomunicaciones . . . . .	30
3.1.2. Aprendizaje automático en telecomunicaciones . . . . .	31
3.2. Modelos de Aprendizaje Automático . . . . .	31
3.2.1. Construir Modelos Inteligentes para Transformar Datos en Conocimiento . . . . .	31
3.2.2. Tipos de Aprendizaje Automático . . . . .	32
3.2.3. Construcción de Modelos de Aprendizaje Automático . . . . .	33
3.2.4. Python en Aprendizaje Automático . . . . .	35
3.3. Métricas de Evaluación . . . . .	35
3.3.1. Métricas de Precisión . . . . .	35
3.3.2. Métrica ROC - AUC . . . . .	37
3.3.3. Perfiles de Precisión Acumulados (CAP) . . . . .	38

3.3.4.	Perfiles de Lift Acumulados . . . . .	38
3.3.5.	Prueba de Kolmogorov - Smirnov (Estadístico KS) . . . . .	39
3.4.	Principales Modelos de Clasificación . . . . .	39
3.4.1.	Regresión Logística . . . . .	39
3.4.2.	Support Vector Machines . . . . .	40
3.4.3.	Kernel Support Vector Machines . . . . .	41
3.4.4.	Decision Trees . . . . .	41
3.4.5.	Random Forest . . . . .	42
3.4.6.	K Nearest Neighbors . . . . .	42
3.4.7.	Naive Bayes . . . . .	42
3.4.8.	Red Neuronal Artificial . . . . .	43
<b>4.</b>	<b>Selección del Modelo</b>	<b>45</b>
4.1.	Introducción al Modelo . . . . .	45
4.2.	Benchmark de modelos de aprendizaje automático . . . . .	46
4.3.	Evaluación competitiva de modelos de aprendizaje automático . . . . .	48
4.3.1.	Métricas de Precisión . . . . .	48
4.3.2.	Métrica ROC - AUC . . . . .	50
4.3.3.	Tiempos de procesamiento . . . . .	59
4.3.4.	Modelo Seleccionado . . . . .	59
<b>5.</b>	<b>Construcción del modelo</b>	<b>60</b>
5.1.	Obtención y tratamiento de datos de entrenamiento . . . . .	60
5.2.	Tratamiento de Datos Desbalanceado . . . . .	63
5.2.1.	Sobre-muestreo Aleatorio . . . . .	64
5.2.2.	Muestreo Sintético . . . . .	64
5.2.3.	Sub-Muestreo Aleatorio . . . . .	65
5.3.	Selección de Variables . . . . .	66
5.4.	Normalización de Variables . . . . .	67
5.5.	Explicación del modelo . . . . .	68
5.5.1.	Estructura escogida . . . . .	68
5.5.2.	Justificación del uso de la función de unidad lineal rectificadora (rectificadora) . . . . .	71
5.5.3.	Funcionamiento de la red neuronal . . . . .	73
5.5.4.	Programación del modelo . . . . .	74
5.6.	Optimización del modelo . . . . .	76
5.7.	Resultados y medidas de efectividad . . . . .	77
5.7.1.	Altas 3P . . . . .	78
5.7.2.	Altas Cable . . . . .	81
5.7.3.	Altas Internet . . . . .	83
5.7.4.	Altas Fono . . . . .	85
<b>6.</b>	<b>Impacto al negocio</b>	<b>88</b>
6.1.	Altas 3P . . . . .	88
6.1.1.	Cap . . . . .	88
6.1.2.	Lift . . . . .	89
6.2.	Altas cable . . . . .	90

6.2.1. Cap . . . . .	90
6.2.2. Lift . . . . .	91
6.3. Altas Internet . . . . .	92
6.3.1. Cap . . . . .	92
6.3.2. Lift . . . . .	93
6.4. Altas fono . . . . .	94
6.4.1. Cap . . . . .	94
6.4.2. Lift . . . . .	95
6.5. Modelo comparativo de descuentos . . . . .	96
6.5.1. Alta 3P . . . . .	97
<b>7. Extensiones del Modelo</b>	<b>99</b>
7.1. Extensiones de Predictibilidad . . . . .	99
7.2. Extensiones de Dimensionalidad . . . . .	100
7.3. Extensiones de Efectividad . . . . .	100
7.4. Extensiones de interpretabilidad de características y variables . . . . .	100
<b>8. Conclusión</b>	<b>103</b>
<b>Bibliografía</b>	<b>105</b>

# Introducción

VTR es una empresa de telecomunicaciones con operación nacional cuya casa matriz se encuentra situada en Santiago, Chile, y es propiedad de la compañía matriz Liberty Global Inc, un proveedor mundial de servicios de televisión, telefonía y banda ancha, que a su vez tiene operación en más de 30 países. La organización, al igual que el mercado relevante, se encuentra en búsqueda de maximizar la información obtenida a partir de la gran cantidad de datos de clientes que maneja. Dentro de este contexto, se presentan distintos requerimientos y oportunidades, estando entre ellos, el presente trabajo de título.

La oportunidad a trabajar se obtiene de dos principales necesidades: el requerimiento de caracterizar a los clientes a través de un método del tipo cuantitativo, basado en la gran cantidad de datos que se poseen, y el requerimiento de estrategias de campañas basadas en datos y probabilidad de efectividad.

En el pasado las empresas de telecomunicaciones fallaban en estructurar la data que caracterizaba a sus clientes debido a la carencia de poder computacional o de coherencia de datos. Hoy en día, sin embargo, el mundo de los negocios ha llegado a un punto de inflexión que podría hacer de los modelos de aprendizaje automático una herramienta enormemente útil para las empresas de telecomunicaciones interesadas en técnicas más avanzadas para conocer a sus clientes y tomar decisiones. De manera simultánea, el costo de almacenamiento de datos ha disminuido, y un gran set de productos se encuentra presente en la oferta. Añadido a esto, el procesamiento computacional ha crecido a tal punto que permite generar modelos en tiempo real respecto al comportamiento y análisis de inmensas bases de clientes.

En este contexto, el presente trabajo busca generar un modelo capaz de hacerse cargo de esta enorme disponibilidad de datos. Esto es realizado mediante la construcción de una red neuronal artificial con funciones rectificadoras para las capas intermedias, función sigmoide para la capa de resultado y función de dropout para el sobre ajuste. El resultado es la predicción de clasificación de cambios en empaquetamiento por parte de los clientes, la cual tiene como formato la probabilidad individual de cambio.



Los ajustes del modelo serán evaluados en cuanto a precisión y recall, modelos ROC AUC, perfiles de precisión acumulados, perfiles de lift y pruebas de kolmogorov - smirnov.

La empresa hoy en día trabaja con segmentos de clientes de tamaño determinado, en su mayoría univariados y sin asignar probabilidad de cambio en empaquetamiento. Esta segmentación imposibilita la diferenciación dentro de los segmentos y el enfoque de esfuerzos en aquellos clientes con mayor probabilidad de cambio en empaquetamiento.

El valor de la investigación consiste en la obtención de una diferencia en efectividad de campañas lograda por el nuevo modelo que conlleva a una diferencia en el mix de empaquetamiento de clientes y entre el ARPC inicial y final. Añadido a esto, las capacidades de extensión del modelo permiten lograr predicciones futuras aún más precisas, incorporar aumentos de dimensionalidad, lograr mayor efectividad a medida que se poseen más datos respecto a los clientes, extenderse a todo tipo de cambio en empaquetamiento y finalmente generar la oportunidad de interpretabilidad de características y variables de los clientes mediante el uso del resultado probabilístico y el algoritmo de selección de variables utilizados.

# Capítulo 1

## Antecedentes

El presente trabajo expone la definición, construcción y evaluación de un modelo computacional predictivo basado en técnicas de aprendizaje supervisado cuyo objetivo es obtener una mejor predicción respecto a la probabilidad de cambios en nivel de empaquetamiento por parte de los clientes de la empresa de telecomunicaciones VTR.

Ahora bien, como el desarrollo del modelo se encuentra inscrito en el mercado de las telecomunicaciones, su pertinencia y utilidad en cuanto a impacto al negocio debe ser cuestionada y justificada dentro del mismo marco empresarial y utilizando conceptos asociados al rubro.

Así, debido al sector empresarial concomitante al trabajo realizado, el primer capítulo se ocupa de describir el contexto de las telecomunicaciones en Chile, el lugar que ocupa la empresa concerniente al estudio dentro de este contexto, y finalmente la justificación de un modelo predictivo sobre la base conceptual ya desarrollada.

## 1.1. Telecomunicaciones en Chile

El sector de telecomunicaciones en Chile comprende como principales servicios a los siguientes: internet fijo, móvil (3G + 4G), telefonía fija, telefonía móvil, televisión de pago y larga distancia internacional:

Servicio	Unidad	4T-2018	Variación Interanual
Internet fija	Miles de accesos	3.256	6,1 %
Internet móvil 3G+4G	Miles de accesos	18.109	10,9 %
4G	Miles de servicios	14.559	35,2 %
Telefonía fija	Miles de suscriptores	2.997	-6,3 %
Telefonía móvil	Miles de abonados	25.179	9,4 %
TV pago	Miles de servicios	3.329	1,0 %
Larga distancia internacional	Millones de minutos	35,6	-28,7 %

Tabla 1.1: Informe Anual del Sector Telecomunicaciones – Cierre 2018 - SUBTEL.

El conjunto de estos servicios se encuentra regulado por el Ministerio de Transportes y Telecomunicaciones, a través de la Subsecretaría de Telecomunicaciones. A través de la Ley General de Telecomunicaciones (Ley 18.168), la Subtel establece las políticas de regulación tarifaria asociadas a los servicios de telecomunicaciones, así como la fiscalización del acceso libre de todos los chilenos a estos servicios y la no discriminación; determinando estándares de calidad de los servicios y velando por mantener el carácter competitivo del mercado.

Ahora bien, los principales servicios mencionados son, a su vez, parte de la oferta comercial de los principales actores del mercado (quienes poseen la mayor participación). Estos actores corresponden a las empresas: Movistar, Grupo Entel, Claro, Wom, Vtr, Virgin y grupo GTD.

### 1.1.1. Participación de Mercado por Servicio Principal

A continuación se muestran, en términos porcentuales, las participaciones de mercado de mayor tamaño para cada servicio.

Internet móvil:

Los accesos 4G alcanzan un valor total aproximado de 14,6 millones de conexiones para Diciembre 2018, teniendo un crecimiento de 35,2% en los últimos 12 meses, y siendo su distribución por empresa la descrita por las figuras 1.1 y 1.2. Así, las empresas líderes en este servicio resultan ser Entel, Movistar y Claro.

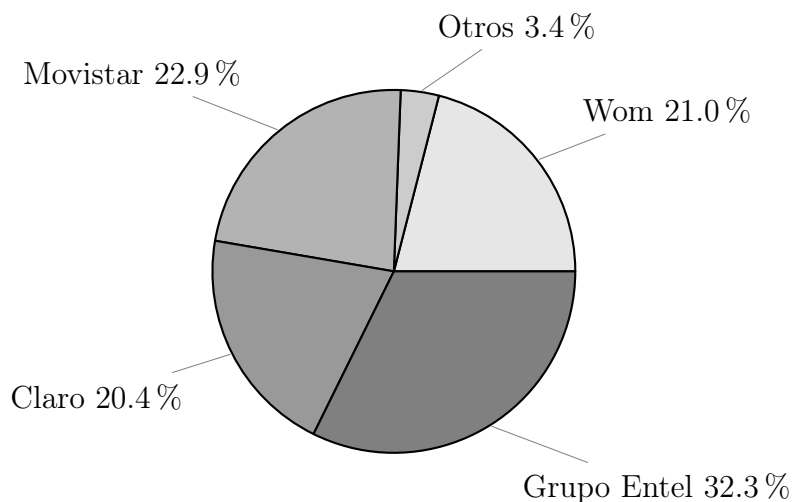


Figura 1.1: Participación de Mercado 4G, Diciembre 2018.

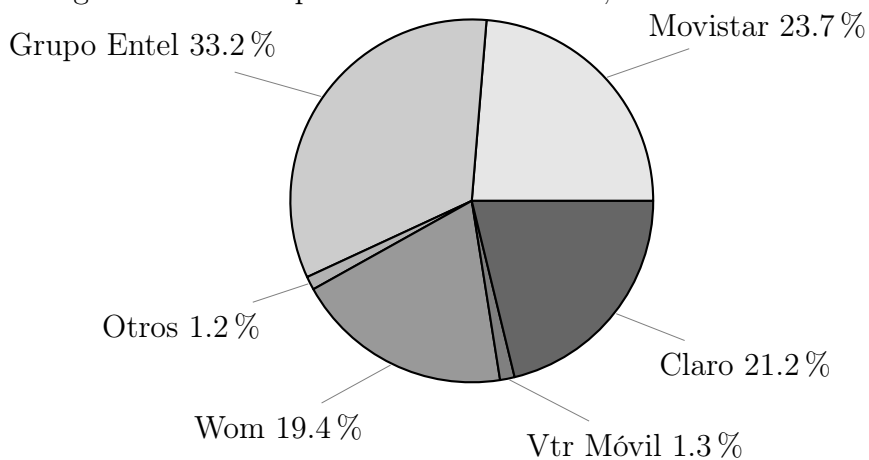


Figura 1.2: Participación de Mercado, conexiones 3G+4G, Diciembre 2018.

#### Internet fijo:

A diciembre del año 2018 el 71,8 % de las conexiones fijas se encuentran comprendidas por las empresas VTR (empresa que presenta un crecimiento de un 6,6 % en los últimos 12 meses) y Movistar. La distribución de estas conexiones se puede observar en la figura 1.3.

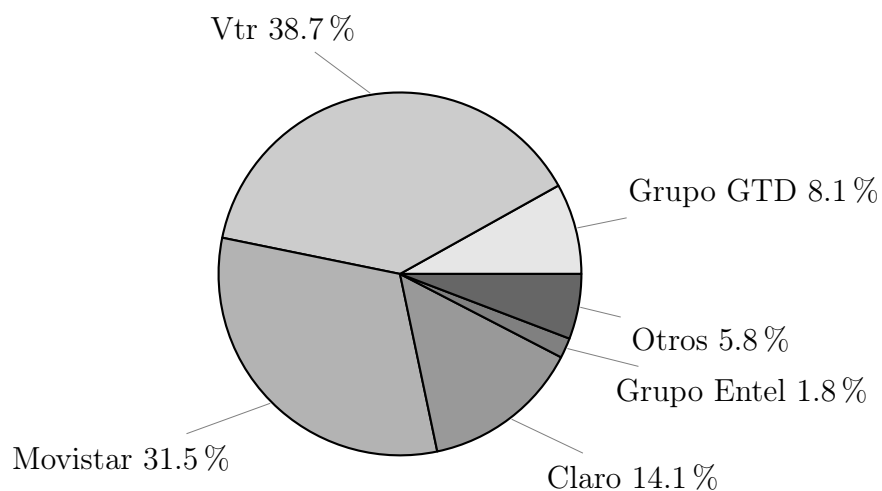


Figura 1.3: Participación de Mercado Internet Fijo, Diciembre 2018.

#### Telefonía móvil:

A diciembre del año 2018 se identifican tres principales operadores, los cuales agrupan un 83,2 % del mercado: Entel, Movistar y Claro. Cabe destacar la evolución de la empresa WOM, con un crecimiento de 49,0 % en los últimos 12 meses. La distribución de participaciones se puede observar en la figura 1.4.

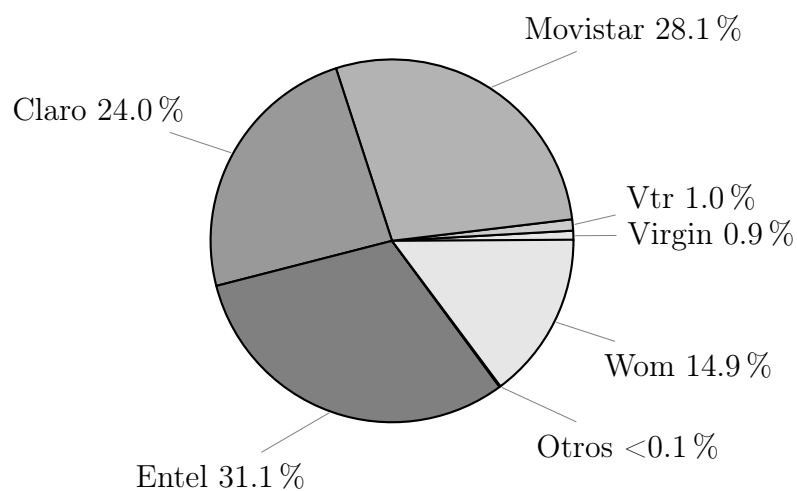


Figura 1.4: Participación de Mercado Telefonía Móvil, Diciembre 2018.

Telefonía fija:

A diciembre del año 2018, Movistar es el principal actor del mercado, con un 40,8% de participación. La distribución de participaciones se puede observar en la figura 1.5.

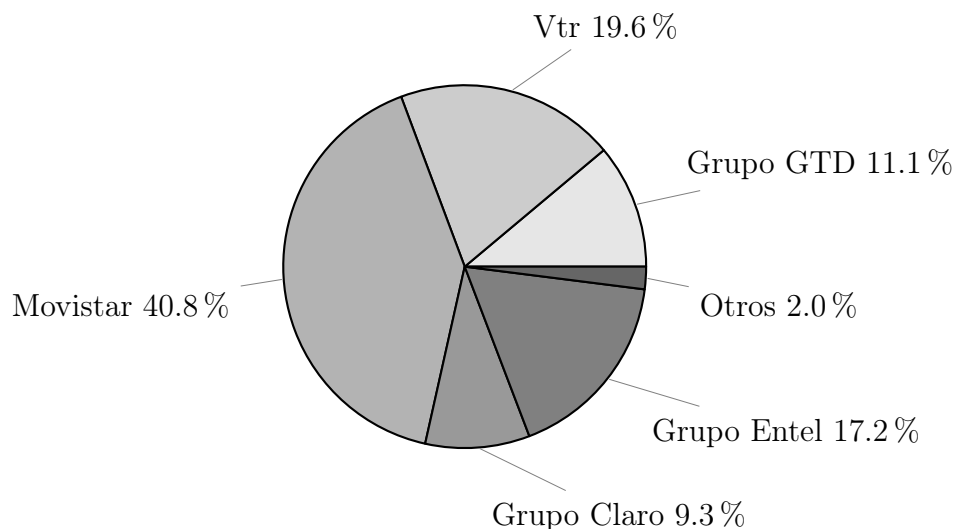


Figura 1.5: Participación de Mercado Telefonía fija, Diciembre 2018.

Televisión pagada:

A septiembre del año 2018, VTR es el principal actor del mercado con una cuota de un 32,3% de la totalidad de suscriptores; le siguen DirecTV (20,3%), Movistar (19,6%) y Claro (12,2%). La distribución de participaciones se puede observar en la figura 1.6.

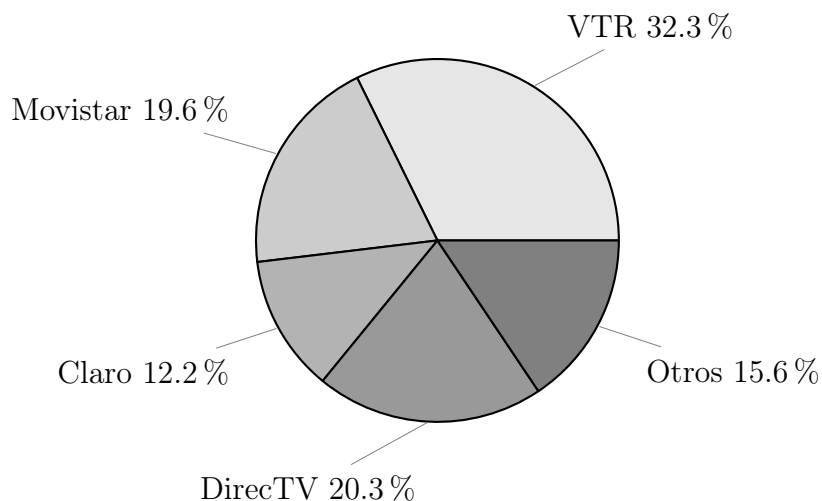


Figura 1.6: Participación de Mercado TV pagada, Diciembre 2018.

### 1.1.2. Empaquetamiento de Servicios en el Mercado

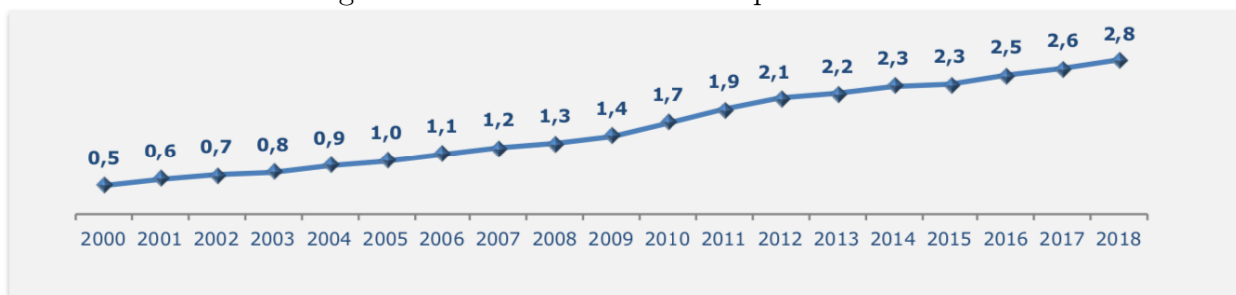
El empaquetamiento define la agrupación de los servicios ya descritos anteriormente en distintas combinaciones para formar conjuntos denominados paquetes, los cuales son presentados al mercado en este formato.

Considerando esto, los clientes de cada empresa pueden estar suscritos a uno o más servicios al mismo tiempo, definiendo un índice de empaquetamiento. De la misma manera, los clientes pueden optar por la diversificación de sus servicios a partir de las distintas empresas, lo cual aún genera un índice de empaquetamiento de servicios, pero en el formato de número de servicios por habitante.

El nivel de empaquetamiento resulta relevante para el estudio de las posibilidades de cross-selling dentro de las empresas, y a su vez, resulta relevante para la justificación de la pertinencia del modelo expuesto en el presente trabajo.

En Chile, a Diciembre del año 2018, el sector muestra un tamaño de 52,9 millones de servicios de telecomunicaciones y su caracterización evolutiva en número de servicios por habitante y total de servicios pueden ser observados en las figuras 1.7 y 1.8 respectivamente.

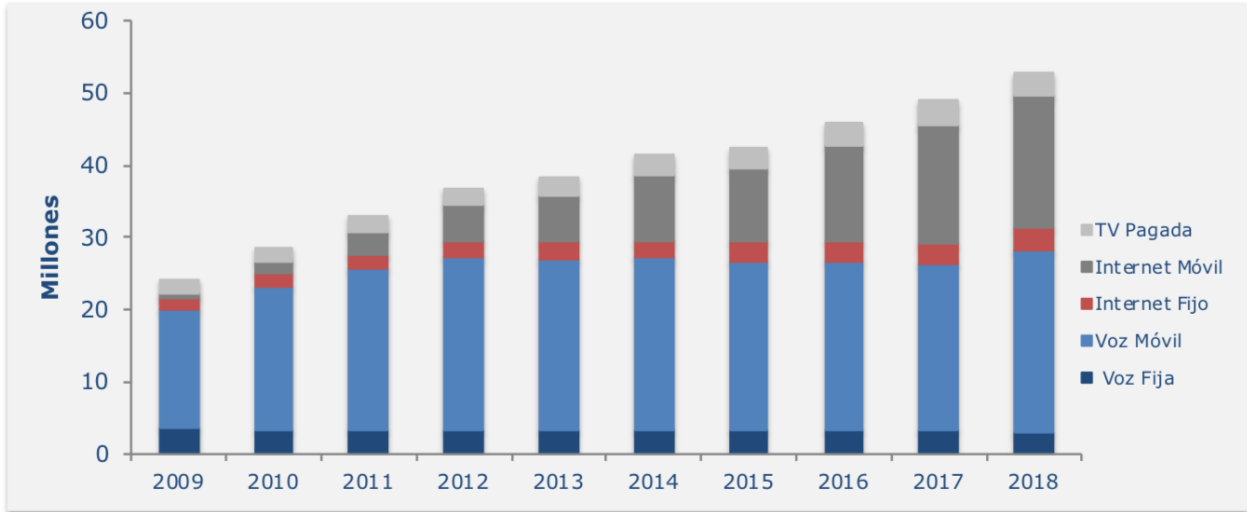
Figura 1.7: Número de servicios por habitante:



El nivel de empaquetamiento por habitante muestra una clara tendencia al alza, alcanzando un nivel de 2,8 servicios por cliente al cierre del año 2018. Respecto al total de servicios, claramente la tendencia mantiene el mismo comportamiento, superando los 51 millones de servicios al cierre del año 2018; y siendo los principales servicios, en términos de cantidad de abonados, internet y voz móvil.

Considerando los niveles de empaquetamiento ya mencionados, es conveniente conocer también la oferta comercial relevante: la cantidad de planes empaquetados existentes y las empresas oferentes de dichos planes. Existe una gran variedad de planes ofertados debido a la gran cantidad de combinaciones entre tipos de servicio y tipo de paquete. Los tipos de servicios varían, por ejemplo, en la velocidad de internet (fijo o móvil), la cartera de

Figura 1.8: Total de servicios:



canales ofrecidas (televisión de pago) y las cantidades de minutos que posee cada plan de telefonía (fija o móvil), entre otras posibles variaciones. A su vez, los tipos de paquete pueden considerar combinaciones dobles o triples de servicios.

Empresa	Planes empaquetados informados
Telefónica Chile	152
VTR	45
CMET	22
CSAT	16
Entel	61
GTD Manquehue	507
Telcoy	4
Telefónica del Sur	14
TOTAL	821

Tabla 1.2: Cantidad de planes empaquetados – 2009 - SUBTEL.



**Empaquetamientos de producto fijo ofrecidos públicamente por las principales empresas del mercado:** A continuación se detallan todos los paquetes ofrecidos públicamente por las principales empresas del mercado al mes de Junio, año 2019.

MOVISTAR:

- Empaquetamiento triple: Pack Trío HD Pro 600 MB, Pack Trío HD Pro 400 MB, Pack Trío HD Pro 200 MB, Pack Trío HD Pro 50 MB.
- Empaquetamiento doble (Internet + Televisión): Pack Dúo HD Pro Fibra 600 Megas, Pack Dúo HD Pro Fibra 400 Megas, Pack Dúo HD Pro Fibra 200 Megas, Pack Dúo HD Pro Fibra 50 Megas.
- Empaquetamiento doble (Internet + Fono): Pack Dúo Fibra 600 Megas, Pack Dúo Fibra 400 Megas, Pack Dúo Fibra 200 Megas, Pack Dúo Fibra 50 Megas, Pack Dúo Fibra 8 Megas.
- Empaquetamiento simple (Internet): Internet Fibra 600 Megas, Internet Fibra 400 Megas, Internet Fibra 200 Megas, Internet ADSL/Fibra Hasta 50 Megas, Internet ADSL Hasta 8 Megas.
- Empaquetamiento simple (Televisión): Plan televisión HD Pro. (Este empaquetamiento permite complementar con 9 canales premium, por lo cual la oferta total corresponde a todas las combinaciones posibles anexas al plan).
- Empaquetamiento simple (Fono): Plan ilimitado local, Plan ilimitado, Plan Control.

VTR:

- Empaquetamiento triple: Triple Pack Vive Pro, Triple Pack Vive Más, Triple Pack Vive.
- Empaquetamiento doble (Internet + Televisión): Doble Pack Vive Pro, Doble Pack Vive Más, Doble Pack Vive.
- Empaquetamiento doble (Televisión + Telefonía): TV Hogar HD + fono ilimitado, TV Full HD + fono ilimitado.
- Empaquetamiento doble (Internet + Fono): Mega 30 + fono ilimitado, Mega 200 + fono ilimitado, Mega 400 + fono ilimitado.
- Empaquetamiento simple (Internet): Mega 30, Mega 200, Mega 400.
- Empaquetamiento simple (Televisión): TV Hogar HD, TV Full HD.

ENTEL:

- Empaquetamiento triple: Triple Pack Simple, Triple Pack Pro, Triple Pack Fibra 100, Triple Pack Fibra 250, Triple Pack Fibra 500, Triple Pack Fibra Giga, Triple Pack Simple Deportes, Triple Pack Pro Deportes, Triple Pack Fibra 100 Deportes, Triple

Pack Fibra 250 Deportes, Triple Pack Fibra 500 Deportes, Triple Pack Fibra Giga Deportes.

- Empaquetamiento doble (Internet + Televisión): Doble Pack Simple, Doble Pack Pro, Doble Pack Fibra 100, Doble Pack Fibra 250, Doble Pack Fibra 500, Doble Pack Fibra Giga, Doble Pack Simple Deportes, Doble Pack Pro Deportes, Doble Pack Fibra 100 Deportes, Doble Pack Fibra 250 Deportes, Doble Pack Fibra 500 Deportes, Doble Pack Fibra Giga Deportes.
- Empaquetamiento doble (Televisión + Fono): Telefonía Súper HD, Telefonía Deportes HD.
- Empaquetamiento doble (Internet + Fono): Telefonía Simple, Telefonía Pro, Telefonía Fibra 100, Telefonía Fibra 250, Telefonía Fibra 500, Telefonía Fibra Giga.
- Empaquetamiento simple (Internet): Internet Plan Simple, Internet Plan Pro, Internet Fibra 100, Internet Fibra 250, Internet Fibra 500, Internet Fibra Giga.
- Empaquetamiento simple (Televisión): Televisión Súper HD, Televisión Deportes HD, Televisión Fibra HD, Televisión Fibra Deportes HD. simple (Telefonía): Televisión Fija, Telefonía Fibra.

#### CLARO:

- Empaquetamiento triple: 3 Play Plan TV Pro HD + Internet 20 Mbps + Telefonía Hogar, 3 Play TV Pro HD Más + Internet 20 Mbps + Telefonía Hogar, Play TV Pro HD + Internet 50 Mbps + Telefonía Hogar, 3 Play TV Pro HD Más + Internet 50 Mbps + Telefonía Hogar, 3 Play TV Pro HD + Internet 200 Mbps + Telefonía Hogar, 3 Play TV Pro HD Más + Internet 200Mbps + Telefonía Hogar, 3 Play TV Pro HD Fútbol + Internet 20 Mbps + Telefonía Hogar, 3 Play TV Pro HD Fútbol + Internet 50 Mbps + Telefonía Hogar, 3 Play TV Pro HD Fútbol + Internet 200 Mbps + Telefonía Hogar.
- Empaquetamiento doble (Internet + Televisión): 2 Play Plan TV Pro HD + Internet 20 Mbps, 2 Play Plan TV Pro HD Más + Internet 20 Mbps, 2 Play Plan TV Pro HD + Internet 50 Megas, 2 Play Plan TV Pro HD Más + Internet 50 Megas, 2 Play Plan TV Pro HD + Internet 200 Megas, 2 Play Plan TV Pro HD Más + Internet 200 Megas, 2 Play TV Pro HD Fútbol + Internet 20 Mbps, 2 Play Plan TV Pro HD Fútbol + Internet 50 Megas, 2 Play TV Pro HD Fútbol + Internet 200 Megas.
- Empaquetamiento doble (Televisión + Fono): 2 Play Plan TV Pro HD + Telefonía Hogar, 2 Play Plan TV Pro HD Más + Telefonía Hogar, 2 Play TV Pro HD Fútbol + Telefonía Hogar.
- Empaquetamiento doble (Internet + Fono): 2 Play Telefonía Hogar + 20 Mbps, 2 Play Telefonía Hogar + 50 Megas, 2 Play Internet 200/10 Mbps + Telefonía Hogar.
- Empaquetamiento simple (Internet): Plan Internet 20 Mbps, Plan Internet 50 Mbps, Plan internet 200/10 Mbps.
- Empaquetamiento simple (Televisión): Plan TV Pro HD Cable, Plan TV Pro HD Más Cable, Plan TV Pro HD Fútbol Cable.
- Empaquetamiento simple (Telefonía): Telefonía Hogar.

## Grupo GTD:

- Empaquetamiento triple: Tripack 200 Mbps, Tripack 300 Mbps, Tripack 400 Mbps.
- Empaquetamiento doble (Internet + Televisión): Duopack 200 Mbps, Duopack 300 Mbps, Duopack 400 Mbps.
- Empaquetamiento doble (Internet + Fono): Duopack 200 Mbps, Duopack 300 Mbps, Duopack 400 Mbps.
- Empaquetamiento simple (Internet): 200 Mbps, 300 Mbps, 400 Mbps.
- Empaquetamiento simple (Televisión): Plan Familia, Canales premium, Canales Temáticos.
- Empaquetamiento simple (Telefonía): Plan Voz 1000 minutos.

## **Resumen:**

Empresa	3PACK	I+TV	I+F	TV+F	I	TV	F	TOTAL
Movistar	4	4	5	0	5	1	3	22
VTR	3	3	3	2	3	2	0	16
Entel	12	12	6	2	6	4	2	44
Claro	9	9	3	3	3	3	1	31
Grupo GTD	3	3	3	0	3	3	1	16
TOTAL	31	31	23	7	23	16	8	138

Tabla 1.3: Tipos de empaquetamientos ofrecidos a Junio 2019 - Información Pública de cada Empresa.

Entel resulta ser la empresa con mayor cantidad de paquetes ofrecidos, seguida por Claro.

## **1.2. Contexto VTR en Chile**

### **1.2.1. Datos estadísticos generales**

Realizada la descripción contextual del mercado relevante de las telecomunicaciones en Chile, se procede a situar a la empresa concerniente al trabajo dentro de este marco.

VTR como marca pertenece a la compañía internacional Liberty Global, una de las empresas de cable más grandes del mundo. La empresa destaca en los sectores de banda ancha residencial con una participación del 38,7%, televisión de pago con un 32,3% y telefonía fija

con un 19.6 %, según los últimos datos estadísticos informados por Subtel.

Durante el año 2018 registró un total de 3,216,600 hogares conectados. Además, entregó un total de 2,795,500 servicios fijos. En adición a esto, tuvo ingresos que alcanzaron aproximadamente los 580,616 miles de millones de pesos, un 6 % más que en el año 2017, y un EBITDA de 228,562 miles de millones de pesos (Reporte de Sostenibilidad 2018, VTR). Liberty Global plc controla el 100 % de VTR a través de sus filiales United Chile Ventures, dueña del 61,8% de sus acciones; United Chile LLC, titular del 18,2% de la propiedad, y VTR Chile Holding SpA, que posee el 20 % restante de las acciones.

De acuerdo a los resultados informados por la empresa al cierre del año 2018, la empresa amortizó el 10 % de los pagarés prioritarios preferentes, por \$1,4 mil millones a octubre.

VTR Finance B.V. ("VTR Finance"), se posiciona como el principal operador de cable en Chile, ofreciendo servicios de entretenimiento, información y comunicación a aproximadamente 1,3 millones de abonados de Internet banda ancha, 1,1 millones de abonados de video, 0,6 millones de abonados de telefonía fija y 0,2 millones de abonados móviles.

#### Aspectos relevantes al 3er trimestre 2018:

El número de clientes tuvo un crecimiento de 67.000 clientes durante los últimos 12 meses, significando un aumento del 5 % de la cartera, debido a nuevas construcciones de red.

Los RGU tuvieron un aumento de 3 mil unidades durante el período de análisis, cuyo principal impulso fue el crecimiento en servicios de banda ancha y video; siendo contrarrestados por la disminución en el número de clientes de telefonía fija.

Se tuvo un aumento, durante el período, de aproximadamente 18.000 nuevos RGU correspondientes al servicio de Banda Ancha, aproximadamente 4.000 nuevos RGU correspondientes al servicio de video y una disminución aproximada de 18.000 RGU en telefonía fija. Respecto a los abonados móviles, estos tuvieron un aumento de 10.000 durante el período de análisis, resultando en 246.000 abonados, un 19 % más respecto al mismo período del año anterior.

Gracias a las nuevas iniciativas de construcción y actualización, se obtuvieron 50.000 instalaciones nuevas durante el período de análisis, resultando en un total de 135.000 hogares antes no contemplados.

### Datos financieros relevantes al 3er trimestre 2018:

Los ingresos del período de análisis alcanzaron el valor de 163 mil millones de pesos chilenos. Este aumento se obtuvo gracias a las suscripciones fijas residenciales (aumento del ARPU por RGU y de la cantidad media de abonados); gracias a las suscripciones Business To Business (RGU SOHO); y gracias a las nuevas suscripciones móviles. Respecto al índice ARPU promedio por cliente, este se mantuvo estable en un valor de 33.670 pesos chilenos.

### **1.2.2. Resultados Liberty Global Latino América al primer trimestre 2019**

La compañía internacional tuvo un crecimiento en ingresos del 4% a \$943.000.000 (USD); logró un record, para ser un primer trimestre, de 73.000 nuevos RGU, mejorando la tendencia de suscripción móvil; y obtuvo una rentabilidad operacional de \$113.000.000 (USD).

Respecto a sus principales líneas de negocio, tuvo los siguientes resultados para el período de análisis:

C&W: Record de crecimiento en RGU, debido a banda ancha, agregando 32.000 nuevos abonados en total.

VTR/Cabletica: El crecimiento de la banda ancha en Chile y Costa Rica generó la adición de 20.000 nuevos RGU. VTR expandió su red líder, agregando sobre 40.000 hogares.

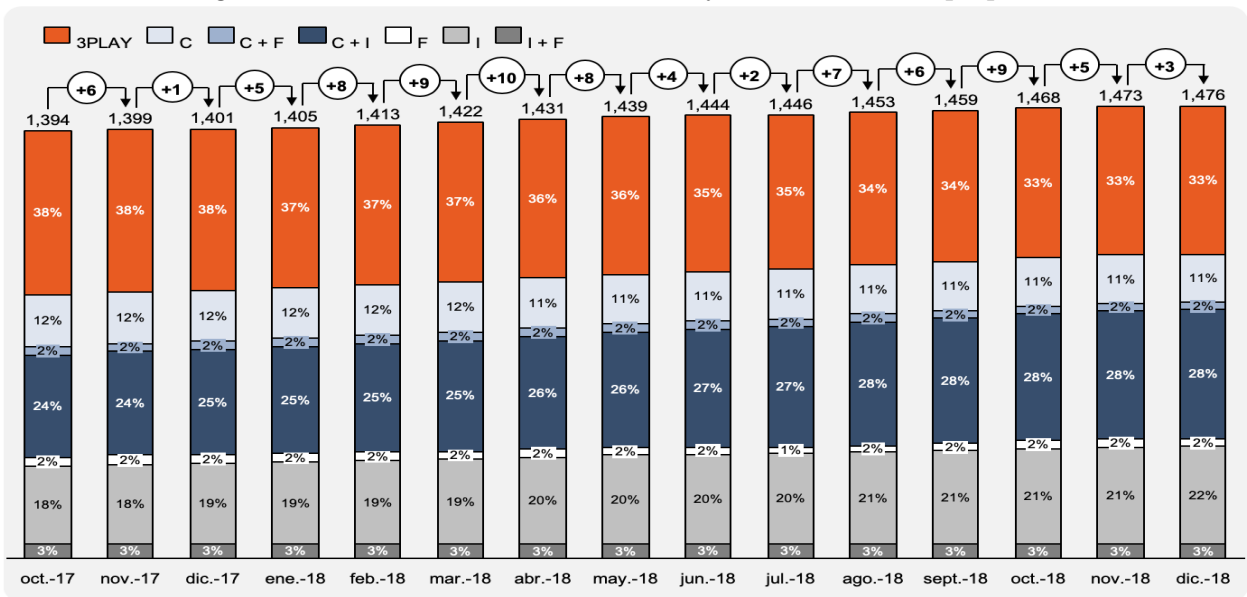
Liberty Puerto Rico: 22.000 nuevos RGU para el primer trimestre, anexo a un crecimiento general para todos los productos, debido principalmente a la obtención del premio Ookla® Speedtest® por segundo año consecutivo. Añadido a esto, tuvo un crecimiento de 5.000 nuevos hogares en la red.

### 1.2.3. Comportamiento de productos y clientes abonados

Actualmente VTR cuenta con casi 1.5 millones de clientes a lo largo de Chile. Ahora bien, el comportamiento de estos clientes y la adquisición de nuevos clientes siguen patrones de crecimiento para algunos productos, estabilidad para otros, y mayormente fuga en el caso de algunos. Debido a esto, se puede proceder a caracterizar el desempeño de la organización en cuanto a sus productos:

Para el total de clientes por tipo de empaquetamiento, desde octubre 2017 hasta diciembre 2018, se tiene el siguiente comportamiento: \*3Play corresponde al triple pack, la letra C al producto “Cable”, F al producto “Fono”, y la letra I al producto “Internet”.

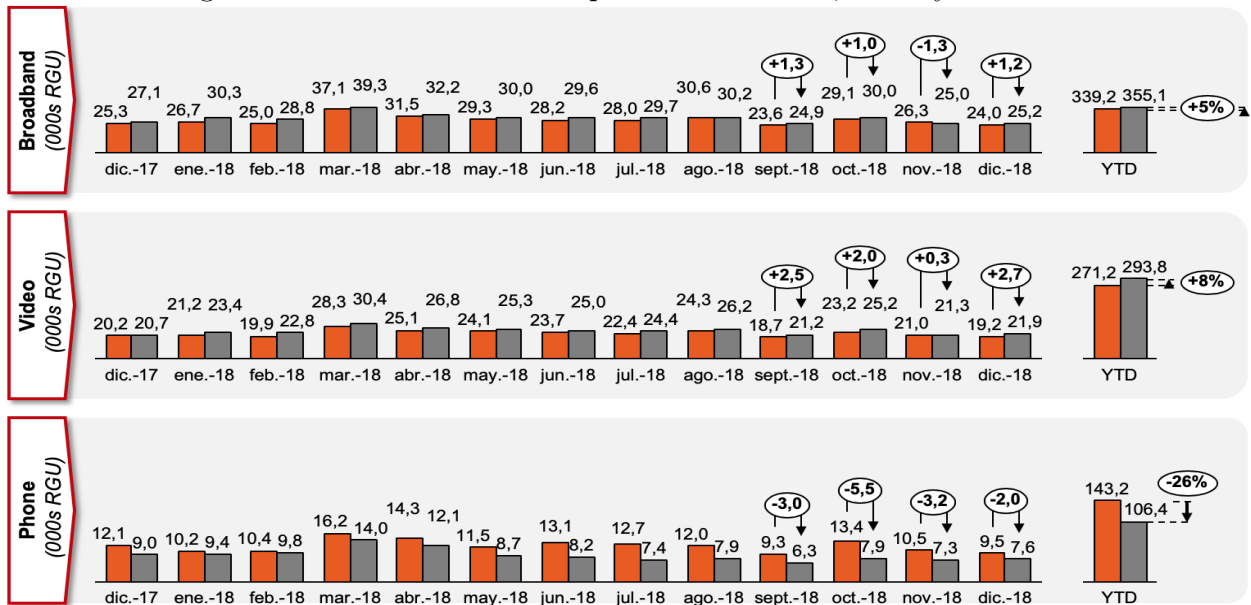
Figura 1.9: Cantidad total de clientes y distribución de paquetes:



Además, se pueden caracterizar los movimientos de clientes de manera mensual:

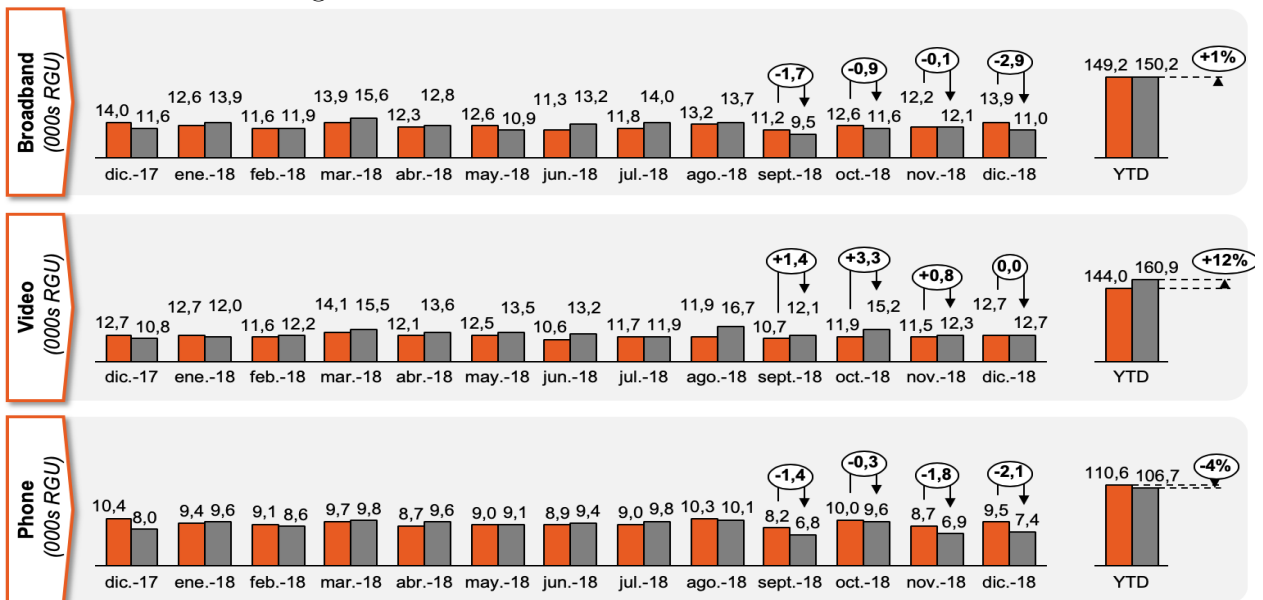
1) Niveles de ventas para banda ancha, video y telefonía. \* El color gris corresponde a la cantidad real, el color naranja corresponde a pronósticos de acuerdo con el comportamiento anterior de los productos.

Figura 1.10: Niveles de ventas para banda ancha, video y telefonía:



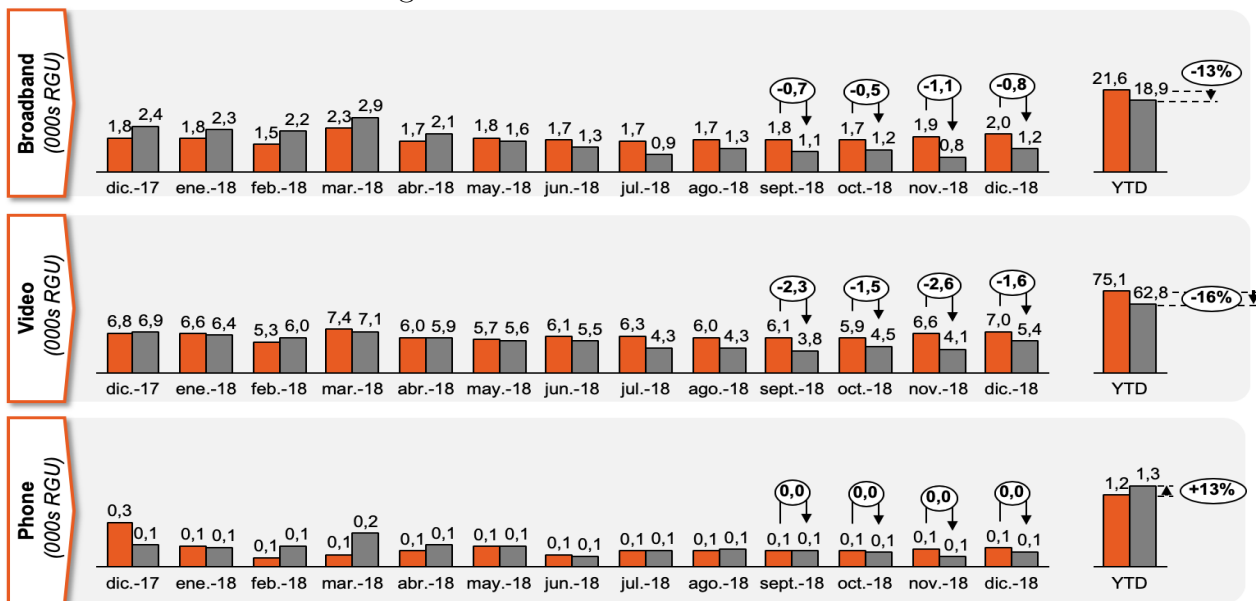
2) Niveles de desconexiones para banda ancha, video y fono. \* El color gris corresponde a la cantidad real, el color naranja corresponde a pronósticos de acuerdo con el comportamiento anterior de los productos.

Figura 1.11: Desconexiones voluntarias finalizadas:



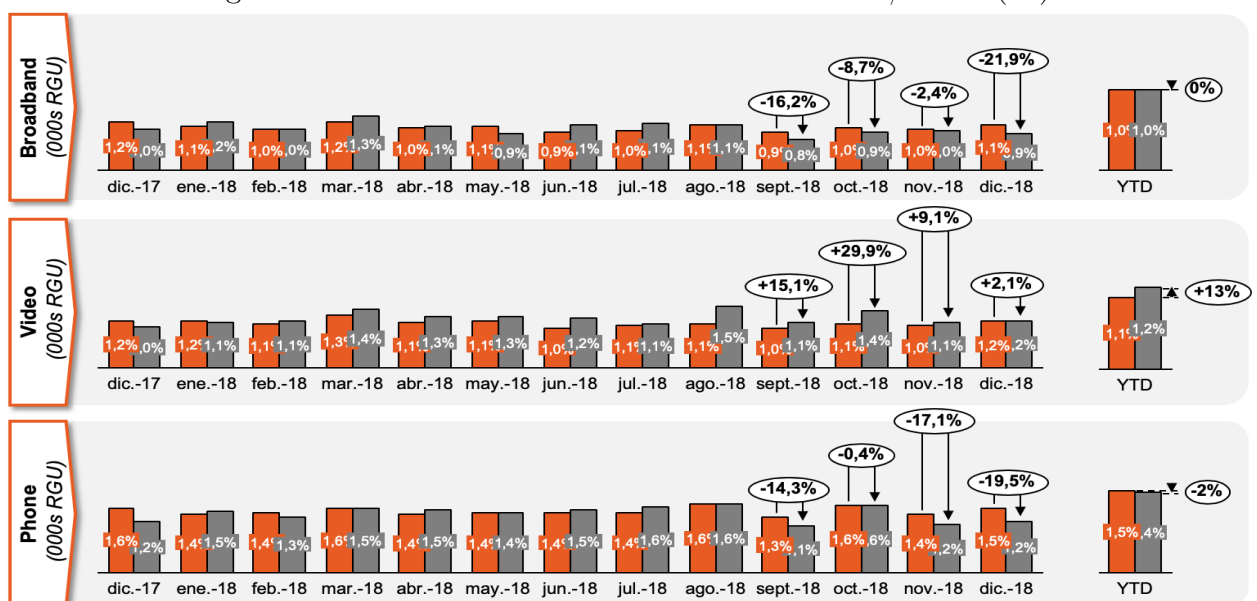
3) Niveles de reconexiones para banda ancha, video y fono. \* El color gris corresponde a la cantidad real, el color naranja corresponde a pronósticos de acuerdo con el comportamiento anterior de los productos.

Figura 1.12: Reconexiones finalizadas:



4) Niveles de fuga de clientes para banda ancha, video y fono. \* El color gris corresponde a la cantidad real, el color naranja corresponde a pronósticos de acuerdo con el comportamiento anterior de los productos.

Figura 1.13: Desconexiones voluntarias finalizadas / RGU (%):





De esta manera, se pueden resumir los movimientos de clientes en cuanto a desconexiones, reconexiones, cross selling y desempaquetamiento de la siguiente manera:

Figura 1.14: Nuevos clientes y reconexiones:

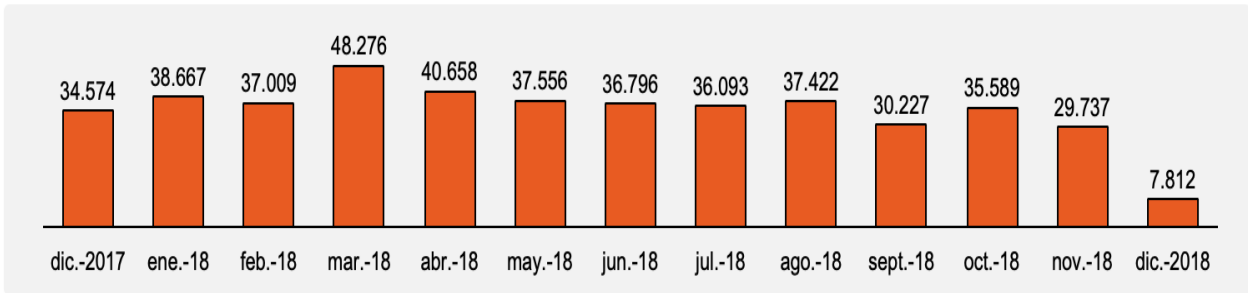


Figura 1.15: Cross-Selling y desempaquetamiento:

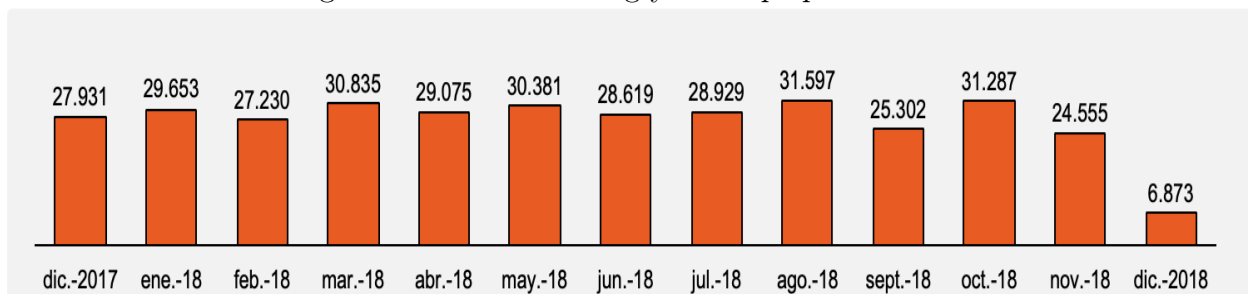
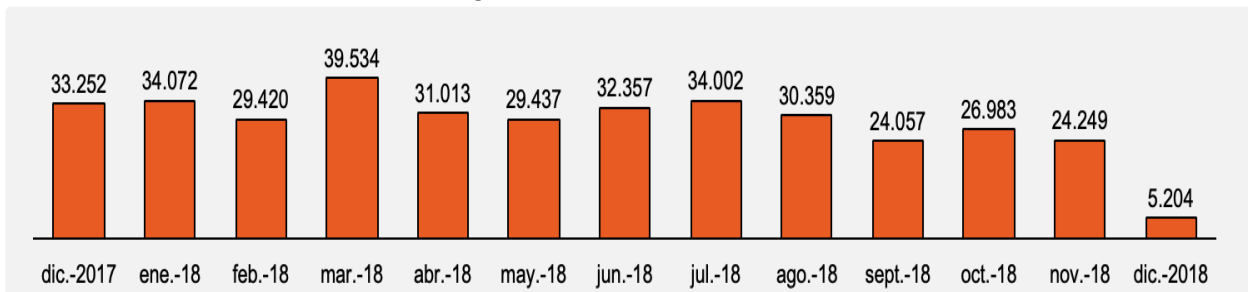


Figura 1.16: Desconexiones:



#### 1.2.4. ARPC por empaquetamiento

Habiendo considerado los datos estadísticos generales y los movimientos respecto a productos y clientes abonados, se procede a caracterizar el ingreso promedio por cliente y las tasas de conversión de clientes.

Ambos datos son relevantes para la posterior justificación de las campañas de cross-selling, que suponen un aumento en los ingresos promedio sujetos al cliente receptor de la campaña; y para la justificación de un modelo predictivo de efectividad, que supone un aumento en efectividad esperada de las mencionadas campañas.

Respecto a los ingresos promedio por cliente, estos pueden observarse en las figuras 1.17, 1.18, 1.19 y 1.20 respectivamente, donde 3P corresponde a los servicios empaquetados triples, 2P a los servicios empaquetados dobles y 1P a los servicios individuales. Se observa que, de manera esperable, un mayor empaquetamiento significa un mayor ingreso promedio por cliente. Además, se observa que el mayor impacto al ingreso lo tiene el servicio de internet, seguido por el servicio de cable.

Figura 1.17: Ingreso promedio por cliente general:

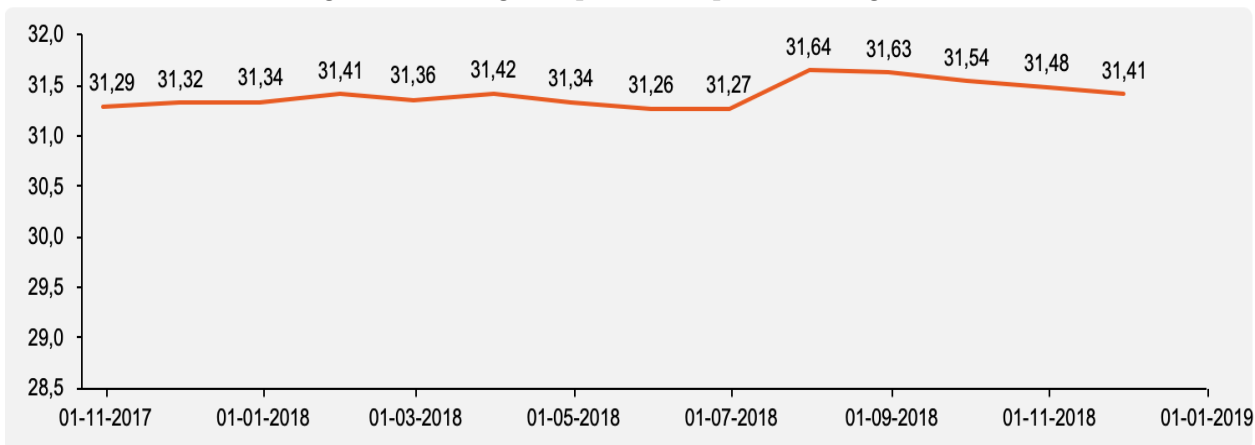


Figura 1.18: Ingreso promedio por cliente 3P:

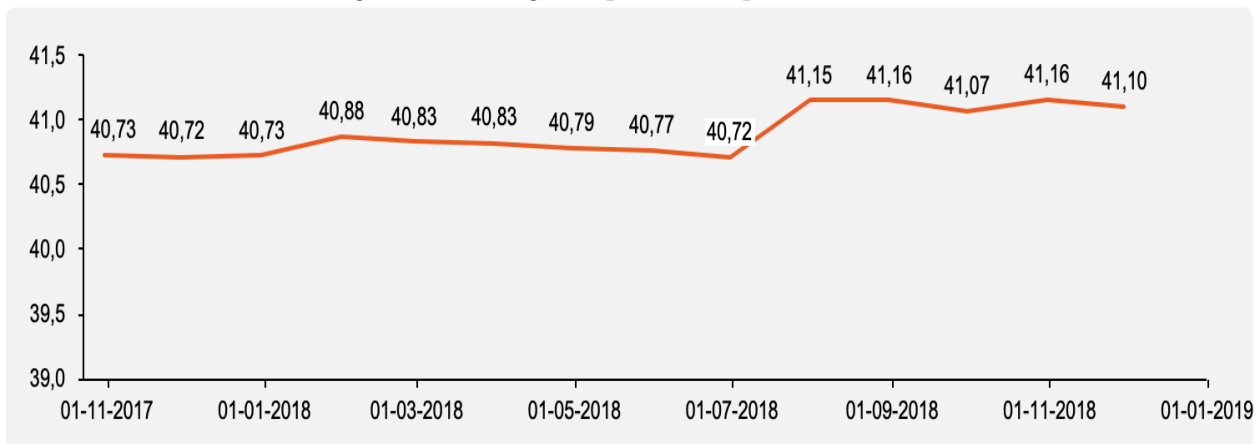


Figura 1.19: Ingreso promedio por cliente 2P:

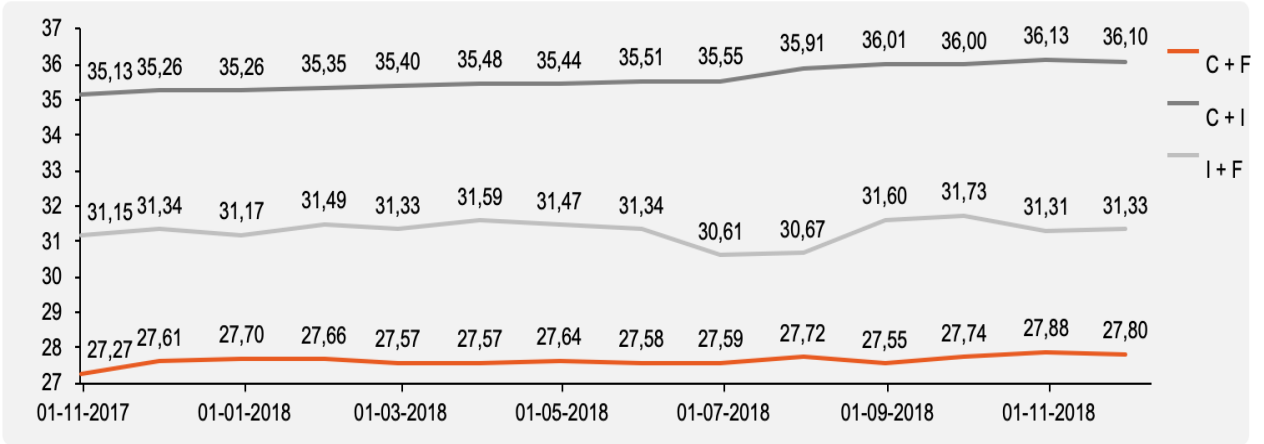
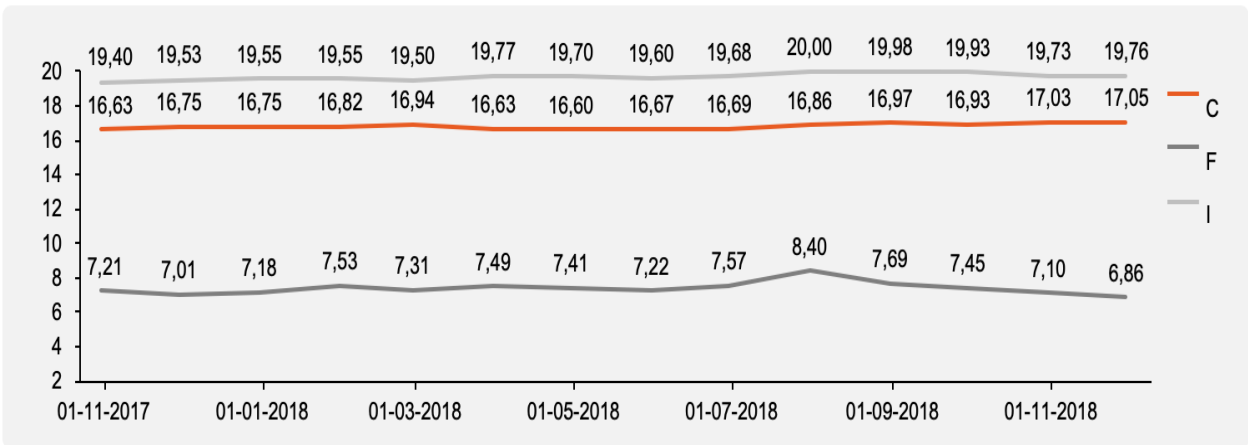


Figura 1.20: Ingreso promedio por cliente 1P:



### 1.3. Tasas de conversión (Cross-selling)

Habiendo considerado los ingresos promedio por cliente, en las figuras 1.21, 1.22 y 1.23 se puede observar la tasa de conversión de los clientes de un paquete a otro, identificadas como movimientos de cross-selling.

Para cada servicio se muestran las combinaciones de posibles movimientos, así, para el caso del servicio de Internet, se observan las tasas de conversión de clientes con cable (C), fono (F) y doble pack (C+F), que contrataron Internet, por lo cual finalmente poseen los paquetes de cable e internet, fono e internet y triple pack respectivamente. Lo mismo ocurre para el resto de servicios en su respectiva gráfica evolutiva.

Figura 1.21: Tasa de conversión, Cross-selling, internet:

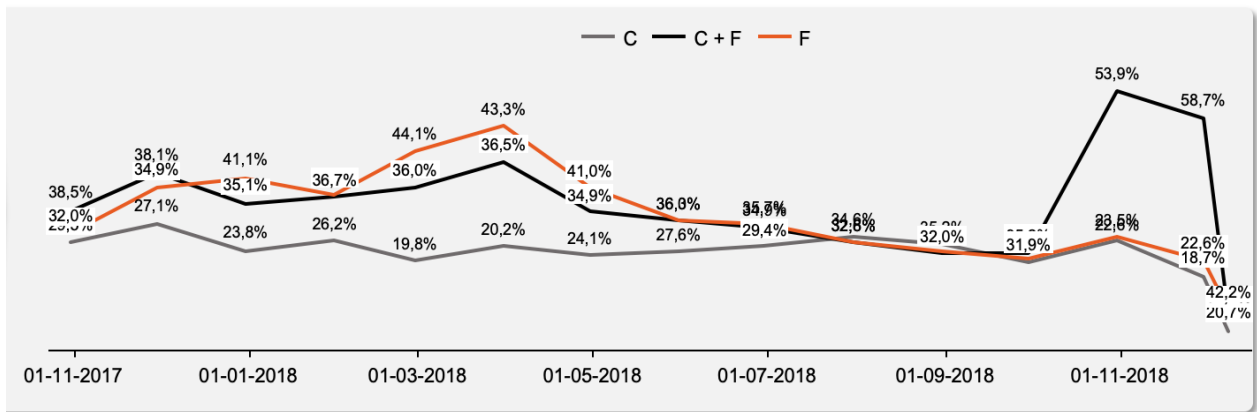


Figura 1.22: Tasa de conversión, Cross-selling, video:

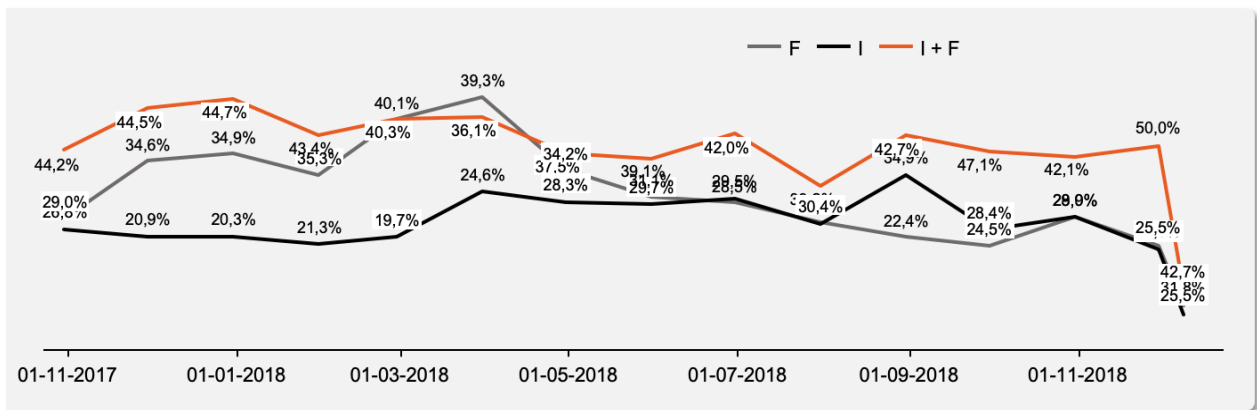
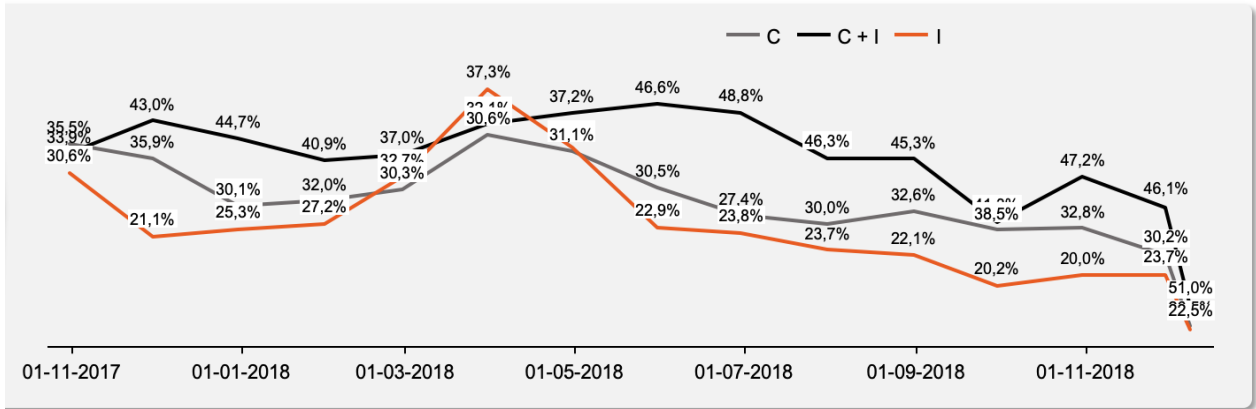


Figura 1.23: Tasa de conversión, Cross-selling, fono:



## 1.4. Justificación de un modelo predictivo

La subgerencia de Pricing y Portfolio tiene como objetivo la proposición y gestión de paquetización comercial y de fijación de precios, lo cual logra mediante el estudio de bases de datos referidas a los clientes de la organización. Específicamente, la subgerencia busca determinar patrones en los comportamientos de los clientes, generar proyectos de venta para los ejecutivos en búsqueda de una mayor captura y retención de clientes, y generar y regular los proyectos de fijación de precios en base a los comportamientos observados a partir del análisis ETL de las bases de datos de clientes. Además, el área trabaja en base a un Long Reach Plan, que tiene como objetivo la proyección del presupuesto destinado a cada cliente para satisfacción de necesidades y superación de expectativas respecto a los servicios contratados.

El problema afecta al solicitante en el aumento de rentabilidad y efectividad de las campañas de promoción de cross selling. Actualmente se cuenta con un sistema simplificado de priorización de clientes basado en el tipo de cliente que se tiene por empaquetamiento, y los productos que componen ese empaquetamiento, priorizando así las campañas únicamente hacia aquellos clientes de mayor valor. Ahora, de acuerdo con el subgerente, esto ya genera un aumento de alrededor de un 20 % en efectividad, sin embargo, requiere de una mejor segmentación y de un mejor conocimiento cuantitativo de la efectividad de las campañas, dado que el modelo actual es sumamente simplificado.

Identificación del problema:

El problema u oportunidad identificado corresponde a la propuesta de un mejor modelo para el sistema general de campañas de Cross-selling, debido a que actualmente no existe una segmentación o priorización de clientes cuantitativa respecto a quienes se les entregan las campañas, así como tampoco existen datos de predicción de efectividad basados en probabilidad de éxito y en atributos propios de los segmentos de clientes.

La oportunidad es clara debido a que se espera que cualquier modelo basado en datos debiese ser mejor que una priorización univariada basada en ingresos y paquete por clientes. Añadido a esto, cualquier propuesta que implique una mejora en la efectividad de las campañas de Cross-selling genera un aumento en ingresos y posiblemente en rentabilidad (es necesario considerar el comportamiento de clientes posterior a la campaña, efectos en fuga de clientes y costos a largo plazo del nuevo empaquetamiento).

### 1.4.1. Hipótesis de causa y alternativas de solución

Existen tres principales causas de la oportunidad identificada en el tema de trabajo de título, y estas apuntan a distintos contextos:

1) Modelo de segmentación simplificado e ingenuo:

Actualmente VTR no posee un modelo cuantitativo de segmentación de clientes a quienes se les aplica las campañas de Cross-selling (esto se extiende también a otras campañas). El modelo actual es altamente simplificado en el sentido que considera modelos univariados y no extensivos: se segmenta por tipo de producto, tipo de empaquetamiento, ARPC, pero siempre considerando solo una de estas variables para cada modelo de segmentación; además, el sistema de priorización para aplicación de campañas considera únicamente parámetros monetarios: se le ofrece la campaña a los clientes de acuerdo al aumento en ARPC o ARPU, de manera decreciente (no se considera valor esperado, probabilidad de efectividad, probabilidad de fuga o reconversión al paquete anterior, ni otros valores cuantitativos esenciales para predecir y medir la efectividad de una campaña).

2) Falta de predictibilidad del modelo:

El modelo actual no es capaz de predecir qué campaña será más efectiva para qué cliente, ni cuál cliente es más susceptible a algún tipo de producto o cambio en empaquetamiento específico.

3) Falta de incorporación de grandes cantidades de datos:

Actualmente se tienen enormes cantidades de datos de los clientes, no solo en cuanto a cantidad de variables, si no también en cuanto a evolución temporal. La cantidad de datos permite no solo conocer y segmentar mejor a los clientes, si no también estudiar su comportamiento temporal e identificar patrones dentro de él. Ahora, para generar modelos que consideren e incorporen esta enorme cantidad de datos en sus análisis, es necesario utilizar técnicas de aprendizaje artificial.

Ahora, en cuanto a alternativas de solución para estas causas, se propone:

1) Generar un modelo que entregue probabilidades específicas que, dado un set de atributos, segmento y comportamiento temporal del cliente, entregue la probabilidad de que este cliente realice un cambio de empaquetamiento específico (pudiéndose variar el nivel de empaquetamiento manteniendo el resto de valores fijos).

2) Generar un modelo de aprendizaje supervisado en redes neuronales, capaz de incorporar enormes cantidades de datos y aprender de los patrones de los clientes a medida que se

ingresan nuevos valores al modelo.

### 1.4.2. Estimación del valor agregado del modelo

El valor agregado de la investigación es claro: el diferencial de efectividad de campañas logrado por un modelo frente a otro conlleva a una diferencia en el mix de empaquetamiento de clientes, luego existen diferencias entre el ARPC inicial y el ARPC final, siendo esta última diferencia un aumento en ingreso por cliente, debido a su nuevo nivel de empaquetamiento.

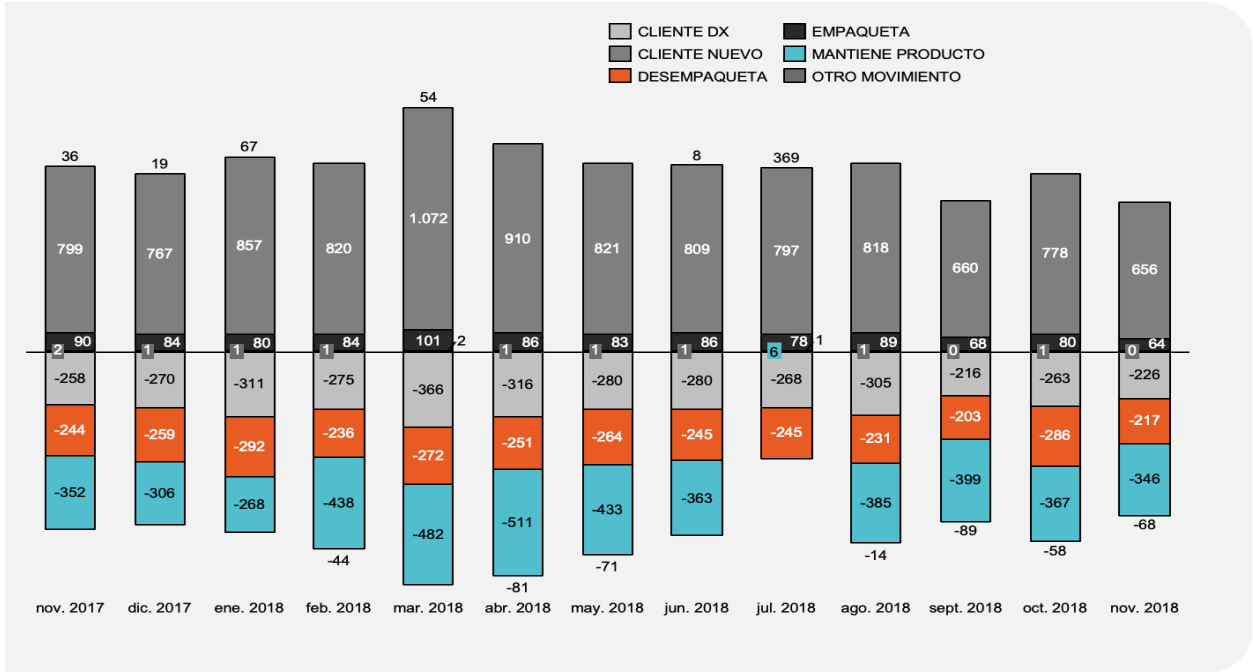
Ahora bien, para realizar el cálculo se debe considerar este aumento en ARPC como valor esperado, debido a que el modelo entregará una probabilidad de efectividad. Así mismo, ambos modelos no pueden ser comparados posterior a la aplicación de campañas, debido a que estas siguen uno u otro orden de priorización, pero no el de ambos al mismo tiempo. Así, para el cálculo del valor agregado de la investigación se debe realizar la ponderación del diferencial de ARPC por probabilidad de efectividad de cada modelo, y luego obtener la diferencia entre estos resultados.

Añadido a esto, se debe considerar la posibilidad de reconversión a empaquetamientos anteriores una vez finalizado el descuento o campaña, y el impacto en la fuga de clientes (promedio actual es de 2%).

Los impactos actuales manejados por la organización por migración entre niveles de empaquetamiento por cliente resultan ser los siguientes:



Figura 1.24: Impacto en ARPC (CLP):



# Capítulo 2

## Objetivos y Alcance

### 2.1. Objetivo General

Generar un modelo capaz de entregar probabilidades de cambio en empaquetamiento específicas por cliente, dentro del mercado de las telecomunicaciones, para la empresa VTR Comunicaciones SpA, enmarcado en empaquetamiento de productos fijos.

El modelo debe ser capaz de incorporar todas las variables que se requiera respecto al stock completo de clientes y entregar la probabilidad de cambio para cada cliente específico. La medición del logro de este objetivo serán las métricas de precisión y recall, métricas de comparación de ajuste del nuevo modelo respecto a un modelo aleatorio e impacto al negocio.

### 2.2. Objetivos Específicos

1. Generar un algoritmo de identificación de nivel de empaquetamiento por cliente para cada mes de permanencia, almacenando todas sus variaciones mensuales.
2. Analizar el rendimiento de modelos análogos de clasificación y su desempeño respecto al modelo escogido.
3. Estimar el efecto de un modelo predictivo capaz de generar un orden de clientes basados en su probabilidad de cambio en empaquetamiento.

## 2.3. Alcances

- El presente trabajo abarca desde la generación del algoritmo de obtención de datos respecto a empaquetamientos mensuales, hasta la obtención de las probabilidades específicas por cliente.
- El modelo tiene como producto final una lista de probabilidades estimadas por cliente.
- Se estudiarán únicamente los cambios en empaquetamiento que finalicen en un empaquetamiento de mayor ARPU respecto al inicial. Ahora bien, el modelo desarrollado es capaz de predecir cualquier cambio en empaquetamiento.
- El rendimiento del modelo será evaluado en cuanto a su capacidad de predicción y el impacto de posibilitar el orden de clientes por probabilidad de cambio en empaquetamiento.
- Las posibles segmentaciones realizadas a clientes no serán tratadas, dado que el modelo será generalizado y la discriminación de clientes será simplemente respecto a las variables de estudio y la probabilidad resultante. Añadido a esto, el modelo puede ser aplicado sobre cualquier set de clientes, luego también puede ser aplicado a futuro sobre segmentos específicos.
- No se tratarán recomendaciones o campañas de cross selling específicas, el modelo es aplicable de manera general y las recomendaciones serán únicamente respecto al orden de probabilidad propuesto.

## 2.4. Resultados esperados

El modelo debe ser capaz de incorporar todas las variables que se requiera respecto al stock completo de clientes y entregar la probabilidad de cambio para cada cliente específico. La medición del logro de este objetivo serán las métricas de precisión y recall, métricas de comparación de ajuste del nuevo modelo respecto a un modelo aleatorio e impacto al negocio.

Se espera lograr una predicción individual para cada cliente que entregue la probabilidad de que este, basado en una serie de variables que lo caracterizan, pertenezca a una cierta clase. Cada clase corresponderá a un empaquetamiento específico contratado por el cliente. Esta predicción será específica respecto a productos fijos y respecto a cambios en empaquetamientos que terminen en un empaquetamiento con mayor cantidad de productos en relación al empaquetamiento inicial.

Se espera también lograr una capacidad de predicción que supere a un modelo aleatorio, lo cual implica superar una precisión y un recall promedio de 50%. Añadido a esto, se define como corte esperado al menos un 65% de precisión promedio de clases (promediado entre

clase 0 sin cambio en empaquetamiento y clase 1 con cambio en empaquetamiento específico) y un recall de la misma magnitud, igual o mayor al 65 %. Esta definición es propuesta en base a la precisión del modelo actual de fuga de clientes utilizado por la organización. Dado que se busca predecir cambios en empaquetamiento y la fuga de clientes es un tipo de cambio de empaquetamiento que finaliza como desconexión del cliente, se espera al menos obtener una precisión similar para el caso de cualquier variación en empaquetamiento del cliente.

En cuanto a las métricas de ajuste frente a un modelo aleatorio, se espera que al momento de graficar la razón de verdaderos positivos respecto a la razón de falsos positivos el área bajo la curva resulte en un valor mayor a un modelo aleatorio en aproximadamente 25 puntos porcentuales. Añadido a esto, se espera que los perfiles de precisión acumulados muestren una clara ventaja al momento de priorizar clientes por probabilidad, obteniendo una distribución de probabilidad que permita mayor grado de separabilidad que un modelo aleatorio. De acuerdo a lo mencionado, si el modelo resulta superar a un modelo aleatorio y a otros modelos, su perfil de lift acumulado debería ser al menos dos veces mejor en cuanto al resultado de predicción.

Finalmente, también se espera que la extensión del modelo permita utilizar las probabilidades obtenidas para lograr posibles identificaciones de características y variables relevantes en los clientes y así tomar decisiones de priorización respecto a los resultados. De la misma manera, se espera mostrar como un modelo de probabilidad individual por cliente permite la generación de campañas de descuentos ajustadas de manera individual y permite a su vez una reducción en costos junto a una mayor efectividad. Añadido a esto, se presentará una extensión al modelo que permite identificar patrones en las variables utilizadas y su valor para cada cliente basándose en las probabilidades obtenidas y el entrenamiento realizado; todo esto con la finalidad de escapar el entorno de caja negra que presenta el modelo como tal.

# Capítulo 3

## Marco Teórico

### 3.1. Estado del arte

#### 3.1.1. Modelos tradicionales en telecomunicaciones

Las empresas de telecomunicaciones utilizan diversas bases de datos para crear perfiles de clientes robustos. Generando un perfil altamente detallado del cliente y una vista segmentada de la base de clientes permite elaborar mejores servicios al cliente, diferenciación, y personalización de recomendaciones enfocadas en distintas campañas, entre ellas: cross-selling. Mientras más robusto el perfil del cliente, mayor es el impacto de los modelos analíticos. Los portafolios de productos de las empresas de telecomunicaciones se están volviendo cada vez más complejos, evidenciado por el incremento en el número de ofertas de ancho de banda, planes móviles, contenido de los paquetes, entre otros. Esto crea un problema para las ventas y los analistas al momento de analizar las técnicas de cross-selling. Las reglas tradicionales de cross-sell se basan en la popularidad de los paquetes, basado en marketing mayoritariamente cualitativo.

Ahora, el marco del presente trabajo de título es aquel de la generación de un modelo estadística y analíticamente preciso para la predicción de probabilidad de cambios en empaquetamiento de clientes basado en técnicas computacionales de inteligencia artificial y minería de datos con el fin de manejar grandes cantidades de datos, caracterizar de manera certera a los clientes, y poder reconocer patrones de cambios en empaquetamiento que permitan extender el modelo a una mejor priorización de oferta en campañas.

### 3.1.2. Aprendizaje automático en telecomunicaciones

En el curso del tiempo, el data science ha probado su gran valor y eficiencia. Los profesionales data scientists encuentran cada vez más maneras de implementar soluciones de big data para una gran variedad de problemas industriales. Actualmente, los datos son un combustible necesario para cualquier compañía que busque tener éxito en el mercado, y las compañías de telecomunicaciones no son una excepción. Debido a estas circunstancias, las empresas de telecomunicaciones no pueden darse el lujo de no utilizar data science en la toma de sus decisiones.

Actualmente existen múltiples aplicaciones de aprendizaje automático en telecomunicaciones, estando entre ellas principalmente la detección de fraude, análisis predictivos, segmentación de clientes, prevención de fuga de clientes, predicción del lifetime value de los clientes, gestión de redes y optimización, desarrollo de productos, motores de recomendación, análisis de sentimiento de clientes, análisis en tiempo real y optimización de precios, entre otros.

## 3.2. Modelos de Aprendizaje Automático

### 3.2.1. Construir Modelos Inteligentes para Transformar Datos en Conocimiento

Actualmente existe una gran abundancia de datos, tanto estructurados como no estructurados, para prácticamente todo ámbito de la ciencia. En la segunda mitad del siglo veinte, nace el término de aprendizaje automático o machine learning, como un subgrupo de la categoría de Inteligencia Artificial. Este término involucra algoritmos de auto-aprendizaje que derivan conocimiento a partir de los datos permitiendo generar predicciones.

En vez de requerir humanos que manualmente deriven reglas y construyan modelos analizando grandes grupos de datos, el aprendizaje automático ofrece una alternativa de mayor eficiencia para capturar el conocimiento de los datos y así gradualmente mejorar el rendimiento de los modelos predictivos y permitir la toma de decisiones basadas en datos. El aprendizaje automático no solo se ha vuelto incrementalmente importante en las investigaciones de la ciencia de la computación; también juega un rol de vital importancia en el diario vivir de los habitantes. Gracias al aprendizaje automático existen filtros de spam robustos, softwares de reconocimiento de texto y voz, motores de búsqueda confiables, modelos de optimización de tráfico, modelos predictivos de accidentes y robos; entre muchas otras aplicaciones.

### 3.2.2. Tipos de Aprendizaje Automático

Fundamentalmente existen tres tipos de aprendizaje automático: aprendizaje supervisado, aprendizaje no supervisado, y aprendizaje con reforzamiento.

#### Aprendizaje Supervisado:

Este tipo de aprendizaje posee datos con etiquetas, posee retroalimentación directa y predice futuros resultados. El principal objetivo de esta técnica es aprender un modelo a partir de datos de entrenamiento etiquetados que permiten realizar predicciones sobre datos futuros que no se poseen. El término supervisado hace referencia al set de muestras en el cual las señales de salida (etiquetas) son conocidas.

La clasificación es una subcategoría del aprendizaje automático donde el objetivo es predecir las etiquetas categóricas de las clases en instancias nuevas de datos; utilizando como base las observaciones pasadas. Estas etiquetas de clase son valores discretos no ordenados que pueden ser entendidos como la agrupación de miembros de una instancia particular.

#### Aprendizaje con Reforzamiento:

Este tipo de aprendizaje es un proceso de decisión, se compone de un sistema de recompensas y busca aprender una serie de acciones.

En el aprendizaje con reforzamiento el objetivo es desarrollar un sistema (agente) que mejore su propio rendimiento basado en interacciones con el entorno. Dado que la información sobre el estado actual del entorno típicamente incluye una señal de recompensa, se puede pensar el aprendizaje con reforzamiento como un campo relacionado al aprendizaje supervisado. Sin embargo, en el aprendizaje con reforzamiento la retroalimentación no es el valor correcto o la etiqueta, sino una medida de cuán positiva fue la acción medida a través de una función de recompensa. A través de su interacción con el entorno, un agente puede entonces usar el reforzamiento para aprender una serie de acciones que maximicen esta función de recompensa a través de un acercamiento de prueba y error exploratorio o a través de planificación deliberada.

Un ejemplo típico del aprendizaje con reforzamiento es un motor de ajedrez. Aquí, el agente decide sobre una serie de movidas dependiendo del estado actual del tablero (el entorno), y la recompensa puede ser definida como ganar o perder al final del juego.

### Aprendizaje no Supervisado:

Este tipo de aprendizaje no posee etiquetas, no integra retroalimentación y busca encontrar estructuras escondidas en los datos.

En el aprendizaje supervisado se conoce la respuesta correcta de antemano al momento de entrenar el modelo y en el aprendizaje con reforzamiento se define una medida de recompensa para acciones particulares realizadas por el agente. Ahora bien, en el aprendizaje no supervisado se trabaja con datos no etiquetados y estructuras de datos no conocidas. Así, la idea es explorar la estructura de los datos y extraer información significativa sin la guía de una variable de resultado conocido o una función de recompensa.

Un ejemplo de aprendizaje no supervisado es el Clustering o agrupamiento de clases. Esta es una técnica que permite organizar un set de información en subgrupos (clusters) significativos, sin tener ningún conocimiento previo acerca de su pertenencia a uno u otro grupo. Cada cluster que nace a partir del análisis define un grupo de objetos que comparten un cierto grado de similitud pero que son más disímiles respecto a objetos en otros clusters; por ejemplo, en marketing permite descubrir grupos de clientes basado en sus intereses.

### **3.2.3. Construcción de Modelos de Aprendizaje Automático**

La construcción de los modelos de aprendizaje automático sigue la siguiente lógica:

- 1) Preprocesamiento: Extracción de características y tratamiento de las mismas, selección de Variables, reducción de Dimensionalidad y muestreo.
- 2) Aprendizaje: Selección del modelo, validación cruzada, métricas de rendimiento y optimización de hiperparámetros.
- 3) Evaluación del modelo final.
- 4) Predicción de nuevas instancias o etiquetas a partir de nuevos datos.

#### Pre Procesamiento:

El pre procesamiento de los datos busca permitir que estos se caractericen en un formato a partir del cual los distintos modelos puedan aprender de ellos. Así, lo primero que se hace es seleccionar o crear las variables: si los datos sin procesar fueran imágenes de un objeto, el pre procesamiento primero generaría variables que caractericen al objeto: intensidad del color, largo, ancho, curvas, entre otras. Añadido a esto, muchos modelos requieren que las características seleccionadas se encuentren en la misma escala para un rendimiento óptimo, lo cual se logra transformando las características en el rango  $[0,1]$  o mediante la normalización con media cero y varianza unitaria.



Algunas de las características seleccionadas pueden estar altamente correlacionadas, por lo cual la reducción de dimensionalidad puede ser útil para comprimir las variables a un sub-espacio característico de menor dimensión.

### Entrenamiento y Selección de Modelo:

Una importante conclusión en el ámbito del aprendizaje automático, extraída a partir de una famosa publicación de David Wolpert: "No free lunch theorems", es que no existe aprendizaje "gratuito". (The Lack of A Priori Distinctions Between Learning Algorithms, D.H. Wolpert, 1996, No free lunch theorems for optimization, D.H. Wolpert & W.G. Macready, 1997).

Intuitivamente, cada algoritmo de clasificación tiene sus sesgos inherentes, y ningún modelo de clasificación posee superioridad si no se realizan ciertos supuestos sobre la tarea. En la práctica es esencial comparar al menos una serie de diferentes algoritmos de tal manera de poder entrenar y seleccionar el mejor modelo.

Ahora bien, antes de seleccionar el modelo, es necesario definir la métrica que se utilizará para medir el rendimiento del mismo. Una métrica común utilizada en los modelos de clasificación corresponde a la precisión, que es definida como la proporción de instancias correctamente clasificadas.

Finalmente, no se puede esperar que los parámetros por default de los distintos algoritmos de aprendizaje provistos por las distintas librerías sean los óptimos para el problema específico a tratar. Debido a esto, se deben utilizar constantemente técnicas de optimización de hiperparámetros que permitan ajustar el rendimiento del modelo específico.

### Evaluación y Predicción de Futuras Instancias:

Luego de haber seleccionado un modelo y haberlo ajustado en el set de datos de entrenamiento específico, se procede a utilizar el set de datos de prueba para evaluar cuál es el rendimiento del modelo en datos que no ha visto antes; estimando de esta manera el error de generalización del mismo.

Si el rendimiento del modelo está dentro de los márgenes esperados, se puede utilizar para predecir nuevos datos. Ahora bien, es necesario aclarar que el set de datos de prueba debe pasar por el mismo tratamiento que los datos de entrenamiento; de otra manera el rendimiento del modelo puede ser demasiado optimista.

### 3.2.4. Python en Aprendizaje Automático

Python es uno de los lenguajes de programación más populares utilizado en el área de la ciencia de los datos y debido a esto posee una gran cantidad de librerías de gran utilidad y de uso libre. Para todos los modelos de aprendizaje automático tratados, el presente trabajo referirá principalmente a las librerías scikit-learn, tensorflow y keras, reconocidas como las más populares, accesibles y de mejor ajuste para este tipo de modelos de acuerdo a Sebastian Raschka y Vahid Mirjalili, en "Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow".

Añadido a esto, se utilizará una serie de paquetes para el trabajo y manipulación de datos; y para la construcción y ajuste de los modelos:

- NumPy: Paquete útil para la manipulación de datos en arreglos multidimensionales.
- Pandas: Librería construída sobre NumPy que provee herramientas de alto nivel para el trabajo y manipulación de datos tabulados.
- Matplotlib: Librería modificable enfocada en visualización cuantitativa de datos.

## 3.3. Métricas de Evaluación

### 3.3.1. Métricas de Precisión

En primera instancia se evaluará el modelo respecto a tres principales métricas. Ahora, la realización del análisis que procede requiere la previa definición de los principales términos a utilizar y la caracterización de su pertinencia en evaluación.

- True Positive / Verdadero Positivo: El modelo predice un resultado positivo y el estado real es efectivamente positivo.
- False Positive / Falso Positivo: El modelo predice un resultado positivo y el estado real es negativo.
- True Negative / Verdadero Negativo: El modelo predice un resultado negativo y el estado real es efectivamente negativo.
- False Negative / Falso Negativo: El modelo predice un resultado negativo y el estado real es positivo.

- Precisión: Corresponde al ratio entre verdaderos positivos y el número total de resultados positivos:

$$\frac{VerdaderosPositivos}{VerdaderosPositivos + FalsosPositivos}$$

. Intuitivamente es la habilidad que posee el clasificador para clasificar un resultado como negativo dado que en la muestra real es negativo. El mejor valor que alcanza es 1, el peor valor que alcanza es 0.

- Recall: Corresponde al ratio entre verdaderos positivos y el número de verdaderos positivos sumado al número de falsos negativos:

$$\frac{VerdaderosPositivos}{VerdaderosPositivos + FalsosNegativos}$$

. Intuitivamente es la habilidad que posee el clasificador para encontrar todas las muestras positivas. El mejor valor que alcanza es 1, el peor valor que alcanza es 0.

- F1 Score:

El F1 Score puede ser interpretado como el promedio ponderado entre la precisión y el recall, alcanzando su mejor valor en 1 y su peor valor en 0. La contribución relativa entre la precisión y el recall al F1 Score es igual. La fórmula es la siguiente:

$$2 * \frac{Precision * Recall}{Precision + Recall}$$

### 3.3.2. Métrica ROC - AUC

Adicional a la precisión, el recall y el F1 Score, existen otras medidas de evaluación especialmente útiles para medir el rendimiento de los modelos de clasificación.

Entre estas medidas se encuentra la curva AUC - ROC (Área bajo la curva - Característica Operativa del Receptor). Esta métrica es una medida de rendimiento para los problemas de clasificación a distintos límites. ROC es una curva de probabilidad y AUC representa el grado o la medida de separabilidad. Esta curva muestra cuán hábil es un modelo para distinguir entre clases. Mientras más alta sea el área bajo la curva, mejor es el modelo al momento de identificar a los 0's como 0's y a los 1's como 1's. Análogamente, mientras más alto el AUC, mejor es el modelo para distinguir entre clientes con cambio de empaquetamiento y cliente sin cambios en empaquetamiento.

La curva ROC es graficada utilizando la razón de verdaderos positivos (eje y) respecto a la razón de falsos positivos (eje x).

Razón de verdaderos positivos = Recall = Sensitividad =

$$\frac{\textit{VerdaderosPositivos}}{\textit{VerdaderosPositivos} + \textit{FalsosNegativos}}$$

Especificidad =

$$\frac{\textit{VerdaderosNegativos}}{\textit{VerdaderosNegativos} + \textit{FalsosPositivos}}$$

Razón de falsos positivos = 1 - Especificidad =

$$\frac{\textit{FalsosPositivos}}{\textit{VerdaderosNegativos} + \textit{FalsosPositivos}}$$

Un modelo excelente tiene un AUC cercana a 1, lo cual significa que tiene una buena medida de separabilidad. Un modelo pobre tiene un AUC cercana a 0, lo cual significa que tiene la peor medida de separabilidad, de hecho, significa que está tomando el resultado recíproco: está prediciendo los 0's como 1's y los 1's como 0's. Así, cuando el AUC es 0,5, significa que el modelo no tiene capacidad de separación alguna.

### 3.3.3. Perfiles de Precisión Acumulados (CAP)

El gráfico comparativo de perfiles de precisión acumulados funciona de la siguiente manera: se tienen como resultados positivos los cambios en empaquetamiento de los clientes, basado en una serie de variables. Si un grupo de clientes fuera contactado de manera aleatoria, el número acumulado de cambios en empaquetamiento seguiría una tendencia lineal hasta un valor máximo correspondiente al total de clientes dentro del mismo grupo. Esta distribución tiene el nombre de CAP aleatoria. Por otro lado, una predicción perfecta determina exactamente qué miembros del grupo cambiarán de empaquetamiento, de tal manera que el número máximo de cambios en empaquetamiento será alcanzado con un mínimo de clientes contactados. Esto produce una curva CAP de alta pendiente que se mantiene plana una vez alcanzado el máximo (contactar a todos los otros miembros del grupo no producirá más cambios en empaquetamiento), correspondiendo así a la curva CAP perfecta.

Ahora, el modelo programado predice la verosimilitud de que los individuos cambien de empaquetamiento y rankea las probabilidades de tal manera de producir una lista de potenciales clientes a contactar primero. El número acumulado de cambios de empaquetamiento resultante incrementará rápidamente y eventualmente tendrá una tendencia relativamente plana a medida que el número de clientes contactados se acerca al número de miembros del grupo de análisis. Esto resulta en una distribución que se posiciona entre la curva aleatoria CAP y la curva perfecta CAP.

### 3.3.4. Perfiles de Lift Acumulados

El gráfico de perfiles de Lift acumulados permite responder a la siguiente pregunta: cuando se aplica el modelo y se seleccionan los mejores deciles, ¿cuántas veces mejora el resultado respecto a no tener ningún modelo?

Esta gráfica ayuda a explicar cuán mejor es un modelo al momento de seleccionar clases, respecto a realizar selecciones aleatorias. Es simplemente el ratio entre el porcentaje de clasificaciones positivas respecto al porcentaje de observaciones consideradas; en el caso específico del presente modelo, sería el porcentaje de clientes contactados que cambiaron su empaquetamiento dividido por la cantidad de clientes considerados en la muestra.

Así, un lift de 1 es equivalente a ninguna ganancia comparada a la selección de clientes de manera aleatoria; un lift de 2, por ejemplo, corresponde a obtener el doble de cambios en empaquetamiento comparado con la selección de la misma cantidad de clientes de manera aleatoria.

La gráfica de lift tiene solo una línea de referencia: el modelo aleatorio.

El modelo aleatorio significa que cada observación obtiene una clasificación aleatoria, luego todos los casos son divididos en deciles basándose en esa clasificación aleatoria. Así, el lift correspondiente será una línea horizontal con el valor de 1.

### 3.3.5. Prueba de Kolmogorov - Smirnov (Estadístico KS)

La prueba de Kolmogórov-Smirnov es una prueba no paramétrica que determina la bondad de ajuste de dos distribuciones de probabilidad entre sí: cuantifica la distancia entre la función de distribución positiva de la muestra de clasificación y la función de distribución negativa de la muestra de clasificación. Es simplemente la máxima diferencia entre el ratio de verdaderos positivos y el ratio de falsos positivos. Un mayor estadístico KS significa un mejor modelo debido a su mejor capacidad de clasificación.

## 3.4. Principales Modelos de Clasificación

### 3.4.1. Regresión Logística

La regresión logística es un modelo de clasificación de fácil implementación que tiene un muy buen rendimiento en clases linealmente separables. Para explicar la idea detrás de la regresión logística como modelo probabilístico primero se debe introducir el concepto de razón de probabilidades: las probabilidades a favor de un evento en particular, que puede ser escrito como:

$$\frac{p}{(1 - p)}$$

donde  $p$  es la probabilidad de un evento positivo. El término evento positivo no necesariamente significa "bueno", si no que refiere al evento que se busca predecir. Así, se puede pensar al evento positivo como una etiqueta de clase del tipo  $y = 1$ . Más aún, se puede definir la función logit, que es simplemente el logaritmo de la razón de probabilidades (log refiere al logaritmo natural):

$$\text{logit}(p) = \log\left(\frac{p}{(1 - p)}\right)$$

La función logit toma valores de input en el rango 0 a 1, los transforma a valores en

el rango entero de números reales, y permite expresar una relación lineal entre los valores característicos y las probabilidades logarítmicas:

$$\text{logit}(p(y = 1, x)) = w_0x_0 + w_1x_1 + \dots + w_mx_m$$

La probabilidad corresponde al caso en que una muestra pertenezca a la clase 1, dado un set de características  $x$ .

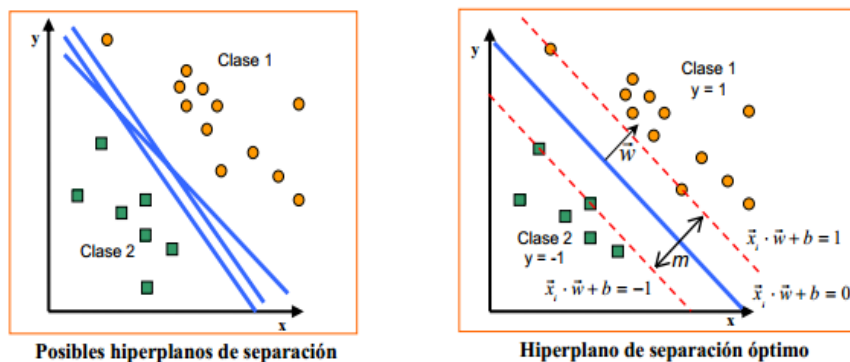
Ahora, el interés se encuentra en predecir la probabilidad de que una cierta muestra pertenezca a una clase particular, que corresponde a la inversa de la función logit, también llamada función logística sigmoide debido a su forma de S:

$$\phi(Z) = \frac{1}{1 + \exp -z}$$

### 3.4.2. Support Vector Machines

Este algoritmo es un potente modelo de clasificación enfocado en maximizar un margen: la distancia entre el hiperplano de separación (límite de decisión) y las muestras de entrenamiento que se encuentran más cercanas a este hiperplano, los cuales son denominados vectores de soporte.

Figura 3.1: Margen:



La racionalidad detrás de la creación de límites de decisión con grandes márgenes refiere al bajo error de generalización: modelos con bajos márgenes son más susceptibles a sobreajuste.

### 3.4.3. Kernel Support Vector Machines

Este algoritmo de clasificación corresponde a una modificación al modelo de vectores de soporte que permite resolver problemas de clasificación no lineales. Los métodos de Kernel tratan los datos linealmente inseparables mediante la creación de combinaciones no lineales de las características originales para así proyectarlas a un espacio de mayor dimensión mediante una función de mapeo  $\phi$ , espacio en el cual se vuelven linealmente separables.

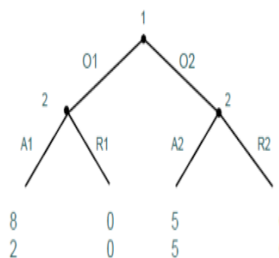
Por ejemplo, es posible transformar un set de datos bidimensional en un espacio característico tridimensional donde las clases se vuelven linealmente separables mediante la siguiente proyección:

$$\phi(x_1, x_2) = (z_1, z_2, z_3) = (x_1, x_2, (x_1)^2 + (x_2)^2)$$

### 3.4.4. Decision Trees

Los clasificadores del tipo Decision Tree son modelos atractivos si se busca interpretabilidad. Este modelo puede ser pensado como la descomposición de los datos mediante la toma de una decisión basada una serie de preguntas.

Figura 3.2: Árbol de Decisión:



Utilizando un algoritmo de decisión, el modelo comienza en la raíz del árbol y divide los datos en la característica que genere la mayor ganancia de información (dependiendo del modelo). En un proceso iterativo, se repite la división de los datos en cada nodo resultante hasta que los nodos sean puros, esto significa que todas las muestras, en cada nodo, pertenezcan a una misma clase. En la práctica, esto puede significar un árbol con mucha profundidad y muchos nodos, lo cual fácilmente puede llevar a sobre ajuste. Debido a esto, típicamente se define un límite para la profundidad máxima del árbol.



### 3.4.5. Random Forest

Este modelo corresponde a un conjunto de modelos del tipo Decision Tree. La lógica detrás del modelo es generar promedios entre múltiples y profundos árboles de decisión que individualmente sugren de una muy alta varianza. Haciendo esto, se logra construir un modelo más robusto que posee un mejor rendimiento de generalización y es menos susceptible al sobre ajuste.

### 3.4.6. K Nearest Neighbors

El algoritmo de K vecinos más cercanos es fundamentalmente diferente a los modelos ya descritos. Este modelo es un ejemplo de un aprendizaje "flojo", denominado así debido no debido a su simplicidad, más bien debido a que no aprende una función discriminadora a partir de los datos de entrenamiento: en vez de esto memoriza los datos de entrenamiento. De esta manera, su funcionamiento es el siguiente:

- Se define un número k y una métrica de distancia.
- Se encuentran los k vecinos más cercanos a la muestra que se busca clasificar.
- Se asigna una etiqueta de clase que refiere al tipo de dato con mayor cantidad de instancias dentro de ese vecindario.

### 3.4.7. Naive Bayes

También denominado clasificador bayesiano ingenuo, este modelo es un clasificador de probabilidades que tiene como base el teorema de bayes:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

El algoritmo asume que la presencia de una característica o su ausencia no está relacionada con la misma presencia de otra característica presente en la clase variable. Así, si la renta, la profesión y el grado socio económico de un cliente lo identifican como una oportunidad de venta cruzada, el modelo considera que cada uno de estos atributos contribuye de manera independiente a la probabilidad de que el cliente sea una oportunidad de venta cruzada, independientemente de la presencia de los otros atributos.

Una ventaja de este tipo de clasificadores es que requieren una muy pequeña cantidad de datos de entrenamiento para la estimación de parámetros requeridos para la clasificación (solo requiere las varianzas de las variables de cada clase).

### 3.4.8. Red Neuronal Artificial

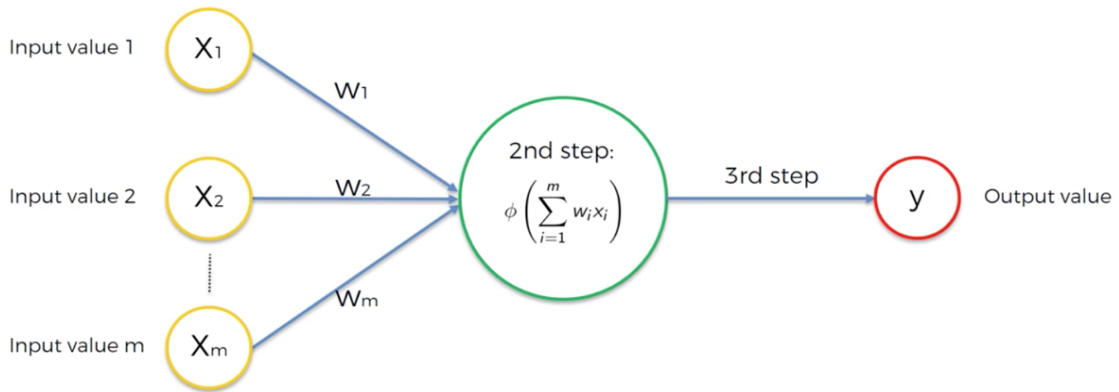
El principal componente de la red son las neuronas, las cuales al igual que las neuronas orgánicas, reciben señales de entrada (valor numérico) desde otras neuronas, y emiten una señal de salida (valor numérico).

Un segundo componente es la función de activación. Esta función debe ser escogida y corresponde a una función matemática que asigna un valor 'y' para cada valor 'x' de entrada. Ahora bien, como se observa en la figura 4.1, la neurona recibe la suma de las variables de entrada ponderada cada una por un peso específico, luego el resultado es evaluado en la función de activación y se obtiene un valor de salida:

$$salida = \phi\left(\sum_{i=1}^m w_i x_i\right)$$

De esta manera, una red neuronal se compone de una serie de neuronas agrupadas por capa

Figura 3.3: Función de Activación:



y cada capa recibe señales de entrada de la capa anterior, ponderadas por un peso numérico presente en la unión entre neuronas.

La manera en que la red aprende sobre los datos de entrenamiento es la siguiente:

1. Se inicializan de manera aleatoria los pesos utilizando valores cercanos a 0, pero no exactamente 0.
2. Se ingresa la primera observación del set de datos en la capa de entrada, cada variable asignada a un nodo de entrada.
3. Comienza la propagación hacia delante: desde izquierda a derecha, las neuronas son activadas de tal manera que el impacto de cada neurona activada sea limitado por los pesos. Se propagan las activaciones hasta llegar al resultado predicho y.
4. Se realiza la comparación entre el resultado predicho y el resultado real, se mide el error generado.
5. Comienza la propagación inversa: desde derecha a izquierda el error es propagado de manera inversa. Se actualizan los pesos de acuerdo a cuánto impacto tienen en el error y cuánta responsabilidad por el mismo. La razón de aprendizaje decide en qué magnitud se actualizan los pesos.
6. Se repiten los pasos anteriores y se actualizan los pesos luego de cada observación (aprendizaje forzado), o bien se repiten los pasos anteriores y se actualizan los pesos solo luego de un lote de observaciones (aprendizaje por lotes).
7. Finalmente, cuando todo el set de entrenamiento ha pasado por la red neuronal artificial, se tiene un 'epoch'. Se procede a realizar una serie de 'epochs'(determinada en el código).

# Capítulo 4

## Selección del Modelo

### 4.1. Introducción al Modelo

El modelo busca predecir los cambios en empaquetamiento de los clientes entregando una probabilidad de pertenencia a una clase. En este sentido, el problema tiene un carácter de clasificación, donde la clase a predecir es aquella que pertenece a un cierto tipo de empaquetamiento.

Ahora bien, existe una gran cantidad de clasificadores de aprendizaje automático supervisado, por lo cual la selección de un modelo apropiado para el problema requiere previamente estudiar el rendimiento de los distintos modelos sobre los datos, y sus características propias respecto al tratamiento matemático de los datos.

Para realizar una selección adecuada del modelo, se procede a estudiar la literatura empírica existente, se identifican posibles comparaciones del tipo "benchmark" respecto a una serie de variables y finalmente se comparan los principales modelos de manera competitiva sobre un set de datos adecuado y referente al presente problema de estudio.

## 4.2. Benchmark de modelos de aprendizaje automático

A continuación se muestran las principales características de los modelos respecto al tratamiento matemático de los datos y la manera de procesamiento y optimización del modelo como tal.

- Generativo (modela explícitamente la distribución de cada clase) o discriminativo (modela los límites de decisión entre las clases):
- Loss Function: nombre de la función de pérdida optimizada por el algoritmo.
- Límites de Decisión / Forma de la función de regresión: se describe la forma de la superficie de decisión o de la función de regresión.
- Algoritmo de estimación de parámetros / Algoritmo de predicción: Se nombra o describe concisamente el algoritmo utilizado para estimar los parámetros o para predecir el valor de la nueva instancia.
- Reducción de Complejidad del Modelo: Se nombra una técnica para limitar la complejidad del modelo y prevenir su sobre ajuste.

Modelo	Carácter	Función de Pérdida	Límite de Decisión
Naive Bayes	Generativo	$\log(P(X, Y))$	Igual varianza: límite lineal, Varianza desigual: Límite cuadrático
Regresión Logística	Discriminativo	$\log(P(X, Y))$	Lineal
Decision Trees	Discriminativo	$\log(P(X, Y))$ o Función 0,1	División del espacio característico alineando por ejes
K Nearest Neighbors	Discriminativo	Función 0,1	Arbitrariamente compleja
Support Vector Machines	Discriminativo	Función de máximo margen $ 1 - y(w^T x) _+$	Lineal (depende del kernel)
Redes Neuronales	Discriminativo	Suma de errores cuadráticos	Relaciones tanto lineales como no lineales.

Tabla 4.1: Comparación de Algoritmos de Clasificación

Modelo	Algoritmo de Estimación	Reducción de Complejidad
Naive Bayes	Estimar $\hat{\mu}$ , $\hat{\sigma}^2$ y $P(Y)$ utilizando máxima verosimilitud	Estimación máxima a posteriori
Regresión Logística	No posee forma cerrada utiliza gradiente descendiente	Regularización $L_2$
Decision Trees	Múltiples algoritmos: ID3, CART, C4.5	Cortar árbol o limitar la profundidad del mismo
K Nearest Neighbors	Almacena todos los datos de entrenamiento para clasificar nuevos puntos. Escoge K utilizando validación cruzada	Incrementar K
Support Vector Machines	Resuelve programa cuadrático para encontrar el límite que maximiza el margen.	Reducir el espacio dimensional
Redes Neuronales	Propagación hacia atrás, Propagación hacia delante.	Reducir la cantidad de capas. Regularización.

Tabla 4.2: Comparación de Algoritmos de Clasificación

Las funciones de pérdida resultan ser en su mayoría similares, sin embargo existen principales diferencias entre aquellos que modelan utilizando funciones logarítmicas y aquellos que modelan siguiendo un camino que minimza los errores.

Para el presente problema se busca clasificar categorías o niveles de empaquetamiento, luego se requiere una función de pérdida que permita discriminar a partir de los errores. Además, el carácter debe ser claramente discriminativo, debido a que se busca modelar los límites de decisión entre las distintas clases de empaquetamiento en clientes.

Considerando las características descritas y dejando a un lado la reducción de complejidad debido a que se busca el modelo que entregue la mayor precisión posible, los modelos a estudiar que se ajustan en mayor medida al problema de clasificación descrito en el presente trabajo resultan ser: regresión logística, decision trees, k nearest neighbors, support vector machines y redes neuronales.

### 4.3. Evaluación competitiva de modelos de aprendizaje automático

Previa a la optimización y programación del modelo final, se extrae la base de cambios en empaquetamiento que finalizan en empaquetamientos triples para el año 2018, y se realizan una serie de pruebas de efectividad de predicción utilizando distintos modelos de aprendizaje automático. Esto se realiza con la finalidad de justificar empírica y competitivamente el modelo escogido (Red Neuronal Artificial).

#### 4.3.1. Métricas de Precisión

A continuación se procede a evaluar los siguientes modelos en las métricas de precisión, recall y F1 Score (utilizando la base de cambios en empaquetamiento que finalizan en empaquetamientos triples en el transcurso del año 2018): red neuronal artificial, regresión logística, support vector machines, kernel support vector machines, k nearest neighbors, random forest, decision tree, naive bayes. Los resultados se detallan a continuación.

Estado/Métrica	Precisión	Recall	F1 Score	Datos de soporte
0	0.59	0.62	0.61	6.001
1	0.61	0.58	0.59	6.126
Promedio ponderado	0.60	0.60	0.60	12.127

Tabla 4.3: Métricas de precisión, Red Neuronal Artificial (Entrenada sobre 100 epochs).

Estado/Métrica	Precisión	Recall	F1 Score	Datos de soporte
0	0.58	0.56	0.57	6.023
1	0.57	0.59	0.58	6.104
Promedio ponderado	0.58	0.58	0.58	170.000

Tabla 4.4: Métricas de precisión, Regresión Logística.

Estado/Métrica	Precisión	Recall	F1 Score	Datos de soporte
0	0.59	0.62	0.60	5.954
1	0.61	0.57	0.59	6.173
Promedio ponderado	0.60	0.60	0.60	12.127

Tabla 4.5: Métricas de precisión, Random Forest.

Estado/Métrica	Precisión	Recall	F1 Score	Datos de soporte
0	0.59	0.57	0.58	6.025
1	0.59	0.61	0.60	6.102
Promedio ponderado	0.59	0.59	0.59	12.127

Tabla 4.6: Métricas de precisión, K-Nearest Neighbors.

Estado/Métrica	Precisión	Recall	F1 Score	Datos de soporte
0	0.58	0.60	0.59	5.901
1	0.59	0.56	0.58	6.226
Promedio ponderado	0.58	0.58	0.59	12.127

Tabla 4.7: Métricas de precisión, Decision Tree.

Estado/Métrica	Precisión	Recall	F1 Score	Datos de soporte
0	0.50	1.0	0.67	6.053
1	0.00	0.00	0.00	6.074
Promedio ponderado	0.50	0.33	0.33	12.127

Tabla 4.8: Métricas de precisión, Naive Bayes.

Estado/Métrica	Precisión	Recall	F1 Score	Datos de soporte
0	0.57	0.60	0.59	6.023
1	0.59	0.56	0.58	6.104
Promedio ponderado	0.58	0.58	0.59	12.127

Tabla 4.9: Métricas de precisión, Kernel Support Vector Machines.

Estado/Métrica	Precisión	Recall	F1 Score	Datos de soporte
0	0.70	0.65	0.65	85.000
1	0.30	0.30	0.35	85.000
Promedio ponderado	0.50	0.47	0.50	170.000

Tabla 4.10: Métricas de precisión, Support Vector Machines.

#### Análisis:

Se observa que los modelos que logran mejor precisión, recall, y consecuentemente F1 Score, corresponden a la red neuronal artificial, random forest, decision tree (sub-grupo de random forest) y K-Nearest Neighbors.

Ahora bien, en primera instancia, y considerando el set de datos descritos, tanto la red neuronal artificial como el modelo random forest presentan la mayor precisión al momento de clasificar los cambios en empaquetamiento.



Estado/Métrica	Red Neuronal	Random Forest	Decision Tree	K-NN
0	0.59	0.59	0.58	0.59
1	0.61	0.61	0.59	0.59
Promedio ponderado	0.60	0.60	0.59	0.59

Tabla 4.11: Resumen de Precisión de Mejores Modelos.

### 4.3.2. Métrica ROC - AUC

Adicional a la precisión, el recall y el F1 Score, existen otras medidas de evaluación especialmente útiles para medir el rendimiento de los modelos de clasificación.

Entre estas medidas se encuentra la curva AUC - ROC (Área bajo la curva - Característica Operativa del Receptor). Esta métrica es una medida de rendimiento para los problemas de clasificación a distintos límites. La curva característica operativa del receptor (ROC) es una curva de probabilidad, por lo cual es conveniente graficar la distribución de esas probabilidades. A continuación se grafican las distribuciones de probabilidad para los modelos descritos anteriormente.

Ahora bien, no es posible graficar las curvas de probabilidad para los siguientes modelos:

- Kernel Support Vector Machines: No permite rescatar las probabilidades del modelo (entrega true o false).
- Support Vector Machines: No permite rescatar las probabilidades del modelo (entrega true o false).
- Naive Bayes: Las probabilidades que entrega únicamente clasifican resultados positivos, por lo cual no permite realizar comparación de curvas.

\*Cabe destacar que se grafican las probabilidades para el modelo Decision Tree, pero estas son siempre 1 o 0 debido a que el modelo únicamente clasifica en estas dos categorías, sin considerar puntos intermedios.

Figura 4.1: Distribución de probabilidad de las curvas de clasificación, red neuronal artificial:

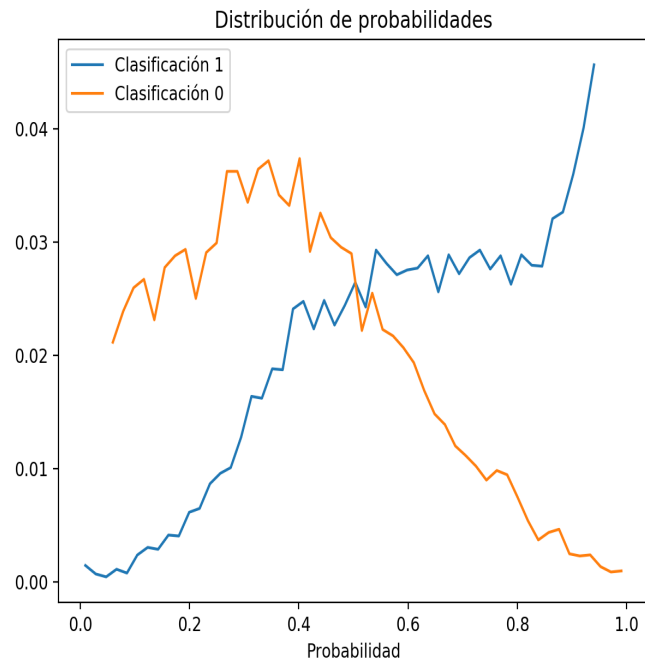


Figura 4.2: Distribución de probabilidad de las curvas de clasificación, random forest:

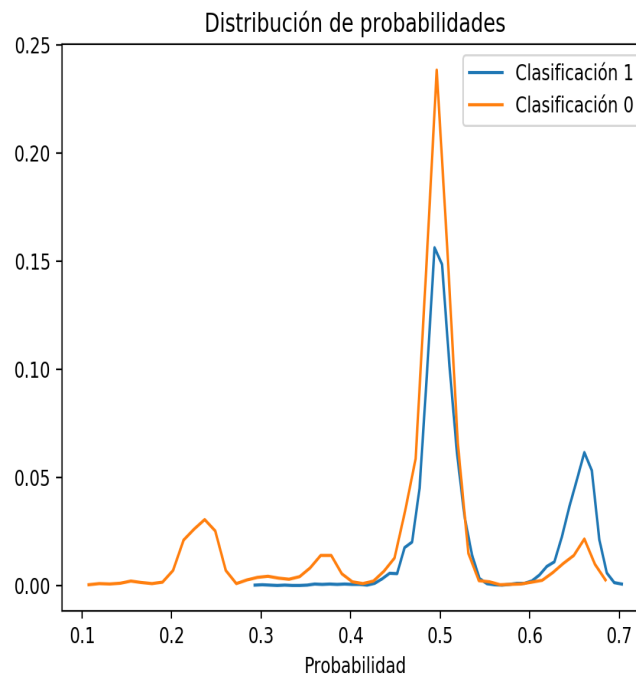


Figura 4.3: Distribución de probabilidad de las curvas de clasificación, decision tree:  
Distribución de probabilidades

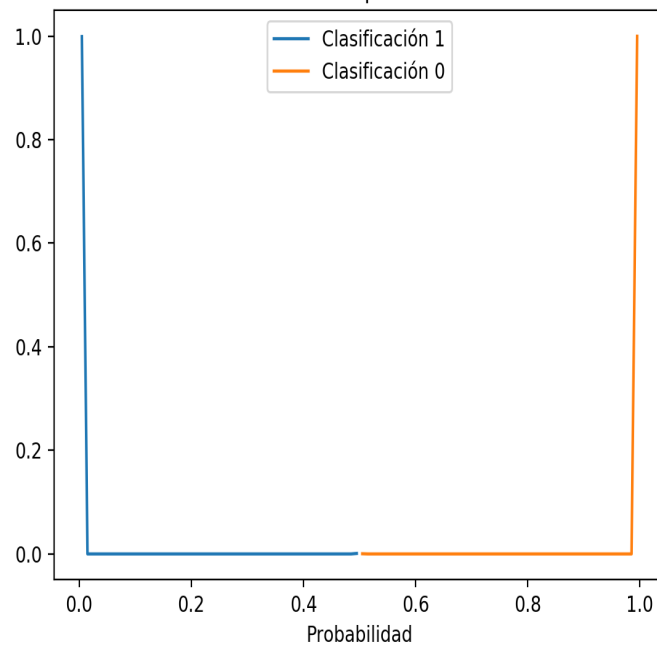


Figura 4.4: Distribución de probabilidad de las curvas de clasificación, K Nearest Neighbors:

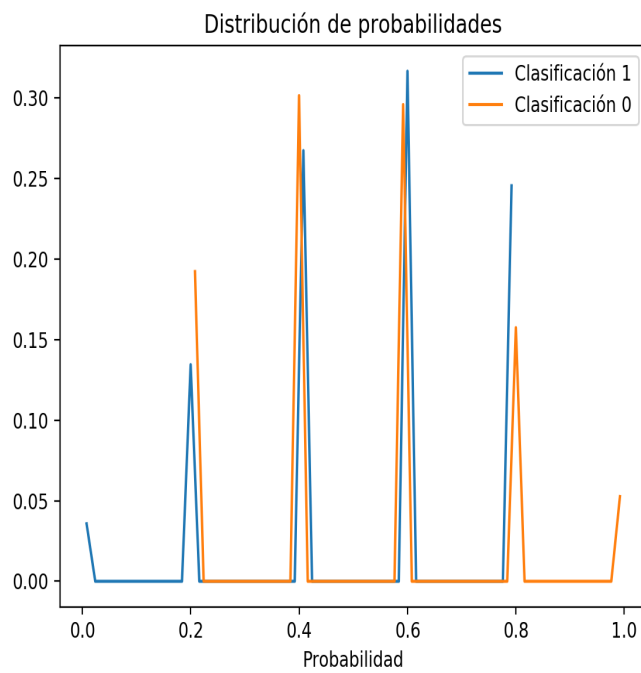
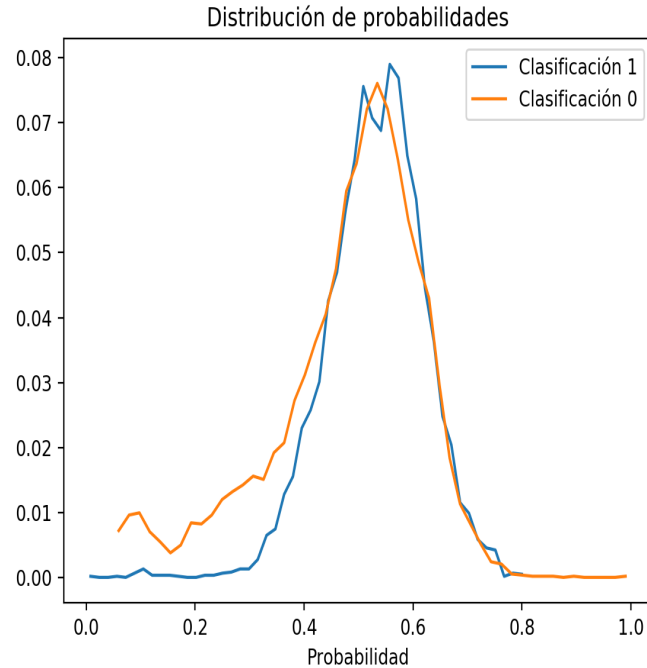


Figura 4.5: Distribución de probabilidad de las curvas de clasificación, Regresión Logística:



Si las distribuciones no se cruzan se tiene un modelo con una medida de separabilidad perfecta para distinguir entre clases 1 y clases 0. Por otra parte, si las curvas se cruzan, se introducen los errores de tipo 1 y errores de tipo 2. Cuando el AUC es de 0.7, significa que el modelo tendrá una probabilidad de 70 % de distinguir entre la clase positiva y la negativa. Finalmente, si el AUC es de 0.5, el modelo es incapaz de distinguir una clase de otra.

#### Análisis:

Se observa que los modelos con mejores curvas de probabilidad, respecto a clasificación de cambios en empaquetamiento, resultan ser: red neuronal artificial, random forest, k nearest neighbors y regresión logística.

Ahora bien, el análisis de distribución de probabilidades permite rescatar una serie de conclusiones importantes para la selección del modelo final:

- El modelo decision tree no entrega probabilidades intermedias, por lo cual no permite la realización de un ranking de probabilidades de manera descendente.
- El modelo k nearest neighbors muestra rangos de probabilidad, presentando una discontinuidad en las mismas, la cual obliga a generar un ranking discreto de probabilidades de cambio de empaquetamiento al momento de ordenar de manera descendente la probabilidad de cada cliente.
- Los modelos red neuronal artificial, regresión logística y random forest son aquellos que presentan una mejor continuidad en cuanto a probabilidad.
- Los modelos red neuronal artificial y random forest resultan ser los que presentan la mejor curva de distribución de probabilidades para la generación de un posible ranking descendente.

## Curva ROC:

Ahora, habiendo considerado las distribuciones de probabilidad, se procede a graficar la curva ROC para los modelos, esta vez se incluyen los modelos Kernel Support Vector Machines, Support Vector Machines, y Decision Tree, debido a que la curva ROC se compone de razones de clasificación, las cuales se pueden generar con las probabilidades enteras que entregan los modelos antes descartados. Los resultados son los siguientes:

Figura 4.6: ROC - Red Neuronal Artificial:

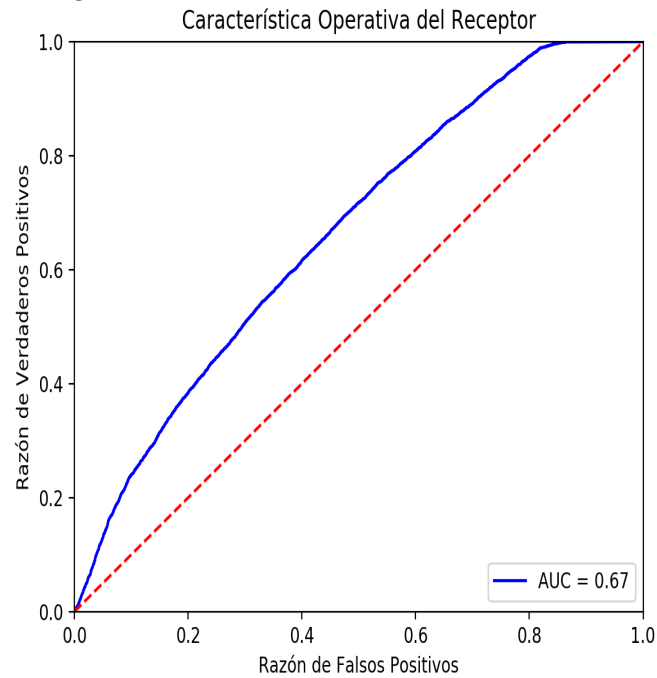


Figura 4.7: ROC - Regresión Logística:  
Característica Operativa del Receptor

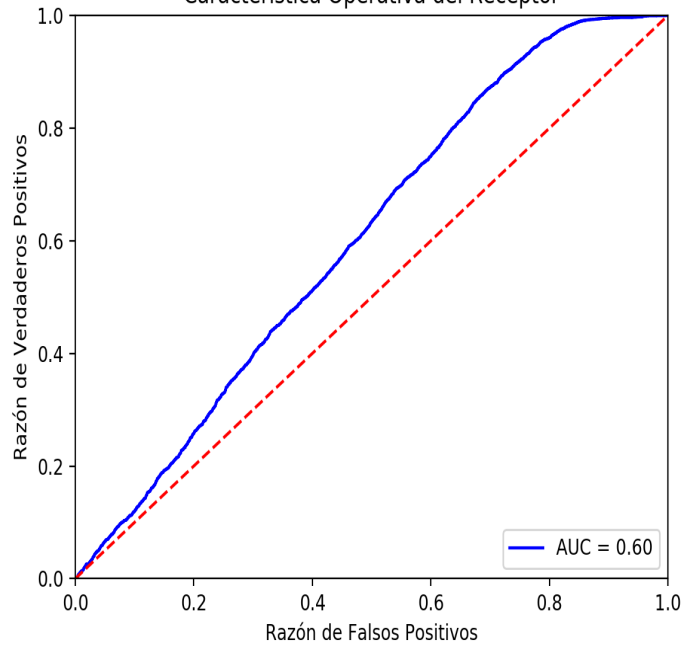


Figura 4.8: ROC - Random Forest:  
Característica Operativa del Receptor

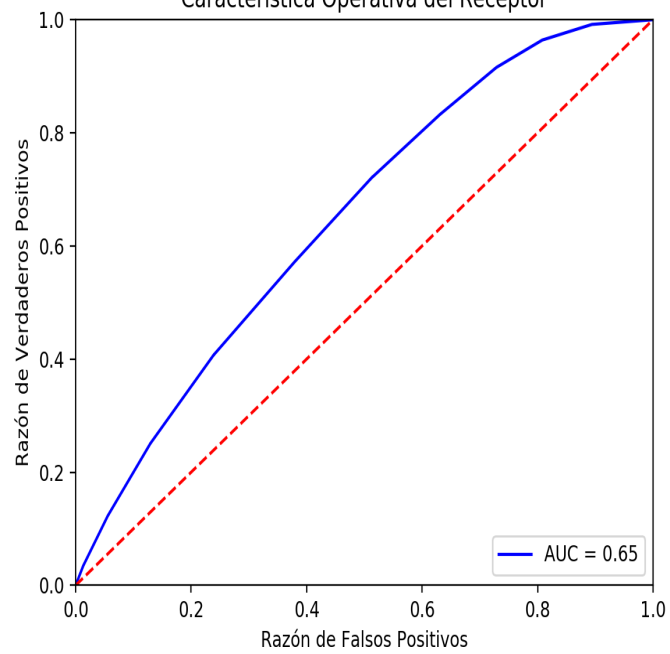


Figura 4.9: ROC - K Nearest Neighbors:  
Característica Operativa del Receptor

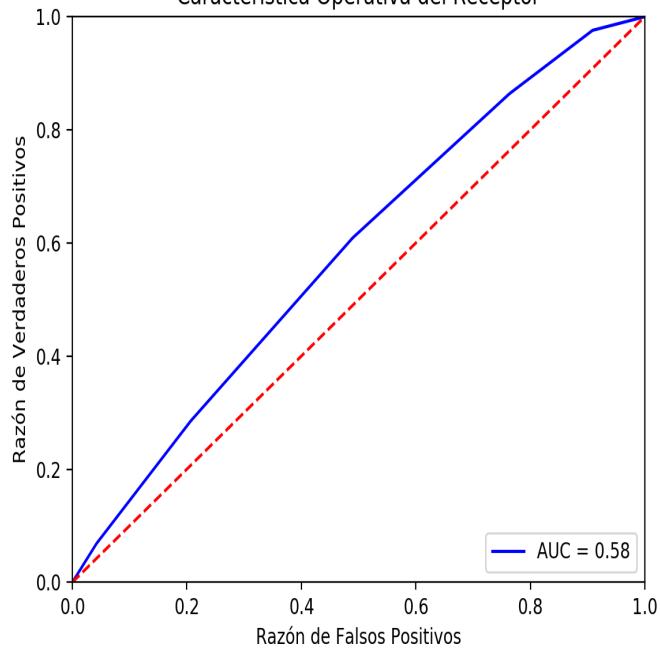


Figura 4.10: ROC - Decision Tree:  
Característica Operativa del Receptor

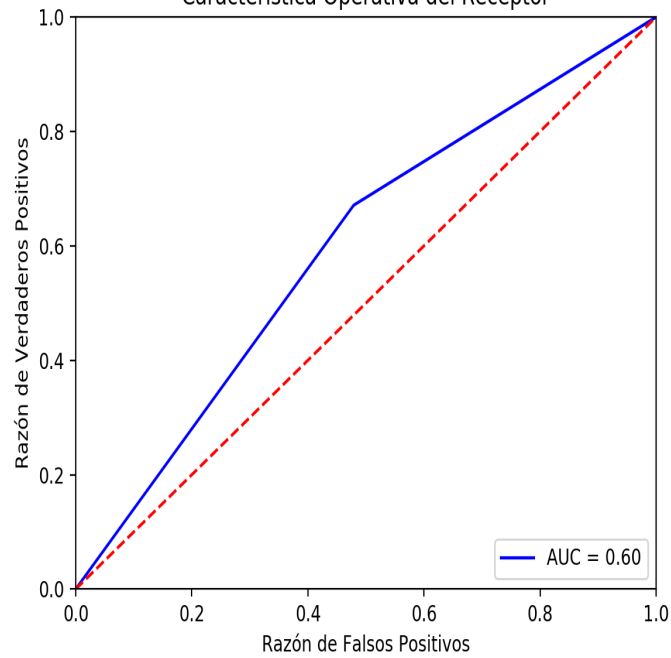


Figura 4.11: ROC - Naive Bayes:  
Característica Operativa del Receptor

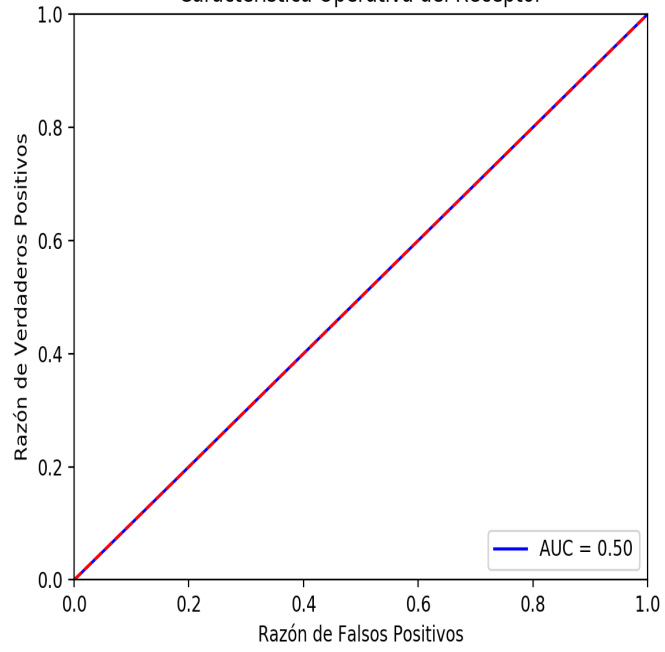


Figura 4.12: ROC - Kernel Support Vector Machines :

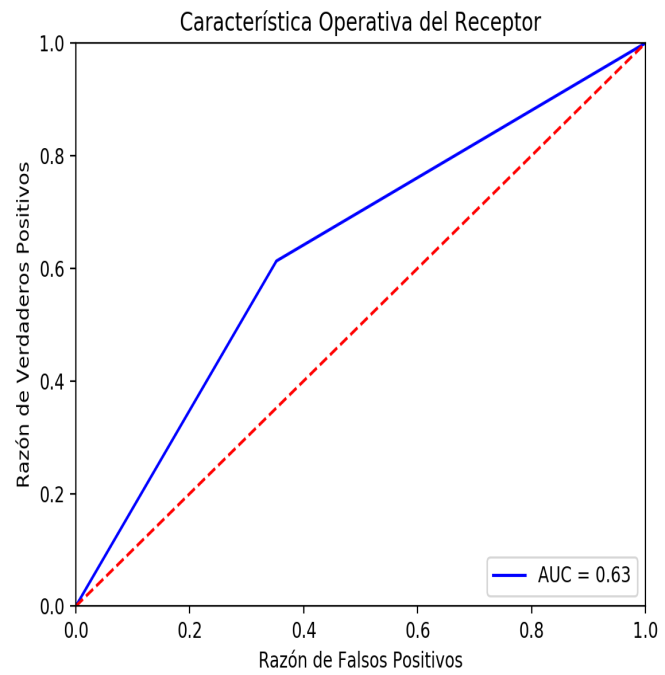
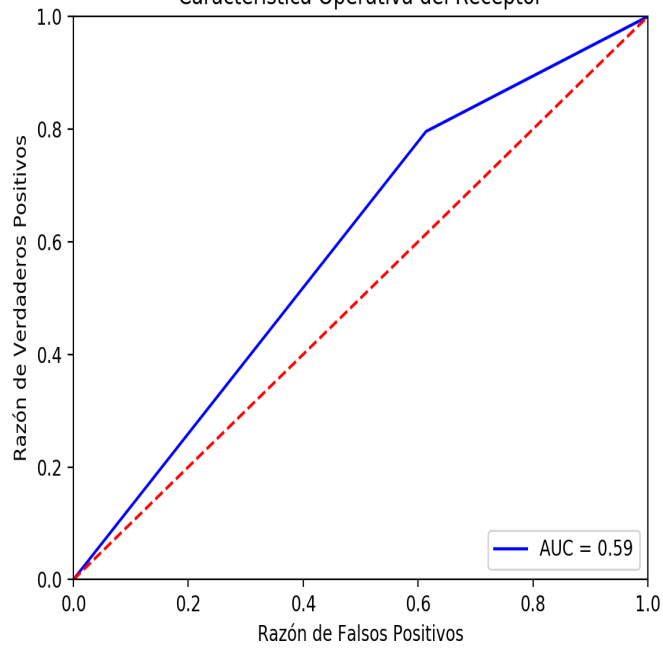




Figura 4.13: ROC - Support Vector Machines :  
Característica Operativa del Receptor



Análisis:

Se observa que los modelos con más alto AUC corresponden a Random Forest y Red Neuronal Artificial. A continuación se presenta una tabla resumen respecto al AUC:

Modelo	AUC
Red Neuronal Artificial	0.67
Random Forest	0.65
Kernel Support Vector Machines	0.63
Regresión Logística	0.60
Decision Tree	0.60
Support Vector Machine	0.59
K Nearest Neighbors	0.58
Naive Bayes	0.50

Tabla 4.12: Resumen AUC.

### 4.3.3. Tiempos de procesamiento

A continuación se muestran los tiempos de duración para el entrenamiento de cada modelo aplicado al set de datos ya mencionado (cambios en empaquetamiento que finalizan en un empaquetamiento triple, durante el transcurso del año 2018).

En general los tiempos de duración son bajos, excepto para los modelos support vector machines, kernel support vector machines y red neuronal artificial. Ahora bien, siguen estando dentro de los márgenes aceptables respecto a la periodicidad con que deben ser entrenados: todos pueden ser corridos en un día.

### 4.3.4. Modelo Seleccionado

Finalmente, debido a las métricas y al análisis realizado, los modelos con mejores resultados fueron:

- Red Neuronal Artificial.
- Random Forest.

Ahora bien, dadas las proyecciones del modelo (mayor dimensionalidad futura), y mayor cantidad de datos y clientes futuros; el modelo escogido finalmente es la red neuronal artificial. Esto debido a que, gracias a su capacidad de deep learning, y lo descrito anteriormente respecto a la literatura relevante, una mayor cantidad de datos futuros aumentará la precisión de clasificación del modelo de red neuronal artificial en mayor medida frente a los modelos de machine learning tradicionales. Aún así, se destaca que el modelo de random forest podría generar un resultado similar al modelo escogido, pero en términos de escalabilidad la red neuronal artificial resulta más pertinente.

Modelo	Tiempo de procesamiento (segundos)
Red Neuronal Artificial	20735,59
Regresión Logística	4,61
K nearest neighbors	58,43
Random forest	3,85
Decision tree	4,12
Naive bayes	2,58
Kernel support vector machines	642,61
Support vector machines	4543,06

Tabla 4.13: Tiempo de procesamiento por modelo.

# Capítulo 5

## Construcción del modelo

La construcción del modelo predictivo de red neuronal artificial requiere una serie de pasos:

- Confección de una base de datos de entrenamiento que incorpore las variables relevantes de estudio y posea un campo binario que caracterice la actividad a predecir.
- Programación y entrenamiento de la red neuronal.
- Análisis de resultados y modificaciones al modelo.

### 5.1. Obtención y tratamiento de datos de entrenamiento

El entrenamiento de la red neuronal artificial busca predecir la actividad de los clientes en cuanto a cambios en su tipo de empaquetamiento, específicamente en aquellos cambios asociados a una venta cruzada: aquellos clientes que poseen un servicio y optan por un empaquetamiento que agrega un servicio adicional al ya asociado.

Considerando esto, el primer paso es la extracción, a partir de la base total de clientes, de aquellos datos relevantes. Para esto se realizan las consultas necesarias a la tabla de empaquetamiento por cliente, la cual es actualizada diariamente durante el mes que corresponde,

y al terminar el mes almacena, de manera permanente y sobre-escribiendo al dato anterior, el tipo de empaquetamiento del último día para cada cliente.

La primera consulta considera la extracción, en formato csv, de la totalidad de clientes, desde el año 2017 a la fecha, con su empaquetamiento mensual y las siguientes variables: día de análisis, rut de la vivienda, tipo de empaquetamiento, localidad, rut del cliente, id de la vivienda, grupo tarifario por servicio (cable, internet y fono), y finalmente las rentas asociadas a cable, fono e internet.

Hecho esto, se inicia el tratamiento de los datos:

- Se agrega, mediante el uso de un diccionario, la descripción del tipo de empaquetamiento asociada al id de empaquetamiento para cada cliente: cable digital full, cable digital light, cable análogo full, cable análogo light, y de la misma manera para todos los empaquetamientos posibles. Añadido a esto, se genera una nomenclatura que agrupa los distintos empaquetamientos por servicio: cable (C), internet (I), fono (F) y todas las combinaciones correspondientes; a su vez, la nomenclatura identifica como 'No cliente' a aquellos con campos nulos o sin información.

- A través de un código en javascript, se crean las variables para cada mes de análisis con el empaquetamiento que tuvo cada cliente durante ese mes (utilizando la nomenclatura mencionada).

- Ahora bien, se tiene una base en la cual pueden existir clientes que aparecen de manera reiterada en múltiples filas (debido a que tuvieron un tipo de empaquetamiento durante más de un mes), estas corresponden a filas que solo varían en la variable del día de análisis (y por consiguiente el mes de análisis). Como el código anterior crea variables mensuales que almacenan el tipo de empaquetamiento para cada fila, se tienen múltiples filas por cliente donde todas las variables, menos la que corresponde al mes de análisis, son nulas. Para lidiar con esto, se ordena la base de manera ascendente por rut del cliente, y se agrupan nuevamente por rut, agregando las variables mensuales de empaquetamiento de manera que conserven aquel campo que no es nulo. Así, el resultado final es una única fila por cliente que posee las variables inicialmente extraídas, y una variable por cada mes en la cual se indica el tipo de empaquetamiento que tuvo en ese periodo.

- Una vez obtenida esta base, se realizan las comparaciones de tipo de empaquetamiento de manera mensual y consecutiva: se compara enero con febrero, febrero con marzo, y de la misma manera para todos los periodos. Se crean variables de diferencia de empaquetamiento para cada iteración. En caso de existir una diferencia, esta nueva variable almacena la concatenación de empaquetamiento (mes inicial y mes consecutivo) separada por un guión; en caso contrario, la variable permanece como campo nulo.

- Mediante un código javascript, se genera una nueva variable por cliente que almacena únicamente la variación en tipos de empaquetamiento, denominada 'tipo de venta'. Utilizando

esta variable, mediante un diccionario, se genera una nueva variable que identifica a las variaciones en empaquetamiento de acuerdo al tipo de movimiento que ocurrió: un campo de tipo de venta 'CF-3P' significa un paso de empaquetamiento cable-fono a empaquetamiento 3 play, por lo cual la nueva variable almacena a este movimiento como una 'Alta internet'; de esta manera, el diccionario clasifica a todos los tipos de movimiento posibles.

- Realizado esto, surgen inconsistencias en las bases iniciales que requieren tratamiento: se filtran los casos de rentas cuyo empaquetamiento no corresponde, debido a que ensucian los datos de entrenamiento. A su vez, se eliminan todos los casos que tengan renta de un servicio pero que no posean ese servicio en su empaquetamiento.

- Se procede a eliminar los casos que no aportan o generan posibles sobre ajustes:  
- Tipos de venta nulos. - Empaquetamientos iniciales del tipo 'no cliente', debido a que se busca tratar únicamente casos de cross-sell (clientes ya existentes). - Empaquetamientos iniciales y finales 'sin información'. - Se elimina la variable grupo tarifario que corresponda en cada caso: para 'Alta internet' se elimina el grupo tarifario internet; lo mismo se realiza para cada tipo de venta. Esto debe realizarse debido a que en las tablas originales ocurre que el grupo tarifario se encuentra escrito en el periodo anterior a la venta (lo cual no debería ocurrir): si no se elimina esta variable el modelo simplemente puede verificar si el grupo tarifario correspondiente es no-nulo e identificar a cada caso verdadero como una venta cruzada, generando un 100 % de precisión.

Una vez obtenida la base anterior, es necesario realizar una nueva transformación de los datos filtrando tipo de venta, debido a que para cada caso la red neuronal artificial debe entrenarse con los datos correspondientes: si se busca predecir una venta cruzada de internet, se deben tener los datos que incluyan solo altas de internet (incluir el resto de datos no aporta al modelo y solo ensucia la predicción). Luego, se procede:

- Se generan variables de estado binarias que almacenan el tipo de venta a analizar: alta internet se asocia a un 1, sin alta se asocia a un 0; se procede igual para todos los tipos de alta.

- Como la cantidad de clientes con variable estado 0 es mucho mayor, se genera una muestra aleatoria del mismo tamaño que el número de clientes con variable de estado 1 para cada caso.

Finalmente, se utiliza la base final para consultar y extraer, mediante el uso del rut del cliente y el identificador de la vivienda, otras variables de análisis que se encuentran en otras tablas y pueden aportar al modelo: grupo socio económico, grado de competencia de otras empresas en el sector correspondiente, meses con descuento del cliente, entre otras.

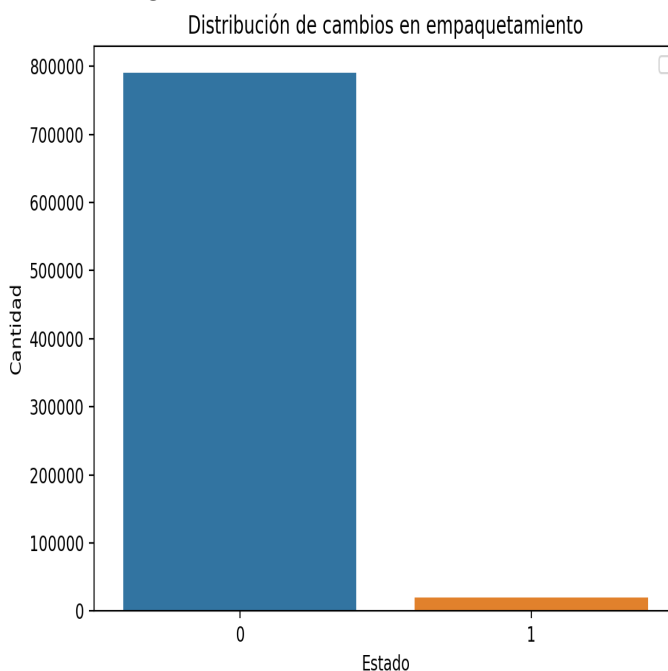
El producto final es un archivo csv con los datos de análisis, el cual será utilizado para entrenar el modelo y generar la predicción.

## 5.2. Tratamiento de Datos Desbalanceado

La base de datos extraída anteriormente posee una característica que complica la clasificación de clases: existe una gran cantidad de clientes que no cambia de empaquetamiento, mientras que la cantidad de clientes que realiza un cambio de empaquetamiento (finalizando con un empaquetamiento específico) es muy pequeña en comparación.

De hecho, explorando una muestra del set de datos de empaquetamientos que finalizan en empaquetamiento triple, filtrado para un mes y para los paquetes simples iniciales de cable y fono, se observa que el número de clientes que mantienen su empaquetamiento (clase mayoritaria) es 790.351, que corresponde al 97.48 % del set de datos y el número de altas 3P (clase minoritaria) es 20.422, que corresponde al 2.52 % del set de datos.

Figura 5.1: Distribución de datos:



Ahora bien, este desbalanceo de datos ocasiona un claro problema: cuando el set de datos posee datos no representados, la distribución de clases comienza sesgada. En el caso específico de los cambios en empaquetamiento, la cantidad de clientes que no cambian de empaqueta-

miento es tan grande que es muy difícil predecir la clasificación debido a que los modelos en general buscan ajustarse a todo el set de datos. Añadido a esto, la gran diferencia en cantidad de datos por clase oculta las correlaciones entre variables para la clase a clasificar.

Habiendo considerado esto, existen una serie de técnicas que se pueden aplicar para solucionar el problema, entre ellas: sobre-muestreo aleatorio, sub-muestro aleatorio y muestreo sintético.

### 5.2.1. Sobre-muestreo Aleatorio

Esta técnica consiste en generar réplicas de la clase minoritaria ponderadas por ciertos factores o dimensionadas a cierta constante.

- Ventajas de esta técnica: No genera pérdida de información.
- Desventajas de esta técnica: incrementa la verosimilitud del sobre ajuste, dado que replica eventos de la clase minoritaria.

Ahora, esta técnica no fue escogida debido a que el sobre ajuste sería muy alto, y la creación de demasiadas copias de uno u otro cliente aumentaría la probabilidad de estos casos, y sesgaría el resultado final del modelo.

### 5.2.2. Muestreo Sintético

Una de las variantes de esta técnica (la que fue probada en el presente modelo), es la técnica SMOTE: Synthetic Minority Over-Sampling Technique. Lo que hace este método es la creación de datos sintéticos en vez de sobre-muestrear con reemplazo. La clase minoritaria se sobre-muestrea tomando cada instancia de esta e introduciendo ejemplos sintéticos a lo largo de los segmentos que unen a los  $K$  vecinos más cercanos de las clases minoritarias; dependiendo de la cantidad de sobre-muestreo requerido, se escoge una cantidad de vecinos de manera aleatoria.

El procedimiento es el siguiente:  
Considerando los datos de entrenamiento con  $s$  muestras y  $f$  características en el espacio característico de los datos. Estas características, por simplicidad, son continuas. Así, el espacio característico para la clase minoritaria que se busca sobre-muestrear puede ser: renta,

coordenadas  $x$  e  $y$ , y tiempo de existencia del cliente (todas continuas por simplicidad del ejemplo). Para sobre-muestrear, se toma una muestra del set de datos y se consideran sus  $k$  vecinos más cercanos (en el espacio característico). Para crear una instancia sintética, se toma el vector entre esos  $k$  vecinos, y la instancia evaluada, y posteriormente se multiplica el vector por un número aleatorio  $x$  que se encuentra entre 0 y 1. Una vez realizado esto, se agrega la instancia sintética al set de datos.

De esta manera, esta técnica tiene ventajas y desventajas:

- Ventajas: Alivia el sobre ajuste del sobre muestreo por replicación de datos, debido a que se generan instancias sintéticas. No existe pérdida de información y es simple y fácil de entender.
- Desventajas: Mientras genera instancias sintéticas, la técnica SMOTE no considera que los ejemplos vecinos pueden pertenecer a otras clases, creando un problema de traslape entre clases que puede introducir ruido adicional. Además, la técnica no es muy práctica para datos de alta dimensionalidad.

Teniendo esto en cuenta, este método fue programado y probado en el set de datos, para visualizar si provoca una mejora en las métricas de precisión, recall y F1 score. El resultado no fue efectivo, debido principalmente a dos razones:

- La alta dimensionalidad del set de datos y la discontinuidad y no linealidad de las mismas genera la producción de datos sintéticos que ocasionan tanto ruido que entorpecen la clasificación.
- De la misma manera, se generaron muchas clasificaciones erróneas debido al traslape de clases generadas por los  $k$  vecinos más cercanos.

Habiendo considerado esto, la técnica SMOTE fue descartada para el tratamiento del desbalanceo de datos.

### 5.2.3. Sub-Muestreo Aleatorio

Esta técnica consiste en la extracción de datos aleatorios a partir de la clase mayoritaria para igualar la cantidad de datos de la clase minoritaria. Si bien esta técnica tiene como principal desventaja la pérdida de potencial información dentro de la clase mayoritaria, esta pérdida no es tan relevante debido a que la información que se busca rescatar es la de la clase efectiva: no interesa predecir los 0's, pero sí es importante para el modelo predecir los 1's.

Considerando esto, y dado que al probar los modelos entrenados con sub-muestreo, sobre muestras de datos ajenas a los datos de entrenamiento, se obtuvieron mejores métricas de efectividad, esta técnica fue la escogida para el tratamiento del desbalanceo de los datos.



### 5.3. Selección de Variables

La base final utiliza el rut del cliente y el id de la vivienda para extraer el resto de las variables independientes, que se utilizarán para el entrenamiento del modelo, a partir de una serie de consultas a las tablas de análisis relevantes.

Para seleccionar las variables se extrae primero una gran base con las variables relativas al cliente que podrían o no ser importantes para la predicción de los cambios en empaquetamiento, la totalidad de estas variables se detalla a continuación:

Las primeras variables obtenidas, a partir del proceso de obtención de datos de entrenamiento detallado anteriormente refieren al cliente y corresponden a: rut, id de localidad, id de vivienda, empaquetamiento inicial, renta relevante al mes de estudio; y estado (variable creada para identificar los respectivos casos positivos versus los casos negativos).

Realizado esto, se procede a extraer, mediante el rut del cliente, y desde la tabla que identifica al cliente, las siguientes variables: time stamp de creación del registro, número de versión del registro, flag de convenio tipo comercio, fecha de registro del cliente, fecha de la última modificación, apellido paterno, apellido materno, nombres, razón social, flag identificador de persona natural o jurídica, código de categoría del cliente, código de profesión del cliente, fecha de nacimiento, fono de contacto, flag VIP, código del usuario, descripción del email, otros fonos de contacto, preferencia de servicio, código de localidad preferente, fono y fax, preferencia de fono y fax, fecha de contacto, flag identificador de sexo, rut del representante legal, flag de uso web del cliente y flag de cantidad de bloqueos.

Posteriormente, se procede a extraer, mediante el identificador de la vivienda, las variables referentes a la vivienda en la cual habita el cliente o para la cual contrató el servicio. Las variables extraídas en este paso corresponden a: código de localidad, identificador de la calle, prefijo de municipalidad, tipo de agrupación, número de agrupación, número de nodo, número de subnodo, tipo de cableado, fecha de creación de la vivienda, fecha de creación del usuario, fecha de validación del ingreso, tipo de vivienda, flag de exclusión, código de grado socio económico, número de coordenada x, número de coordenada y.

Finalmente, se extraen, nuevamente mediante el identificador de la vivienda, las variables referentes al nivel de competencia y factibilidades: claro hfc, fibra movistar, adsl movistar, fibra entel, fibra telsur, cobre telsur, fecha de actualización de cada uno de los registros anteriormente mencionados, distribución de los competidores, composición de la fibra, grado de competencia, dirección, tipo de cableado y región.

Finalmente, teniendo todas las variables de estudio mencionadas, se procede a la selección de aquellas variables que aportan más al modelo. Debido a que el funcionamiento de la red neuronal corresponde al modelo de una caja negra (en el cual solo se conocen inputs y outputs, pero no se pueden caracterizar las decisiones y procesos intermedios de manera numérica), la mejor manera de seleccionar variables corresponde a la realización de un grid search con la totalidad de las variables.

El grid search corresponde a la declaración de una serie de parámetros y la posterior prueba de cada uno de ellos, generando métricas de rendimiento útiles para la selección.

En este caso se generaron métricas correspondientes a precisión, recall y f1 score. El conjunto de variables que mejores resultados entregara en base a estas métricas, fue el escogido para el modelo.

Ahora bien, antes de realizar el grid search, y para optimizar el tiempo de duración del mismo, se excluyeron ciertas variables que claramente no agregan valor al modelo por su especificidad relativa al cliente, estas variables son:

## 5.4. Normalización de Variables

Previo a la programación del modelo es necesario realizar el pre-procesamiento de los datos de entrenamiento, para ello se programan, utilizando Python, una serie de funciones para el tratamiento de las variables obtenidas como resultado de las transformaciones ETL mencionadas en la sección anterior.

El pre-procesamiento de los datos involucra la transformación de aquellos datos de carácter categórico y aquellos datos numéricos que poseen dimensiones y unidades muy distintas.

En primera instancia, se busca transformar aquellas variables categóricas: grado socio-económico, nivel profesional, género, localidad, empaquetamiento inicial, grado de competencia y las factibilidades respectivas a la competencia.

Para que estas variables sean tratables numéricamente en el modelo se deben transformar a variables binarias capaces de caracterizar a los clientes. Esta transformación comúnmente se realiza de manera manual mediante la identificación de las variables categóricas. Ahora bien, debido a que los procesos de extracción, transformación y carga que originan los datos

de entrenamiento a utilizar en el modelo, es posible automatizar el proceso de identificación y transformación de variables categóricas mediante una apropiada programación del pre-procesamiento.

De esta manera, se crea en el código una variable con estructura de lista que incorpora los nombres de las columnas identificadas como categóricas en los valores del documento resultante de los procesos de ETL mencionados anteriormente. Una vez creada esta lista, se utiliza para crear variables binarias para cada categoría y posteriormente, mediante la programación de ciertas funciones, se elimina una variable para cada categoría, a fin de evitar la trampa de las variables binarias haciendo uso de la completitud lineal de las mismas: un ejemplo de esto es la variable género, donde la identificación negativa de un cliente como masculino implica su pertenencia al género femenino (teniendo como supuesto la existencia única de 2 géneros).

Una vez que todas las variables poseen un caracter numérico, es necesario procesarlas para que sean comparables numérica y estadísticamente. Para ello se programa el tratamiento de cada una de las variables escalando sus valores para lograr distribuciones con media cero y varianza unitaria en cada variable incorporada al modelo.

## 5.5. Explicación del modelo

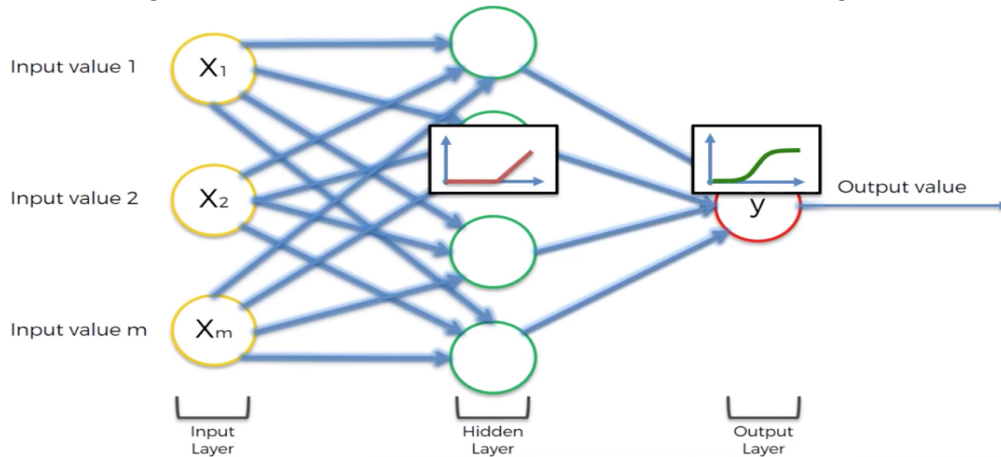
El modelo escogido corresponde a una red neuronal artificial, compuesta por una serie de estructuras, y entrenada de una manera específica. En las siguientes secciones se detallará el funcionamiento y procedimiento utilizado.

### 5.5.1. Estructura escogida

La red neuronal escogida como modelo en el presente trabajo tiene la siguiente estructura: una capa de variables independientes de entrada (compuesta por las variables mencionadas en la sección anterior), una capa intermedia, y una capa de variables de salida, la cual entregará la probabilidad de venta cruzada para cada cliente(figura 4.2).

Ahora bien, para la capa intermedia se escogió la función de activación rectificadora debido a su pertinencia para el modelo, la cual se basa en un sustento matemático que será tratado en la sección siguiente.

Figura 5.2: Estructura de red neuronal artificial escogida:



La función rectificadora permite tener un concepto fundamental en el modelamiento predictivo: baja densidad. Este concepto evita a su vez el problema del sobre ajuste en los modelos predictivos.

El problema del sobre ajuste se puede explicar con un simple ejemplo: Se tiene un modelo de clasificación y la idea es enseñarle a clasificar caracteres chinos mirando únicamente al siguiente ejemplo: Los primeros 5 caracteres pertenecen a la primera categoría

Figura 5.3: Datos de entrenamiento (ejemplo):

把打扒捕拉 被劈盐派喔

y los siguientes 5 caracteres pertenecen a la segunda categoría. Los 10 caracteres son los únicos datos de entrenamiento que se tienen.

Ahora, si el modelo es demasiado inteligente, o bien no posee 'baja densidad', tendrá suficiente memoria como para recordar 5 caracteres. Luego, al observar los 10 datos de entrenamiento, el modelo aprende una manera de categorizarlos: recuerda los 5 primeros caracteres de manera exacta. Siempre que un caracter no sea uno de los 5 primeros, el modelo pondrá a ese dato en la segunda categoría. Claramente este método funciona muy bien en los 10 datos de entrenamiento, debido a que el modelo puede obtener un 100% de precisión.

Sin embargo, ahora se muestra un nuevo dato al modelo: Este nuevo caracter debería

Figura 5.4: Nuevo dato (ejemplo):

揪

pertenecer a la primera categoría, pero debido a que nunca estuvo en los datos de entrena-

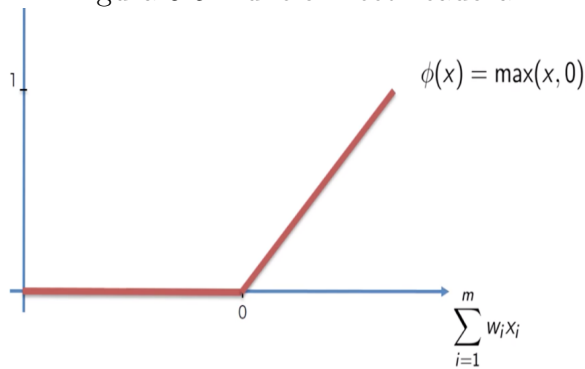
miento, el modelo nunca lo ha visto. Así, basado en su algoritmo, el modelo clasificará a este caracter dentro de la segunda categoría, lo cual es un error. El patrón es bastante claro y se

Figura 5.5: Clasificación (ejemplo):

把打扒捕拉被劈盐派喔

puede observar en la figura 4.5: todos los caracteres que pertenecen a la primera categoría tienen una característica en común. El modelo falló en la tarea porque es 'muy inteligente' y los datos de entrenamiento son muy pequeños. Este es el problema del sobre ajuste, y una manera de evitarlo es mediante la "baja densidad". La baja densidad puede lograrse forzando al modelo a ser 'menos inteligente' haciendo que solo memorice una parte del caracter, logrando así reconocer el patrón ya mencionado. Esto es esencialmente lo que hace la función rectificadora: genera vectores que se activan únicamente luego de cierto valor 'x', ayudando así a la baja densidad del modelo (figura 4.6).

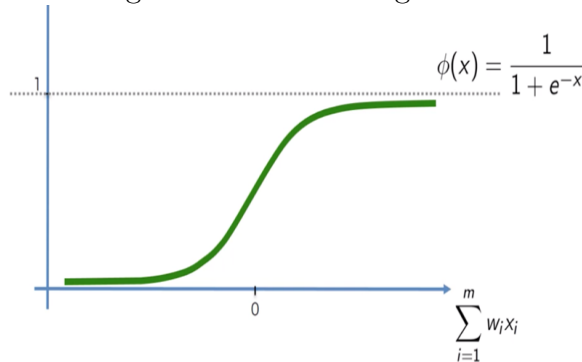
Figura 5.6: Función rectificadora:



Respecto a la función de activación que se escogió para la capa de salida, esta corresponde a la función sigmoide (figura 4.7). Esta función fue escogida debido a que el carácter binario de la variable dependiente permite modelar la probabilidad de venta cruzada de la siguiente manera:

$$P(y = 1) = \phi\left(\sum_{i=1}^m w_i x_i\right)$$

Figura 5.7: Función sigmoide:



Finalmente, se añaden capas intermedias denominadas "dropout" que buscan evitar el sobreajuste del modelo. Esto lo logran activando y desactivando neuronas de manera aleatoria de tal manera que el modelo no tenga la capacidad de memorizar la totalidad de los datos y deba encontrar los patrones principales que permitan la clasificación.

### 5.5.2. Justificación del uso de la función de unidad lineal rectificadora (rectificadora)

La justificación del uso de esta función, como fue mencionado anteriormente, nace del concepto de baja densidad (sparsity en inglés). Debido a esto, primero es necesario tratar este concepto y sus ventajas.

La baja densidad fue introducida por primera vez en el contexto de la codificación de baja densidad en el área de la neurociencia (Olshausen y Field, 1997) y a partir de ese momento, ha sido un elemento fundamental para las redes neuronales, las cuales explotan el concepto de representaciones distribuidas de baja densidad (Ranzato et al, 2007, 2008; Mairal et al, 2009).

Utilizar una función rectificadora no lineal permite obtener ceros reales en los pesos de

activación y por ende permite una baja densidad más efectiva:

- Separar información: Uno de los objetivos de los algoritmos de aprendizaje profundo (Bengio, 2009) es separar los factores que explican la variación en los datos. Una representación densa no posee separabilidad debido a que cualquier cambio en el input modifica la mayor parte de las entradas en el vector de representación. Por su parte, si una representación es tanto de baja densidad como robusta a pequeños cambios en el input, el set de características distintas de cero es altamente conservada por los pequeños cambios en el input.
- Representación eficiente del tamaño de las variables: Diferentes inputs pueden contener diferentes cantidades de información y su representación sería más conveniente utilizando una estructura de datos de tamaño variable, la cual es común en las representaciones computacionales de la información. La variación en el número de neuronas activas permite al modelo controlar la dimensionalidad efectiva de la representación para un input dado y una precisión específica.
- Separabilidad lineal: las representaciones de baja densidad son también más susceptibles a ser linealmente separables, o más fácilmente separables simplemente debido a que la información es representada en un espacio de alta dimensionalidad. Además, esto puede reflejar el formato original de los datos. En aplicaciones relacionadas a texto, por ejemplo, los datos originales sin procesar poseen, desde ya, baja densidad.
- Distribuido pero de baja densidad: Las representaciones distribuidas de alta densidad son potencial y exponencialmente más eficientes que aquellas representaciones puramente locales (Bengio, 2009). Ahora, la eficiencia de las representaciones de baja densidad es aún exponencialmente mayor, donde la potencia del exponente corresponde a las variables identificadas como no nulas.

Teniendo en cuenta lo anteriormente mencionado, una función de activación rectificadora permite una fácil obtención de representaciones de baja densidad:

$$\text{rectificadora}(x) = \max(0, x)$$

Por ejemplo, luego de la inicialización uniforme de los pesos en la red, alrededor del 50% de los valores de salida continuos son ceros reales, y esta fracción puede aumentar fácilmente con la regularización de las variables.

La única no linealidad en la red nace de la selección del camino asociado con neuronas que están o no activas. Para un input dado, solo un subconjunto de neuronas están activas. La computación es lineal en este subconjunto: una vez que se selecciona este subconjunto de neuronas, el output es una función lineal del input (a pesar de que un cambio lo suficientemente grande puede generar un cambio discreto en el set de neuronas activas). Así, la función calculada por cada neurona o por el output de la red en términos del input, resulta ser lineal por partes.

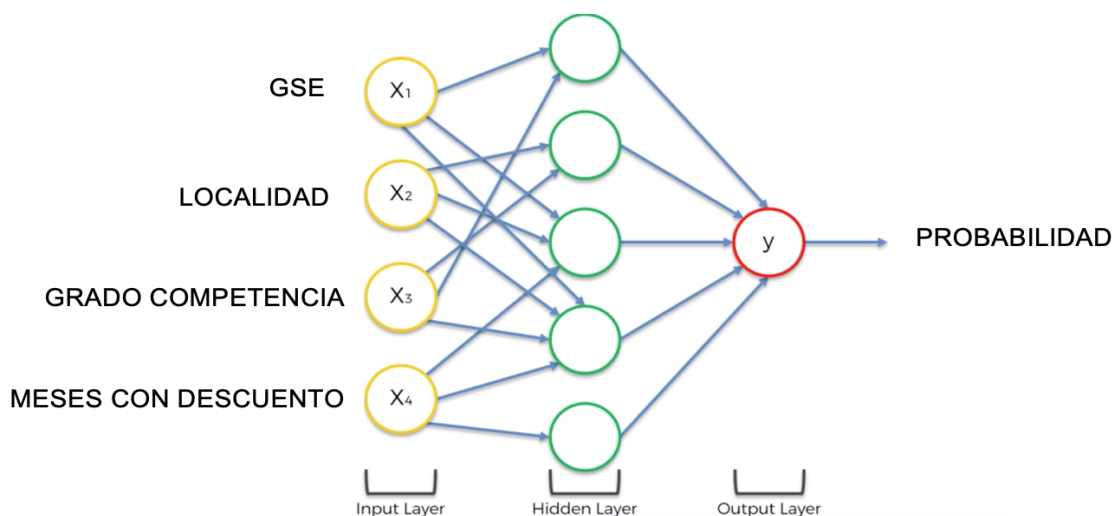
De esta manera, el modelo puede ser pensado como un número exponencial de modelos

lineales que comparten parámetros (Nair y Hinton, 2010).

### 5.5.3. Funcionamiento de la red neuronal

En la capa de variables de entrada se encuentran aquellas variables mencionadas al final de la sección anterior. Sin embargo, para ejemplificar el funcionamiento se utilizarán solo las siguientes variables (el funcionamiento es el mismo para el modelo completo): grado socio económico (GSE), localidad, grado de competencia y meses con descuento. Cada neurona

Figura 5.8: Funcionamiento de la red:



de la capa intermedia se enfocará en determinadas variables de entrada dependiendo del efecto que busca capturar. Por ejemplo, si existiera un efecto asociado a la localidad y el grado de competencia, ocurriría lo siguiente: los pesos serían significativos únicamente para las conexiones entre las neuronas de entrada de localidad y grado de competencia, luego la función de activación de aquella neurona (función rectificadora) solo generaría un valor de salida al alcanzar ciertos valores para la localidad y los meses con descuento.

Una vez detallado el funcionamiento respecto a la interacción de los elementos de la red, la manera en que el modelo aprende sobre los datos de entrenamiento es aquella descrita anteriormente en el marco teórico, desde la inicialización de manera aleatoria de los pesos utilizando valores cercanos a 0 (no exactamente 0) pasando por la propagación hacia delante e inversa, y finalizando con la repetición de una serie de epochs.



## 5.5.4. Programación del modelo

A continuación se detalla la programación del modelo, incluyendo desde la normalización de las variables hasta el resultado final.

El primer paso corresponde a la importación de las librerías para tratamiento de los marcos de datos (dataframes); estas corresponden a las librerías Pandas (librería de alto nivel escrita en python para tratamiento de estructura y análisis de datos), y NumPy (librería para tratamiento numérico de datos a través de arreglos N-Dimensionales).

Una vez importadas las librerías, se lee el set de datos en el formato de marco de datos de python y se completan los valores NaN (valores no numéricos faltantes), debido a que no permiten ser categorizados o transformados al formato de número flotante (decimal) correspondiente.

Hecho esto, se transforman todas las variables numéricas no categóricas a datos de formato flotante, debido a que el posterior tratamiento de los mismos requiere de operaciones numéricas y debido a que el formato en que fueron leídos es de estructura de escritura simple (trata los números al igual que letras).

Posteriormente se crea un diccionario que identifica todas las variables categóricas a utilizar. Habiendo creado el diccionario se procede a transformar todas las variables categóricas a variables binarias (dummies) y se elimina una por cada categoría para evitar la trampa de las variables binarias (colinealidad perfecta que se da por la completitud de las categorías: si un cliente es hombre y se tienen únicamente 2 géneros en la categoría basta con una variable binaria debido a que un 1 en categoría mujer implica un 0 en categoría hombre y vice versa).

Hecho esto se crean dos sets de datos, uno que incorpora todas las variables independientes (x) y otro set que incorpora la variable clasificadora a predecir (y).

Posteriormente, se genera la división de los conjuntos de datos en proporciones 80-20 de manera aleatoria, se mantiene un 80 % de los datos para el entrenamiento del modelo y un 20 % de datos que se utilizarán para testear el modelo.

Una vez realizada la división de los datos, se procede a normalizar las variables utilizando librerías de escaladores estándar para lograr que los datos tengan media 0 y desviación estándar 1. Esto se realiza para todas las variables independientes.

Teniendo los datos normalizados y divididos de esta manera, se procede a programar el modelo de clasificación utilizando librerías de aprendizaje automático como Keras y Tensorflow y siguiendo la estructura anteriormente mencionada:

1. Crear el clasificador de red secuencial.
2. Incorporar las capas descritas anteriormente, identificando las funciones de activación a utilizar, el número de dimensiones de entrada y el tipo de inicializador uniforme para los pesos.
3. Crear dos capas denominadas dropout para evitar el sobreajuste del modelo.
4. Incorporar la capa de salida e identificar la función sigmoide junto a la dimensión de salida que en este caso corresponde a una única variable binaria.
5. Definir el optimizador y gradiente de descenso a utilizar de acuerdo al ajuste de parámetros que logró mejor precisión en la selección del modelo. Para el modelo los parámetros de mejor ajuste fueron el optimizador adam (extensión al gradiente descendiente estocástico), la precisión como métrica de ajuste de entrenamiento y la función de pérdida binaria logarítmica:

$$H_p(Q) = -\frac{1}{N} \sum_{i=1}^N y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i))$$

6. Entrenar la red definiendo el número óptimo de datos por lote y de epochs. En el presente modelo se escogieron 35 líneas por lote y 100 epochs.
7. Realizar las predicciones utilizando el modelo entrenado y aproximar las probabilidades con límite de clasificación 0,5.
8. Generar la matriz de confusión y las métricas de precisión, recall y f1 score para la evaluación.
9. Guardar las probabilidades para proceder con las otras métricas de ajuste del modelo.

## 5.6. Optimización del modelo

Para obtener mejores niveles de precisión, rendimiento (en procesamiento y tiempo), y evitar el sobre ajuste del modelo, se programan una serie de ajustes:

1) Un primer ajuste es el testeo de la probabilidad límite para la clasificación, se prueban distintos niveles límite y se comparan las medidas de efectividad. Debido a que el presente modelo busca generar un ranking de probabilidad de cambio de empaquetamientos, se escoge como probabilidad límite el valor de 0.5.

2) Un segundo ajuste es la incorporación de una capa denominada dropout, entre las capas ya existentes de la red neuronal artificial. La labor que cumple esta capa es activar y desactivar, de manera aleatoria (pudiendo escoger el parámetro de aleatoriedad), las distintas neuronas presentes en el modelo. Mediante esta desactivación aleatoria, lo que se busca es que el modelo sea incapaz de memorizar todos los datos de entrenamiento, y deba aprender a identificar las distintas clasificaciones utilizando la menor cantidad de atributos posibles, generando así la deseada escasez capaz de evitar el sobre ajuste de los modelos.

3) Un tercer y último ajuste es el ajuste de hiper-parámetros del modelo. Mediante el testeo de distintos parámetros y la utilización de funciones del estilo grid-search se prueba una serie de parámetros en búsqueda del óptimo entre precisión y rendimiento: tamaño de los lotes de datos procesados, cantidad de epochs utilizados para el entrenamiento del modelo, y los distintos tipos de optimizadores existentes para el gradiente descendiente.

## 5.7. Resultados y medidas de efectividad

### Precisión, recall y F1-Score

Una vez entrenada la red neuronal artificial, utilizando el 80 % de los datos, se procede a predecir la probabilidad de venta cruzada para el 20 % de clientes restante. Posteriormente, se obtiene la precisión del modelo comparando los datos reales con los predichos por la red neuronal; para esto se realiza una aproximación de la probabilidad a 1 para aquellas mayores a 0,5 y a 0 para los casos contrarios.

Los resultados obtenidos tienen el siguiente formato:

Estado Real	Probabilidad Predicha	Clasificación Predicha
1	0.9286	True
0	0.1594	False
1	0.9697	True
0	0.0004	False
1	0.9286	True
0	0.6796	True
1	0.4985	False

Tabla 5.1: Resultados de la predicción del modelo. (Estado real - Probabilidad predicha - Clasificación predicha)

### 5.7.1. Altas 3P

Se incluyen todos los empaquetamientos iniciales simples y dobles que finalizan en una alta triple. Las métricas de precisión, recall y f1 score resultan ser las siguientes para el modelo:

Estado/Métrica	Precisión	Recall	F1 Score	Datos de soporte
0	0.64	0.79	0.71	388.685
1	0.73	0.56	0.63	388.964
Promedio ponderado	0.69	0.68	0.67	777.649

Tabla 5.2: Métricas de precisión, Red Neuronal Artificial, datos de entrenamiento (Entrenada sobre 100 epochs).

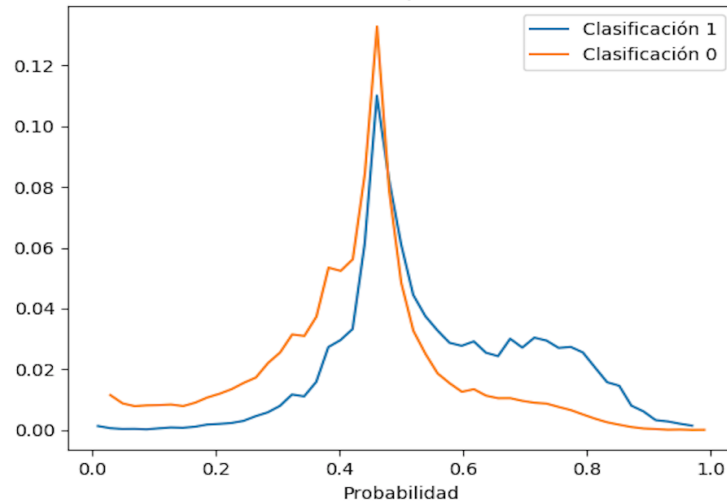
Estado/Métrica	Precisión	Recall	F1 Score	Datos de soporte
0	0.64	0.78	0.70	166.315
1	0.72	0.55	0.62	166.964
Promedio ponderado	0.68	0.67	0.66	333.279

Tabla 5.3: Métricas de precisión, Red Neuronal Artificial, datos de testeo (Entrenada sobre 100 epochs).

De acuerdo a los resultados obtenidos, se observa que no existen indicios de sobre ajuste respecto a los resultados obtenidos en datos de testeo contra resultados obtenidos en datos de entrenamiento.

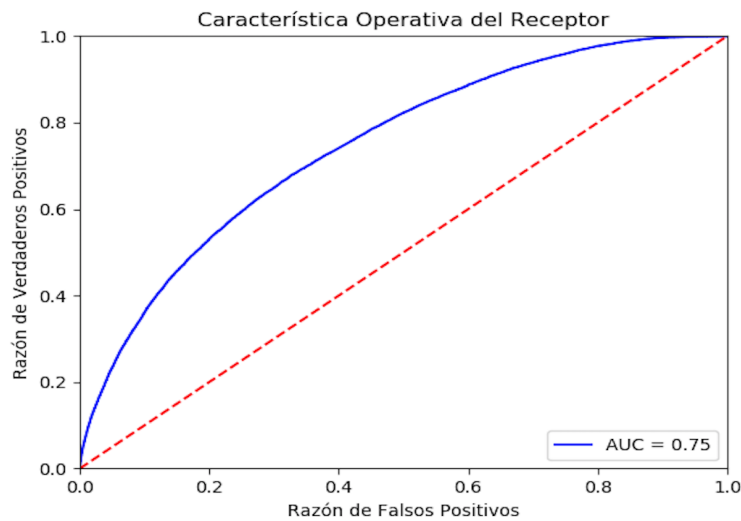
Añadido a esto, para los datos de testeo se logra una precisión promedio de 0.68: un 68 % de las identificaciones positivas fueron correctas; y un recall promedio de 0.67: un 67 % de las clases positivas reales fueron identificadas correctamente.

Figura 5.9: Distribución de Probabilidades  
Distribución de probabilidades



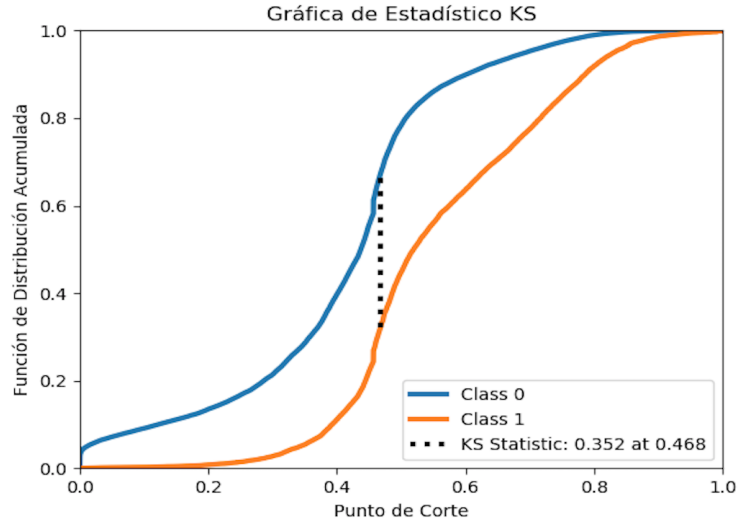
Como se trata de un modelo de predicción conductual, el resultado esperable es una distribución de probabilidades con alta concentración en los valores centrales [0.4 - 0.6]. El resultado obtenido tiene justamente esta forma, donde la curva de clasificación tipo 1 (Alta 3P) se ajusta hacia el resultado de probabilidad más alta y la curva de clasificación tipo 0 (Mantiene empaquetamiento) se ajusta hacia el resultado de probabilidad más baja.

Figura 5.10: ROC



Respecto a la curva característica operativa del receptor, se obtiene una razón de verdaderos positivos siempre mayor a la de falsos positivos, resultando en un área bajo la curva de 0.75, 25 puntos porcentuales sobre un modelo aleatorio.

Figura 5.11: Estadístico KS



Finalmente, la prueba de Kolmogorov - Smirnov resulta en un estadístico KS de 0.352. Este punto corresponde al máximo nivel de separación entre las distribuciones de probabilidad de las clases. El estadístico no resulta ser tan alto debido a la distribución de probabilidades similar entre clases; ahora bien, siempre existe una separación entre las curvas de distribución acumuladas y estas nunca se cruzan, por lo cual el modelo logra una buena distinción entre clases.

## 5.7.2. Altas Cable

Se incluyen los empaquetamientos iniciales simples que finalizan en una alta de cable. Las métricas de precisión, recall y f1 score resultan ser las siguientes para el modelo:

Estado/Métrica	Precisión	Recall	F1 Score	Datos de soporte
0	0.67	0.65	0.66	209.739
1	0.64	0.65	0.65	198.510
Promedio ponderado	0.65	0.65	0.65	408.249

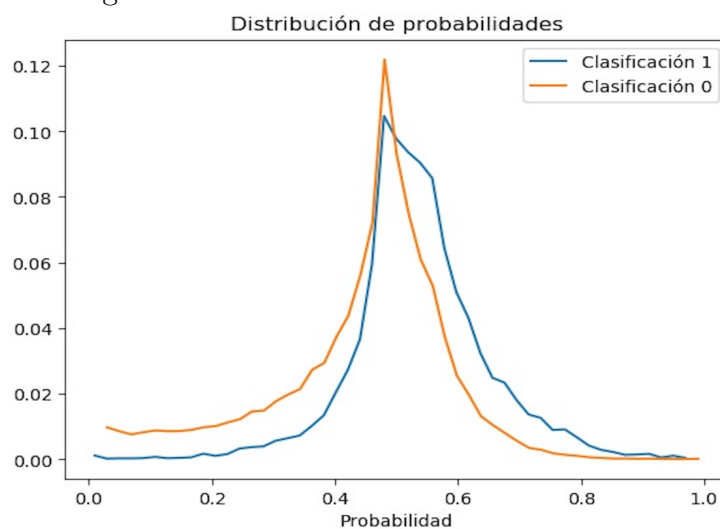
Tabla 5.4: Métricas de precisión, Red Neuronal Artificial, datos de entrenamiento (Entrenada sobre 100 epochs).

Estado/Métrica	Precisión	Recall	F1 Score	Datos de soporte
0	0.66	0.65	0.65	90.261
1	0.63	0.64	0.63	84.703
Promedio ponderado	0.64	0.64	0.64	174.964

Tabla 5.5: Métricas de precisión, Red Neuronal Artificial, datos de testeo (Entrenada sobre 100 epochs).

Para los datos de testeo se logra una precisión promedio de 0.64: un 64 % de las identificaciones positivas fueron correctas; y un recall promedio de 0.64: un 64 % de las clases positivas reales fueron identificadas correctamente.

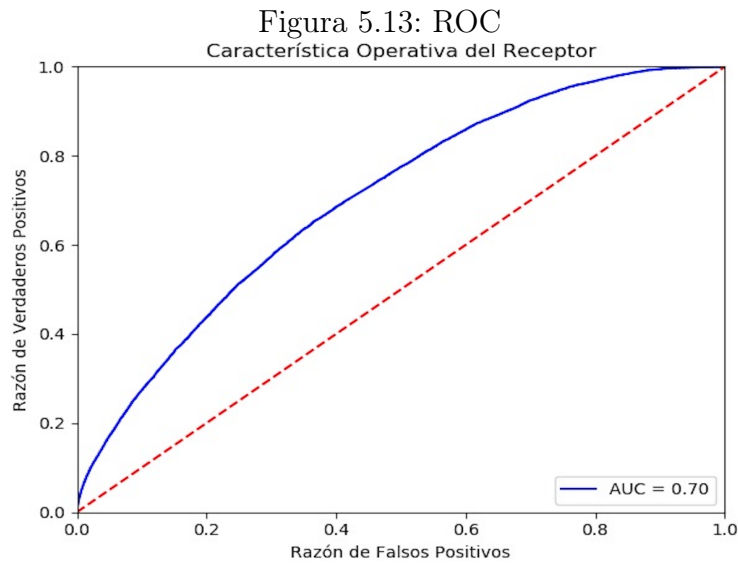
Figura 5.12: Distribución de Probabilidades



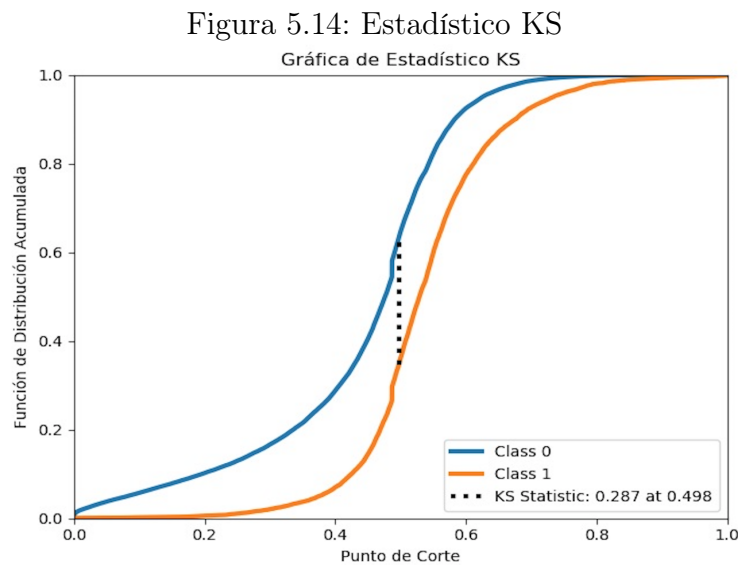
La distribución de probabilidades muestra el mismo comportamiento anterior, de acuerdo



a un modelo de predicción conductual.



Respecto a la curva característica operativa del receptor, se obtiene una razón de verdaderos positivos siempre mayor a la de falsos positivos, resultando en un área bajo la curva de 0.70, 20 puntos porcentuales sobre un modelo aleatorio.



Finalmente, la prueba de Kolmogorov - Smirnov resulta en un estadístico KS de 0.287. El estadístico nuevamente no resulta ser tan alto debido a la distribución de probabilidades similar entre clases, sin embargo siempre existe una separación entre las curvas.

### 5.7.3. Altas Internet

Se incluyen todos los empaquetamientos iniciales simples que finalizan en una alta internet. Las métricas de precisión, recall y f1 score resultan ser las siguientes para el modelo:

Estado/Métrica	Precisión	Recall	F1 Score	Datos de soporte
0	0.85	0.58	0.69	139.985
1	0.69	0.90	0.78	142.291
Promedio ponderado	0.77	0.74	0.74	282.276

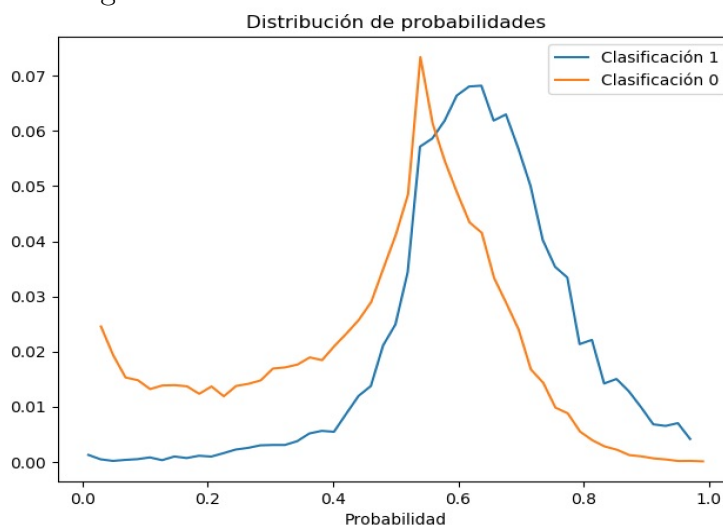
Tabla 5.6: Métricas de precisión, Red Neuronal Artificial, datos de entrenamiento (Entrenada sobre 100 epochs).

Estado/Métrica	Precisión	Recall	F1 Score	Datos de soporte
0	0.84	0.58	0.68	60.015
1	0.68	0.89	0.77	60.961
Promedio ponderado	0.76	0.73	0.73	120.976

Tabla 5.7: Métricas de precisión, Red Neuronal Artificial, datos de testeo (Entrenada sobre 100 epochs).

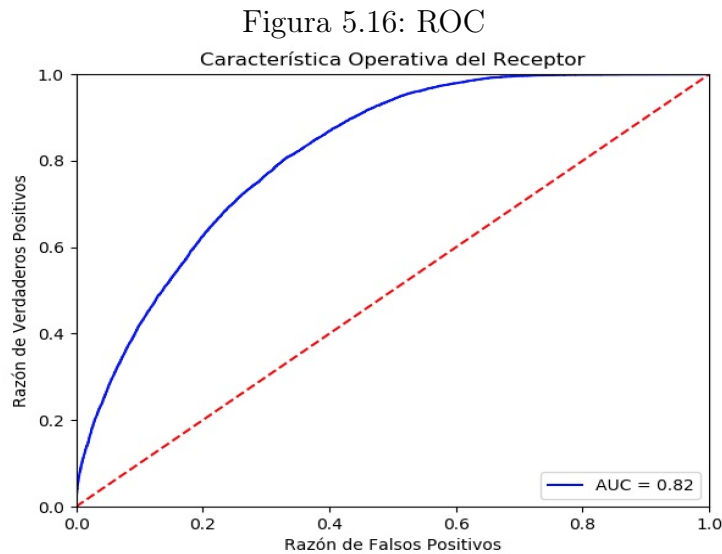
Para los datos de testeo se logra una precisión promedio de 0.76: un 76 % de las identificaciones positivas fueron correctas; y un recall promedio de 0.73: un 73 % de las clases positivas reales fueron identificadas correctamente.

Figura 5.15: Distribución de Probabilidades

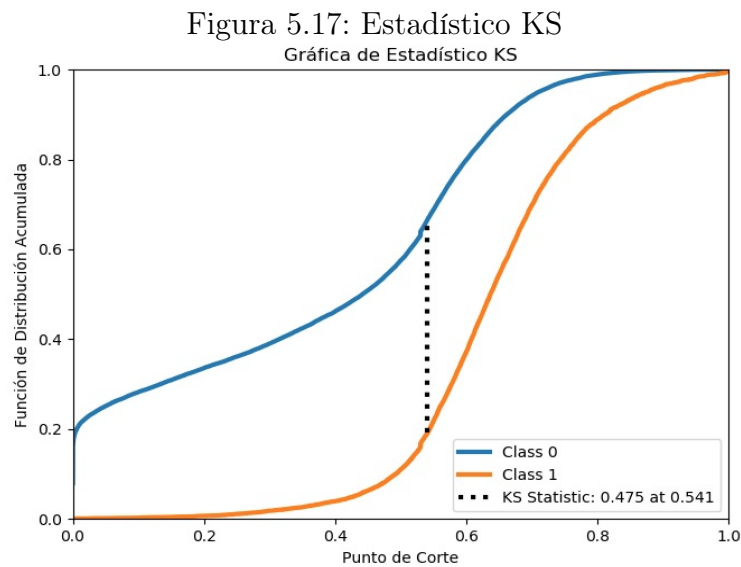


Nuevamente, y por las razones ya descritas, se obtiene una forma acorde al modelo para

la distribución de probabilidades.



Respecto a la curva característica operativa del receptor, se obtiene una razón de verdaderos positivos siempre mayor a la de falsos positivos, resultando en un área bajo la curva de 0.82, 32 puntos porcentuales sobre un modelo aleatorio.



Finalmente, la prueba de Kolmogorov - Smirnov resulta en un estadístico KS de 0.475. El estadístico resulta ser más alto que los resultados anteriores y siempre existe una separación entre las curvas.

### 5.7.4. Altas Fono

Se incluyen todos los empaquetamientos iniciales simples que finalizan en una alta fono. Las métricas de precisión, recall y f1 score resultan ser las siguientes para el modelo:

Estado/Métrica	Precisión	Recall	F1 Score	Datos de soporte
0	0.93	0.86	0.89	20.969
1	0.84	0.92	0.88	17.200
Promedio ponderado	0.89	0.89	0.89	38.169

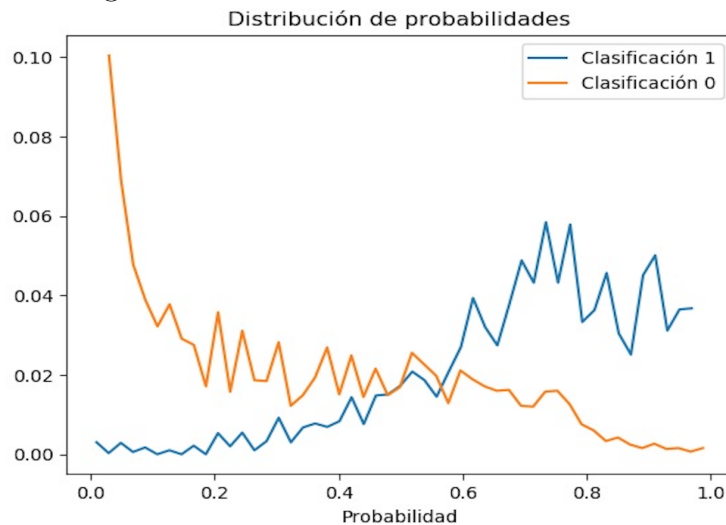
Tabla 5.8: Métricas de precisión, Red Neuronal Artificial, datos de entrenamiento (Entrenada sobre 100 epochs).

Estado/Métrica	Precisión	Recall	F1 Score	Datos de soporte
0	0.90	0.86	0.88	9.031
1	0.84	0.88	0.86	7.328
Promedio ponderado	0.87	0.87	0.87	16.359

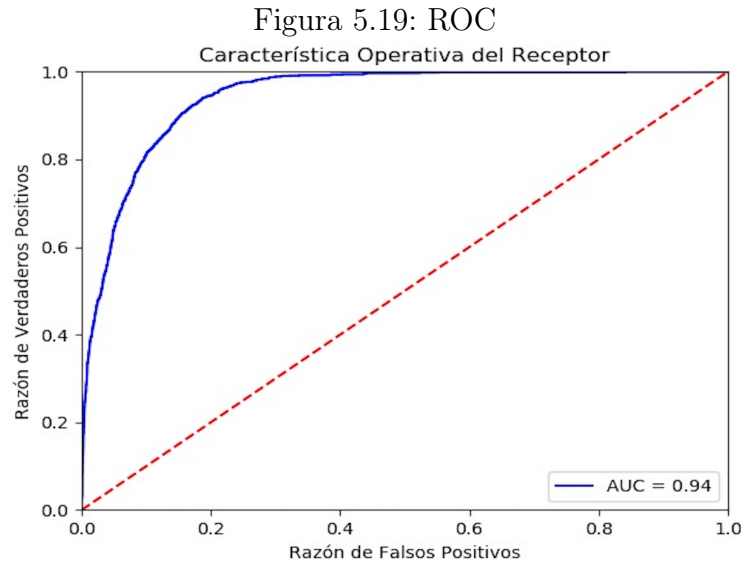
Tabla 5.9: Métricas de precisión, Red Neuronal Artificial, datos de testeo (Entrenada sobre 100 epochs).

Para los datos de testeo se logra una precisión promedio de 0.87: un 87 % de las identificaciones positivas fueron correctas; y un recall promedio de 0.87: un 87 % de las clases positivas reales fueron identificadas correctamente.

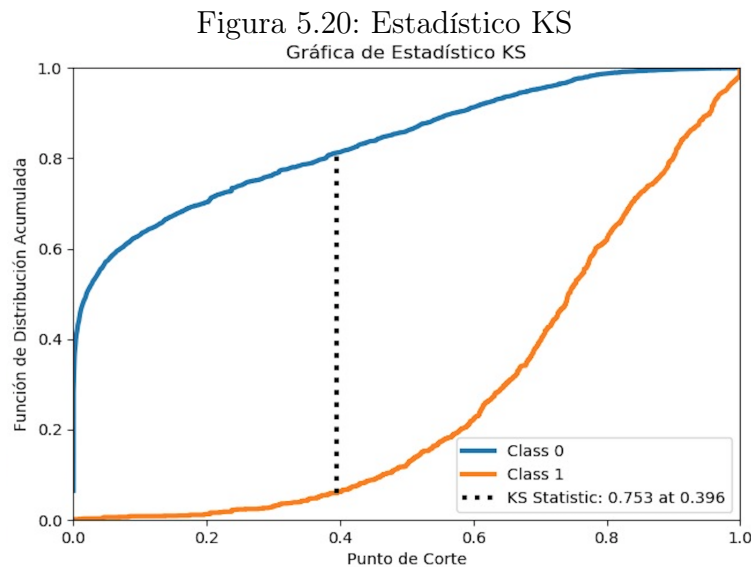
Figura 5.18: Distribución de Probabilidades



Las curvas de distribución de probabilidades para este caso resultan ser mucho más discriminativas entre la clase 1 y 0.



Respecto a la curva característica operativa del receptor, se obtiene una razón de verdaderos positivos siempre mayor a la de falsos positivos, resultando en un área bajo la curva de 0.94, 94 puntos porcentuales sobre un modelo aleatorio.



Finalmente, la prueba de Kolmogorov - Smirnov resulta en un estadístico KS de 0.753. El estadístico resulta muy alto y muestra gran separabilidad de clases.

Ahora bien, la gran diferencia que existe entre las métricas de este último caso (Fono) respecto a los demás casos refiere a un comportamiento específico para este tipo de empaquetamiento. Los clientes que acceden a fono poseen un comportamiento mucho más discriminable debido a que pertenecen a segmentos que requieren el uso de un fono fijo o bien poseen una tendencia al empaquetamiento triple debido a su poca sensibilidad al precio.

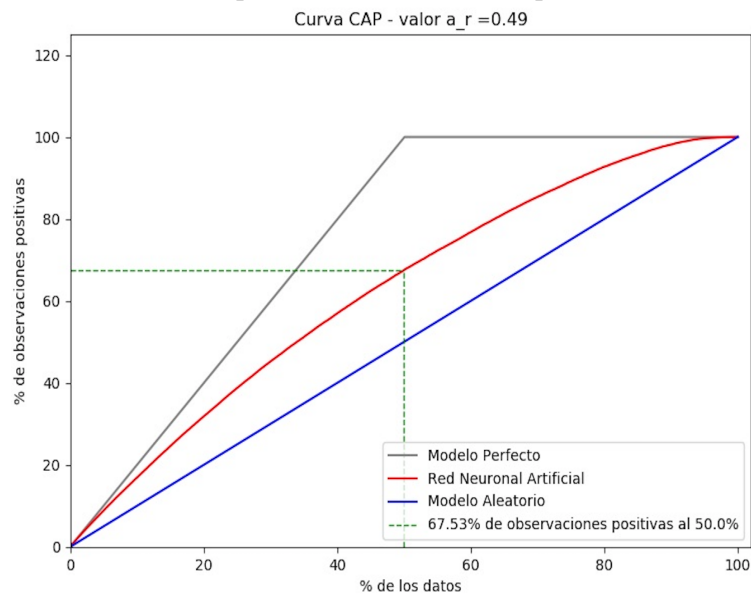
# Capítulo 6

## Impacto al negocio

### 6.1. Altas 3P

#### 6.1.1. Cap

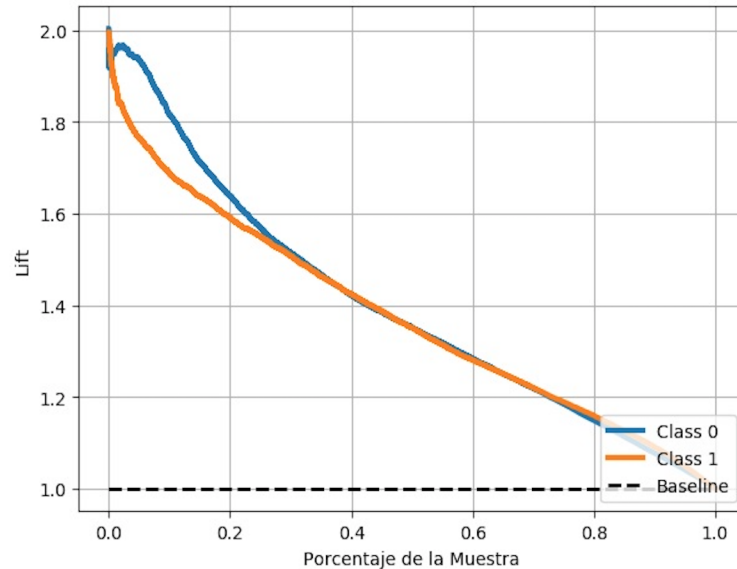
Figura 6.1: Perfiles de precisión acumulados para datos balanceados



Se observa que al 50 % de datos, se tiene un 67.53 % de observaciones positivas, estando 17.53 puntos porcentuales sobre un modelo aleatorio.

### 6.1.2. Lift

Figura 6.2: Curva Lift para datos balanceados  
Curva de Lift



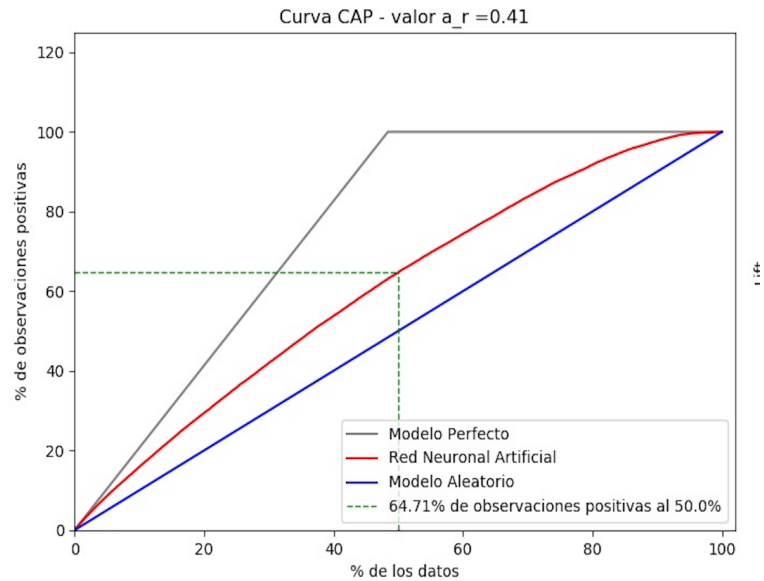
Las curvas de clasificación del modelo muestran un mejor rendimiento respecto a un modelo aleatorio para cualquier porcentaje de la muestra, logran superar un factor de 1.8 veces el poder clasificador de un modelo aleatorio y al 40 % de la muestra poseen un factor de 1.4 veces el poder clasificador de un modelo aleatorio.



## 6.2. Altas cable

### 6.2.1. Cap

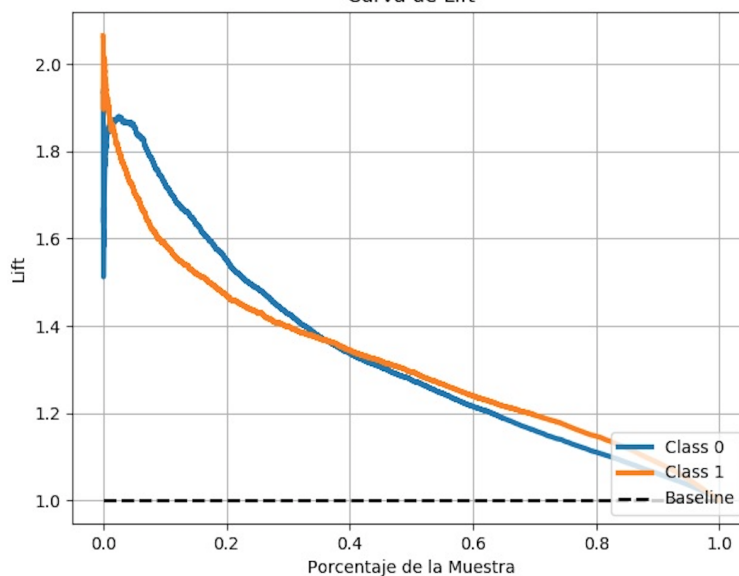
Figura 6.3: Perfiles de precisión acumulados para datos balanceados



Se observa que al 50% de datos, se tiene un 64.71% de observaciones positivas, estando 14.71 puntos porcentuales sobre un modelo aleatorio.

## 6.2.2. Lift

Figura 6.4: Curva Lift para datos balanceados  
Curva de Lift

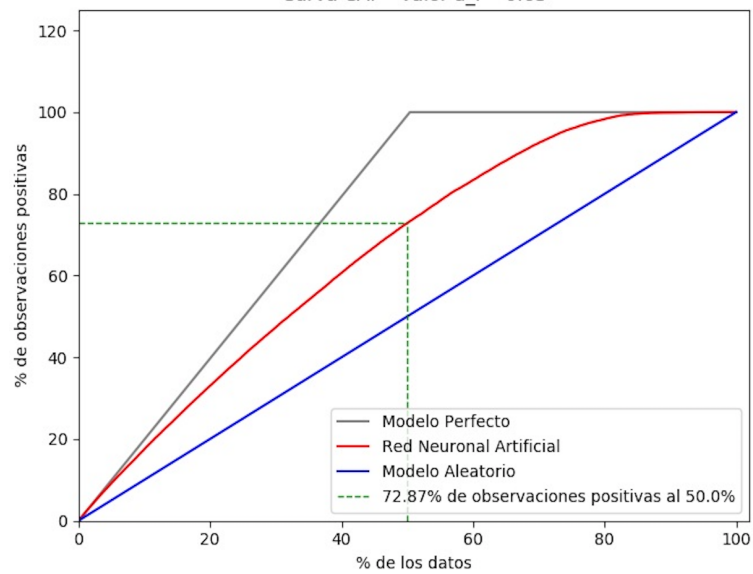


Las curvas de clasificación del modelo muestran un mejor rendimiento respecto a un modelo aleatorio para cualquier porcentaje de la muestra, logran superar un modelo aleatorio en un factor de 2 veces el poder clasificador para la clase 1 y 1.8 veces el poder clasificador para la clase 0 y al 40 % de la muestra poseen un factor aproximado de 1.3 veces el poder clasificador de un modelo aleatorio.

## 6.3. Altas Internet

### 6.3.1. Cap

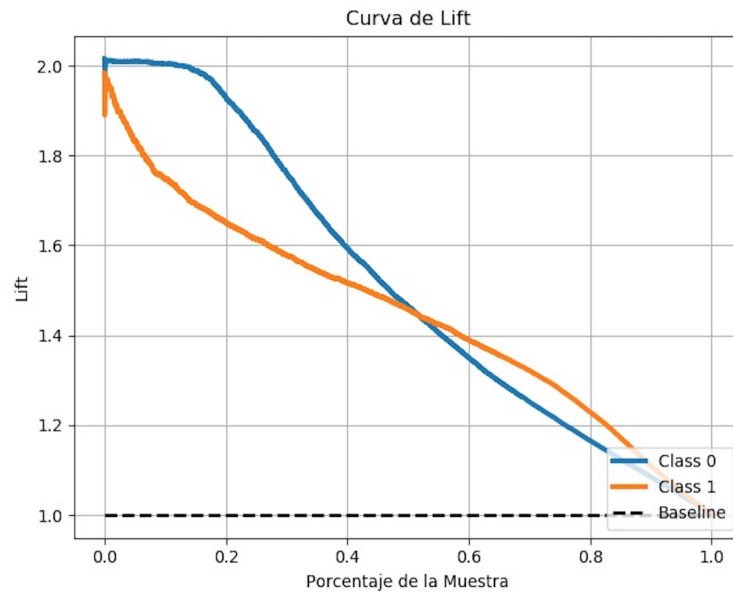
Figura 6.5: Perfiles de precisión acumulados para datos balanceados  
Curva CAP - valor  $a_r = 0.63$



Se observa que al 50 % de datos, se tiene un 72.87 % de observaciones positivas, estando 22.87 puntos porcentuales sobre un modelo aleatorio.

### 6.3.2. Lift

Figura 6.6: Curva Lift para datos balanceados

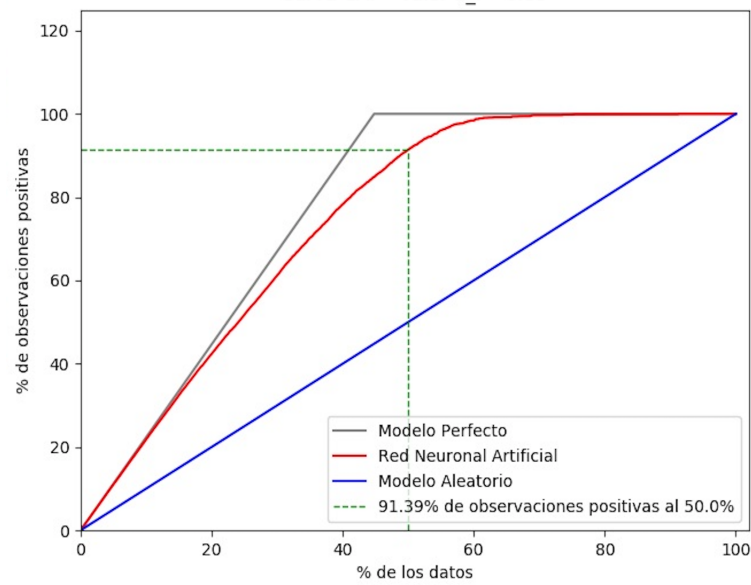


Las curvas de clasificación del modelo muestran un mejor rendimiento respecto a un modelo aleatorio para cualquier porcentaje de la muestra, logran superar un modelo aleatorio en un factor de 2 veces el poder clasificador para la clase 0 y 1.9 veces el poder clasificador para la clase 1; y al 40 % de la muestra poseen un factor aproximado de 1.5 veces el poder clasificador de un modelo aleatorio para la clase 1 y 1.6 veces para la clase 0.

## 6.4. Altas fono

### 6.4.1. Cap

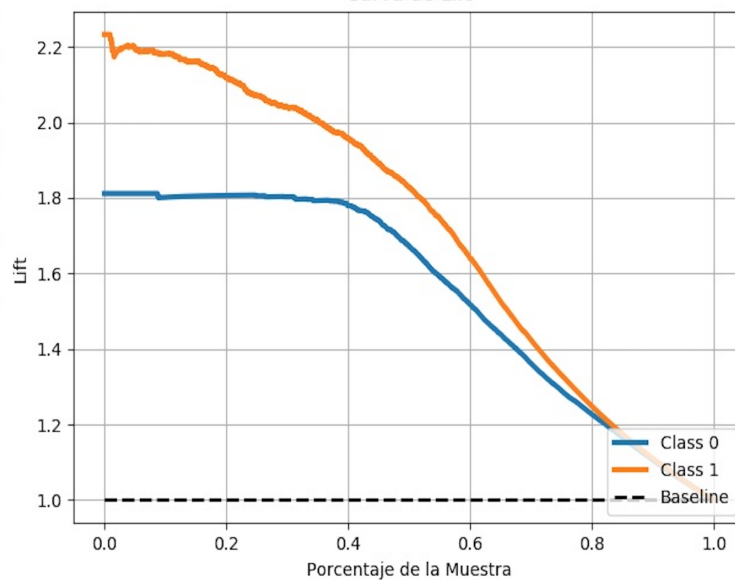
Figura 6.7: Perfiles de precisión acumulados para datos balanceados  
Curva CAP - valor  $a_r = 0.89$



Se observa que al 50 % de datos, se tiene un 91.39 % de observaciones positivas, estando 41.39 puntos porcentuales sobre un modelo aleatorio.

## 6.4.2. Lift

Figura 6.8: Curva Lift para datos balanceados  
Curva de Lift



Las curvas de clasificación del modelo muestran un mejor rendimiento respecto a un modelo aleatorio para cualquier porcentaje de la muestra, logran superar un modelo aleatorio en un factor de 2.2 veces el poder clasificador para la clase 1 y 1.8 veces el poder clasificador para la clase 0; y al 40 % de la muestra poseen un factor aproximado de 1.8 veces el poder clasificador de un modelo aleatorio para la clase 0 y 2.0 veces para la clase 1.

## 6.5. Modelo comparativo de descuentos

A continuación se presenta un modelo comparativo que busca mostrar la diferencia entre una campaña con política de igual descuento para la totalidad de los clientes objetivo y una campaña con política basada en el modelo propuesto: descuentos en base a la probabilidad de cambio de empaquetamiento.

Se tomará el supuesto de igual cantidad de clientes por campaña (1000 clientes), por lo cual los costos asociados a las campañas se tratarán como equivalentes, considerando que se contacta a la misma cantidad de clientes. De esta manera, el único costo que se considerará para el análisis comparativo será el del descuento proporcionado al cliente como parte de la campaña.

Respecto al beneficio, se tomará el supuesto de igual probabilidad de fuga para los empaquetamientos, debido a que se utilizara una probabilidad de fuga promedio y debido a que los cambios en empaquetamiento son equivalentes, por lo cual el empaquetamiento final es el mismo, y sin importar la probabilidad que se utilice, esta debería ser la misma para ambos. Al mismo tiempo, se realizará el análisis para 6 meses respecto a cada modelo.

De esta manera, se procede a realizar el análisis de costo beneficio para cada uno de los cambios en empaquetamiento, tanto para un modelo con igual descuento para todos y un modelo con descuentos multiplicados por un factor relativo a la probabilidad de cambio de empaquetamiento para cada cliente (mayor probabilidad significa menor descuento).

Tabla de ARPC's y delta ARPU

$\Delta Paquetes$	ARPC final(CLP)	ARPC inicial(CLP)	$\Delta ARPC promedio(CLP)$
3P-CF	41.100	27.800	13.300
3P-CI	41.100	36.100	5.000
3P-FI	41.100	31.300	9.800
3P-C	41.100	17.000	24.100
3P-F	41.100	6.900	34.200
3P-I	41.100	19.800	21.300
CI-I	36.100	19.800	16.300
CI-C	36.100	17.000	19.100
CF-F	27.800	6.900	20.900
CF-C	27.800	17.000	10.800
FI-I	31.300	19.800	11.500
FI-F	31.300	6.900	24.400

Tabla 6.1: Diferencia en ARPC promedio para cambios en empaquetamiento

### 6.5.1. Alta 3P

Descuento igual para todos: Supuestos: Descuento de 5k, 30% de conversión de clientes, 2% de fuga de clientes mensual, delta de ARPU.

Se asume que 300 clientes de 1000 aceptan la campaña, luego los costos y beneficios son para esos 300 clientes.

Mes	Costo(CLP)	Beneficio(CLP)	Utilidad(CLP)
1er mes	1.500.000	3.990.000	2.490.000
2do mes	1.470.000	3.910.200	2.440.200
3er mes	1.440.600	3.831.996	2.391.396
4to mes	1.411.788	3.755.356	2.343.568
5to mes	1.383.552	3.680.249	2.296.697
6to mes	1.355.881	3.606.644	2.250.763

Tabla 6.2: Descuento igual para todos

Descuento en base a probabilidades para 300 clientes. Supuestos: se utilizan las probabiliades para calcular un descuento en base a ellas (\$1000 de base + \$5000 multiplicado por la probabilidad de cambio de empaquetamiento). Se ordena de manera descendente la probabilidad y se escogen los primeros 300.

Mes	Costo(CLP)	Beneficio(CLP)	Utilidad(CLP)
1er mes	583.615	3.990.000	3.406.385
2do mes	571.943	3.910.200	3.338.257
3er mes	560.504	3.831.996	3.271.492
4to mes	549.294	3.755.356	3.141.941
5to mes	538.308	3.680.249	2.296.697
6to mes	527.542	3.606.644	3.079.102

Tabla 6.3: Descuento en base a probabilidad de cambio en empaquetamiento (considerando 300 clientes)

Descuento en base a probabilidades para 562 clientes. Supuestos: se utilizan las probabilidades para calcular un descuento en base a ellas (\$1000 de base + \$5000 multiplicado por probabilidad) se ordena de manera descendente la probabilidad y se escogen los primeros 562 para igualar el costo inicial de 300 clientes con descuento igual para todos (primera tabla).



Mes	Costo(CLP)	Beneficio(CLP)	Utilidad(CLP)
1er mes	1.503.370	7.474.600	5.971.230
2do mes	1.473.303	7.325.108	5.851.805
3er mes	1.443.837	7.178.606	5.734.769
4to mes	1.414.960	7.035.034	5.620.074
5to mes	1.386.661	6.894.333	5.507.672
6to mes	1.358.927	6.756.446	5.397.519

Tabla 6.4: Descuento en base a probabilidad de cambio en empaquetamiento (Considerando 562 clientes)

Para cualquier caso el beneficio en términos de utilidad resulta ser mayor al tener la posibilidad de discriminar probabilísticamente a cada cliente y poder entregar políticas de descuento discretizadas en base a la probabilidad de cambio en empaquetamiento entregada por el modelo de clasificación.

# Capítulo 7

## Extensiones del Modelo

### 7.1. Extensiones de Predictibilidad

El modelo propuesto fue evaluado en la predicción específica de ciertos tipos de cambio en empaquetamiento de los clientes. Ahora bien, este puede ser extendido debido a su diseño.

El diseño tiene como primer paso un algoritmo de extracción de cambios en empaquetamiento mensuales durante un año de análisis. Ahora, las variables utilizadas son extraídas mediante ciertos identificadores clave que pueden ser extendidos a cualquier tipo de base utilizada.

De esta manera, la creación de la variable de estado a clasificar (el empaquetamiento final), puede ser cambiada por cualquier tipo de empaquetamiento, generando así la posibilidad de evaluar cualquier cambio en empaquetamientos, incluso el empaquetamiento que tiene como forma final la fuga del cliente. Las probabilidades pueden ser generadas para evaluar cualquier clasificación final y el modelo es capaz de predecir todas las opciones de cambio en empaquetamiento posibles.

Sin embargo, las métricas de precisión, recall, F1 Score, curva ROC y análisis CAP tendrán variaciones; esto principalmente debido a que las variables que identifican a un cambio de empaquetamiento del tipo cross-sell pueden no ser las mismas que identifican el cambio de empaquetamiento del tipo fuga.

## 7.2. Extensiones de Dimensionalidad

Teniendo en cuenta la última consideración de la extensión en predictibilidad del modelo, esta puede ajustarse haciendo uso de una segunda extensión: dimensionalidad.

Debido a que el modelo inicia con un algoritmo de extracción de cambios en empaquetamiento que extrae tanto el rut del cliente como el identificador de vivienda, el algoritmo de extracción de variables de estudio puede ser extendido.

Solo se requiere modificar las variables seleccionadas en el algoritmo que escribe el archivo final: se pueden incluir todas las variables a las que se tenga acceso. Añadido a esto, la optimización del modelo que prueba las distintas combinaciones de variables puede nuevamente ser utilizada para seleccionar las variables de mayor impacto en la predicción que se busca realizar.

## 7.3. Extensiones de Efectividad

Finalmente, la efectividad o el rendimiento del modelo tiene una importante posibilidad de extensión: una mayor existencia de datos de entrenamiento genera un mejor rendimiento en cuanto a predictibilidad del modelo. Mientras mayor sea la cantidad de datos que se manejan respecto a los clientes, y mayor sea la base sobre la cual aprende el modelo, mejores serán las predicciones y la efectividad del algoritmo de aprendizaje automático.

## 7.4. Extensiones de interpretabilidad de características y variables

Añadido a las extensiones mencionadas anteriormente, también se espera que el modelo permita utilizar las probabilidades obtenidas para lograr identificaciones de características y variables relevantes en los clientes que permitan tomar decisiones de priorización respecto a los resultados.

Ahora bien, como el modelo tiene un funcionamiento del tipo caja negra, este solo produce

las probabilidades y no los pesos específicos de cada neurona o de cada variable presente al momento de generar la probabilidad obtenida. Debido a esto, una conclusión recurrente identifica a este tipo de modelos como útiles en cuanto a predicción pero poco útiles en cuanto al negocio y a la identificación de patrones en los clientes: solo se obtiene una probabilidad que permite actuar de manera reactiva, pero no se conoce la importancia de cada variable en la probabilidad generada.

Sin embargo, esta conclusión no es completamente correcta. Debido a que la probabilidad obtenida es predicha a partir de una serie de variables que ingresan al modelo y estas variables se encuentran identificadas y detalladas para cada cliente; por lo cual pueden ser utilizadas en conjunto para interpretar el valor de cada una de ellas.

La herramienta utilizada para esta interpretación lleva el nombre de valores SHAP (SHapley Additive exPlanation) y proviene del estudio de la teoría de juegos. Fue publicada el año 2017 por Scott M. Lundberg y Su-In Lee.

Entender por qué un modelo realiza una cierta predicción puede ser tan crucial como la precisión de la predicción en muchas aplicaciones, y para el presente modelo resulta de gran relevancia.

El modelo SHAP asigna a cada característica (variable) un valor de importancia para cada predicción particular. La idea es utilizar la teoría de juegos cooperativa para interpretar el objetivo del modelo. Todas las características son contribuyentes; están intentando predecir la clasificación, que corresponde al juego; así, la recompensa resulta ser la clasificación real menos el resultado de predicción del modelo.

El valor Shapley es definido a través de una función de valor para el conjunto de jugadores  $S$ . El valor Shapley de una característica corresponde a su contribución al pago del juego, promediada y sumada sobre todas las posibles combinaciones de valores para la característica:

$$\phi_j(val) = \sum_{S \subseteq x_1, \dots, x_p / x_j} |S|!(p - |S| - 1)!(val(S \cup x_j) - val(S))$$

Donde  $S$  es un subconjunto de las características utilizadas en el modelo,  $x$  es el vector de valores para las características de la instancia a ser explicada y  $p$  es el número total de características.

$$Val_x(S)$$

es la predicción para los valores de las características en el conjunto  $S$  que están marginadas

sobre las características que no se encuentran en el conjunto  $S$ .

Una manera intuitiva de entender el valor Shapley es la siguiente ilustración: los valores de la característica entran a una pieza en orden aleatorio. Todos los valores de las características en la pieza participan en el juego (contribuyen a la predicción). El valor Shapley del valor de la característica es el cambio promedio en la predicción que la coalición (que ya se encuentra en la pieza) recibe cuando el valor de la característica se les une.

De esta manera, utilizando los valores Shapley se puede conocer la importancia de cada característica en la probabilidad obtenida para cada cliente individual, lo cual permite interpretabilidad del modelo e incluso permite extrapolar esta importancia respecto a segmentos de clientes.

# Capítulo 8

## Conclusión

El presente trabajo de título logra generar un modelo capaz de predecir la probabilidad de cambio en empaquetamiento por cliente basándose en variables de entrenamiento que caracterizan al mismo, mostrando un rendimiento de ajuste por sobre una serie de modelos y superando los resultados esperados.

Se logra la definición y selección de cambios en empaquetamiento específicos para el estudio, resultando aquellos que generan un aumento en el ARPC por nivel de empaquetamiento: altas triple, altas internet, altas cable y altas fono.

Debido a la generación de un algoritmo de obtención de datos respecto a cambios en empaquetamiento mensuales y el estudio de posibles modelos análogos de clasificación, se logra seleccionar efectiva y justificadamente el modelo de red neuronal artificial, capaz de ser entrenado en una gran cantidad de datos y de entregar probabilidades específicas con buen resultado en cuanto a métricas de ajuste.

Los resultados y medidas de efectividad de predicción muestran un mínimo de capacidad discriminativa del 64 % en cuanto a precisión y recall para los empaquetamientos que finalizan en una alta de cable y un máximo del 87 % de precisión y recall para los empaquetamientos que finalizan en una alta de fono. Estando siempre igual o sobre los resultados esperados.

Añadido a esto, el modelo presenta un mínimo de 20 puntos porcentuales en cuanto a razón de verdaderos positivos respecto a falsos positivos y una mínima separación en la distribución de clases de 0.287 puntos en la prueba de Kolmogorov - Smirnov. Por otra parte, el mejor rendimiento lo logra con una curva ROC que se encuentra 84 puntos porcentuales sobre un

modelo aleatorio y una separación de distribución de clases de 0.753 puntos en la prueba de Kolmogorov - Smirnov.

Como último punto en cuanto al ajuste del modelo, los perfiles de precisión acumulados muestran un mínimo de 14.71 puntos porcentuales de observaciones positivas al 50 % sobre un modelo aleatorio y un máximo de 41.39 puntos porcentuales. Añadido a esto, el modelo muestra curvas de poder predictivo (LIFT) que se encuentran siempre por sobre los modelos aleatorios (2 veces en la mayoría de los cambios en empaquetamiento).

De esta manera, el modelo permite realizar campañas basadas en probabilidades cuantitativas y ordenar a los clientes en base a su probabilidad de cambio en empaquetamiento, generando la posibilidad de discriminar en cuanto a descuentos discretizados entregados y superar a cualquier modelo aleatorio o univariado como los actualmente utilizados en la empresa.

Finalmente, la extensión del modelo permite utilizar las probabilidades obtenidas para lograr posibles identificaciones de características y variables relevantes en los clientes y así tomar decisiones de priorización respecto a los resultados. De la misma manera, en el presente trabajo se muestra como un modelo de probabilidad individual por cliente permite la generación de campañas de ajustadas de manera individual y permite a su vez una reducción en costos junto a una mayor efectividad. Añadido a esto, se presenta una extensión al modelo que permite identificar patrones en las variables utilizadas y su valor para cada cliente basándose en las probabilidades obtenidas y el entrenamiento realizado; todo esto con la finalidad de escapar el entorno de caja negra que presenta el modelo como tal.

El modelamiento y análisis realizados proporcionan una relevante herramienta de discriminación en cambios de empaquetamiento en el rubro de telecomunicaciones, prediciendo de manera efectiva y ajustada la probabilidad de estos cambios y entregando una fuerte herramienta para la generación de campañas y la evaluación de las mismas para ventas cruzadas; pudiendo ser extendido a cualquier cambio en empaquetamientos (incluso aquellos que finalizan en fuga de clientes), y proporcionando la oportunidad de incluir una mayor variedad y cantidad de variables, junto a un entrenamiento capaz de incorporar una enorme cantidad de datos.

# Bibliografía

Patel, A. A. 2019. Hands-On Unsupervised Learning Using Python: how to build applied machine learning solutions from unlabeled data. Sebastopol, CA: OReilly Media. 362p.

Warr, K. 2019. Strengthening deep neural networks: making Ai less susceptible to adversarial trickery. Beijing: OReilly. 246p.

Plunkett, J. W. (2019). Plunketts artificial intelligence (Ai) machine learning industry almanac 2019: the only comprehensive guide to the artificial intelligence machine learning industry. Houston, TX: Plunkett Research, Ltd. 533p.

Fader, P.S., B.G. Hardie and K. Jerath, 2007. Estimating CLV using aggregated data: The Tuscan lifestyles case revisited. J. Interact. Market., 21: 55-71.

Hecht-Nielsen, R., 1988. Applications of counter propagation network. Neural Network, 1: 131-139.

Hwang, H., T. Jung and E. Suh, 2004. An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry. Expert Syst. Appl., 26: 181-188.

Kaishev, V.K., J.P. Nielsen and F. Thuring, 2013. Optimal customer selection for cross-selling of financial services products. Expert Syst. Appl., 40: 1748-1757.

Kamakura, W.A., M. Wedel, F. De Rosa and J.A. Mazzon, 2003. Cross-selling through database marketing: A mixed data factor analyzer for data augmentation and prediction. Int. J. Res. Market., 20: 45-65.



Li, S., B. Sun and A.L. Montgomery, 2011. Cross-selling the right product to the right customer at the right time. *J. Market. Res.*, 48: 683-700.

Li, S., B. Sun and R.T. Wilcox, 2005. Cross-selling sequentially ordered products: An application to consumer banking services. *J. Market. Res.*, 42: 233-239.

Liu, H.Y., 2007. Development of a framework for customer relationship management (CRM) in the banking industry. *Int. J. Manag.*, 24: 15-32.

Thuring, F., J.P. Nielsen, M. Guillen and C. Bolance, 2012. Selecting prospects for cross-selling financial products using multivariate credibility. *Expert Syst. Appl.*, 39: 8809-8816.

Verhoef, P.C. and B. Donkers, 2001. Predicting customer potential value an application in the insurance industry. *Decis. Support Syst.*, 32: 189-199.

Zboja, J.J. and M.D. Hartline, 2012. An examination of high-frequency cross-selling. *J. Relationship Market.*, 11: 41-55.

Zhang, Z., N. Zheng and T. Wang, 1998. Fuzzy counter propagation neural network, universal approximation and application to time series prediction. *Proceedings of the International Conference on Neural Networks and Brain*, October 27-30, 1998, Beijing, China, pp: 130-133.

Rojas, I., Joya, G., Catala, A. 2019. *Advances in computational intelligence: 15th International Work-Conference on Artificial Neural Networks*, Iwann 2019, Gran Canaria, Spain, June 12-14, 2019, proceedings. Cham, Switzerland: Springer. Part 1.

Zhang, Y., Chen, D., Ye, C. 2019. *Toward deep neural networks: Wasd neuronet models, algorithms, and applications*. Boca Raton, FL: CRC Press, Taylor Francis Group. 340p.