



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

**MODELO DE CARACTERIZACIÓN DE LA LEALTAD DE CLIENTES  
BANCARIOS INCORPORANDO TÉCNICAS DE TEXT MINING PARA UNA  
EMPRESA DE INVESTIGACIÓN DE MERCADO**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL

**NICOLE PATRICIA MOLINA BAUTISTA**

PROFESOR GUÍA:  
PABLO MARÍN VICUÑA

MIEMBROS DE LA COMISIÓN:  
ALEJANDRA PUENTE CHANDÍA  
NICOLÁS FRITIS COFRÉ

SANTIAGO DE CHILE  
2021

## MODELO DE CARACTERIZACIÓN DE LA LEALTAD DE CLIENTES BANCARIOS INCORPORANDO TÉCNICAS DE TEXT MINING PARA UNA EMPRESA DE INVESTIGACIÓN DE MERCADO

Ipsos, empresa líder en investigación de mercado en Chile, busca mantener su relevancia adaptándose a las necesidades de su mercado. En esa búsqueda, se vió que existe la oportunidad de agregar valor a sus clientes de la industria bancaria, mediante la caracterización de la lealtad de los usuarios incorporando sus opiniones expresadas en encuestas. De esta forma, los bancos podrán entender a partir de la experiencia de sus usuarios, qué es lo que impacta positiva o negativamente su lealtad.

En Ipsos el procesamiento de textos se realiza -principalmente- de forma manual lo que implica un gasto aproximado de \$32.400.000 al año por la contratación de equipo externo para tal labor. Actualmente, existen herramientas de *Text Mining* que facilitan la tediosa labor de categorizar las respuestas abiertas de los clientes y que no han sido del todo exploradas por la empresa.

Para llevar a cabo este objetivo se trabajó con las encuestas realizadas en un banco en particular y el trabajo se dividió en dos partes. En primer lugar, el minado de textos, en el que se implementaron distintas herramientas para normalizar, limpiar y transformar las verbalizaciones. Luego, se implementó un modelo de detección de tópicos LDA para conocer los temas tratados por los comentarios de las encuestas mensuales del banco.

Después, la categorización de los comentarios se incorporó como variable adicional en el modelo de caracterización de lealtad. Se replicó el modelo de regresión actual de la empresa, para compararlo con cuatro modelos *XGBoost* que difieren en la inclusión o no de los tópicos entre sus variables independientes y si es que la variable dependiente es la nota de recomendación o el tipo de cliente: promotor, neutro o detractor.

Comparando los resultados de los modelos se concluyó que el modelo *XGBoost* que incorpora tópicos entre sus variables independientes y agrupa la variable dependiente entre clientes promotores, detractores y neutros, supera en un 21 % el AUC del modelo base. Por otro lado, el mismo modelo sin incluir tópicos mejora el modelo base en un 16 % .

Gracias al análisis realizado a partir de ambos modelos, se sugiere a la empresa utilizar ambos modelos para que se complementen entre sí. Por un lado, aprovechar el modelo que no incluye tópicos para monitorear la evolución de la satisfacción y la incidencia medidos objetivamente y, por otra parte, implementar el modelo con tópicos para explorar nuevas fortalezas y debilidades y determinar cómo influyen estas en su fidelización.

*A mi abuelo  
por su apoyo incondicional*

***Gracias***

# Agradecimientos

Se termina este camino universitario con la finalización de este trabajo. Fue un año difícil, no solo por el desafío de escribir una memoria, sino que también se sumó la muerte de mi abuelo y la pandemia. Un año en que pude ver poco a mi familia y amigos, la vida cotidiana ya no era lo mismo y todos tuvimos que adaptarnos.

Primero, me gustaría agradecer a quienes tuvieron relación directa con mi memoria. A mi profesor guía, Pablo Marín, por siempre tener la mejor disponibilidad y entregar feedback a la velocidad de la luz, muchas veces eran duros, pero hicieron que este trabajo sea lo que es hoy. También agradecer a Ipsos por darme la oportunidad de trabajar con ellos. Más en específico, gracias a mi contraparte, Diego Araya, por su empatía y apoyo a lo largo de todo el proceso.

Como no agradecer a mis padres, quienes me dejaron alejarme del nido para volar a Santiago e independizarme más pronto de lo esperado. Gracias por su esfuerzo, por su apoyo e infinito amor que siempre han tenido para mi. Siempre están en mi corazón y siempre estaré para ustedes.

Gracias a mi hermano menor Daniel, por su apoyo, cariño y por también entender mi decisión de marcharme a otra ciudad. Gracias a mis hermanas, Josy y Katy, que siempre han sido mis modelos a seguir, por motivarme y apoyarme, sobretodo en la toma de decisiones importantes. Gracias en especial a Katy, por permitirme convivir con ella durante estos 6 años, nunca me faltó nada y junto con la Amandita siempre fueron una buena compañía.

Quiero agradecer a mi abuelo, que ya no está entre nosotros, por siempre haber confiado en mis capacidades, siempre apoyarme incondicionalmente y por sobretodo, por haberme querido tanto desde que tengo memoria.

Gracias a mis amigos de la universidad, desde que los tengo en mi vida, la universidad se volvió una experiencia que disfruté un montón. Somos un grupo grande, pero no por eso menos cercano, sé que estaremos los unos para el otro por mucho tiempo más.

Me faltan personas por nombrar pero lamentablemente el espacio es limitado, amigos de la vida u otros familiares que de alguna u otra forma también fueron parte de este camino, gracias a todos.

Por último, pero no menos importante, quiero agradecer a Lucas, por su apoyo constante a lo largo de todo este proceso universitario, por siempre estar ahí, aguantarme en mis momentos de estrés o angustia y quererme por sobre todas las cosas. Al final del día, eras el único que sabía de los malos momentos, y siempre tuviste una palabra y un abrazo para subirme el ánimo y alentarme a seguir adelante.

# Tabla de Contenido

<b>1. Introducción</b>	<b>1</b>
<b>2. Antecedentes generales</b>	<b>2</b>
2.1. Caracterización de la empresa . . . . .	2
2.2. Información del área de la empresa . . . . .	3
2.3. Industria bancaria . . . . .	3
<b>3. Descripción del proyecto y justificación</b>	<b>7</b>
<b>4. Objetivos y alcance</b>	<b>10</b>
4.1. Objetivo General . . . . .	10
4.2. Objetivos específicos . . . . .	10
4.3. Alcances . . . . .	10
<b>5. Marco conceptual</b>	<b>12</b>
5.1. CRISP-DM . . . . .	12
5.2. Minería de textos . . . . .	13
5.2.1. Enfoques . . . . .	14
5.3. Algoritmos de Information Retrieval (IR) . . . . .	14
5.4. Árboles de decisión . . . . .	15
5.5. Gradient tree boosting . . . . .	15
5.6. XGBoost . . . . .	17
5.7. Métricas de evaluación . . . . .	17
5.8. NPS, Impacto y Criticidad . . . . .	18
5.9. Imputación múltiple . . . . .	19
<b>6. Metodología</b>	<b>21</b>
6.1. Entender el negocio . . . . .	21
6.2. Entender los datos . . . . .	21
6.3. Preparación de los datos . . . . .	22
6.4. Modelamiento . . . . .	23
6.5. Evaluación . . . . .	23
6.6. Despliegue . . . . .	23
<b>7. Desarrollo metodológico</b>	<b>24</b>
7.1. Entender los datos . . . . .	24
7.2. Preparación de los datos . . . . .	29
7.3. Modelamiento . . . . .	33

7.3.1. Modelo de Tópicos . . . . .	34
7.3.2. Modelo de Caracterización de Lealtad . . . . .	37
7.4. Evaluación . . . . .	39
7.5. Despliegue de resultados . . . . .	41
<b>8. Análisis de resultados</b>	<b>42</b>
8.1. Análisis de tópicos . . . . .	42
8.2. Modelo de caracterización . . . . .	45
<b>9. Conclusiones</b>	<b>57</b>
9.1. Conclusiones generales . . . . .	57
9.2. Recomendaciones para Ipsos . . . . .	59
9.3. Recomendaciones bancarias para la muestra analizada . . . . .	60
9.4. Trabajos futuros . . . . .	61
<b>Bibliografía</b>	<b>63</b>

# Índice de Tablas

2.1.	Composición de la industria bancaria chilena . . . . .	4
7.1.	Ficha metodológica cuestionario . . . . .	24
7.2.	Distribución en evaluaciones de clientes . . . . .	28
7.3.	Cantidad y porcentaje de nulos y “NS/NR” . . . . .	28
7.4.	Descripción de caracteres en comentarios positivos . . . . .	31
7.5.	Descripción de caracteres en comentarios negativos . . . . .	31
7.6.	Comparación de algoritmos de stemming en un comentario . . . . .	33
7.7.	Variables independientes modelo actual . . . . .	37
7.8.	Hiperparámetros testeados en modelos XGBoost . . . . .	39
7.9.	Métricas de evaluación de desempeño en datos de prueba . . . . .	40
7.10.	AUC de entrenamiento vs prueba . . . . .	40
8.1.	Tópicos y su respectiva polaridad . . . . .	42
8.2.	Impacto y Criticidad de los tópicos . . . . .	45

# Índice de Ilustraciones

2.1.	Ranking sectorial de satisfacción en Chile . . . . .	5
2.2.	Ranking sectorial de recomendación en Chile . . . . .	5
2.3.	Participación en colocaciones de los principales bancos del Sistema Bancario . . . . .	6
5.1.	Metodología CRISP-DM . . . . .	12
7.1.	Recuento de encuestas por región . . . . .	25
7.2.	Edades encuestados año a año . . . . .	26
7.3.	Distribución de género en encuestas . . . . .	26
7.4.	Distribuciones de recomendación y satisfacción global por mes . . . . .	27
7.5.	Ejemplos de textos limpios . . . . .	32
7.6.	Comparación de palabras frecuentes luego de procesamientos . . . . .	32
7.7.	Nubes de palabras de comentarios positivos y negativos respectivamente . . . . .	33
7.8.	Coherencia según número de tópicos en comentarios positivos . . . . .	35
7.9.	Coherencia según n° tópicos en comentarios negativos . . . . .	35
7.10.	Nubes de palabras de tópicos comentarios positivos . . . . .	36
7.11.	Nubes de palabras de tópicos comentarios negativos . . . . .	36
7.12.	Base de datos unificada . . . . .	37
8.1.	: N° de comentarios negativos por tópico . . . . .	43
8.2.	N° comentarios positivos por tópico . . . . .	43
8.3.	NPS asociado a cada tópico . . . . .	44
8.4.	Impacto de las variables en los outputs del modelo sin tópicos . . . . .	46
8.5.	Caracterización de la recomendación sin incluir tópicos . . . . .	47
8.6.	Caracterización de la recomendación entre clientes promotores . . . . .	48
8.7.	Caracterización de la recomendación entre clientes neutros . . . . .	49
8.8.	Caracterización de la recomendación entre clientes detractores . . . . .	50
8.9.	Incidencia de variables en el modelo con tópicos . . . . .	51
8.10.	Incidencia de atributos entre clientes detractores . . . . .	52
8.11.	Incidencia de atributos entre clientes promotores . . . . .	53
8.12.	Incidencia de atributos entre clientes neutros . . . . .	54
8.13.	Caracterización de la lealtad incluyendo tópicos . . . . .	55

# Capítulo 1

## Introducción

José Manuel Mena, presidente de la Asociación de Bancos e Instituciones Financieras (ABIF), señaló en la memoria anual 2019 de la organización que la industria bancaria siempre pone al cliente en el centro de sus decisiones, con el objetivo de mejorar su calidad de vida y hacer un impacto positivo en sus proyectos.

Para entender mejor a sus clientes, varios bancos en Chile trabajan en colaboración con Ipsos, una empresa multinacional, líder en investigación de mercado, encargada de implementar un programa de medición de experiencia de cliente, el cual se basa en medir los diferentes puntos de contacto y líneas de negocio del banco.

El desafío que se presenta es diseñar un modelo para el segmento personas que permita determinar los factores que influyen en la caracterización de la lealtad de sus clientes mensualmente, sin dejar de lado el principio de parsimonia ni mermar la experiencia del cuestionario (en flujo y tiempo). Por esto, surge la oportunidad de evaluar el efecto que posee incorporar el análisis de la data no estructurada (entregada en las preguntas abiertas de los cuestionarios) en el modelo.

Actualmente, las técnicas que utiliza Ipsos para procesar los textos de las encuestas son caras o poco eficientes. Es por esto, que surge una oportunidad para mejorar el análisis. Ipsos quiere integrar las respuestas abiertas de los clientes de la industria bancaria en la determinación de los elementos que determinan su lealtad, para aportarles mayor valor y que puedan gestionar esfuerzos que repercutan en la experiencia de los clientes.

El objetivo principal de esta memoria es entender mejor la experiencia que perciben los clientes de un banco particular en Chile para determinar acciones que podrían tomar para mejorar sus servicios. Para esto se creará una metodología que optimice el análisis de las opiniones abiertas de los clientes e integrarlo a la caracterización de la lealtad. Así, a lo largo de este trabajo se evaluarán distintos modelos para cumplir con el objetivo final y dar paso a que se implemente y replique en el futuro.

# Capítulo 2

## Antecedentes generales

### 2.1. Caracterización de la empresa

Ipsos Groups S.A. es una empresa multinacional -presente en más de 90 mercados- que se encarga de realizar investigación de mercado y consultoría en el sector público y privado. Ofrece sus servicios a través de 17 unidades especialistas que responden a distintos objetivos de marketing, de las cuales, 7 operan en Chile.

Fue fundada en 1975 (París) por el economista Didier Truchot (CEO). En 1982 se incorpora a la compañía el sociólogo Jean-Marc, quien junto a Truchot fija un objetivo principal para la compañía: “Hacer de la investigación basada en encuestas una de las principales maneras de entender la sociedad contemporánea” [1].

La investigación de mercado surge como una rama de la investigación social [2] y se define como “la reunión, el registro y el análisis de todos los hechos acerca de los problemas relacionados con las actividades de las personas, las empresas y las instituciones en general” [3]. Esto permite -en líneas generales- que las instituciones públicas y privadas tomen mejores decisiones.

La misión de Ipsos es: “Ayudar a los clientes a comprender el mundo tal como es con el fin de operar de manera eficaz y sostenible mediante la aportación de información precisa y relevante que sea fácil de usar” . Sus expectativas se concentran en la visión de la empresa: “Ser el socio preferido de nuestros clientes en nuestras áreas de especialización” .

Ipsos opera en el país con un plantel profesional de 160 colaboradores a través de 7 líneas de servicio: Brand Health Tracking, Public Affairs, UU Cualitativo, Mystery Shopper, Audience Measurement, Mkt Strategy & Understanding y Customer Experience. Estos diferentes servicios le permiten actualmente ser el número 1 en ventas de market research en Chile y contar con alrededor de 70 clientes, ligados a distintas industrias, en donde las más relevantes son: Servicios Financieros (Banca y Seguros), Telecomunicaciones, Retail y Automotriz.

Jorge López, country manager de Ipsos Chile, señaló a Diario Financiero en 2019 que la empresa se convirtió en el principal actor de la investigación de mercado en Chile, con un 30 % de participación.

Una ventaja competitiva de Ipsos -que le permite diferenciarse de sus competidores- es justamente el amplio abanico de líneas de servicios que ofrece a sus clientes. Por otro lado, ser una empresa global con presencia en más de 80 países permite tener un gran respaldo de marca y conocimiento. Poseen centros de conocimiento globales, que realizan innovaciones en investigaciones y productos de los cuales los equipos locales se pueden nutrir y aportar valor. Permite el intercambio de conocimiento y experiencias entre países y adoptar mejores prácticas y casos de éxito. Todo esto les ha permitido posicionarse como líderes en el mercado nacional y ocupar el tercer lugar a nivel mundial.

## 2.2. Información del área de la empresa

El trabajo de título se realizará específicamente para la línea de servicio *Customer Experience (CX)*, cuyo enfoque es la medición, análisis y gestión de la experiencia cliente. Es la línea de servicio más grande en Ipsos Chile y también a nivel latinoamericano.

Esta línea de servicio contribuye a que las marcas creen y transformen experiencias para sus clientes, logrando con esto el retorno de la inversión en cada iniciativa en la que trabajan. Son responsables de llevar a cabo Programas de Investigación y Asesoría de la experiencia del cliente basados en tópicos de lealtad, experiencia, satisfacción y en traducir cómo las iniciativas de las marcas están retornando la inversión.

Los principales clientes de esta área se dividen en cuatro grandes sectores: servicios financieros, telecomunicaciones, retail y automotriz. Donde el primero es el que aporta el mayor porcentaje de los ingresos de Ipsos.

El trabajo de memoria se realizará también en colaboración directa con el área de Analytics. Esta área es una de las más recientes de Ipsos Chile, con aproximadamente un año de antigüedad. El foco es prestar soporte metodológico a todas las Líneas de Servicio de Ipsos, principalmente en términos de análisis estadísticos y últimamente en *Text Analytics*.

## 2.3. Industria bancaria

Según la Ley General de Bancos: “Banco es toda sociedad anónima especial que, autorizada en la forma prescrita por esta ley y con sujeción a la misma, se dedique a captar o recibir en forma habitual dinero o fondos del público, con el objeto de darlos en préstamo, descontar documentos, realizar inversiones, proceder a la intermediación financiera, hacer rentar estos dineros y, en general, realizar toda otra operación que la ley le permita” .

Desde el año 1925, los bancos son regulados por la Superintendencia de Bancos que en 1975 cambió su nombre a Superintendencia de Bancos e Instituciones Financieras (SBIF), encargados de la supervisión financiera del país.

El 2019 se marca un momento histórico en el que la SBIF se integró a la Comisión para el Mercado Financiero (CMF). El fin es que el país cuente con un supervisor financiero dotado de mayores herramientas, facultades y recursos de supervisión, con una mirada global del sector financiero, dado que, la CMF integra también las funciones atribuidas a las Superintendencia

de Valores y Seguros hasta 2018.

La industria bancaria actualmente está compuesta por 14 bancos establecidos en Chile, un banco estatal, 6 sucursales de bancos extranjeros y 3 sucursales de bancos locales en el extranjero. A continuación, se puede ver la composición actual de la industria.

Tabla 2.1: Composición de la industria bancaria chilena

<b>Tipo de banco</b>	<b>Bancos</b>
Establecido en Chile	Banco de Chile, Banco Internacional, Scotiabank Chile, Banco de Crédito e Inversiones, Banco Bice, HSBC Bank, Banco Santander Chile, Itaú Corpbanca, Banco Security, Banco Falabella, Banco Ripley, Banco Consorcio, Banco BTG Pactual Chile
Sucursales de Bancos Extranjeros	Banco do Brasil S.A., JP Morgan Chase Bank N.A., Banco de la Nación Argentina, MUFG Bank Ltd. China Construction Bank, Bank of China
Sucursal de Banco Local en el Exterior	BCI-Sucursal Miami, Banco del Estado de Chile – Sucursal Nueva York, Itaú Corpbanca New York Branch- Sucursal Nueva York
Banco Estatal	Banco del Estado de Chile

Fuente: Elaboración propia con datos de la CMF

El presidente de la Asociación de Bancos e Instituciones Financieras define 6 lineamientos estratégicos que constituyen los pilares del quehacer gremial de los próximos años: Solvencia, Seguridad, Clientes, Sistema de Pagos, Inclusión Financiera y Banca y sociedad; en donde destaca el relacionado a los clientes: “La relación de la banca con sus clientes es permanente y de largo plazo. Si los clientes crecen en solidez y patrimonialmente, el banco también crece.” En esta aclaración se puede notar el compromiso de los bancos en generar relaciones de largo plazo con sus clientes y para llevar a cabo este objetivo es necesario ofrecerles la mejor calidad de servicio de acuerdo con las necesidades de estos.

Dentro de las industrias que existen en Chile la que corresponde a los bancos se destaca por poseer buenos índices de satisfacción, ubicándose en el séptimo lugar de 33 sectores según el estudio Procalidad 2019 (ver Figura 2.1). Mientras que, en cuanto al nivel de recomendación de los clientes, que se traduce en lealtad, la posición baja a la décima con un NPS de 65 2.2.

## Satisfacción general

Ranking sectorial

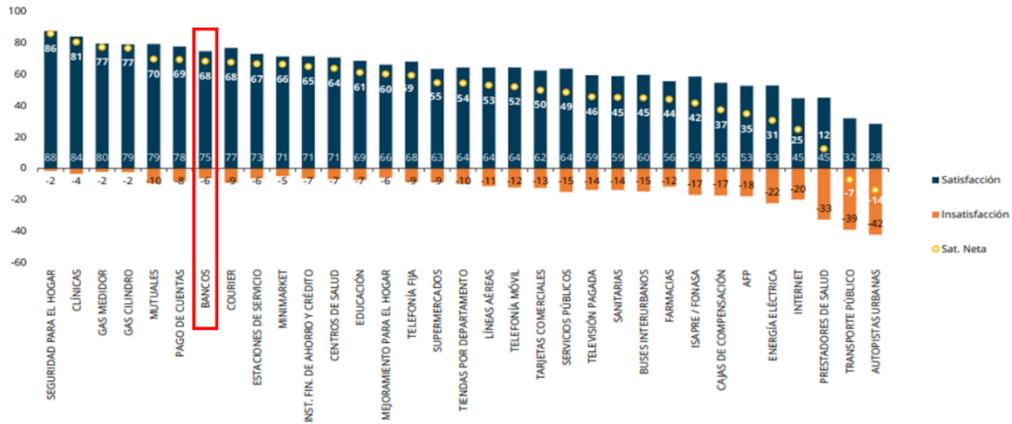


Figura 2.1: Ranking sectorial de satisfacción en Chile. Fuente: Estudio Procalidad 2019

## Recomendación

Ranking sectorial

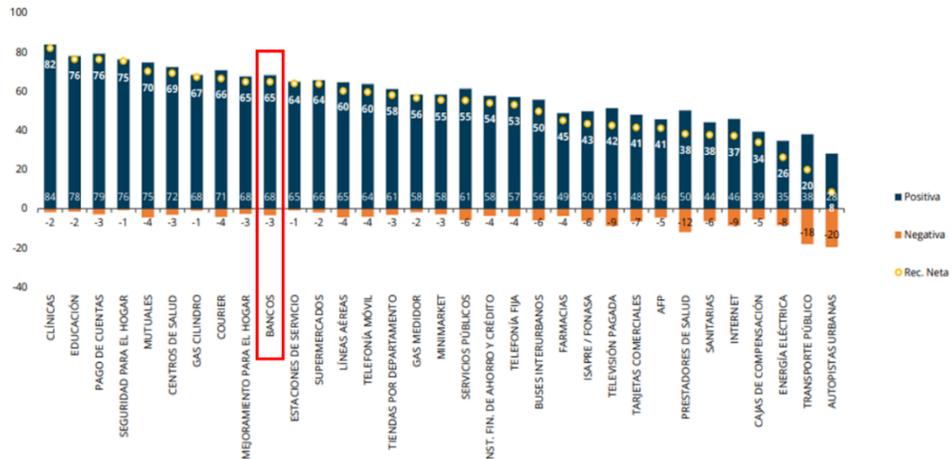


Figura 2.2: Ranking sectorial de recomendación en Chile. Fuente: Estudio Procalidad 2019

Se puede notar que la industria bancaria si bien posee una buena posición en cuanto a satisfacción y recomendación, queda un espacio de mejora, sobre todo en el último aspecto. Esto cobra mayor relevancia en una industria que es altamente competitiva, más aún, considerando que en junio del presente año se lanzó la nueva ley de Portabilidad Financiera que permite a los clientes cambiarse de proveedor de productos financieros más fácilmente similar a la portabilidad numérica en los servicios de telefonía móvil.

En la figura 2.3 se pueden ver los niveles de participación medido en colocaciones de los distintos bancos que componen la industria bancaria hasta junio 2019. Se puede notar que los bancos con mayor participación son el Banco de Crédito e Inversiones (BCI), Banco

Santander y el Banco de Chile.

**Cuadro III.1** Participación en colocaciones de los principales bancos del Sistema Bancario. Información consolidada, %.

<b>Institución</b>	<b>jun'18</b>	<b>sept'18</b>	<b>dic'18</b>	<b>mar'19</b>	<b>abr'19</b>	<b>may'19</b>	<b>jun'19</b>
Banco de Crédito e Inversiones	17,15	16,99	17,04	17,03	16,93	17,04	16,95
Banco Santander-Chile	17,39	17,52	16,96	16,90	17,06	16,94	16,83
Banco de Chile	16,01	16,06	15,85	15,74	15,81	15,92	15,89
Banco del Estado de Chile	14,07	14,05	13,76	13,69	13,72	13,63	13,75
Scotiabank Chile	6,71	12,62	12,78	12,84	12,89	12,87	12,84
Itaú Corpbanca	12,63	12,43	12,12	12,12	11,97	12,00	12,02
Banco Bice	3,02	3,05	3,09	3,18	3,09	3,06	3,06
Banco Security	3,02	3,03	2,99	2,96	2,97	2,93	2,97
Banco Falabella	1,01	0,99	2,03	2,02	2,03	2,02	2,01
Banco Consorcio	1,35	1,43	1,48	1,54	1,53	1,56	1,60
Banco Bilbao Vizcaya Argentaria, Chile	5,89	---	---	---	---	---	---
<b>Sistema Bancario</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>

Figura 2.3: Participación en colocaciones de los principales bancos del Sistema Bancario. Fuente: Informe Panorama Bancario Trimestral Junio 2019 de la CMF

# Capítulo 3

## Descripción del proyecto y justificación

La solicitud del trabajo de memoria surge inicialmente a partir de la apreciación del director de CX sobre la irrupción del *data analytics* en las empresas ligadas a la investigación de mercado y como cada una lo ha ido adoptando con mayor fuerza en los últimos años. Actualmente no basta con sólo ofrecer los medios para la realización de encuestas masivas, ya que cada vez surgen más recursos que facilitan esta tarea y son más accesibles. Es por esto que las empresas de investigación de mercado se están enfrentando al desafío de aportar valor a sus clientes mediante mejores análisis de los resultados.

“En el mundo global y competitivo actual, los clientes tienen más expectativas, más posibilidades de elección y son menos leales a las marcas” [4]. Se puede observar que los consumidores de las distintas industrias poseen mayor variedad de opciones y herramientas para escoger el servicio que se adapte mejor a sus necesidades y gustos específicos, por lo que adquiere mayor importancia para las organizaciones recoger las opiniones de los clientes, y así tomar acción en base al conocimiento de sus experiencias, necesidades y evaluaciones. De esta labor se encarga Ipsos para varios de los bancos más relevantes en Chile y otras industrias.

Los programas de Ipsos en los que invierten los bancos se ven mayoritariamente representados en los de medición de experiencia cliente, que se basa en medir los diferentes puntos de contacto y líneas de negocio de cada banco. Estos programas son utilizados como uno de los principales insumos para la toma de decisiones estratégicas: en una primera instancia estos programas tienden a soportar parte importante del pago de incentivos de los Bancos, poniendo el KPI de experiencia de cliente dentro de su *balanced scorecard*, impactando financieramente tanto a ejecutivos como gerentes. En segunda instancia, se reportan los resultados bimensuales cuyo objetivo es dirigir la estrategia y modelos de atención de clientes. Por último, de manera semestral, se presentan al directorio los focos de mejora del modelo de atención del banco en particular, donde el insumo principal de la percepción de la experiencia de clientes es el mencionado programa de medición.

Uno de los principales segmentos monitoreados son los clientes Personas. Para este estudio se tomaran encuestas provenientes de un estudio tracking de carácter relacional que levanta más de 36 mil encuestas anuales. Este estudio se implementa mensualmente y desde ahí

se obtienen los principales drivers a trabajar de manera táctica (seguimiento de un kpi de performance mensual) y estratégica de un banco en particular, que representan cambios estructurales identificados a partir de modelos estadísticos calculados con dicha data.

En el presente año se han realizado cambios y ajustes a los modelos de medición del segmento personas, principalmente en beneficio de la reducción del tiempo del instrumento, mejorar las tasas de abandono del cuestionario y mejorar la experiencia al contestar la encuesta. Si bien, el esfuerzo logró el objetivo buscado, se cree que pueden existir mejoras al modelo que permitan reforzar el poder explicativo de este sin afectar el principio de parsimonia ni la experiencia del usuario.

Los cuestionarios que se realizan están hechos bajo un marco que recoge principalmente variables estructuradas limitadas a las preguntas que los analistas deciden o no incorporar, por lo que puede haber variables que no están siendo consideradas. El único espacio en las encuestas para salir de esos límites son las preguntas abiertas, en las que los usuarios se explayan y justifican sus evaluaciones.

La oportunidad es evaluar la capacidad explicativa del modelo basado en clientes del tipo personas con la contribución que podría hacer el análisis de la data no estructurada (entregada en las preguntas abiertas del cuestionario) en los modelos de caracterización de la lealtad de los clientes.

Es relevante para los distintos clientes de Ipsos poder determinar cuáles son las debilidades y fortalezas que apalancan la lealtad de sus usuarios. El nivel de satisfacción del cliente tras una compra o la utilización de un servicio depende de la medida en que los resultados del producto o servicio cumplan las expectativas del consumidor. La satisfacción del cliente ejerce una influencia primordial en el comportamiento de compra futuro. Los clientes satisfechos volverán a utilizar la marca y contarán a otros sus experiencias positivas. Los clientes insatisfechos cambiarán a la competencia y menospreciarán a la marca frente a otros clientes potenciales [5].

“El beneficio principal del enfoque en el cliente y de la consecución de altos niveles de satisfacción reside en conseguir un alto nivel de lealtad. (. . . ) Cuando un cliente recomienda un producto o servicio a otros, significa que tiene la máxima confianza en el valor creado y entregado por la marca o compañía que recomienda” [4]. Por lo que Ipsos, en la gran mayoría de sus encuestas incluye una pregunta relacionada a con que seguridad el usuario recomendaría a otros cercanos el servicio. Adicionalmente, se les pregunta por qué, es ahí en donde se puede explayar y comentar cuál o cuáles de todos los aspectos del servicio son los que influyen en su recomendación o detracción.

En ese sentido, no solo es relevante entender la evaluación con respecto a la satisfacción y la lealtad, si no que tiene más valor determinar cuáles son los aspectos del servicio que afectaron esta evaluación, para que las empresas puedan tomar decisiones y gestionar en base a eso.

Poder desarrollar herramientas de *text mining* que permitan involucrar eficientemente los comentarios de los usuarios permitirá que Ipsos tenga una mayor capacidad para ofrecer a

sus clientes una mejor comprensión y adaptación continua a las nuevas necesidades de sus usuarios, ayudándolos a desarrollar relaciones de largo plazo y de esta forma, como asegura la literatura mejorar la rentabilidad de la empresa.

*Customer Experience* anualmente realiza alrededor de 3 millones de encuestas en donde suele existir al menos una pregunta abierta que aporta la opinión libre del usuario sobre la empresa con respecto a algún tópico y permiten profundizar en los detalles que motivan sus evaluaciones cuantitativas. Hasta el día de hoy, dado el volumen con el que trabajan actualmente, estos comentarios se suelen clasificar manualmente por un equipo externo de personas contratadas exclusivamente para tal labor. Recurrir a estos equipos es costoso, por ejemplo, un programa de 100.000 encuestas tiene un costo de contratación de personal para Ipsos de aproximadamente 9 millones de pesos.

Cabe destacar que, en el último tiempo, la industria bancaria en Chile ha enfrentado una serie de desafíos producto de la pandemia y el estallido social. En esa línea, surge la necesidad de que Ipsos tenga herramientas eficientes que se adapten a la realidad cambiante de los clientes y su entorno. En ese sentido, el trabajo estará dirigido a crear una herramienta dinámica, que no solo sirva para este estudio, sino que se adapte a las circunstancias y permita explorar en las opiniones de las encuestas mensuales que realiza el banco, qué es lo que más se comenta en ese momento.

Según estos antecedentes la oportunidad presentada puede resumirse en que el área de *Customer Experience* de Ipsos necesita diseñar un modelo que permite entender qué factores apalancan la recomendación integrando de forma óptima las opiniones abiertas de los consumidores.

Por último, el área de *Analytics* se vería beneficiada al tener una metodología más eficiente, eficaz y consistente, que permitirá mejorar el valor que tienen sus entregables al banco en cuestión u extenderlo a otros clientes y en el futuro replicarlo a los demás segmentos a un costo menor que el que implica hoy. Hoy en día el procedimiento de generar la clasificación de comentarios para un programa nuevo tarda aproximadamente 2 semanas de trabajo full time de un analista, se estima que con la nueva metodología este proceso no debería tardar más de tres días, lo que se traduce en mayor capacidad de procesamiento de encuestas y por ende, en una mayor rentabilidad.

En un año Ipsos recoge aproximadamente 24 programas de encuestas que requieren el servicio de *text analytics*. Esto se traduce en aproximadamente 360.000 comentarios que procesar. En el caso de recurrir a un equipo externo, el proceso tendrá un costo de \$ 32.400.000.

Con la metodología que se pretende implementar para este trabajo, y que tiene como objetivo ser replicable en el futuro para otros programas o rubros, Ipsos podría procesar las solicitudes de un año en 13 semanas por solo un analista. Actualmente, realizar estos análisis -una vez al año- genera a la empresa un ingreso de 84.000 UF. Con el nuevo procedimiento, se estima que los mismos análisis puedan ser realizado en 4 o 5 meses, duplicando la eficiencia del trabajo y abriendo la posibilidad de obtener este ingreso dos veces al año.

# Capítulo 4

## Objetivos y alcance

### 4.1. Objetivo General

Diseñar y desarrollar un modelo que permita determinar y analizar los factores que apalancan la lealtad de clientes bancarios, integrando preguntas abiertas de las encuestas de experiencia cliente.

### 4.2. Objetivos específicos

1. Desarrollar una metodología para clasificar en tópicos los comentarios de texto libre relacionadas a la sección de recomendación (NPS) de las encuestas
2. Identificar tópicos que se relacionan con un impacto positivo o negativo en el NPS
3. Caracterizar la satisfacción y la recomendación mediante la construcción de diferentes modelos que incorporen los tópicos y las evaluaciones de satisfacción efectuadas por los clientes mensualmente
4. Identificar drivers de la recomendación, cuantificar el impacto que tiene cada uno de ellos para los distintos tipos de clientes
5. Establecer recomendaciones de aplicaciones del modelo a futuro y sugerencias de gestión en base a los resultados

### 4.3. Alcances

El desarrollo del trabajo se limitará a abarcar sólo el banco de estudio y se profundizará el análisis sobre las preguntas abiertas e indicadores sobre satisfacción y recomendación generales, es decir, los demás indicadores serán utilizados sólo para encontrar relaciones que afecten los mencionados anteriormente y las demás preguntas abiertas que pudieran presentarse no serán consideradas en el estudio.

Se limitará a trabajar con las encuestas realizadas entre julio y octubre del 2020 que forman parte de una base de análisis de clientes cuentacorrentistas personas.

No se encuentra en el plan de trabajo desarrollar una plataforma o software de visualización de los resultados para Ipsos. Tampoco se propondrán nuevas métricas de satisfacción o lealtad.

# Capítulo 5

## Marco conceptual

### 5.1. CRISP-DM

*Cross-Industry Standard Process for Data Mining* (CRISP-DM) es una metodología para el proceso de descubrir conocimiento útil a partir de los datos. Lo que diferencia a esta metodología de otras como KDD (*Knowledge Discovery in Databases*) es que incluye la comprensión del negocio como punto inicial del proceso, para poder determinar los objetivos de acuerdo con los antecedentes y contexto en el que se desarrolla el trabajo. Además, las relaciones entre los pasos del proceso no son fijas, ya que pueden existir entre cualquier etapa dependiendo de los objetivos, antecedentes, y el interés del usuario sobre los datos. Bajo esta metodología el ciclo de un proyecto de minería de datos consiste en seis fases, expuesta en la figura [6]:

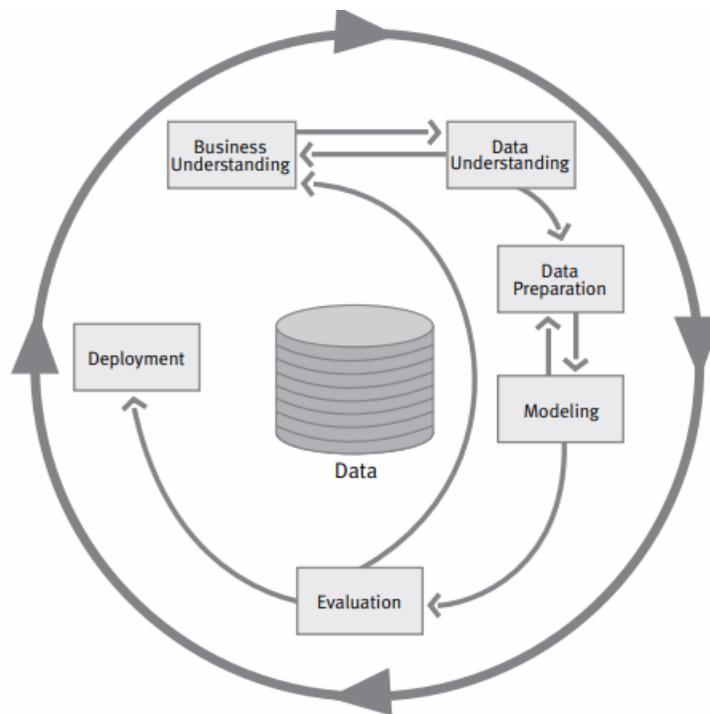


Figura 5.1: Metodología CRISP-DM [6]

## 5.2. Minería de textos

La minería de texto se enfoca en el descubrimiento de patrones interesantes y nuevos conocimientos en un conjunto de textos, es decir, su objetivo es descubrir cosas tales como tendencias, desviaciones y asociaciones entre “gran” cantidad de información textual [7].

Este proceso usualmente consta de una serie de cinco pasos[8]:

- **Adquisición de la data:** Obtener el corpus que será minado en busca de conocimiento nuevo. Para esto se pueden usar API's (*Application Programming Interface*) o *Web Crawlers* en caso de que los textos se encuentren en la web como *reviews* de algún producto en el *e-commerce*. En otros casos, los datos pueden ser extraídos manualmente o generados automáticamente en las bases de datos del negocio[8].
- **Preprocesamiento:** Utilizar herramientas comunes de Procesamiento de Lenguaje Natural asociado al análisis léxico. Algunas de las más comunes son[8]:
  - Tokenización: Separar el texto en una lista de palabras independientes.
  - Stemming: Proceso heurístico que lleva las palabras a su raíz, por ejemplo, las palabras “bibliotecas” , “bibliotecario” se transforman en “bibliotec”
  - Lematización: Es similar al “stemming” pero incorpora análisis morfológico de cada palabra, es decir, las lleva a como las encontraríamos en un diccionario tradicional: singular para sustantivos, masculino singular para adjetivos, infinitivo para verbos.
  - Remover Stopwords: Actividad de remover palabras que son usadas para estructurar el lenguaje que no contribuyen a su contenido. Ejemplos son los artículos, conectores y pronombres.
  - Segmentación de oraciones: Separar los párrafos en oraciones. Este paso posee sus dificultades ya que muchas veces los puntos no son usados para separar las frases, si no que en números o abreviaciones.
  - Part-of-Speech (POS) tagging: Es el paso de etiquetar cada palabra de una oración con su función en el lenguaje: adjetivo, sustantivo, verbo, adverbio y preposición.
- **Minado de textos:** Esta es una de las etapas más amplias, existen distintas técnicas con distintos enfoques, pero lo que caracteriza esta etapa es el descubrimiento de conocimiento, el cual se puede dar a través de detección de patrones, representaciones vectoriales, modelos de aprendizaje supervisado o no supervisado, entre otros. La elección del modelo dependerá del objetivo del investigador[9].
- **Interpretación de resultados:** En esta etapa se da paso a interpretar y validar el conocimiento obtenido tras realizar el proceso, este paso puede concluir en el retroceso en el proceso, iterando en los pasos que sean necesarios para cumplir de mejor forma los objetivos[9].
- **Visualización:** Etapa en que se crean formas de poder visualizar los resultados de acuerdo a las necesidades del cliente, esto puede ir desde un reporte a una aplicación.

### 5.2.1. Enfoques

Hay dos tipos de enfoques establecidos para llevar a cabo un proceso de minado de textos[8]:

- **No supervisado basado en diccionarios léxico-gráficos:** El proceso se asienta en reglas y heurísticas obtenidas del conocimiento del lenguaje
- **Supervisado basado en *machine learning*:** Se crean algoritmos que aprenden información subyacente de datos previamente catalogados, lo que permite clasificar datos nuevos sin etiquetar.

## 5.3. Algoritmos de Information Retrieval (IR)

Los algoritmos de retiro de información (IR) se encargan de resumir la información disponible en el conjunto de documentos para determinar elementos comunes y agruparlos. Algunos de ellos utilizan puntos de vista probabilísticos para determinar estos grupos, como el modelo de *Latent Dirichlet Allocation* (LDA).

- **LDA**

Se basa en la idea de que “cada documento puede ser descrito por una distribución de tópicos y cada tópico puede ser descrito por una distribución de palabras” [10]. Cada tópico es representado por una distribución multinomial por todas las palabras existentes en los documentos (vocabulario). Un diccionario para un tópico se genera en base a la mezcla entre las palabras del vocabulario y probabilidad de pertenecer al tópico [11]. Las notaciones para este modelo son:

$k = \text{Número de tópicos (fijo)}$

$V = \text{Tamaño del vocabulario}$

$M = \text{Número de documentos}$

$N = \text{Número de palabras en cada documento}$

$w = \text{Una palabra en un documento. Representado como un vector de tamaño } V$

$w = \text{Representa un documento vector de } w's, \text{ de tamaño } N$

$D = \text{Corpus, una colección de } M \text{ documentos}$

$z = \text{Un tópico de un set de } k \text{ tópicos. Un tópico es una distribución de palabras.}$

Los parámetros por calcular son:

$\alpha : \text{Relacionado con la distribución de los tópicos para todo el corpus.}$

$\theta : \text{Matriz donde } \theta(i, j) \text{ representa la probabilidad de que el documento } i \text{ pertenezca al tópico } j$

$\eta : \text{Relacionado a la distribución de palabras en cada tópico}$

$\beta : \text{Matriz donde } \beta(i, j) \text{ representa la probabilidad de que el tópico } i \text{ contenga la palabra } j$

Estamos interesados en definir  $P(\theta_{1:M}, z_{1:M}, \beta_{1:k}, |D; \alpha_{1:M}, \eta_{1:k})$  que en palabras simples nos dice que para un set de M documentos, cada uno con N palabras, donde cada una esta generada por un t3pico de un set de K t3picos, estoy buscando la probabilidad conjunta posterior de  $\theta, z, \beta$  dado D (el corpus que disponemos) y usando los par3metros  $\alpha$  y  $\eta$ .

En este modelo tanto  $\theta$  como  $\beta$  distribuyen Dirichlet. Existen muchas formas de resolver y optimizar los par3metros de este modelo, pero queda fuera del alcance de este trabajo de memoria [10].

## 5.4. 3rboles de decisi3n

Los 3rboles de decisi3n para la clasificaci3n y regresi3n (CART) fueron desarrollados por Breiman (1984). Son un tipo de modelo de *machine learning* no param3trico de aprendizaje supervisado. Sirven para predecir una variable continua o categorica, utilizando reglas binarias con las que se consigue repartir en cada nodo las observaciones en funci3n de las variables independientes.

Para la selecci3n de la variable que dividirá el 3rbol en dos hojas se utilizan métricas como la impureza de Gini o la entropía, para comparar los sets de datos resultantes de todas las posibilidades de divisi3n del nodo. [12]

La impureza de Gini indica la probabilidad de equivocarse al asignar una clase al azar a un dato seleccionado también aleatoriamente. La funci3n de partici3n de Gini se define como:

$$\phi(s, t) = I_G(t) - \sum_n^N P_n I_G(n) \tag{5.1}$$

Donde  $s$  es la partici3n que se está evaluando,  $N$  son los nodos,  $P_n$  es la proporci3n de muestras del nodo  $t$  que van al nodo hijo  $n$  e  $I_G(t)$  es la impureza de Gini, dado por:

$$I_G(t) = 1 - \sum_i^I t_i \tag{5.2}$$

En esta formula  $t_i$  es la proporci3n de la clase  $i$  en la muestra del nodo  $t$ . De esta forma, se busca el  $s^*$  para cada nodo tal que:

$$\phi(s, t) = \max_{s \in S} \phi(s, t) \tag{5.3}$$

con  $S$  conjunto de posibles particiones. [12]

## 5.5. Gradient tree boosting

La idea detr3s del boosting es generar m3ltiples modelos de predicci3n “débiles” secuencialmente, y que cada uno de estos tome los resultados del modelo anterior, para generar un modelo más “fuerte”, con mejor poder predictivo y mayor estabilidad en sus resultados. En este caso, para conseguir un modelo más fuerte, se emplea el algoritmo de optimizaci3n

## Gradient Descent.

En este algoritmo cada modelo es comparado con el anterior. Si un nuevo árbol tiene mejores resultados, entonces se toma como base para realizar modificaciones. En caso contrario, se regresa al mejor modelo anterior y se modifica ese de una manera diferente. Qué tan grandes son los ajustes de un modelo a otro es uno de los hiper parámetros que define el usuario previamente.

Este proceso se repite hasta llegar a un punto en el que la diferencia entre modelos consecutivos es insignificante, lo cual nos indica que hemos encontrado el mejor modelo posible, o cuando se llega al número de iteraciones máximas definido por el usuario.

La teoría detrás de esta intuición es la siguiente. Para un *dataset* con  $n$  observaciones y  $m$  variables.

$$\mathbf{D} = \{(x_i, y_i)\} (|\mathbf{D}| = n, x_i \in \mathbb{R}^m, y_i \in \mathbb{R}) \quad (5.4)$$

Un árbol de ensamblaje utiliza  $K$  funciones aditivas para predecir el output.

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i), f_k \in \mathcal{F} \quad (5.5)$$

Donde  $\mathcal{F} = \{f(x) = w_{q(x)}\} (q : \mathbb{R}^m \rightarrow T, w \in \mathbb{R}^m)$  es el espacio de árboles CART,  $q$  representa la estructura de cada árbol y  $T$  es el número de hojas en el árbol.

A diferencia de los árboles de decisión tradicionales, cada CART contiene un puntaje en cada hoja que se representará como  $w_i$ , en donde  $i$  representa una hoja del árbol.

Para determinar qué funciones utilizamos en el modelo, el algoritmo minimiza la siguiente función objetivo.

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$$

donde  $\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$

Donde  $l$  es una función de pérdida que mide la diferencia entre la predicción  $\hat{y}_i$  y el objetivo  $y$ ,  $\Omega$  penaliza la complejidad del modelo y los demás términos ayudan a suavizar las funciones y evitar el sobreajuste.

Sea  $\hat{y}_i$  la predicción de la observación  $i$  en la iteración  $t$ , se necesita sumar  $f_t$  para minimizar el siguiente objetivo.

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{t-1}) + f_t(x_i) + \Omega(f_t) \quad (5.6)$$

Esto quiere decir que agregamos la función  $f_t$  que más mejora el modelo en  $t$ . [13]

## 5.6. XGBoost

XGBoost (Extreme Gradient Boosting) es uno de los algoritmos supervisados de machine learning más utilizados en la actualidad. Utiliza el principio detallado en el punto anterior.

La función que optimiza en cada paso  $t$  toma la forma de una expansión de segundo orden del teorema de Taylor. [14]

$$obj^t = \sum_{i=1}^n \left[ m_i f_t(p_i) + \frac{1}{2} c_i f_t^2(p_i) \right] + R(f_t) \quad (5.7)$$

En donde  $m_i$  y  $c_i$  son inputs del modelo. Por otro lado, la regularización juega un papel importante a la hora de definir la complejidad del modelo. La definición del árbol  $f_t(p_i)$  es la siguiente.

$$f_t(p) = w_{q(p)}, w \in R^L, q : R^d \rightarrow \{1, 2, 3, 4, \dots, L\} \quad (5.8)$$

En donde  $w$  representa un vector de puntajes de las hojas,  $L$  es el número de hojas y  $q$  asigna hojas a los correspondientes datos. En XGBoost la complejidad está dada por:

$$R(f) = \alpha L + \frac{1}{2} \beta \sum_{j=1}^L w_j^2 \quad (5.9)$$

La estructura del árbol se establece calculando la regularización, las puntuaciones de las hojas y la función objetivo en cada nivel, ya que no es posible calcular simultáneamente todas las combinaciones de árboles.[14]

La ganancia se calcula en cada nivel a medida que una hoja se divide en una hoja izquierda y una hoja derecha, y la ganancia se calcula en la hoja actual con la regularización lograda en las posibles hojas adicionales. Si la ganancia no alcanza el valor de regularización adicional, entonces esa rama se abandona (concepto también llamado poda). Así es como XGBoost penetra profundamente en los árboles y clasifica los datos y, por lo tanto, se calcula la precisión y otros parámetros.[14]

## 5.7. Métricas de evaluación

### ■ Coherencia de tópicos

Métrica de evaluación para identificar la coherencia semántica de los modelos de tópicos. Basado en la idea de “las palabras que pertenecen a un solo concepto ocurrirán”, se define la coherencia del tópico como[15]:

$$C(t, V^t) = \sum_{m=2}^M \sum_{l=1}^{m-1} \log \frac{D(v_m^t, v_l^t) + \epsilon}{D(v_l^t)}$$

donde  $D(v)$  es el número de documentos con al menos un token tipo  $v$ ,  $D(v, v')$  es el número de documentos que contienen uno o más de tokens de tipo  $v$  y al menos un token de tipo  $v'$  y  $V^t = (v_1^t, v_2^t, \dots, v_M^t)$  es una lista de las  $M$  palabras más probables en un tópico  $t$ . [15].

- **Matriz de confusión**

Para el caso de los modelos para caracterizar la satisfacción y la lealtad se utilizarán las siguientes métricas de clasificación multiclase, en donde se debe construir la matriz de confusión  $M$  correspondiente al caso en el que las filas serán las clases predichas mientras que las columnas las clases reales. Entonces se define para la clase  $x^1$ :

- Verdadero positivo: Posición  $M(x,x)$  en la matriz
- Falso positivo: Suma de la columna  $x$  sin posición  $M(x,x)$
- Falso negativo: Suma de la fila  $x$  sin  $M(x,x)$

De esta forma se puede calcular:

- Accuracy: Porcentaje total de elementos clasificados correctamente

$$Accuracy = \frac{VerdaderoPositivo + VerdaderoNegativo}{Casos\ totales}$$

- Recall: Número de elementos identificados correctamente como positivos del total de positivos verdaderos

$$Recall = \frac{Verdadero\ Positivo}{VerdaderoPositivo + FalsoNegativo}$$

- Precisión: Número de elementos identificados correctamente como positivo de todos los elementos identificados como tal

$$Precisión = \frac{Verdadero\ Positivo}{VerdaderoPositivo + FalsosPositivo}$$

- Puntuación F1: Combina tanto la precisión como recall para sopesar las desventajas de usarlos por si solos

$$F1Score = 2 \frac{Precisión\ Recall}{Precisión + Recall}$$

## 5.8. NPS, Impacto y Criticidad

El *Net Promoter Score* es una medida muy utilizada en investigación de mercado y en particular en Ipsos. Es un índice que busca reflejar el nivel de lealtad de los clientes sobre algún servicio, producto o entidad. Para esto se les pregunta a los entrevistados en “¿Con qué probabilidad recomendaría “X” a un amigo o persona cercana en una escala del 1 al 7?” en donde los que contestan más alto se les considera más leales.

De esta forma surge el concepto de detractores y promotores, donde los primeros son quienes no recomiendan el objeto de estudio y los otros promueven el uso de este. La escala

<sup>1</sup> Fuente: <https://sitiobigdata.com/2019/01/19/machine-learning-metrica-clasificacion-parte-3/#>

para ajustar a los clientes en cada uno de estos grupos varía, en el caso particular del caso de estudio, quienes contesten con nota 6 - 7 se les considera promotores, 1 -4 son detractores y 5 se les considera consumidores pasivos o neutros.

El NPS se calcula restando el porcentaje de promotores menos el porcentaje de detractores de la misma para alguna categoría o globalmente, por lo que un NPS de -100 se traduce en que solo hay detractores y al contrario, 100 es porque posee solo promotores.

Por otro lado, tenemos el impacto, métrica que usa Ipsos para determinar que tanto afecta a un cliente la poca satisfacción o lealtad hacia cierta categoría, comparando el NPS global con un NPS específico. Por ejemplo, se tiene el NPS de los clientes que a través de los comentarios determinaron conflictos con la app, si restamos el NPS global menos el NPS del tópico da como resultado el impacto, es decir, en cuanto afecta el NPS global que los clientes tengan problemas con la app.

Por último, se encuentra la Criticidad, que toma los dos indicadores anteriores para evaluar la relevancia que tiene cierto tópico. Se calcula multiplicando la ocurrencia (porcentaje de veces que un tópico es mencionado) y el impacto. La intuición detrás de este indicador nos permite por ejemplo determinar si un tópico está ocurriendo mucho, pero tiene bajo impacto, se obtendrá un Criticidad bajo por lo que no se le debe prestar mayor atención versus un tópico que ocurre mucho y además esta afectando mucho el NPS global.

## 5.9. Imputación múltiple

Este método surge como una solución para tratar los *missing values* propuesta por Rubin (1987). Consiste en métodos de simulación de Monte Carlo en los que se sustituyen los datos faltantes a partir de un número  $m$  -mayor a 1- de simulaciones, en general se recomienda entre 3 y 10.[16]

El análisis inicia con los datos observados e incompletos. La imputación múltiple crea varias versiones completas de los datos, reemplazando los valores faltantes con variables factibles, que se obtienen de una distribución modelada específicamente para cada valor faltante. Esto se realiza  $m$  veces, en el que cada *datasets* posee datos imputados distintos.[17]

Luego se estiman los parámetros de interés para cada uno de los *datasets* imputados, mediante la aplicación de métodos analíticos que se hubieran utilizado en el caso que los datos estuvieran completos.[17]

Finalmente, se agrupan las estimaciones para el parámetro  $m$  en una única estimación y se estima la varianza. La varianza combina la varianza dentro de las imputaciones y ente las imputaciones.[17]

Esta metodología asume los siguientes supuestos. (i) El patrón de datos faltantes es aleatorio (MAR), lo cual significa que la probabilidad de que existan datos omitidos en la variable X dependen de otras variables, pero no de X. (ii) Se requiere que el modelo utilizado para generar los datos imputados sea apropiado; es decir, que exista correlación alta entre la variable a imputar y el vector de covariable que se utilizará para modelar los datos que se

utilizarán como sustitutos. (iii) Finalmente, también es necesario que el modelo de análisis guarde relación con el que se utilizó para efectuar el procedimiento de imputación.[16] Bajo las condiciones adecuadas, las estimaciones agrupadas no serán sesgadas y tienen las propiedades estadísticas correctas.

# Capítulo 6

## Metodología

La metodología base a utilizar para este trabajo de memoria es la de CRISP-DM, explicada en el punto 5.1. Dentro de este también se llevarán a cabo las etapas del proceso de minería de texto detallados en 5.2. De esta forma los pasos a seguir son:

### 6.1. Entender el negocio

En esta fase se estudiará la empresa para la que se está trabajando con el fin de entender las necesidades que poseen y los objetivos que desean alcanzar. Por otro lado, se indagará en como llevan los procesos actuales para encontrar las debilidades y oportunidad de mejora.

Luego, se debe estudiar en que se encuentra actualmente la industria bancaria en la que esta inserta el banco de estudio para tener un contexto mucho más aterrizado de los desafíos a los que se enfrentan, entender los servicios generales que ofrecen a los clientes, como está compuesta la industria, entre otros.

Una vez comprendido a cabalidad cual es el problema y el objetivo por perseguir, se incluirá en esta fase el estudio del estado del arte con respecto a las distintas metodologías para poder llevar a cabo el trabajo de memoria.

### 6.2. Entender los datos

En esta fase se debe proceder con la adquisición de los datos, que vendrían siendo encuestas hechas entre julio y octubre del 2020, cada mes consta de aproximadamente 3.000 encuestas y consolidarlas en una base que cumpla con los requisitos de los modelos que se llevaran a cabo después. Luego se procede a entender cómo se obtuvieron estas encuestas: el o los canales por los que se llevó a cabo, en qué lugares, a qué perfil estaban dirigidas, tasas de respuestas, etc. Esto servirá para entender los sesgos que pueden tener y qué criterios se tuvieron en cuenta al hacer la encuesta de estudio.

Finalmente se realiza un análisis exploratorio de los datos que ya están estructurados para tener una noción inicial de las distribuciones que posee, entender relaciones entre variables, descubrir los primeros *insights* de las personas que contestaron y apreciaciones generales sobre el servicio y generar hipótesis a partir de esto.

La encuesta de estudio posee 40 preguntas por lo que se debe hacer una selección justificada de que variables se van a utilizar, cuáles serán dependientes e independientes. Además, se selecciona la pregunta abierta ligada a la recomendación para proceder a tratar con *text mining*.

### 6.3. Preparación de los datos

En esta fase se preparará la base de datos final por lo que se deben tratar tanto los indicadores numéricos como la data desestructurada, es decir, los comentarios. Estos últimos se tratan con las herramientas mencionadas en 5.2 en la etapa de preprocesamiento y es la parte más extensa del trabajo.

En el caso de los textos, para que la siguiente fase sea efectiva se debe poner mucho énfasis en el preprocesamiento. Primero se deben tomar las variables de preguntas abiertas que se van a analizar y eliminar todos aquellos comentarios vacíos o que no aportan información, existen casos de textos ilegibles o que simplemente no aportan, por ejemplo comentarios del tipo: “porque es un buen banco” .

Luego se continua con el preprocesamiento en dos pasos:

- **Limpieza:** Con librerías de Python se pueden efectuar procesos de limpieza como eliminar las mayúsculas, remover los caracteres no alfanuméricos, corregir palabras mal escritas – con ayuda de un diccionario creado manualmente acorde al caso- y remover doble espacios.
- **Transformación:** A priori las transformaciones que se pretenden utilizar son:
  - **Entidades Clave:** Crear una lista de palabras que separadas pierden el sentido que tienen, por ejemplo “Banco de Chile” , el algoritmo lo procesará como tres tokens distintos, por lo que se debe cambiar a algo del estilo “BancoChile” . La idea es luego reemplazar los valores por los correctos para que tengan sentido.
  - **Remover números:** A simple vista, los usuarios tienden a usar números para repetir la evaluación que mencionaron, esto no aporta a el análisis.
  - **Remover StopWords:** Existen muchas palabras que no son útiles en un proceso de semántica como los conectores, artículos, etc. Para esto existen diccionarios en español previamente hechos que se pueden utilizar y adaptar a la base de datos propia.
  - **Stemming y Lemmatización:** Se debe optar por alguna de las dos formas de reducir las palabras a su raíz como se detalla en el marco conceptual.

Con respecto a los datos numéricos se realizará una limpieza de datos eliminando outliers, datos incomprensibles y errores. En el caso de usar variables categóricas éstas deben estar expresada como *dummy's*. Por último, evaluar si es necesario realizar algún tipo de transformación o agrupación de las variables.

## 6.4. Modelamiento

En esta etapa se tratarán inicialmente los textos para que se transformen en data estructurada que pueda ser procesada. Se realizará gracias al paso de Minado de Textos (detallado en la Sección 5.2).

Más específicamente, para realizar la determinación de los tópicos se implementará el algoritmo LDA, en donde primero se debe determinar el valor de los parámetros del modelo -número de tópicos, alpha y beta- con la métrica de coherencia.

Una vez ajustado el modelo, se procederá con la fase de interpretación de los tópicos obtenidos ayudado con herramientas visuales como nubes de palabras.

Luego se continuará con la fase de visualización en el que se implementarán gráficos que permitirán apreciar a grandes rasgos los tópicos, como varían según si son positivos o negativo y su relación con el NPS.

Una vez conformes con los resultados del procesamiento de los textos se procederá a implementar el modelo XBoost (detallado en la Sección 5.6) para predecir la lealtad en base a las variables independientes seleccionadas y con los resultados determinar cuáles son las variables que la determinan (drivers). Se utilizará un 80 % de los datos para entrenar y un 20 % de testeo.

## 6.5. Evaluación

Se evaluarán los modelos resultantes de la sección anterior en base a las características del negocio y las distintas métricas para evaluar el desempeño de los modelos: *accuracy*, *weighted F1 Score*, *macro F1 Score* y AUC (detalladas en Sección 5.7). Si no es lo que se espera se debe iterar en los pasos anteriores que sean necesarios. Cuando se concluye cual es el mejor modelo se procede a interpretar los resultados y analizar la información.

## 6.6. Despliegue

Se determinará la forma más adecuada para que los clientes visualicen los resultados en una presentación o informe, además se realizará una infografía del paso a paso para que Ipsos pueda replicar el proceso a futuro, tanto para el mismo tipo de encuesta como para otros.

# Capítulo 7

## Desarrollo metodológico

En esta sección se presentarán los resultados de experimentos y discusiones obtenidos en cada fase de la metodología CRISP-DM, excluyendo la fase inicial de “Entender el negocio” ya que fue presentado en la Sección 2 de Antecedentes Generales.

### 7.1. Entender los datos

Los datos provienen de las encuestas del programa de medición de experiencia cliente implementada por Ipsos para los clientes del banco de estudio entre julio y octubre de 2020. En general, los bancos entregan una base de datos con los contactos de una muestra amplia de sus clientes, luego Ipsos se encarga de limpiar esa base, reparar los datos que se puedan reparar y contactar aleatoriamente vía teléfono hasta conseguir las cuotas exigidas por el estudio en particular. En este contacto telefónico primero se corrobora que el cliente tenga antigüedad de una cierta cantidad de meses en la cuenta corriente del banco en cuestión, que no posean relación con personas que trabajen en bancos, financieras u empresas de investigación de mercado, y que permita grabar la entrevista. Si cumple con todos esos requisitos se procede con la entrevista telefónica que tiende a durar unos 8 minutos y no existen incentivos (bonificación o premio) para contestarla.

Tabla 7.1: Ficha metodológica cuestionario

Aspecto	Descripción
<b>Público Objetivo</b>	Hombres y mujeres mayores de 18 años que posean cuenta corriente
<b>Ámbito geográfico</b>	Aleatorio
<b>Diseño de Investigación</b>	Investigación Cuantitativa
<b>Técnica de recolección</b>	Encuestas telefónicas (CATI),
<b>Método de muestreo</b>	Aleatorio
<b>Número de encuestas</b>	Se obtienen 13.429 encuestas efectivas
<b>Fecha de Medición</b>	Mes de julio a octubre del año 2020

Fuente: Elaboración propia con datos del caso de estudio

La base entregada posee 13.429 observaciones y 311 columnas que serán posteriormente reducidas en la etapa de preprocesamiento. La encuesta está compuesta por 40 preguntas.

Existen tres tipos de preguntas en el cuestionario:

- **Perfil:** Clasifican a los encuestados para poder generar y comparar resultados por segmento. Incluye la edad, género y región de residencia.
- **Hábitos:** Indagan sobre lo que la gente hace, en el caso de esta encuesta, suelen haber preguntas sobre la frecuencia de uso de los distintos servicios.
- **Actitudes:** Intenta averiguar lo que las personas piensan u opinan sobre algo o alguien. En el caso de esta encuesta, existen muchas preguntas sobre calificar los distintos aspectos del servicio.

Se tienen aproximadamente 3.300 datos para cada mes entre julio y octubre de 2020. En la Figura 7.1 se puede notar que las encuestas son predominantemente de la Región Metropolitana con 8.219 datos, mientras que las demás regiones suman en total 5.209.

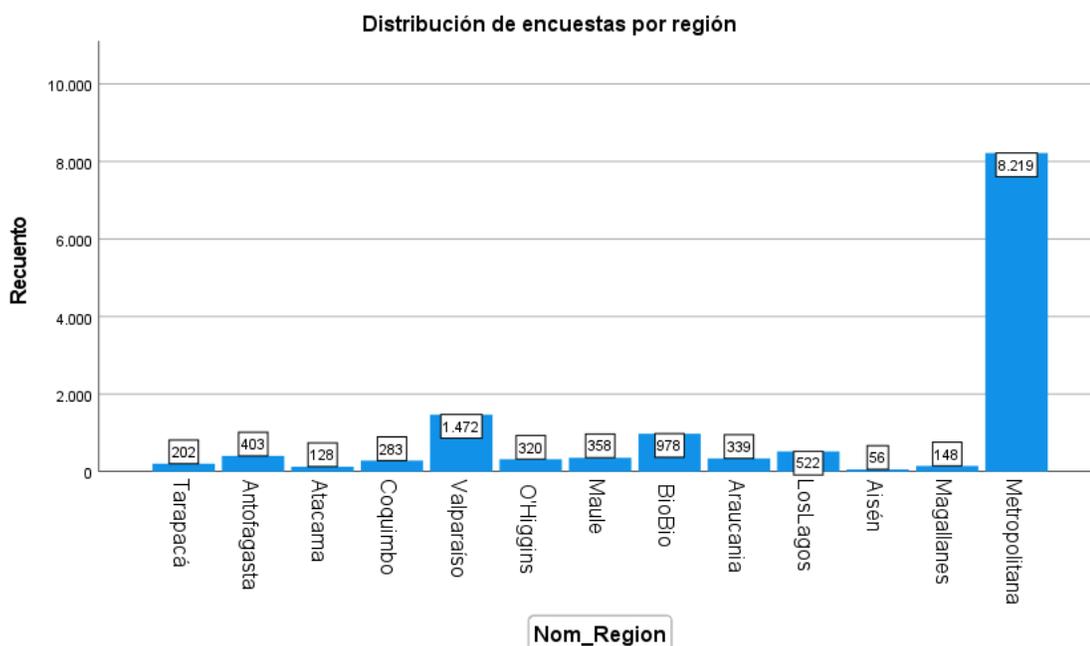


Figura 7.1: Recuento de encuestas por región. Fuente: Elaboración propia con datos de las encuestas

En cuanto a las edades de los encuestados, en el siguiente gráfico se puede ver que la mayor densidad se encuentra entre los 25 y 50 años, perdiendo densidad hacia las edades mayores.

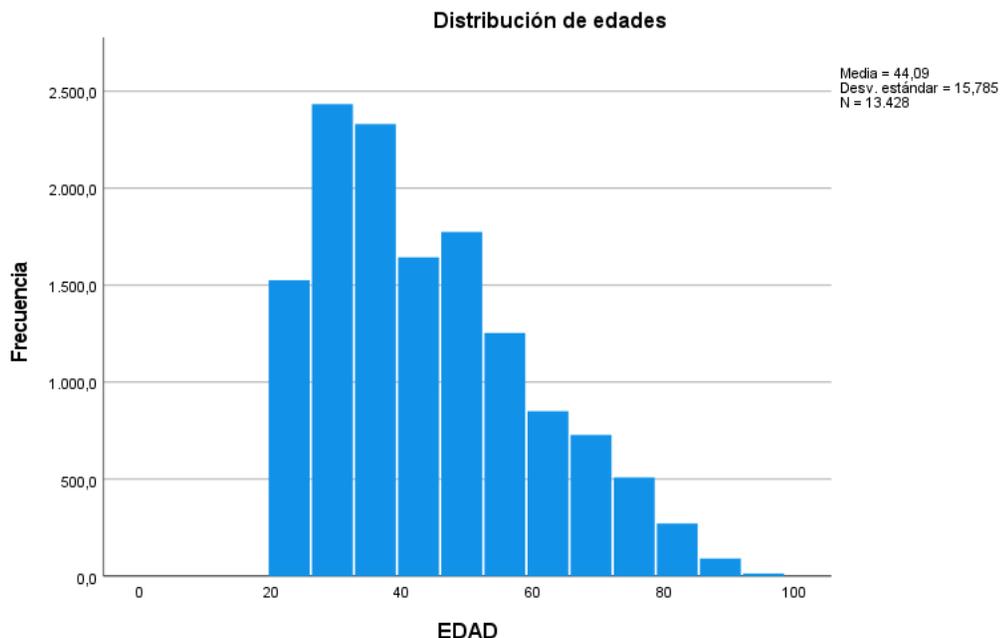


Figura 7.2: Edades encuestados año a año. Fuente: Elaboración propia con datos de las encuestas

En cuanto al sexo en la Figura 7.3 podemos notar que es equitativa la proporción entre mujeres y hombres por lo que en este sentido las encuestas no parecen estar sesgadas.

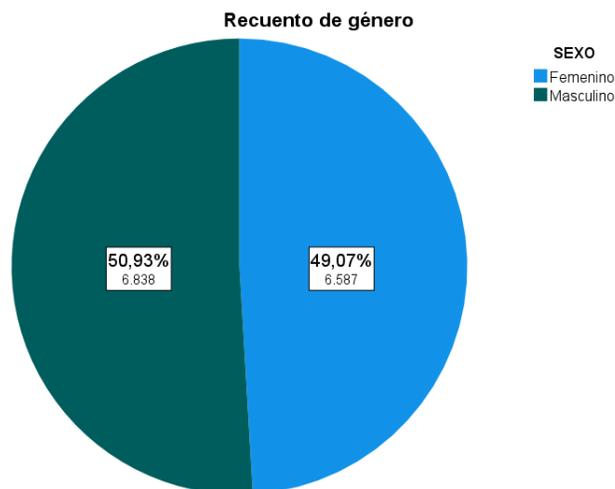


Figura 7.3: Distribución de género en encuestas. Fuente: Elaboración propia con datos de las encuestas

El cuestionario en estudio posee 47 preguntas divididas en 6 secciones fijas: Recomendación, Ejecutivo, Canal Presencial, Canales Remotos, Satisfacción Global y Problemas, y 3 secciones, en las que cada encuestado solo contesta una de estas: Inversiones, Financiamiento y Beneficios. De la muestra estudiada, fueron contestadas por el 12% , 9% y 79% de los encuestados respectivamente.

Las variables más relevantes en este estudio son la de “Recomendación” , que se encuentra al comienzo del cuestionario y da respuesta a la pregunta “*Un amigo o un familiar está*

interesado en contratar los servicios del banco considerando el servicio que le entrega y en base a una escala de 1 a 7 en la que 1 significa “ Definitivamente no lo recomendaría” y 7 significa “ Definitivamente lo recomendaría” , ¿Con qué nota recomendaría Usted al Banco” , la de “Satisfacción Global” (que por simplicidad llamaremos “Satisfacción” ) que responde a “Pensando en el servicio general otorgado por el banco y luego de evaluar todo lo anterior ¿qué tan satisfecho se encuentra usted? considere una escala de 1 a 7 donde 1 es muy insatisfecho y 7 muy satisfecho” .

Se puede ver en la Figura 7.4 las distribuciones por mes de las variables “Recomendación” y “Satisfacción”, se puede notar que no existen diferencias entre los meses y las evaluaciones se concentran en las notas más altas mientras que son pocas las observaciones con evaluaciones bajas. Esto indica que para los modelos predictivos se tendrá que utilizar alguna herramienta de *subsampling* u *oversampling* para balancear las evaluaciones.

También se puede notar que todos los encuestados respondieron la evaluación de recomendación mientras que para la Satisfacción el 0,1 % contestó “No sé o no quiero responder”

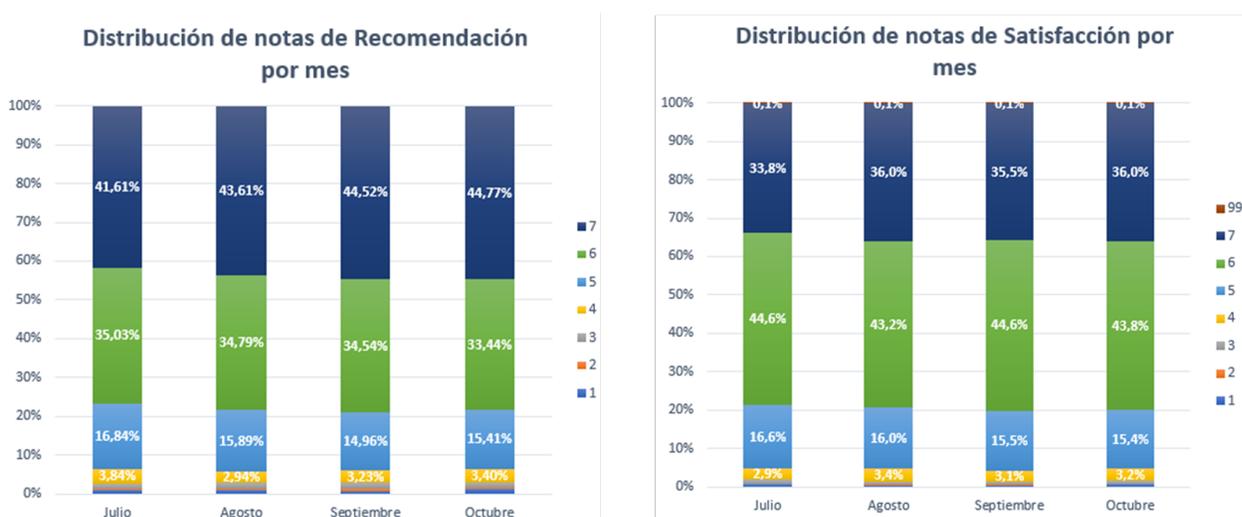


Figura 7.4: Distribuciones de recomendación y satisfacción global por mes.

Fuente: Elaboración propia con datos de las encuestas

También se puede desprender de la Figura 7.4 que en el caso de la recomendación es mayor el porcentaje de nota 7 y en el caso de la satisfacción es mayor el porcentaje de la nota 6. Esta disminución se puede deber a las ubicaciones de las preguntas en el cuestionario, la recomendación es la primera pregunta mientras que la satisfacción se mide luego de haber evaluado uno a uno los puntos de contacto con el banco.

Analizando las variables en que el cliente evalúa cada aspecto del servicio sumado a las evaluaciones de Recomendación y Satisfacción Global, sin tomar en cuenta los nulos ni los “No sabe / No responde” , se puede notar (ver Figura 7.2) que el ejecutivo, los problemas, sucursal, la app y la web poseen la misma evaluación promedio: 6.4, mayor que el promedio de Satisfacción Global y solo difieren entre ellas ligeramente en la desviación estándar. Mientras que los módulos Beneficios, Inversiones y Financiamiento poseen promedios más bajos: 5.6, 6.0 y 5.6 respectivamente.

Tabla 7.2: Distribución en evaluaciones de clientes

	Recomendación	Satisfacción	Ejecutivo	Problemas	Sucursal	App	Web	Beneficios	Inversiones	Financiamiento
<b>count</b>	13428	13428	10368	993	5329	8363	12535	10486	1093	854
<b>mean</b>	6,1	6,2	7,3	9,1	6,9	6,8	6,9	32,1	8,7	7,5
<b>std</b>	1,1	3,1	9,0	17,1	6,7	5,5	6,9	42,2	15,7	13,5
<b>min</b>	1	1	1	1	1	1	1	1	1	1
<b>25 %</b>	6	6	6	5	6	6	6	5	5	5
<b>50 %</b>	6	6	7	6	7	7	7	7	6	6
<b>75 %</b>	7	7	7	7	7	7	7	99	7	7
<b>max</b>	7	99	97	97	99	99	99	99	99	99

Fuente: Elaboración propia con datos de las encuestas

Una conclusión rápida sería que es necesario mejorar los aspectos del servicio con peores evaluaciones, pero no sabemos que incidencia tienen estos aspectos en la evaluación final del cliente, a modo de ejemplo, puede un cliente evaluar muy mal la página web, pero esto puede ser irrelevante a la hora de evaluar al banco en general ya que son otros los aspectos que le importan.

De la Figura 7.2 también se desprende la cantidad de datos nulos o “No sabe/No responde”, pero se puede ver en mayor detalle a continuación:

Tabla 7.3: Cantidad y porcentaje de nulos y “NS/NR”

	Nulos	Porcentaje nulos	NS/NR	Porcentaje NS/NR	Válidos
<b>Financiamiento</b>	12.574	94 %	18	0,1 %	836
<b>Problemas</b>	12.435	93 %	36	0,3 %	957
<b>Inversiones</b>	12.335	92 %	32	0,2 %	1.061
<b>Sucursal</b>	8.099	60 %	27	0,2 %	5.302
<b>Beneficios</b>	2.942	38 %	2.979	22,2 %	7.507
<b>App</b>	5.065	23 %	29	0,2 %	8.334
<b>Ejecutivo</b>	3.060	22 %	101	0,8 %	10.267
<b>Web</b>	893	7 %	8	0,1 %	12.527
<b>Satisfacción</b>	-	0 %	14	0,1 %	13.414
<b>Recomendación</b>	-	0 %	-	0,0 %	13.428

Fuente: Elaboración propia con datos de las encuestas

Los nulos se deben a distintos motivos. En el caso de Financiamiento, Inversiones y Beneficios, ocurren ya que son preguntas excluyentes. Por otro lado, para los puntos de contacto, los nulos se deben a que previamente se le pregunta al entrevistado si es que conoce o ha utilizado el punto de contacto, si es que la respuesta es no, se procede a omitir la pregunta.

Finalmente están las variables escritas, existen 9 preguntas abiertas pero para efectos de este estudio solo se utilizará la pregunta abierta sobre aspectos positivos y negativos del banco en cuestión.

## 7.2. Preparación de los datos

Dado que la encuesta es extensa se realizó primero una selección de variables que permitan acercarse al objetivo del trabajo. Partiendo por eliminar las variables irrelevantes para el estudio como las relacionadas con el proceso de encuesta (hora, entrevistador, fecha, etc.) y datos personales del encuestado o del ejecutivo.

Dado que el enfoque de lealtad está medido por el NPS que corresponde a un enfoque actitudinal de la lealtad de los clientes, se procede a considerar solo las variables actitudinales con respecto a los distintas áreas del servicio para determinar el peso de cada uno en el modelo global que se desarrollará más adelante, es decir, se consideran solo las variables correspondientes a la evaluación del usuario sobre la satisfacción con respecto a cada área de servicio descritas anteriormente en la Tabla 7.2.

Previamente en la Tabla 7.3 se puede apreciar que la cantidad de nulos para Financiamiento, Problemas e Inversiones supera el 90 % por lo que se deciden eliminar estas observaciones. Con respecto a las demás variables con nulos se decide no imputarlos con valores sintéticos ya que se desprenden de una respuesta anterior, por lo que no es azaroso que sean nulos.

Por el contrario, no se puede decir lo mismo para las respuestas “NS/NR” , en este caso se desconoce el por qué el entrevistado no da información. Sin considerar la pregunta de “Beneficios”, este tipo de datos conforma solo el 2 % del total de la base por lo que se puede simplemente eliminar estos datos o imputar con la media y no afectar la muestra. Dado que la nota de los datos observados en estas secciones no distribuye normal, se decide por eliminar estas observaciones que corresponden a 289 datos, quedando finalmente 13.139 observaciones.

En el caso de la evaluación de los beneficios, los “NS/NR” corresponden al 22 % del total de la base por lo que se empleará el método de imputación múltiple que fue el método más efectivo en el estudio de Estrategias para el tratamiento de datos faltantes ("missing data") en estudios con datos longitudinales. [17]. Este método se encuentra detallado en la Sección 5.9.

Luego se procedió con el preprocesamiento de los textos que incluye dos partes: limpieza y transformación, detalladas en la Sección 5.2. Para este trabajo se decidió utilizar solo las preguntas abiertas correspondientes a la evaluación de cosas positivas y negativas asociadas a la recomendación, dado que las otras preguntas son muy específicas y no aportan al objetivo del estudio.

El cuestionario comienza con la pregunta sobre recomendación para luego, según esta respuesta, consultar al encuestado:

- Si la recomendación está entre 2 y 7 se pregunta “¿Qué aspectos positivos destaca de su experiencia con el Banco?”
- Si la recomendación está entre 1 y 6 se pregunta “¿Qué aspectos negativos destaca de su experiencia con el Banco?”

La limpieza de los textos se implementó con varias funciones implementadas en Python y

diccionarios elaborados en Excel <sup>1</sup>. Estas funciones se encuentran detalladas a continuación:

- **Minúsculas:** Dado que los algoritmos de *text mining* consideran distintas las palabras si están escritas en mayúsculas o minúsculas, se procede a normalizar todo a minúsculas.
- **Remover caracteres no Alpha-numéricos y doble espacios:** Se removieron todos los caracteres que no son números ni letras, en el *dataset* de estudio los caracteres de este tipo son los puntos, comas y paréntesis. Además, se removieron los espacios excedentes de modo que las palabras solo estén separadas por un espacio.
- **Palabras mal escritas:** Al diccionario utilizado en [15] se le agregaron los casos identificados más comunes en el *dataset* de estudio, por ejemplo la palabra “problemas” se encontraba escrita de forma errónea en varias ocasiones como “porblemas” .
- **Remover números:** Se identificaron números tanto en los textos positivos como en los negativos, pero al observar detenidamente los casos, tienden a tratarse sobre referencias a la nota con la que evaluaron previamente por lo que no aportan información y fueron removidas tanto su versión numérica como la versión escrita de los números.
- **Entidades clave:** Existen palabras que solo si están juntas adquieren el sentido semántico correcto, por lo que deben ser tratadas, de no ser así, el algoritmo los toma como palabras independientes. Al diccionario implementado en [15] se agregaron los casos propios del *dataset* como la palabra “ejecutivo de cuenta” que gracias al diccionario sería reemplazada por “ejecutivo” o la palabra “página web” se reemplaza por “página” .
- **Remover Stopwords:** Se eliminan las palabras que se suelen repetir mucho en el español pero que no aportan al significado de la oración como los conectores, artículos, pronombre, etc. Además, se agregaron algunas palabras que en este contexto no aportan valor como la palabra “banco” . Por último, palabras como “bueno” , “buen” , “mala” , etc. fueron removidos ya que los comentarios ya vienen clasificados como aspectos positivos o negativos, por lo que no aportan a los resultados considerando que se repiten bastante.
- **Estandarizar acentos:** Al observar los comentarios y algunos resultados preliminares del LDA se pudo notar que existen muchas faltas de ortografía correspondientes a el uso de acentos, por ejemplo, el algoritmo tomaba como palabras distintas “página” y “pagina” . Si bien existen otras palabras que son efectivamente distintas si llevan acento o no, como “sería” y “seria” , en este caso particular otorga un mejor resultado normalizar quitando los tildes que no hacerlo.

<sup>1</sup> Varias de las funciones y diccionarios de stopwords, palabras mal escritas y entidades clave fueron facilitados por Constanza Contreras utilizados en su memoria “Extracción de conocimiento desde los reclamos recibidos en el SERNAC” (12)

Tabla 7.4: Descripción de caracteres en comentarios positivos

	Stopwords	Puntuación	Hashtags	Números	N° palabras
<b>mean</b>	12,4	1,8	0	0	21,6
<b>std</b>	10,8	2,1	0	0,2	16,9
<b>min</b>	-	-	0	0	1
<b>25 %</b>	5	-	0	0	11
<b>50 %</b>	10	1	0	0	17
<b>75 %</b>	16	3	0	0	27
<b>max</b>	194	24	0	3	312

Fuente: Elaboración propia con datos de la empresa

Tabla 7.5: Descripción de caracteres en comentarios negativos

	Stopwords	Puntuación	Hashtags	Números	N° palabras
<b>mean</b>	0	0,9	0	0	14,6
<b>std</b>	1,7	1,9	0	0,3	23,2
<b>min</b>	0	0	0	0	1
<b>25 %</b>	0	0	0	0	1
<b>50 %</b>	0	0	0	0	4
<b>75 %</b>	0	1,0	0	0	20,0
<b>max</b>	162	27	0	0	333

Fuente: Elaboración propia con datos de la empresa

En las tablas previas se puede observar las características de los textos previo al preprocesamiento, tanto de los comentarios positivos (Tabla 7.4) y los comentarios negativos (Tabla 7.5). Se puede ver que no existen hashtags en ninguno de los dos tipos de comentarios. En ambos tipos el 75 % de las observaciones no presenta números. Con respecto a los signos de puntuación el 75 % de los comentarios posee a los más tres signos, pero existen *outliers* que se relacionan directamente con el largo de los textos. En la columna “N° Palabras” se puede ver el número de palabras por comentario y se observa que en general los comentarios tienen a ser reducidos. Existe un 25 % de comentarios con una mayor extensión (más de 20 palabras), estos comentarios *outliers* tienden a aportar más información y por lo mismo es difícil encasillarlo en solo un tema. No existe una tendencia clara que distinga a los comentarios negativos de los positivos.

Luego de aplicar las funciones de limpieza y transformación mencionadas anteriormente se eliminaron 162.754 *stopwords*, 13.114 signos de puntuación y 367 números. En la Figura 7.5 se puede ver un ejemplo de los resultados luego de la aplicación de las funciones anteriores.

Texto positivo	Texto positivo limpio	Texto negativo	Texto negativo limpio
la banca electrónica es muy buena, la atención de los ejecutivos y el personal en las sucursales es muy buena, tiene buenos productos comerciales por esos motivos.	banca electronica atencion ejecutivo personal sucurs productos comerciales motivos	algunos productos son caros como la mantención de la cuenta, las maquina de consulta no funciona bien cuando uno va al banco.	productos caros mantencion cuenta maquina consulta no funciona bien

Figura 7.5: Ejemplos de textos limpios. Fuente: Elaboración propia

Luego se procedió a evaluar como agrupar las palabras con significados similares como las distintas conjugaciones de un verbo o plurales y singulares. Para esto se utilizan algoritmos de *stemming* o *lemmatizacion* descritos en la Sección 5.2. En concordancia con lo planteado en [15] se descarta lematizar los textos ya que palabras como “cuenta” alusivo a cuenta bancaria se transforman a “contar” lo que hace difícil la interpretación a la hora de asignar los tópicos.

Se procedió con dos tipos de algoritmos de *stemming*: *Porter* y *Snowball*, al evaluar las palabras más frecuentes el resultado fue el siguiente:

Sin Stemming		Porter		Snowball	
ejecutivo	550	ejecutivo	737	ejecutiv	935
atención	391	problema	640	buen	750
problema	352	crédito	448	problem	640
ejecutiva	338	servicio	444	pagin	531
crédito	331	atención	391	credit	448
servicio	311	client	388	servici	444
problemas	288	ejecutiva	355	client	397
tema	284	buena	338	atencion	392
buena	279	tema	313	necesit	329
cliente	273	cuenta	286	poc	325
pagina	264	bueno	279	sucurs	317
poco	227	pagina	277	tem	313
veces	226	producto	252	cuent	298
productos	221	poco	244	llam	279
cuenta	220	tarjeta	239	product	252
bueno	218	vece	226	falt	249
necesito	197	necesito	197	mejor	246
falta	193	respuesta	197	solucion	240
ejecutivos	187	falta	193	ofrec	239
sucursal	173	sucurs	173	tarjet	239
sucursales	169	sucursal	169	pag	227
información	166	información	166	vec	226
tarjeta	166	banco	166	cambi	226
bancos	166	encuentro	163	inform	224
encuentro	163	mejorar	158	aplic	220
aplicación	158	aplicación	158	rap	214

Figura 7.6: Comparación de palabras frecuentes luego de procesamientos.

Fuente: Elaboración propia

Se puede notar que el segundo algoritmo (*Snowball*) agrupa en menos categorías las palabras, pero es más difícil de interpretar para el analista, por ejemplo, la abreviación “rap” se refiere a “rapidez” pero no es tan intuitivo- Al contrario del primer algoritmo de *stemming* (*Porter*) en que la interpretación es más sencilla, pero palabras como “bueno” y “buena”

no se agrupan, quedando como palabras distintas. Un ejemplo de esto se puede ver en la siguiente tabla con los distintos resultados tras la limpieza y los dos tipos de *stemming* a un mismo comentario.

Tabla 7.6: Comparación de algoritmos de stemming en un comentario

Texto	Stemming 1	Stemming 2
Porque no	satisface	satisfac
satisface	necesida-	neces pid
mis necesi-	des pido	problem
dades, lo	problema	burocraci
que yo pido	burocracia	
hacen		
mucho		
problema,		
mucha		
burocracia		

Fuente: Elaboración propia

Para probar la interpretabilidad en el algoritmo de *stemming de Snowball* se ejecutó el modelo de LDA para ambos casos donde se comprueba que es más sencillo de interpretar el algoritmo de Porter. Para que este proceso sea dinámico, es importante a juicio de la compañía la interpretabilidad y sencillez del modelo, es por esto que se concluye por trabajar con el algoritmo de Porter.

Finalmente, para tener una primera mirada de las palabras que más se repiten en los comentarios positivos y negativos, se realizó una nube de palabras para cada uno que se puede ver en la Figura a continuación. De estas se desprende que existen temas frecuentes que son comunes para ambos sentimientos como el “ejecutivo”, los “problemas” y la “pagina”, mientras que en el aspecto positivo destaca también la “atención” y el “servicio” y en el negativo destaca “crédito”, “seguro” y “cuenta”.



Figura 7.7: Nubes de palabras de comentarios positivos y negativos. Fuente: Elaboración propia

### 7.3. Modelamiento

### 7.3.1. Modelo de Tópicos

La etapa de modelado comienza con los relativos a los textos y el análisis de tópicos. Esta etapa se llevó a cabo en R debido a lo potente de sus librerías y las herramientas de visualización de resultados del algoritmo LDA.

El primer paso es crear un diccionario con el vocabulario extraído de los comentarios, luego se implementa una matriz término-documento (DTM) que contiene en las filas los comentarios y en las columnas el vocabulario. Con esto podemos representar los comentarios como vectores y ver cuáles y cuántos términos tiene cada uno.

El algoritmo para crear el DTM recibe como input cómo se quiere *tokenizar* los textos, esto quiere decir separar una oración, en una palabra (unigramas), dos palabras (bigramas), etc. En este caso de estudio se procederá con unigramas y bigramas, ya que facilitan la interpretabilidad de los tópicos al agrupar palabras automáticamente como “atención\_ejecutivo” y “fácil\_acceso”. Es el algoritmo el que decide si concatenar o no las palabras según lo usual que sea que dos tokens se encuentren continuos.

Luego se examinó la frecuencia de las palabras y se eliminaron las que solo aparecían una sola vez en todo el DTM, ya que no aportan a la conformación de un tópico y se eliminaron las palabras que aparecen en más de la mitad de los comentarios, ya que significa que la palabra se repite mucho y tampoco aporta a la conformación de agrupaciones.

Con el DTM listo se procede a ejecutar el algoritmo LDA para modelar los tópicos. Para esto se utilizó la librería *"topicmodels"* en el programa R. El *input* de este modelo es el DTM previamente creado y los parámetros son  $\alpha$  y  $\beta$  que por *default* son 0.1, el número de iteraciones se fijó en 2.000 y el número de tópicos se irá iterando para determinarlo.

Una de las dificultades de los algoritmos no supervisados es que el analista es quien debe escoger el número de *clusters* en que se agrupan los datos. Para facilitar esta tarea, se utilizará la métrica de coherencia de tópicos (detallada en la Sección 5.7) para ayudar la elección del número de estos.

A juicio de la compañía más de 25 tópicos dificulta los análisis y menos de 5 tópicos aporta muy poca información, es por esto que se calculó el nivel de coherencia para 1 a 25 tópicos, los resultados fueron los siguientes para los comentarios positivos y negativos:

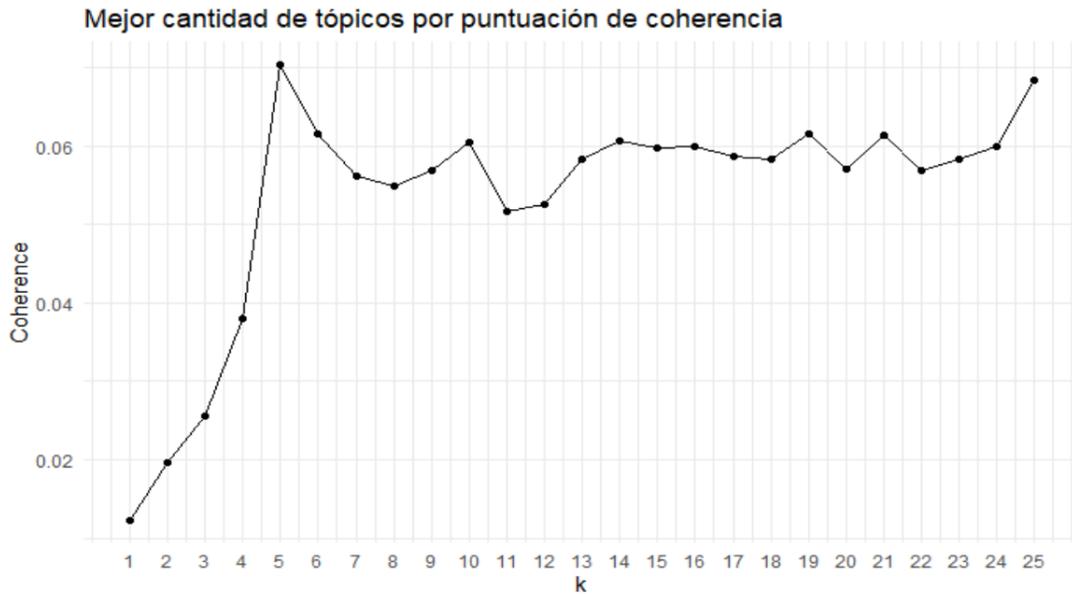


Figura 7.8: Coherencia según número de tópicos en comentarios positivos.  
Fuente: Elaboración propia

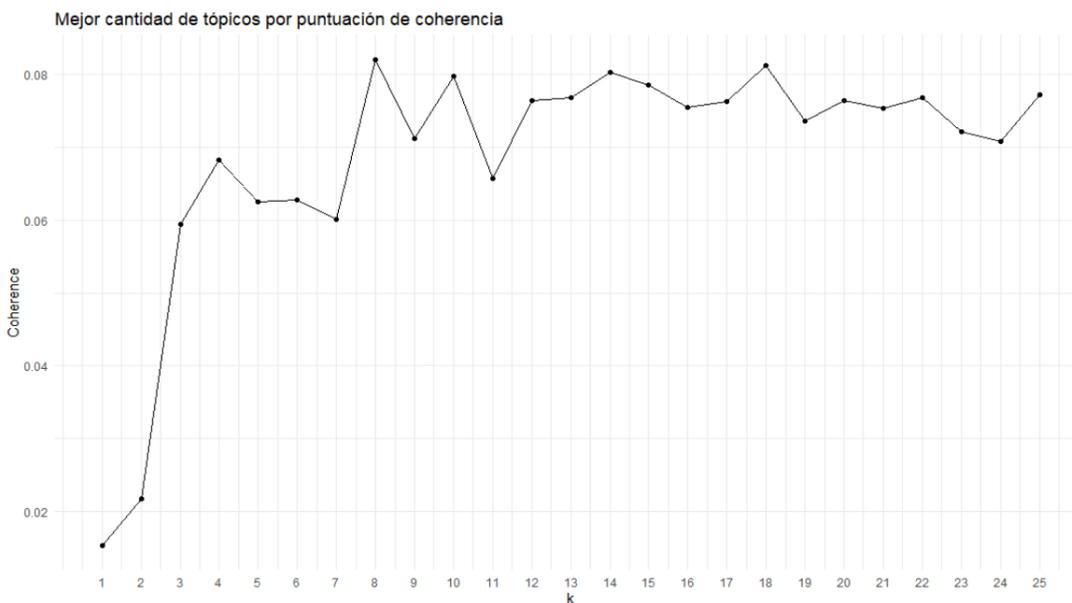


Figura 7.9: Coherencia según número de tópicos en comentarios negativos.  
Fuente: Elaboración propia

En el caso de los comentarios positivos, notamos que el máximo nivel de coherencia calculada son 5 y 25 tópicos, números considerados límites a juicios de la compañía. Por otro lado, en los comentarios negativos el mayor nivel de coherencia se da en 11 y 17 tópicos. Finalmente, se decide por 10 tópicos en ambos ya que no es mucha la variación con los “óptimos” y para que se puedan comparar mejor con la misma cantidad de tópicos. Esto fue validado también observando la conformación de los tópicos dado este número.

En la Figura 7.10 se pueden observar los tópicos de los comentarios positivos con las 10

palabras más influyentes de cada uno. Esto junto a la revisión de los textos asociados a cada tópico dio origen a la interpretación del nombre de cada uno.



Figura 7.10: Nubes de palabras de tópicos comentarios positivos. Fuente: Elaboración propia

Se realizó el mismo ejercicio para los comentarios negativos y los resultados fueron los siguientes:



Figura 7.11: Nubes de palabras de tópicos comentarios negativos. Fuente: Elaboración propia

### 7.3.2. Modelo de Caracterización de Lealtad

Una vez obtenido el t3pico asociado a cada comentario, se incorporan como variable *dummy*, donde el valor 1 representa si el comentario (ya sea positivo o negativo) pertenece al t3pico de esa columna y cero si no (imagen referencial de la base en Figura 7.12).

Recomendacion	Ejecutivo	P7	P7_2	Web	...	Top_neg_2	Top_neg_3	Top_neg_4	Top_neg_5	Top_neg_6	Top_neg_7	Top_neg_8	Top_neg_9
6	7.0	1	1.0	7.0	...	0	0	0	0	0	0	0	0
6	7.0	1	1.0	6.0	...	0	0	1	0	0	0	0	0
7	7.0	1	1.0	6.0	...	0	0	0	0	0	0	0	0
6	7.0	1	1.0	5.0	...	0	0	0	0	0	0	1	0
4	NaN	2	NaN	4.0	...	1	0	0	0	0	0	0	0

Figura 7.12: Base de datos unificada. Fuente: Elaboraci3n propia

#### Replicaci3n de metodolog3a base

Se implement3 la metodolog3a est3ndar que usa la industria de investigaci3n de mercado para caracterizar la lealtad de los clientes que consiste en utilizar una regresi3n lineal m3ltiple. Para el caso de estudio, dadas las variables que disponemos la formulaci3n es la siguiente:

$$Recomendaci3n = \beta_0 + \beta_1 Ejecutivo + \beta_2 Sucursal + \beta_3 Web + \beta_4 App + \beta_5 Beneficios \quad (7.1)$$

Las variables independientes responden a preguntas de la encuesta detalladas en la Tabla 7.7:

Tabla 7.7: Variables independientes modelo actual

Variable	Pregunta
Ejecutivo	¿Cu3l es su nivel de satisfacci3n con su ejecutivo de cuenta? Considere una escala de 1 a 7, donde 1 es “muy insatisfecho” y 7 es “muy satisfecho”
Sucursal	¿Cu3l es su nivel de satisfacci3n con el servicio que recibió en su(s) última(s) visitas a la sucursal? Utilizando la misma escala de 1 a 7
Web	¿C3mo evalúa la satisfacci3n con el sitio web del banco? Considerando una escala de 1 a 7, donde 1 es “muy insatisfecho” y 7 es “muy satisfecho”
App	Considerando una escala de 1 a 7, ¿donde 1 es “muy insatisfecho” y 7 es “muy satisfecho” ¿C3mo evalúa su satisfacci3n con respecto a la aplicaci3n?
Beneficios	Considerando el programa de beneficios que el banco ofrece ¿c3mo los evalúa utilizando una escala de 1 a 7?, donde 1 es “muy insatisfecho” y 7 es “muy satisfecho”

Fuente: Elaboraci3n propia con datos de la encuesta

El modelo es ejecutado en el software SPSS y utiliza el m3todo para tratar los nulos deno-

minado “Excluir casos según pareja” que consiste en que los casos con datos completos para la pareja de variables correlacionadas se utilizan para calcular el coeficiente de correlación en el cual se basa el análisis de regresión <sup>2</sup>. Este método para tratar los nulos sesga los resultados ya que los coeficientes obtenidos para cada variable consideran distintos tamaños de muestra por lo que no son comparables, además asume un patrón MCAR (*Missing Completely at Random*) en los valores perdidos lo que difiere del caso de estudio, ya que como se mencionó anteriormente, los nulos se deben a una respuesta previa. Por lo que, si bien, este método tiene la ventaja de hacer uso de toda la información disponible, la literatura no recomienda su uso, además de que los datos no cumplen con los supuestos básicos de este.

Si bien el método utilizado para caracterizar la lealtad en la empresa es la regresión lineal múltiple, no es conceptualmente correcto, ya que tiene como supuesto que la variable dependiente es continua, cuando en realidad se trata de una variable multiclase ordenada. Por lo que, de querer utilizar un método de regresión, lo conceptualmente correcto es utilizar regresión logística ordinal, pero finalmente se descarta utilizar regresiones dado que requiere que los datos nulos sean imputados o eliminados, quedando disponible solo 2.027 datos (15 % del total de la base original) para ejecutar el modelo, lo que se considera muy bajo para construir un modelo robusto.

De todas maneras, se implementó el modelo base de regresión lineal para poder comparar el estándar actual en la industria de investigación de mercados con los modelos propuestos. Para esto se simuló la misma metodología utilizando una base de entrenamiento del 80 % de los datos y una base de testeo del 20 % . Para todos los modelos se utilizará la misma proporción.

Cómo se pudo ver en el análisis exploratorio de los datos, el 84 % de las evaluaciones de recomendación de los clientes son 6 o 7. Para poder predecir la lealtad con mayor precisión, se balancearon las clases en la base de entrenamiento utilizando la técnica de *random over-sampling*. Esta técnica consiste en duplicar aleatoriamente las observaciones de la clase minoritaria hasta alcanzar el número que posee la clase predominante, con esto la base final pasó de tener 13.115 a 32.004 datos.

Para evaluar el modelo se redondean las predicciones obtenidas en la base de testeo para luego calcular la matriz de confusión y las métricas que se desprenden de este.

Luego de ejecutar el modelo base, se procedió a implementar el algoritmo XGBoost debido -principalmente- a que puede manejar valores perdidos, además de ser un algoritmo robusto, preciso, veloz y simple de implementar. Los distintos modelos de XGBoost que serán evaluados se construyeron utilizando la librería *Sklearn* para Python. Como se mencionó anteriormente, todos los modelos son entrenados con un 80 % de los datos, y testeados con el 20 % restante. Para el ajuste de hiperparámetros se utilizó la técnica de búsqueda en rejilla (*Grid Search*) con validación cruzada de 5 iteraciones. Los hiperparámetros testeados fueron:

<sup>2</sup> Documentación SPSS: [Click para acceder al sitio](#) [consulta: 3 enero 2021]

Tabla 7.8: Hiperparámetros testeados en modelos XGBoost

Hiperparámetro	Valores
Tasa de aprendizaje (Learning_rate)	0.05, 0.10, 0.15, 0.20, 0.25, 0.30
Ponderación hijo mínima (Min_child_weight)	1, 3, 5, 7
$\Gamma$	0.0, 0.1, 0.2, 0.3, 0.4
Muestra de Columna por árbol (Colsample_bytree)	0.3, 0.4, 0.5, 0.7

Fuente: Elaboración propia

XGBoost es un algoritmo que puede entrenarse para que clasifique correctamente todos los datos que se le proporcionan, pero esto lleva a que al testarlo en una muestra nueva el rendimiento disminuye ya que no aprendió a generalizar sobre los datos. A este problema se le conoce como *overfitting* o sobreajuste. Para evitarlo, existe el parámetro “*max depht*” o “profundidad”, que determina el número de nodos de bifurcación de los árboles de decisión usados en el entrenamiento. Para fijar este parámetro se comparó la diferencia entre la métrica AUC testeando el modelo en la base de entrenamiento y el AUC testeando en la base de prueba, variando el parámetro desde 1 a 6, con un salto de 1. El valor que generó menor diferencia entre los dos valores AUC fue 2, por lo que fue el utilizado en los modelos finales.

En primer lugar, se implementó la formulación base en XGBoost, es decir, con variable dependiente la nota de Recomendación de 1 a 7 y variables independientes la satisfacción sobre: Beneficios, Ejecutivo, App, Web y Sucursal. Luego a este mismo modelo, se le incorporó los 20 tópicos encontrados en forma de *dummies* relacionados a los comentarios positivos y negativos sobre el banco.

Finalmente, se decidió probar otro modelo en XGBoost en que la variable dependiente este agrupada en 3 clases, según detractores (notas de 1 a 4), neutros (nota 5) y promotores (nota 6 y 7). Se tomó la decisión de comparar con este agrupamiento ya que se relacionan directamente con el NPS y no es relevante para el estudio predecir exactamente la evaluación del cliente, sino más bien, la percepción general con respecto al banco. Es por esto, que se construyeron 2 modelos con variable dependiente de 3 clases, uno que solo posea las evaluaciones de satisfacción y otro que incorpore los resultados de las variables de texto.

## 7.4. Evaluación

Para evaluar los modelos se calcularon las métricas descritas en la Sección 5.7. Para seleccionar el mejor modelo disponible, se rescatan las métricas *Accuracy*, *weighted F1-Score*, *macro F1-score* y el AUC de cada modelo. Los resultados se pueden ver a continuación.

Tabla 7.9: Métricas de evaluación de desempeño en datos de prueba

	Accuracy	Weighted F1 Score	Macro F1 Score	AUC
Modelo base	0,16	0,2	0,12	0,53
XGBoost 7 clases sin tópicos	0,48	0,51	0,26	0,61
XGBoost 7 clases con tópicos	<b>0,7</b>	0,71	0,42	0,7
XGBoost 3 clases sin tópicos	0,67	0,71	0,49	0,69
XGBoost 3 clases con tópicos	0,69	<b>0,73</b>	<b>0,55</b>	<b>0,74</b>

Fuente: Elaboración propia

Comparando los resultados de los modelos se puede notar que el modelo base, posee un considerable menor desempeño que los demás. Por otro lado, los modelos que incluyen las variables sobre los comentarios de los clientes poseen un mayor desempeño en todas las métricas que su análogo sin los tópicos, lo que indica que la incorporación de la información obtenida por los textos ayuda a poder predecir con mayor exactitud los niveles de recomendación del cliente.

Entre los modelos XGBoost sin tópicos la diferencia entre el que predice entre 7 clases versus 3 clases es notable, siendo un 19 % más exacto este último. Por lo que, si se decide finalmente no incluir los tópicos, el mejor modelo será el que clasifica entre promotor, neutro y detractor.

Luego, entre los resultados de los modelos XGBoost que incluyen tópicos, las diferencias son sutiles, pero se puede notar que es superior el modelo que predice si el cliente es detractor, neutro o promotor, obteniendo el mayor AUC de 0,74, es decir, la probabilidad de que el modelo prediga mejor que lanzar una moneda es de un 74 % .

Para evaluar si el modelo con mayor desempeño se encuentra sobre ajustado o no, se comparó con el AUC en la base de entrenamiento.

Tabla 7.10: AUC de entrenamiento vs prueba

	AUC entrenamiento	AUC prueba
Modelo XGBoost 3 clases con tópicos	0,75	0,74

Fuente: Elaboración propia

En la Tabla 7.10 se puede ver que la diferencia entre el AUC del modelo testeado en la base de prueba es 1 % menor que el testeado en la base de entrenamiento lo que nos indica que el buen desempeño no se debe a un sobreajuste.

## 7.5. Despliegue de resultados

Para presentar los resultados de la memoria a Ipsos, se preparó una presentación que resume el desarrollo metodológico y los principales resultados y conclusiones del trabajo. Además, se realizó una infografía sobre la metodología recomendada para ir evaluando las percepciones que entregan los clientes en las encuestas. Por último, se implementó un instructivo que se complementa con un archivo de *Notebook Jupyter* para los algoritmos implementados en Python (preprocesamiento de textos y modelos de caracterización de lealtad), y un *R Markdown* para los ejecutados en R (modelamiento de tópicos), además de una capacitación a el área de Analytics de la empresa para aclarar el procedimiento y posibles dudas que surjan.

# Capítulo 8

## Análisis de resultados

Se dividió el trabajo en dos fases. La primera trata sobre la extracción de tópicos a raíz de los comentarios que emiten los encuestados mediante el algoritmo LDA. La segunda fase se enfocó en seleccionar el modelo que mejor caracterice la lealtad de los clientes que resultó ser el modelo XGBoost que predice si el cliente es detractor, promotor o neutro y que incorpora la información de la extracción de tópicos.

Esta sección trata sobre el análisis posterior a las conclusiones sobre el modelado de cada una de las fases.

### 8.1. Análisis de tópicos

Como resultado de lo desarrollado en la Sección 7.3, mediante el algoritmo LDA se clasificaron los comentarios negativos y positivos en 10 tópicos cada grupo (Tabla 8.1).

Tabla 8.1: Tópicos y su respectiva polaridad

Negativos	Ambas Polaridades	Positivos
Seguridad		Comunicación Ejecutivo
Beneficios		Trato del Ejecutivo
Experiencia	Aplicación y Sitio Web	Sucursales
Tasas créditos	Tarjeta de crédito	Otros
Cambio Ejecutivo	Tiempo de respuesta	Accesibilidad Sitio Web
Costos Tarjeta		Atención
Esperas en Sucursal		Solución Problemas

Fuente: Elaboración propia

Existen temas comunes entre los aspectos positivos y negativos, con lo que podemos inferir que estos son relevantes en la percepción general del banco. Por otro lado, existen temas que solo se mencionan con una polaridad, por ejemplo, las tasas de los créditos, los beneficios y seguridad solo se mencionan en los comentarios negativos, por lo que son atributos que solo generan críticas y se les debe poner atención y profundizar sobre el impacto que tiene esto en las evaluaciones, para determinar si es recomendable tomar acciones al respecto.

Se debe tener en consideración que entre los tópicos existen temas amplios y otros más específicos sobre lo mismo, por ejemplo, en los comentarios negativos existen los comentarios

relacionados a la tarjeta de crédito y existen otros comentarios negativos asociados específicamente a los costos de la tarjeta. Esto no quiere decir que un comentario este asignado a ambos temas, sino más bien, el tópico más amplio abarca todos los comentarios con respecto a ese tema, menos los específicos sobre el otro tópico.

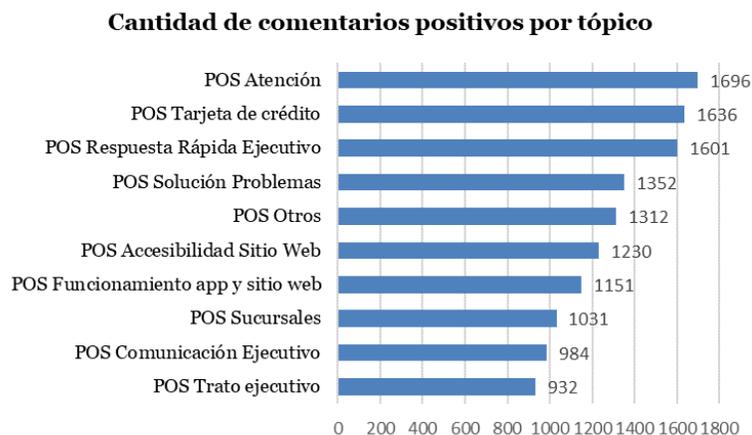


Figura 8.1: : N° de comentarios negativos por tópico. Fuente: Elaboración propia

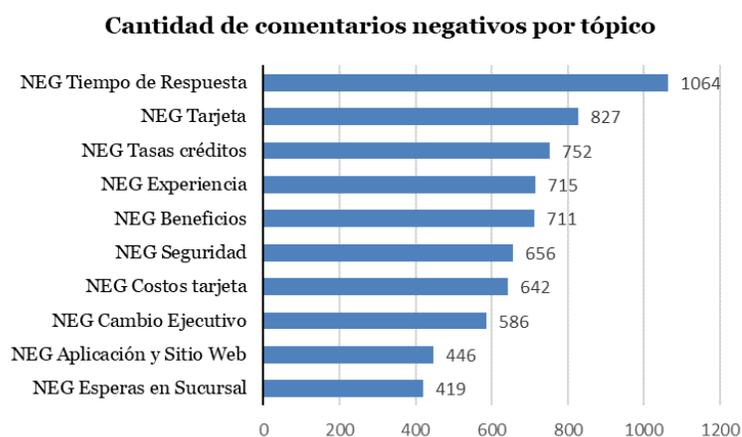


Figura 8.2: N° comentarios positivos por tópico. Fuente: Elaboración propia

La distribución de comentarios por cada tópico se muestra en las Figuras 8.1 y 8.2. De aquí se pueden desprender cuales son los temas más recurrentes entre los clientes. Por el lado de los aspectos negativos se puede ver que los tiempos de respuestas, problemas con las tarjetas y las tasas de los créditos se mencionan bastante. Por el lado de los aspectos positivos se habla bastante sobre la buena atención del banco, la tarjeta de crédito y la respuesta rápida del ejecutivo de cuenta.

En ese mismo sentido, es relevante no solo observar la cantidad de comentarios por tópicos, ya que puede ocurrir que, pese a que un aspecto se mencione bastante, no tenga real incidencia en la evaluación general del banco. Para evaluar esto, se tomaron los niveles de recomendación de los clientes con comentarios categorizados bajo cada tópico para obtener el NPS asociado a

cada uno. En la Figura 8.3 se puede ver el resultado, en donde las barras verdes corresponden a los comentarios positivos mientras que las rojas a los negativos.

Se da la tendencia lógica de que los tópicos positivos se relacionan con un mejor NPS que los tópicos negativos, con la excepción del tópico negativo “Esperas en Sucursal” que está asociado a un NPS de 78. Esto se puede deber a que las personas que mencionan este aspecto como algo negativo no es realmente algo que afecte su evaluación del banco, es decir, su incidencia es baja, lo que se podrá corroborar con la Criticidad (detallado en la Sección 5.8) y el modelo de caracterización más adelante.



Figura 8.3: NPS asociado a cada tópico. Fuente: Elaboración propia

El NPS global del periodo estudiado es igual a 72, con este número se calcula la diferencia con el NPS que tienen los comentarios atribuidos a cada tópico, para ver con más claridad el impacto que tiene cada tema en la evaluación global del banco. En la Tabla 8.2 se pueden ver los resultados de esta diferencia, más conocida como el impacto. Gracias a esta métrica se puede desprender con mayor claridad conclusiones como que los comentarios negativos con respecto a la seguridad son los que tienen mayor impacto en el NPS, mientras que aspectos como dar solución a los problemas que tiene la gente puede generar lealtad en los clientes.

Pero a la hora de tomar decisiones es importante tener en cuenta tanto la frecuencia de ocurrencia de un tópico como el impacto que tienen en el NPS. Si bien, el modelo de caracterización es la herramienta más robusta y clara para determinar esa incidencia, se puede explorar previamente el impacto que tienen con la Criticidad, resultado de la multiplicación del impacto y la frecuencia. En la Tabla 8.2 se pueden ver los resultados de esta multiplicación. Gracias a esta heurística se obtiene un orden para priorizar y poner foco a aquellos que tienen

menor Criticidad, que vendrían siendo la seguridad, beneficios y tiempo de respuesta.

Tabla 8.2: Impacto y Criticidad de los tópicos

Tópico	Impacto	Criticidad
NEG Seguridad	-53	-5
NEG Beneficios	-39	-4
NEG Tiempo de Respuesta	-30	-4
NEG Experiencia	-30	-3
NEG Tasas créditos	-28	-3
NEG Aplicación y Sitio Web	-37	-2
NEG Cambio Ejecutivo	-27	-2
NEG Tarjeta	-14	-2
NEG Costos tarjeta	-15	-1
POS Comunicación Ejecutivo	-11	-1
POS Trato ejecutivo	-11	-1
POS Tarjeta de crédito	-4	-1
POS Sucursales	-6	0
POS Otros	-4	0
POS Funcionamiento app y sitio web	3	0
POS Accesibilidad Sitio Web	8	1
NEG Esperas en Sucursal	6	1
POS Atención	10	1
POS Solución Problemas	14	1
POS Respuesta Rápida Ejecutivo	12	2

Fuente: Elaboración propia

## 8.2. Modelo de caracterización

Se pudo ver en la Sección 7.4 la comparación entre los modelos construidos y se determinó que los que mejor caracterizan la lealtad de los clientes son los XGBoost de 3 clases con y sin tópicos. En esta sección se analizará estos modelos, para evaluar cómo se pueden analizar cada uno, para luego concluir cual es el mejor uso que puede hacer la compañía de estos.

El objetivo final de este trabajo no es que la empresa pueda predecir la lealtad de los clientes, si no, crear una metodología para describir y analizar cuáles son las variables que influyen en la lealtad y cómo estas influyen, mejorando o profundizando el modelo base que usa la compañía. Por lo que es relevante desprender de los modelos la importancia de cada variable.

En el caso de XGBoost la métrica que se utilizará es el *SHAP Value*. Esta métrica que se destaca por sobre el resto, por cumplir con los principios de consistencia y exactitud, permite descomponer cualquier predicción en la suma de los efectos de cada valor de las variables. Para obtener la importancia global de las variables se suman los promedios de los *SHAP Value* de cada clase (detractor, promotor, neutro). Cabe destacar que los *SHAP Value* indican la magnitud del impacto de una variable, no si el impacto es positivo o negativo.

## Modelo sin incluir tópicos

Para el modelo XGBoost sin tópicos, se obtuvo la visualización presentada en la Figura 8.4. Este gráfico se interpreta como que el ejecutivo es la variable más relevante a nivel global a la hora de predecir si un cliente es promotor, detractor o neutro. Al contrario, la evaluación que los clientes emitan de la sucursal es la que menos incidencia tiene en la predicción.

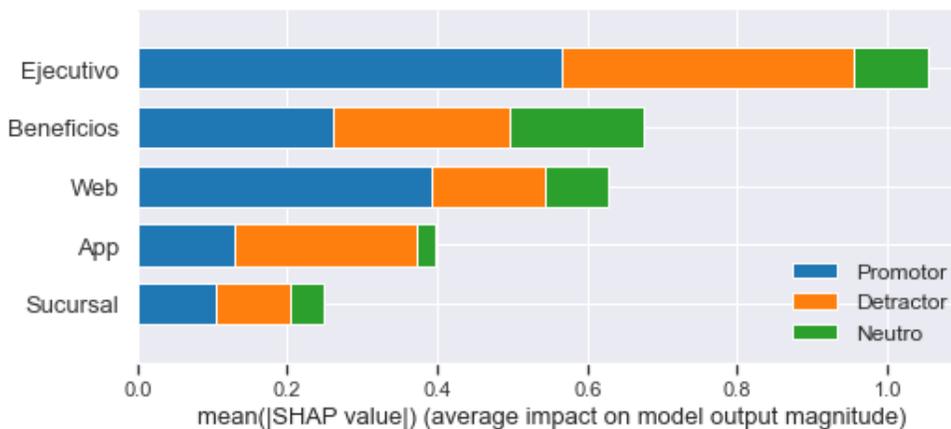


Figura 8.4: Impacto de las variables en los outputs del modelo sin tópicos .  
Fuente: Elaboración propia

Por otro lado, si se enfoca el análisis en los promotores, se puede ver que la página web es un factor que tiene mucha más relevancia en estos que para el resto de los tipos de clientes, esto se puede interpretar como que, el sitio web es muy relevante para las personas que están conformes con el banco, no así con los indiferentes o insatisfechos. Luego, siguen los beneficios, la aplicación y por último la sucursal como el que menos incidencia tiene.

En el caso de los detractores, al igual que los promotores, el ejecutivo es un aspecto clave, pero, llama la atención que la aplicación es el segundo factor con mayor relevancia para este segmento. Luego, siguen los beneficios, la web y al igual que para los promotores, la sucursal es el aspecto menos relevante para este segmento.

Mientras que para los neutros el ejecutivo no es lo más relevante si no que son los beneficios, esto es de utilidad ya que los neutros son los que evalúan con nota 5, es decir, un mínimo cambio en su percepción del servicio impacta directamente en la clasificación de este, y como consecuencia, afecta directamente el NPS. En el caso de este grupo, lo que menos impacto tiene es la sucursal y la aplicación.

De este análisis podemos desprender *insights* comerciales como monitorizar constantemente el desempeño de los ejecutivos bancarios debido a su alta incidencia en todos los grupos de clientes. La sucursal es un aspecto poco relevante para todos los segmentos por lo que no es prioritario tomar acciones sobre este aspecto.

Luego, si cruzamos la incidencia la satisfacción neta de cada atributo, se obtiene un gráfico muy utilizado por la empresa, pero en este caso ya no se trata del resultado de una regresión lineal con variables dependiente la evaluación de 1 a 7, si no que del modelo XGBoost con

variable dependiente de 3 clases. Además, para facilitar el análisis se incorporaron algunas guías para determinar cuadrantes, estas son: el NPS global de los datos considerados en este estudio que corresponde a 72 y la línea vertical corresponde al promedio de las incidencias de los factores considerados en ese gráfico. Los cuadrantes que se definieron son:

- Fortalezas: Cuadrante superior derecho, factores con alta incidencia y alto índice de satisfacción neta
- Debilidades: Cuadrante inferior derecho, factores con alta incidencia y bajo índice de satisfacción neta
- Mantener: Cuadrante superior izquierdo, factores con baja incidencia y alto índice de satisfacción neta
- Prioridades de Segundo Orden: Cuadrante inferior izquierdo, factores con baja incidencia y bajo índice de satisfacción neta

Para el modelo sin incluir aún los tópicos se obtiene el gráfico presentado en la Figura 8.5. Se puede concluir cuáles son las debilidades, en este caso, los beneficios dada su alta incidencia y baja evaluación de satisfacción, por lo que se deben tomar acciones prioritarias con respecto a este factor. Por otro lado, el ejecutivo y la página web son fortalezas globales, por su alta incidencia y buenas evaluaciones.

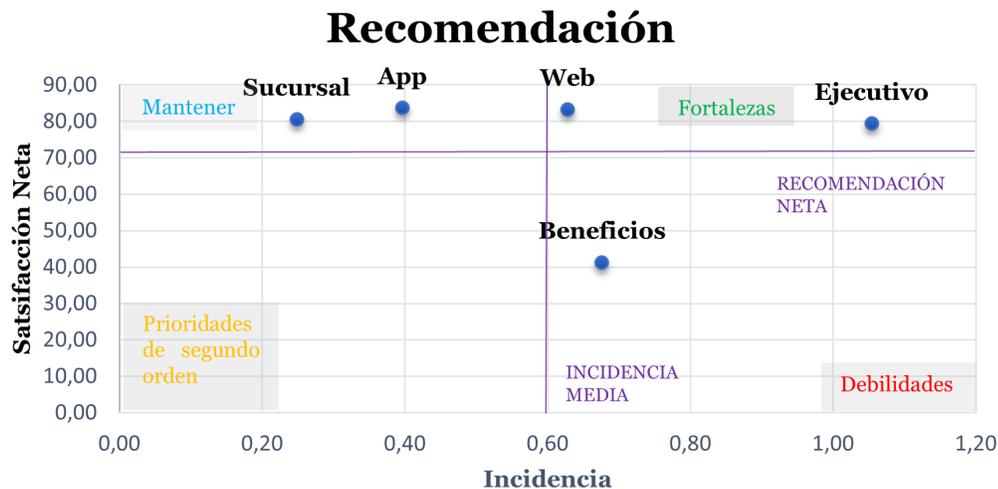


Figura 8.5: Caracterización de la recomendación sin incluir tópicos. Fuente: Elaboración propia

Luego, se puede realizar el análisis separando a los clientes según el tipo de cliente que sea: promotor, detractor y neutro, obteniendo la satisfacción neta para cada atributo de ese grupo junto a su incidencia respectiva. Los resultados de este desglose se encuentran en las Figuras 8.6, 8.7 y 8.8.

## Recomendación (Promotores)

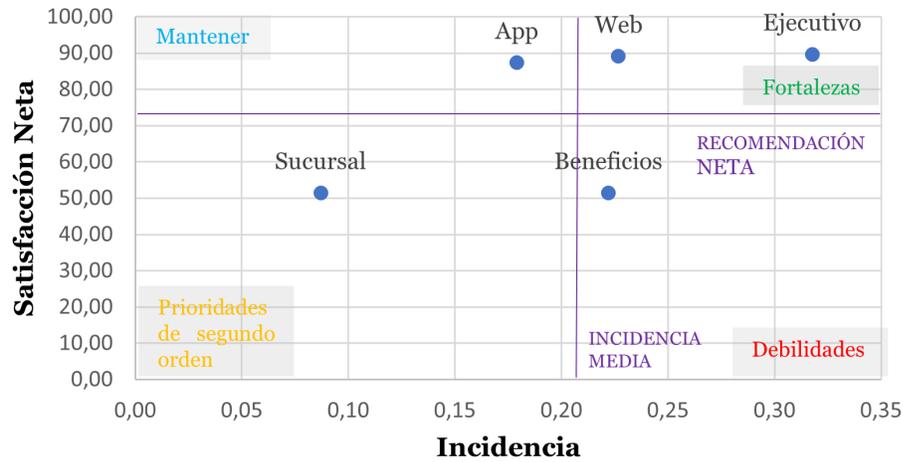


Figura 8.6: Caracterización de la recomendación entre clientes promotores  
 . Fuente: Elaboración propia

Del gráfico de promotores, que corresponde al grueso de la muestra, se nota que tanto las debilidades como las fortalezas se mantienen con respecto a la visión global. Por otro lado, llama la atención que la evaluación de la sucursal cae aproximadamente 30 puntos en este grupo. De esto se infiere que el grupo de personas que recomienda el banco tiene una percepción media de las sucursales, pero no es relevante en su evaluación global, esto se puede deber a que se trata de personas más cercanas a lo digital sin mucha interacción física con el banco. Por ende, es un aspecto que, si bien no se debe ignorar, no es prioritario ya que no genera mal estar en los clientes, por ende, se recomienda abordarlo solo cuando los esfuerzos en los otros aspectos hayan sido considerados.

## Recomendaciones (Neutros)

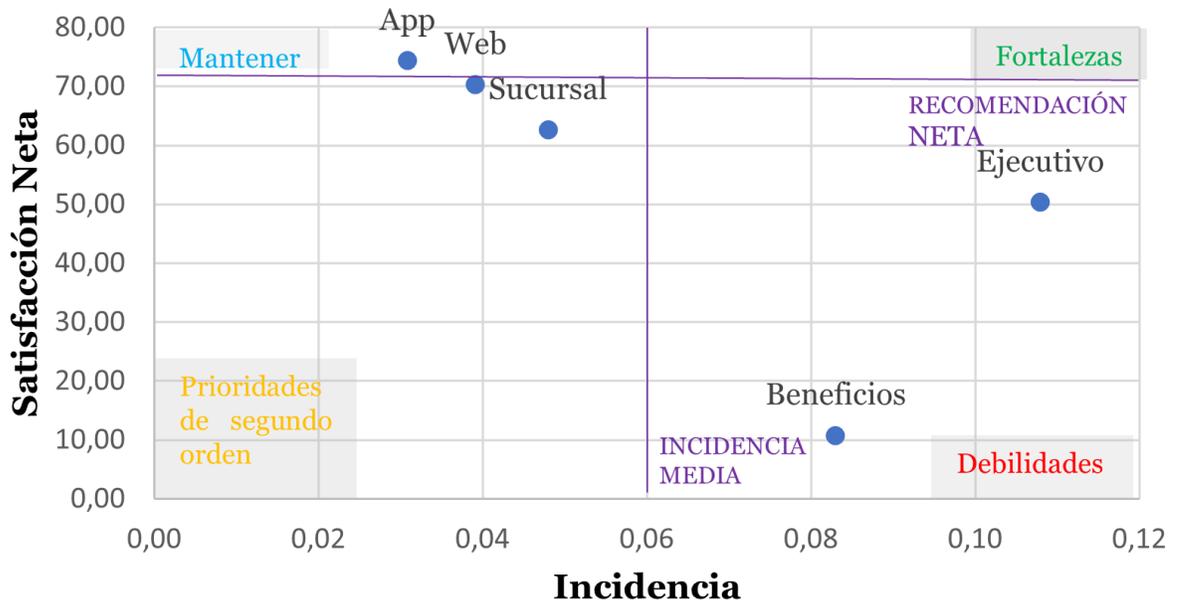


Figura 8.7: Caracterización de la recomendación entre clientes neutros.  
Fuente: Elaboración propia

En el grupo de clientes neutros se puede notar una mayor diferencia con respecto a la visión global, destacando la baja en la evaluación del ejecutivo de cuenta, considerándose una debilidad para este grupo, en conjunto con los beneficios cuyo nivel de satisfacción está muy por debajo de la media. También destaca que las sucursales poseen una leve mayor relevancia para este grupo que los sitios digitales, por lo que podría tratarse de un segmento más tradicional que digital, existiendo una oportunidad de mejora de segundo orden para este aspecto.

## Recomendación (Detractores)

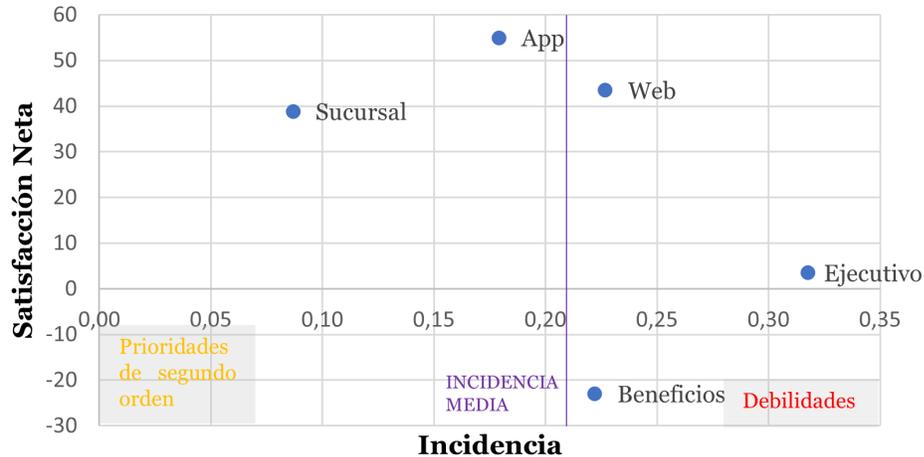


Figura 8.8: Caracterización de la recomendación entre clientes detractores.  
Fuente: Elaboración propia

No es de extrañar que en la visualización del grupo de detractores las evaluaciones de satisfacción neta para cada aspecto bajen, pero llama la atención que las evaluaciones son críticas para los beneficios, en donde la cantidad de notas de 1 a 4 supera a las notas 6 o 7. Cercano a esto se encuentra el ejecutivo de cuenta cuyas evaluaciones son considerablemente bajas. Por último, notamos que para este grupo el sitio web es también un aspecto a mejorar dada su alta incidencia. Al igual que lo promotores, la sucursal es un aspecto poco relevante por lo que también se puede tratar de un grupo más cercano al mundo digital.

Si consideramos las tres visualizaciones se pueden extraer *insights* como que la aplicación es un aspecto relativamente bien evaluado para todos los grupos, pero para ninguno es considerado como prioritario, por lo que es un aspecto que se recomienda mantener como está. Por otro lado, entre los grupos menos conformes, es decir neutros y detractores, el ejecutivo es el aspecto más relevante y es considerado una debilidad crítica que se debe mejorar. Destaca el hecho de que sin la visión de grupos simplemente se hubiese considerado al ejecutivo como una fortaleza y por lo mismo no enfocar esfuerzos serios en mejorarlo. En el caso de los beneficios, es un aspecto considerado una debilidad para todos los grupos, con evaluaciones mínimas en los grupos menos conformes, es por esto que se recomienda poner énfasis también en indagar qué acciones podrían mejorar la evaluación de este aspecto.

### Modelo incluyendo tópicos

Primero se analizará la incidencia de las variables en el modelo que incluye los tópicos (Figura 8.9). A grandes rasgos se puede ver que al igual que en el modelo anterior el ejecutivo es el factor más relevante a nivel global, pero difieren en que a continuación es el sitio web el que adopta una mayor relevancia y no los beneficios. Esto se puede deber a que, al incluir más factores, la incidencia que tenía un aspecto en el modelo anterior, en este nuevo modelo se ve descompuesta en más factores, enriqueciendo el análisis.

Se puede destacar que existen varios tópicos que son más relevantes que las evaluaciones

efectuadas de la app y las sucursales, en donde, podemos decir por ejemplo que un cliente con una mala percepción de los tiempos de respuesta, la seguridad o las tasas en créditos, pueden traer peores resultados que una mala evaluación de la app y las sucursales. Esto adquiere mayor relevancia considerando que estos aspectos no son considerados en el modelo original dado que no son preguntados en la encuesta y solo los llegamos a abordar al tomar en cuenta lo que comentan los clientes.

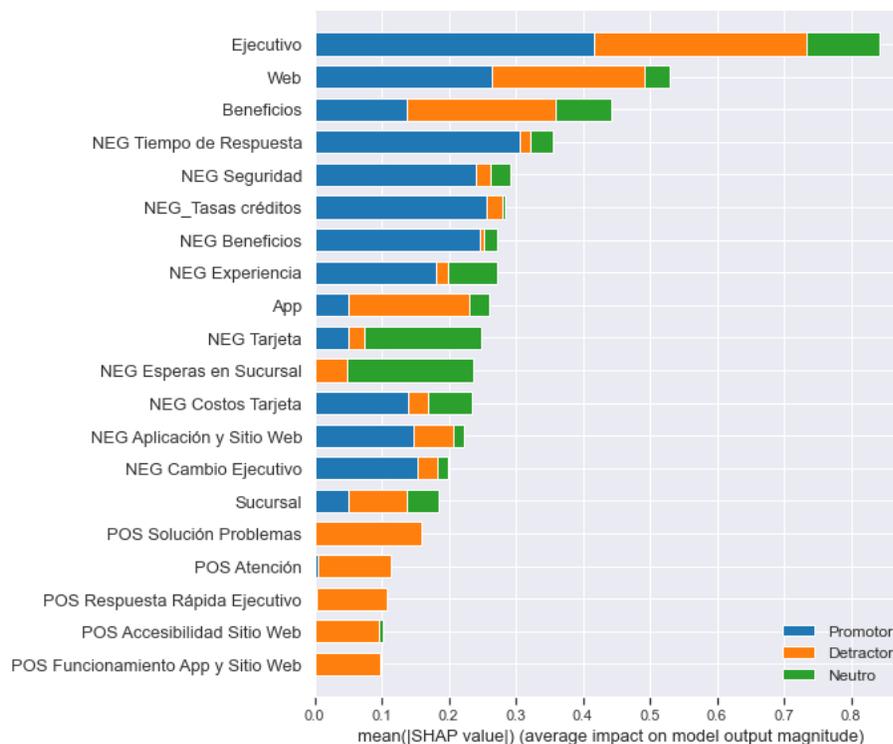


Figura 8.9: Incidencia de variables en el modelo con tópicos. Fuente: Elaboración propia

Es destacable que los comentarios negativos tienen mayor importancia para los promotores y análogamente los positivos para los detractores. Un ejemplo de la interpretación de esto es: tener una opinión negativa sobre los tiempos de respuesta es fundamental en la determinación de si un cliente es o no promotor, no así para los neutros o detractores.

Se puede visualizar con mayor claridad la importancia para cada grupo de clientes en las Figuras 8.10, 8.11 y 8.12.

## Incidencia (Detractores)

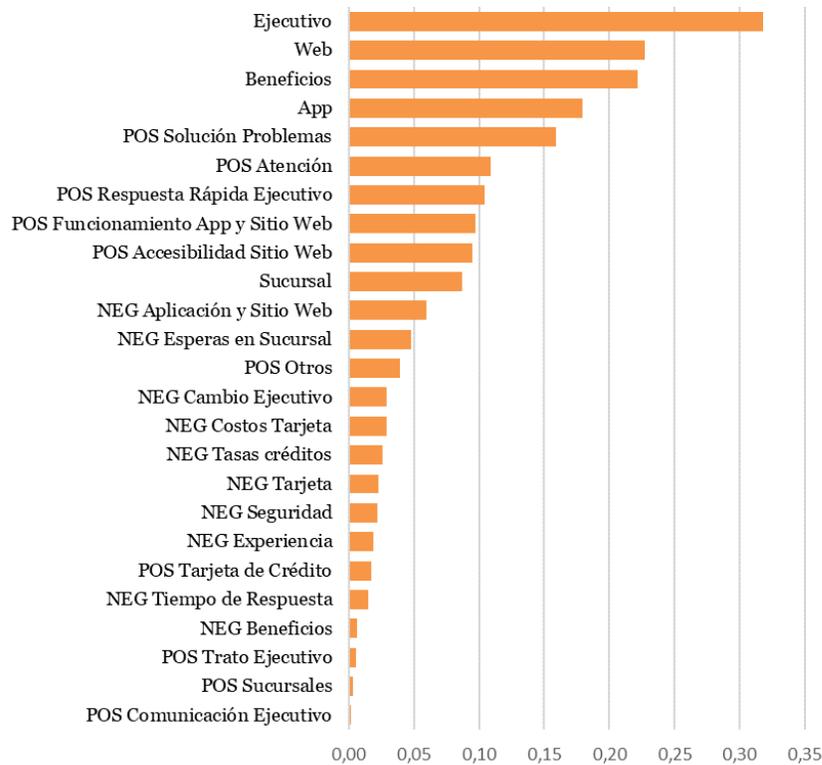


Figura 8.10: Incidencia de atributos entre clientes detractores. Fuente: Elaboración propia

Es importante notar que en el caso de los detractores son más relevantes las evaluaciones realizadas del ejecutivo, la web, aplicación y beneficios, que los comentarios que emiten con respecto al banco. Esto se puede deber a que las notas en esos aspectos tienden a ser bajas, como se pudo observar en el apartado anterior, por lo que al modelo le basta con tomar las evaluaciones realizadas a los 4 aspectos consultados en la encuesta para predecir que el cliente es un detractor. Por ejemplo, si un cliente evalúa en la encuesta alguno de estos aspectos con nota 2, para el modelo aumenta considerablemente la probabilidad de que se trate de un detractor sin necesidad de considerar tanto los resultados obtenidos a partir de los comentarios. El único aspecto evaluado que se escapa de esto es la sucursal, que posee una incidencia media.

## Incidencia (Promotores)

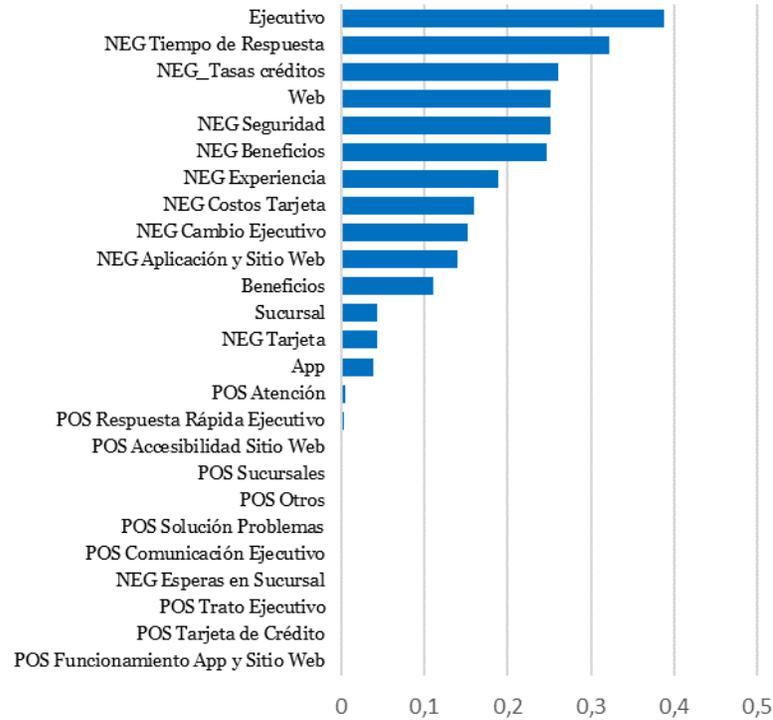


Figura 8.11: Incidencia de atributos entre clientes promotores. Fuente: Elaboración propia

Para el grupo de clientes promotores existen una serie de atributos relacionados a los comentarios que no tienen ninguna relevancia en la caracterización de este tipo de clientes. Pero, a diferencia de los detractores, existen algunos comentarios que sí tienen alta incidencia en su caracterización como lo son los comentarios negativos sobre tiempos de respuesta, tasas de los créditos, la seguridad, experiencia, costos de la tarjeta y cambios de ejecutivo, siendo más relevante incluso que los aspectos evaluados de beneficios, sucursal y aplicación.

Llama la atención que los comentarios negativos sean los que caracterice a los promotores y no los positivos, pero se debe recordar que la métrica *SHAP Value* mide magnitud absoluta y no el sentido de esta, por ende, se puede interpretar como que un cliente es considerado promotor a menos que esté asociado a un comentario negativo, lo que impacta directamente en su clasificación. Lo relevante es saber que son aspectos importantes para este grupo. Dado que se trata de clientes conformes con el banco, no es necesario poner énfasis en tomar acción sobre estos aspectos.

### Incidencia (Neutros)

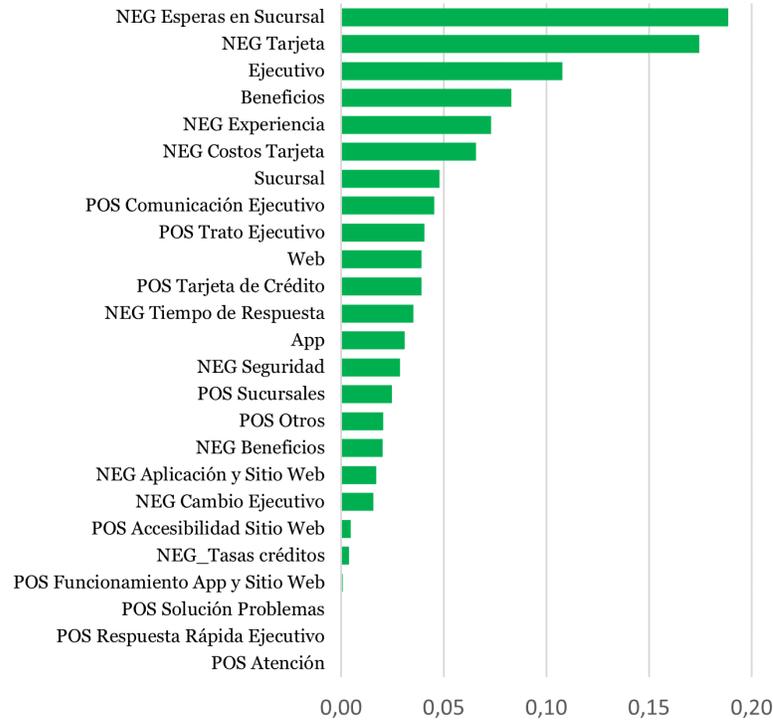


Figura 8.12: Incidencia de atributos entre clientes neutros. Fuente: Elaboración propia

En el caso de los clientes neutros es interesante que no es el ejecutivo el aspecto más importante, si no que los comentarios atribuidos a las esperas en sucursal y a la tarjeta de crédito. Como se mencionó anteriormente, la información que obtengamos de los clientes neutros es clave para mejorar el NPS global, por lo que se recomienda poner foco a las evaluaciones que se realicen de las esperas en sucursal y del uso de la tarjeta de crédito. Sería bueno poder incluir estos aspectos en una próxima encuesta para obtener *insights* más claros sobre estos.

Luego, se simula el gráfico de satisfacción neta vs incidencia, pero se debe tener en consideración que los tópicos no tienen una evaluación de satisfacción directamente relacionada, sino que se toman las notas de recomendación de las personas que mencionaron ese tópico, tratándose entonces de recomendación neta. Es por esto que en esta visualización comparamos indicadores netos para caracterizar de forma más clara y a nivel global la recomendación de los clientes, así se puede tener una visión más amplia de la opinión de estos, lo que fue validado por la compañía. El gráfico se puede observar en la Figura 8.13. Para facilitar el análisis se destacó con colores la polaridad del tópico, verde para los positivos y rojo para los negativos, mientras que los puntos azules hacen alusión a los factores evaluados en la encuesta.

De la vista global se pueden concluir cosas como que el ejecutivo es por mucho el factor más relevante a la hora de recomendar el banco, y está muy bien evaluado por lo que es ideal seguir conservando las buenas prácticas e incluso gestionar para mejorar aún más este

aspecto. Por otro lado, se ve que los beneficios, tanto la evaluación neta como los comentarios negativos con respecto a estos entregan resultados similares, en ambos casos el NPS es bajo y la incidencia es media, por lo que es un aspecto prioritario de accionar para mejorar sus resultados.

Otros aspectos que se encuentran dentro de las debilidades del banco son la seguridad, los tiempos de respuestas, las tasas de los créditos, la tarjeta de crédito y la experiencia. Con respecto a este último, es un tópico que puede ser poco accionable ya que el concepto de “experiencia” se puede atribuir a uno o varios puntos de contacto del banco. Con este tipo de ejemplos, podemos ver que existen distintos tópicos que no van a aportar mucha información ya que solo agrupan los comentarios vagos o poco específicos de algunos clientes, otro ejemplo de esto es el tópico positivo “Otros”.

## Recomendación



Figura 8.13: Caracterización de la lealtad incluyendo tópicos. Fuente: Elaboración propia

Se puede ver que, en el gráfico, los comentarios negativos sobre las esperas en sucursal quedan en el cuadrante de fortalezas, esto ya que las personas que comentan esto tienen buenas evaluaciones sobre el banco. Se debe recordar que los comentarios negativos son realizados solo por las personas que evalúan con nota de 1 a 6, por ende, podemos inferir

que las personas que indican como negativas las esperas en sucursal, son personas cuya recomendación tiende a ser 5 o 6, y es por esto que la recomendación neta es alta. De esto, se puede concluir que las esperas en sucursal son un detalle que, si bien no es crítico, se puede mejorar para que estas personas pasen a ser promotores o entreguen la calificación más alta.

Finalmente concluimos del mapa que los aspectos prioritarios a mejorar son la seguridad, los beneficios, las tasas de los créditos, los tiempos de respuesta y costos de la tarjeta.

Al incluir los tópicos en el modelo, no se recomienda realizar el mapa de incidencia vs indicadores netos para cada grupo de clientes ya que en el caso de los comentarios solo variará la incidencia y no la recomendación neta.

# Capítulo 9

## Conclusiones

### 9.1. Conclusiones generales

La industria bancaria es un mundo altamente competitivo, aún más considerando la facilidad actual que tienen los clientes para cambiarse de un banco a otro. Cada vez tienen más información sobre los productos o servicios que ofrece cada uno, por lo que son más exigentes respecto al servicio que se les ofrece. Las empresas deben hilar fino para encontrar *insights* comerciales que les permitan superar a la competencia, satisfacer y fidelizar a sus clientes.

Es complejo llegar a aprovechar el 100 % de la información con la que cuenta una empresa, ya que la cantidad puede ser abrumadora y la dificultad de procesarlo puede ser un freno. En el caso de Ipsos, se realizan más de 3 millones de encuestas al año, cada una con una serie de datos de cada cliente o usuario, las que usualmente incluyen preguntas abiertas con la opinión personal de estos.

Bajo este contexto, la voz de los clientes se vuelve cada vez más relevante. Es ahí donde se pueden obtener las razones de su satisfacción o insatisfacción y sacar a la luz elementos que la empresa no esté considerando entre sus cartas de acción.

Esta memoria estudió el valor que puede agregar el incorporar la data no estructurada de las encuestas en la caracterización de la recomendación de los clientes y cómo se puede mejorar y afinar los modelos que solo incluyen data estructurada, es decir, los modelos base utilizados en la industria.

Para cumplir el objetivo, en primer lugar, se crearon diccionarios enfocados en el lenguaje que se usa entre los clientes de la industria bancaria y se implementó una metodología que permitió clasificar los comentarios de las encuestas en 20 tópicos. Este número significa una importante reducción con respecto a las 300 categorías que se obtienen del análisis manual. Esto claramente tiene ventajas, considerando que actualmente no existe un método que supere la precisión del humano clasificando manualmente. Además, con más categorías la clasificación puede ser mucho más exacta. Pero clasificar uno a uno los comentarios requiere altos recursos y tiempo, y tratar con tantas categorías dificulta el análisis posterior, por lo que es la opción menos eficiente. La metodología propuesta en esta memoria es capaz de reducir el tiempo de análisis de una encuesta de 2 semanas a 2 días aproximadamente.

La extracción de tópicos es un mecanismo no supervisado que cumple con la necesidad de la empresa de que sean los mismos datos no estructurados los que sugieran temas de los que hablan los clientes y no el analista. Esto permite descubrir si existen elementos que no se están considerando por sesgar según lo que la compañía determina previamente.

Se debe tener en consideración que la extracción de tópicos al ser un método de aprendizaje automático no supervisado y -más aún- considerando que trabaja sobre el lenguaje humano, está sujeto a la existencia de ruido en los datos que tiene impacto sobre la precisión de los resultados del modelo. Por ejemplo, dentro de un tópico puede haber una serie de comentarios que no se ajusten necesariamente a la clasificación asignada, pero sí el grueso de estos.

Bajo estas consideraciones, el análisis de tópicos es un complemento enriquecedor para el análisis exploratorio, permitiendo descubrir información nueva, clasificar los comentarios de forma rápida y eficiente y tener una visión global sobre qué es lo que hablan los usuarios.

Además, sólo extrayendo los tópicos, estos se pueden relacionar con una mayor o menor recomendación por parte de los clientes, permitiendo entender cuáles son más críticos. Para el caso de estudio de esta memoria -entre los tópicos asociados a un bajo NPS- se consideran críticos los comentarios negativos sobre la seguridad del banco y los beneficios que ofrecen. En el caso contrario, los comentarios relacionados a una buena recomendación son los comentarios positivos sobre dar solución a los problemas que tienen los clientes y las respuestas rápidas del ejecutivo.

Luego al calcular la Criticidad -que considera tanto el impacto de cada tópico sobre el NPS global y la frecuencia con la que se menciona cada uno- podemos ver un orden de priorización sobre qué temas tratar con mayor énfasis, coincidiendo con los mencionados en el párrafo anterior relacionados a un peor NPS. Según el API, los temas a tratar con mayor prioridad son la seguridad, los beneficios y los tiempos de respuesta.

Esto es lo que se obtiene solo de la extracción y análisis de tópicos. Pero esto se pudo enriquecer aún más implementando un modelo que nos permita caracterizar la lealtad con alta precisión. Tras la implementación de 5 tipos de modelos con distintas variables, el primero imita el modelo estándar que usa la industria con regresión lineal y variable dependiente las notas de 1 a 7, y los otros 4 modelos son las distintas combinaciones del modelo XGBoost con y sin tópicos, y clasificando con nota de 1 a 7 o entre tres clases: neutro, promotor y detractor. De esto se pudo concluir que el modelo que agrupa los clientes en tres grupos e incluye los tópicos es el que posee mejor AUC, superando en un 21 % al modelo base. Además, el mismo modelo sin tópicos igual tuvo buenos resultados mejorando un 16 % el AUC versus el modelo base.

Con esto se pudo concluir que la información obtenida desde los comentarios de los usuarios efectivamente ayuda a predecir y caracterizar la lealtad de los clientes. Por lo tanto, es información que se debe aprovechar mejor por la compañía.

Una ventaja importante del modelo XGBoost es que la métrica - SHAP Value- para medir la importancia de cada variable permite descomponer la importancia de cada atributo para cada grupo. Esto nos ayuda a caracterizar no solo la lealtad global, sino que también

caracterizar cada grupo: detractores, neutros y promotores, y tomar decisiones en base a la visión de cada uno.

La desventaja es que, a diferencia de las regresiones, *SHAP value* no indica si la incidencia de cada factor es positiva o negativa, y no se tiene una formula concreta para medir como variaría la recomendación global si algún atributo cambiara, porque, al ser un modelo de árboles ensamblados, el *SHAP value* varía para cada observación.

Si bien, el mejor modelo fue el de XGBoost de 3 clases incluyendo tópicos, se debe considerar que los tópicos extraídos pueden ir variando en el tiempo, ya que se trata de un modelo no supervisado. Es por esto, que no se puede garantizar tener tendencias y evoluciones que sean precisas sobre los temas que tratan los comentarios. Por lo mismo, no se recomienda descartar el modelo XGBoost de 3 clases sin tópicos, ya que este sí permite tener una visión objetiva y certera de las evaluaciones que realizan los usuarios sobre los distintos aspectos considerados en la encuesta y poder trazar la evolución de la incidencia y de la satisfacción neta a lo largo del tiempo.

Además, este modelo permite desglosar el mapa de incidencia versus satisfacción neta para los 3 tipos de clientes. Esto es una ventaja considerando que el grueso de la muestra son promotores, por lo que la vista global puede diferir bastante de las percepciones de neutros y detractores como se vio en el análisis previo.

La desventaja de este modelo es que se miden pocas variables, y agregar más implicaría aumentar la cantidad de preguntas de la encuesta, lo que suele no ser bien recibido por el encuestado, ya que la propensión a contestar una encuesta es mucho más alta si la encuesta toma poco tiempo. Es por esto, que incorporar una pregunta adicional a una encuesta debe estar bien justificado. En ese sentido, el modelo de extracción de tópicos y el modelo de predicción que incluye tópicos se recomienda utilizarlo mensualmente como herramienta exploratoria. Su objetivo es analizar qué elementos son los más criticados o promovidos durante la medición, compararlo con el periodo anterior, tomar los aspectos que constantemente son considerados como debilidades y evaluar incorporarlo como pregunta evaluada en las encuestas futuras.

## 9.2. Recomendaciones para Ipsos

La recomendación para Ipsos es mensualmente implementar ambos modelos XGBoost de 3 clases con y sin tópicos. Del modelo sin tópicos visualizar la vista global de la recomendación y además las correspondientes a cada uno de los tipos de clientes, para identificar las fortalezas y debilidades de cada uno. Luego, construir un gráfico que muestre la evolución de la incidencia y la satisfacción neta para cada uno de los aspectos evaluados, para así monitorear los cambios en el comportamiento de los clientes o estar alerta por si existen cambios bruscos en alguno de los valores.

Por otro lado, se recomienda utilizar el modelo que incluye tópicos para complementar y enriquecer la información que se obtiene a partir de la misma encuesta. De estos podemos concluir qué variables se pueden incorporar en el futuro a la encuesta. Para determinar esto

se recomienda enfocarse en los que se encuentren en el cuadrante de debilidades en la visión global. También se recomienda considerar los resultados que se obtienen en las mediciones, y desglosar la incidencia entre los tres tipos de clientes. Si es que existen tópicos con resultados críticos serán recomendados como foco a accionar.

### **9.3. Recomendaciones bancarias para la muestra analizada**

A partir del análisis realizado sobre las encuestas obtenidas las recomendaciones son las siguientes:

- La seguridad del banco y las tasas de créditos son aspectos que solo se encuentran en los comentarios negativos, por ende, son elementos que no generan buenas opiniones entre los clientes. Además, se consideran debilidades a nivel global, dado que son incidentes y están asociados a bajas recomendaciones. No se podría haber vislumbrado esto sin el análisis de tópicos, por lo que se recomienda incorporar preguntas sobre el nivel de satisfacción de lo usuarios con respecto a estos dos puntos, para en el futuro monitorizarlos mejor.
- Los beneficios son un aspecto con bajas evaluaciones y relevante para todos los grupos de clientes, por lo que es un factor que debe priorizarse y tomar acciones gestionables al respecto. Se puede implementar una campaña específica de encuestas para descubrir qué tipos de beneficios buscan los clientes actuales del banco para luego tomar acciones con estos resultados.
- Sin duda, el ejecutivo de cuentas es el aspecto más relevante para todos los grupos de usuarios, y si bien, en la visión global se considera una fortaleza, desde la perspectiva de los clientes neutros y detractores es una debilidad, porque las notas de satisfacción caen abruptamente. A partir de esto, se recomienda desglosar la opinión de los clientes sobre el ejecutivo, preguntando cómo evalúan los tiempos de respuesta, la amabilidad, la facilidad de comunicación y la solución a los problemas que se le presentan. Todos son aspectos que se encontraron gracias a la extracción de tópicos a partir de los comentarios. Una vez se obtengan datos sobre estas preguntas de satisfacción, se sugiere crear un mapa de caracterización de la satisfacción del ejecutivo, similar a lo que se realizó con la recomendación a nivel global.
- Las esperas en sucursal son un aspecto que afecta principalmente a los clientes neutros, por lo que puede ser un aspecto clave a mejorar para que estos clientes pasen a ser promotores y no bajen sus evaluaciones. De esta forma, si muchos clientes neutros mejoran su consideración sobre las esperas, puede impactar en el NPS global del banco.
- La aplicación y las sucursales son factores que tienden a estar bien evaluados y no son especialmente relevantes para ningún caso, por lo que se recomienda no implementar esfuerzos especiales en mejorar estos puntos de contacto.
- Otro aspecto que se menciona bastante en los comentarios es la tarjeta de crédito. Este producto reúne tanto comentarios positivos como negativos y se encuentra en el cuadrante de debilidades en la visión global de la recomendación. Por esto, se recomienda

también incluirlo en encuestas futuras para que el usuario evalué directamente su satisfacción con respecto a los productos que le entrega el banco.

## 9.4. Trabajos futuros

Además de las recomendaciones de uso que se plantearon en el apartado anterior, se plantean trabajos sugeridos para que la compañía enriquezca y complemente los análisis que elaboran para sus clientes.

Se corroboró en esta memoria el aporte que generan los comentarios en la caracterización de la lealtad de los clientes, pero el método de extracción de tópicos no supervisado tiene la desventaja de ser poco medible y preciso. Se sugiere implementar un modelo de clasificación de comentarios supervisado, con las etiquetas que se obtengan de los tópicos u otras que sugiera la compañía. Para esto, se debe etiquetar manualmente una muestra importante de datos para que el entrenamiento del modelo sea lo más preciso posible. Con esto, los comentarios podrán ser clasificados con mayor exactitud y se tendrían las mismas etiquetas todos los meses, por lo que se podría monitorizar la evolución de la incidencia y evaluaciones de cada uno.

Con esto, no se recomienda descartar el modelo no supervisado, ya que será un apoyo para el descubrimiento de nuevos aportes que vengan desde los mismos usuarios. Quizás teniendo el modelo supervisado, la implementación del modelo desarrollado en esta memoria sea recomendable usarlo trimestralmente para tener una visión global de qué es lo que se está hablando en el último tiempo.

Esta memoria se realizó para los clientes *personas*, dejando de lado las encuestas de *pymes* y *grandes empresas*, por lo que se sugiere evaluar si los diccionarios creados se adaptan a estos, y si no, modificarlos para implementar el mismo estudio a estos segmentos, conocer qué comentan estos usuarios y así entregarle mayor valor al cliente.

Ipsos es una compañía que trabaja para varios bancos e industrias, por lo que inicialmente se recomienda replicar el trabajo realizado para todos los clientes banco que se encuentran en su cartera de clientes. Luego, crear diccionarios que se adapten a cada una de las industrias para expandirlo a todos los clientes que requieran mayor profundidad en su análisis, con el objetivo de no requerir grandes inversiones en equipos humanos que clasifiquen las opiniones de los usuarios.

Se sugiere a la empresa no filtrar las preguntas de la encuesta con una pregunta previa sobre si ha usado o no cierto punto de contacto, ya que esto genera una serie de nulos que son problemáticos en los análisis. Se recomienda solo dejar la opción de “No sabe/ no responde” , para evitar -por ejemplo- que un usuario con mala evaluación de las sucursales no entregue su percepción sobre estas, ya que por la misma razón no la ha visitado en los últimos 6 meses.

Si se tienen menos nulos, se pueden realizar imputaciones más limpias que permitan probar otros modelos de predicción de lealtad como regresiones logísticas ordinales o modelos estructurales, que fueron descartados en este estudio por la abrumadora cantidad de nulos

que existían. Además, esto permitiría rescatar otras variables que fueron descartadas por el mismo motivo, como la evaluación del cliente con respecto al financiamiento o la atención telefónica.

# Bibliografía

- [1] Ipsos, “Orgullosos de ser ipsos.” [https://www.ipsos.com/sites/default/files/2017-02/Proud\\_to\\_be\\_Ipsos-Peru.pdf](https://www.ipsos.com/sites/default/files/2017-02/Proud_to_be_Ipsos-Peru.pdf), Mayo 2015.
- [2] S. Vivanco Z., *La investigación de mercados. Un análisis de sus principales características como herramienta de información sobre la conducta social en contextos de consumo y espacio de desarrollo profesional para los/as sociólogos*. PhD thesis, Universidad de Chile, 2010.
- [3] M. Benassini, *Introducción a la Investigación de Mercados: Enfoque para América Latina*. Pearson, 2009.
- [4] R. J. Best, *Marketing estratégico*. Pearson Prentice Hall, 2007.
- [5] P. Kotler, D. Cámara, and I. Cruz Roche, *Marketing*. Pearson: Prentice Hall, 2004.
- [6] P. Chapman, *CRISP-DM 1.0: Step-by-step Data Mining Guide*. SPSS, 2000.
- [7] M. Montes y Gómez, “Minería de texto: Un nuevo reto computacional,” *Instituto Nacional de Astrofísica, Óptica y Electrónica*, 2001.
- [8] J. A. Balazs and J. Velásquez D., “Opinion mining and information fusion: A survey,” Junio 2015.
- [9] J. Mansilla, “Knowledge discovery in text.” <https://www.linkedin.com/pulse/knowledge-discovery-text-javier-mansilla/>, Diciembre 2018.
- [10] T. Ganegedara, “Intuitive guide to latent dirichlet allocation.” <https://towardsdatascience.com/light-on-math-machine-learning-intuitive-guide-to-latent-dirichlet-allocation-437c81220158>, Agosto 2018.
- [11] J. Aldunate Castillo, “Modelo de fuga con minería de texto en diferentes canales para empresa de retail financiero,” 2018.
- [12] D. I. Iribarra, *Predicción y descripción de la exclusión educativa del sistema escolar regular chileno, ciencia de datos para la innovación pública*. Universidad de Chile, 2018.
- [13] T. Chen and C. Guestrin, *XGBoost: A scalable Tree Boosting System*. University of Washington, 2016.
- [14] D. Singh, Sukhpreet, N. Abdullah-Al, and R. Abbas, *Effective intrusion detection system using XGBoost*. MDPI, 2018.
- [15] C. D. Contreras, “Extracción de conocimiento nuevo desde los reclamos recibidos en el servicio ancional del consumidor mediante tecnicas de text mining,” 2014.
- [16] F. Medina and M. Galván, *Imputación de datos: Teoría y práctica*. CEPAL, 2007.

- [17] R. O. Calafati, *Estrategias para el tratamiento de datos faltantes ("missing data") en estudios con datos longitudinales*. Universitat Oberta de Catalunya, 2017.