



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

**Identificación de factores a mejorar para aumentar la  
recomendación, análisis de NPS para clientes neutros a partir de  
procesamiento de lenguaje natural**

**Leydi Agudelo Restrepo**

Universidad Nacional de Colombia  
Facultad de Minas, Departamento de Ciencias de la Computación y Decisión  
Medellín, Colombia  
2022

**Identificación de factores a mejorar para aumentar la  
recomendación, análisis de NPS para clientes neutros a partir de  
procesamiento de lenguaje natural**

**Leydi Agudelo Restrepo**

Trabajo de grado presentado como requisito parcial para optar al título de:

**Magister en Ingeniería Analítica  
Profundización**

Director:

Ph.D. Juan David Ospina Arango

Línea de Investigación:

Procesamiento de Lenguaje Natural

Universidad Nacional de Colombia

Facultad de Minas, Departamento de Ciencias de la Computación y Decisión

Medellín, Colombia

2022

## Resumen

Este trabajo presenta una metodología para identificación de factores a mejorar para aumentar la intención de recomendación de los clientes de una entidad financiera. Se utilizaron 7.000 verbatims de los clientes clasificados como neutros en la encuesta net promoter score, que corresponden a la información de los años 2019, 2020 y 2021. Se realizó procesamiento de esta información para obtener un corpus estructurado que representa el contenido de todos los verbatims. Se aplicó análisis de sentimientos con el objetivo de analizar los sentimientos subyacentes en los verbatims de los clientes, el cual permitió concluir que en cada uno de los años analizados más de un 80% de los clientes tenían un sentimiento negativo asociado a su verbatim. Se entrenaron diferentes algoritmos con la intención de encontrar los temas más relevantes para los clientes, los modelos entrenados fueron, asignación latente de dirichlet, indexación semántica latente y factorización de matriz no negativa, los modelos fueron evaluados utilizando el enfoque de la maximización de la coherencia, combinado con juicio humano para validar los tópicos en el contexto del negocio. Entre los modelos entrenados el que mejores resultados arrojó fue la asignación latente de dirichlet, alcanzando la coherencia máxima para dos tópicos, los cuáles fueron caracterizados como costos y servicio en los canales de atención. Combinando el resultado del modelado de tópicos con el análisis de sentimientos, se descubrió que el sentimiento que predomina por parte de los clientes en los dos tópicos es negativo, y para algunos casos es positivo o neutro.

**Palabras clave:** procesamiento de lenguaje natural, análisis de sentimientos, modelado de tópicos, net promoter score, asignación latente de dirichlet, indexación semántica latente, factorización de matriz no negativa.

## Abstract

### Identification of factors to be improved to increase recommendation, NPS analysis for neutral customers based on natural language processing

This paper presents a methodology for identifying factors to be improved in order to increase the recommendation intention of customers of a financial institution. We used 7,000 verbatims from customers classified as neutral in the net promoter score survey, corresponding to information from the years 2019, 2020 and 2021. This information was processed to obtain a structured corpus representing the content of all verbatims. Sentiment analysis was applied with the objective of analyzing the underlying sentiments in the customer verbatim, which allowed concluding that in each of the years analyzed more than 80% of the clients had a negative sentiment associated with their verbatim. Different algorithms were trained with the intention of finding the most relevant topics for the customers, the trained models were, latent dirichlet assignment, latent semantic indexing and non-negative matrix factorization, the models were evaluated using the coherence maximization approach, combined with human judgment to validate the topics in the context of the business. Among the trained models, the one that yielded the best results was the latent dirichlet assignment, reaching the maximum coherence for two topics, which were characterized as costs and service in the service channels. Combining the results of the topic modeling with the sentiment analysis, it was found that the predominant customer sentiment for the two topics is negative, and for some cases it's positive or neutral.

**Keywords:** Natural language processing, sentiment analysis, topic modeling, net promoter score, latent dirichlet assignment, latent semantic indexing, non-negative matrix factorization.

<b>Contenido</b>	<b>pág</b>
Resumen	3
Palabras clave	3
Abstract	4
Keywords	4
Lista de figuras	7
Lista de tablas	8
Lista de abreviaturas	9
Introducción	10
1. Justificación	11
2. Objetivos	14
2.1 Objetivo general	14
2.2 Objetivos específicos	14
3. Estado del arte	15
3.1 Aprendizaje automático y procesamiento de lenguaje natural	15
3.2 Procesamiento de la información	17
3.2.1 Normalización del texto	18
3.2.2 Tokenización	18
3.2.3 Stopwords	19

3.2.4 n-gramas	19
3.2.5 Lematización y derivación	19
3.2.6 Bolsa de palabras	20
3.2.7 Corpus	20
3.3 Análisis de sentimientos	20
3.4 Modelado de tópicos	22
3.5 Técnicas de evaluación de modelos de tópicos	25
3.5.1 Evaluación por juicio humano	26
3.5.2 Evaluación por métricas cuantitativas	28
4.Encuesta Net Promoter Score y muestra	32
5.Metodología	35
5.1 Extracción de información y análisis de sentimientos	37
5.2 Procesamiento de la información	39
5.3 Modelado de tópicos	40
5.3.1 Optimización de hiperparámetros	41
5.4 Cantidad de tópicos óptimos y modelo seleccionado	46
6.Resultados y Análisis	49
6.1 Caracterización de los tópicos	49
6.2 Sentimiento dominante en cada tópico	54
6.3 Vinculación de los resultados con la realidad de los clientes	56
7.Conclusiones y recomendaciones	64
7.1 Conclusiones	64
7.2 Recomendaciones	65
Bibliografía	66

<b>Lista de figuras</b>	<b>Pág.</b>
Figura 1: clasificación de NPS	12
Figura 2: variación de la calificación en el NPS, para clientes en diferentes periodos	14
Figura 3: extracción de información y análisis de sentimientos	37
Figura 4: procesamiento de información	39
Figura 5: modelado de tópicos	40
Figura 6: coherencia vs número de tópicos para LDA	47
Figura 7: coherencia vs número de tópicos para LSI	48
Figura 8: coherencia vs número de tópicos para NMF	48
Figura 9: mapa de distancia entre tópicos	50
Figura 10: términos más frecuentes en el corpus	51
Figura 11: términos más frecuentes en el tópico costos	52
Figura 12: términos más frecuentes del tópico servicio en canales de atención	53
Figura 13: comportamiento del número de productos en la compañía, un año después de la encuesta, para clientes neutros	57
Figura 14: comportamiento del número de productos en la compañía, un año después de la encuesta, para clientes promotores	58
Figura 15: comportamiento en el número de productos en otras entidades financieras, un año después de la encuesta, para clientes neutros	59
Figura 16: comportamiento en el número de productos en otras entidades financieras, un año después de la encuesta, para clientes promotores	59
Figura 17: clientes neutros desertores un año después de la encuesta	60
Figura 18: clientes promotores desertores un año después de la encuesta	61
Figura 19: cumplimiento en el tiempo de promesa, en el año anterior a la encuesta	63

## Lista de tablas

	<b>Pág.</b>
Tabla 1: Distribución de los datos	34
Tabla 2: librerías disponibles para NLP en Python	36
Tabla 3: sentimientos identificados en los verbatims	38
Tabla 4: diccionario de hiperparametros para LDA	42
Tabla 5: resultado obtenido aplicando LDA para el modelado de tópicos. El mejor resultado se resalta en azul	43
Tabla 6: diccionario de hiperparametros para LSI	44
Tabla 7: resultado obtenido aplicando LSI para el modelado de tópicos. El mejor resultado se resalta en azul	44
Tabla 8: diccionario de hiperparametros para NMF	46
Tabla 9: resultado obtenido aplicando MNF para el modelado de tópicos. El mejor resultado se resalta en azul	46
Tabla 10: mejores coherencias obtenidas	49
Tabla 11: distribución de los clientes en los tópicos	53
Tabla 12: sentimiento dominante por tópico	56



## Lista de abreviaturas

<b>Abreviatura</b>	<b>Término</b>
LDA	Asignación latente de Dirichlet
LSI	Indexación semántica latente
NMF	Factorización de matriz no negativa
BTM	Modelo de tópicos bitérmino
NPS	Net Promoter Score
NLP	Natural Language Processing

## Introducción

El net promoter score (NPS) es una de las herramientas usadas para medir la satisfacción de los clientes con los productos y servicios ofrecidos por una compañía y consiste en realizar una encuesta a una muestra de clientes con una periodicidad definida. A los clientes seleccionados en la muestra se les pregunta, en una escala del 0 al 10, “¿Qué tan probable es que recomiende la empresa X a un amigo o colega?” Las respuestas permiten clasificar a los clientes en promotores, neutros y detractores. Además, la calificación se complementa con la justificación del cliente y esa respuesta es almacenada en lenguaje natural.

El NPS es importante porque el voz a voz que generan los clientes con intención de recomendar puede verse reflejado en un crecimiento de la empresa [34], pero el voz a voz de los clientes con malas experiencias también puede afectar el crecimiento de esta. Por eso las empresas deberían preocuparse por indagar sobre las razones de insatisfacción de los clientes detractores y neutros y saber por qué asignan esta calificación, estas razones se pueden hallar con un adecuado análisis de los verbatims.

El análisis de esta información, almacenada como verbatims, requiere técnicas para datos no estructurados, como las técnicas de procesamiento de lenguaje natural (NLP). Entre las que se encuentran la tokenización, la lematización, el análisis de sentimientos para detectar la polaridad en el texto, la segmentación de oraciones, modelado de tópicos, resumen de texto, descomposición, entre otras técnicas que son muy útiles para ayudar a la computadora a comprender el texto con mayor claridad y no solo interpretar por palabras, sino darles un contexto a esas palabras.

En este trabajo se propone una metodología estándar para el análisis de los verbatims, y se aplica a un caso de estudio de una compañía del sector financiero, de la cual se tiene información disponible de 7.000 verbatims de los clientes clasificados como neutros en los años 2019, 2020 y 2021. La aplicación de técnicas de NLP como análisis de sentimientos y modelado de tópicos sobre esta información, permitirá identificar los factores que más influyen en la satisfacción e insatisfacción, para así saber cuáles son los aspectos de la compañía que están impactando negativa y positivamente la opinión de los clientes, la metodología propuesta también permite identificar para cada cliente cuál es el tema que más impacta su opinión y el sentimiento asociado a ese tema.

## 1. Justificación

La satisfacción de los clientes puede ser medida por varios instrumentos, uno de ellos es el Net Promoter Score (NPS). El NPS refleja la intención de recomendación y preferencia por parte de los clientes, entonces si se logra aumentar la preferencia de los clientes, se logra aumentar el NPS de la empresa. Una forma de aumentar la recomendación y la preferencia de los clientes, es identificando factores en los cuales los clientes tienen sentimientos negativos relacionados.

Los detractores son clientes que dan una mala calificación en la encuesta, probablemente motivados por una mala experiencia que han tenido con la empresa, por lo que llevar estos clientes a ser promotores puede ser realmente difícil. Pero los clientes neutros si pueden representar una oportunidad para aumentar la preferencia e intención de recomendación, por las siguientes razones:

1. Es más fácil llevar un cliente neutro a ser promotor que un cliente detractor a ser promotor, porque de acuerdo a la escala de la encuesta, los clientes neutros están más cerca de ser promotores, que los detractores, como se muestra en la figura 1. Además, por definición un cliente neutro es un cliente no leal, del cual la empresa podría ganarse su lealtad, pero un cliente detractor si es un cliente insatisfecho por alguna experiencia negativa.

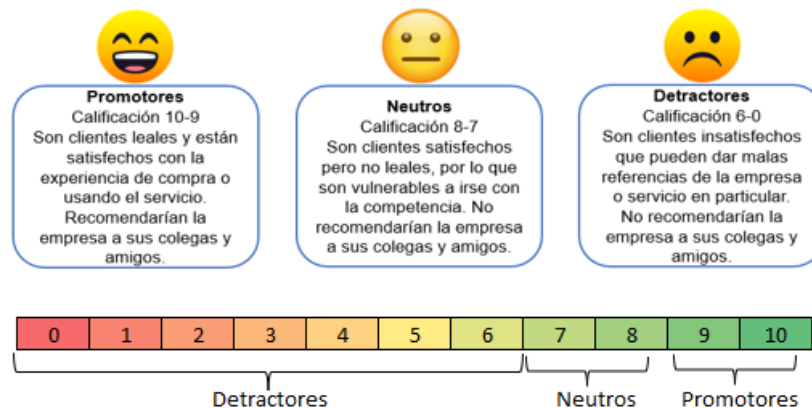


Figura 1: clasificación de NPS

2. Usualmente a las empresas les preocupa más los clientes detractores y sus estrategias están más orientadas a ellos. Tal vez piensan que los clientes neutros al no tener una opinión muy definida hacia lo negativo o lo positivo no tendrán ningún impacto en su crecimiento. Pero cuando se hace una revisión detallada de los verbatims en el NPS de estos clientes se encuentra que algunos realmente tienen una opinión neutra porque no han vivido ninguna experiencia significativa con los servicios que le presta la empresa, pero en otros si se identifican inconformidades. Por ejemplo, los siguientes verbatims, que corresponden a clientes neutros, en encuestas realizadas en el año 2019:
  - “Los costos de mantenimiento son muy elevados. La atención presencial en oficinas es exageradamente demorada.”
  - “El costo de manejo de las cuentas es muy alto”
  - “Más agilidad en la atención, oficinas muy congestionadas”
  
3. En la compañía, se ha estudiado la variación en la calificación de los clientes que han salido en la muestra para la encuesta en diferentes periodos, en donde se ha evidenciado que una parte de clientes neutros en una aplicación de la encuesta para alguna aplicación posterior son detractores, entonces, si la empresa identifica los factores a mejorar en base a los verbatims, probablemente esos clientes que son neutros no se conviertan en detractores sino en promotores. Obtener un mismo cliente en dos o más encuestas es difícil, ya que la selección de estos se hace aleatoriamente, pero entre las encuestas realizadas en los años 2019, 2020 y 2021 han salido 897 clientes en la muestra de al menos dos periodos. A continuación, se representa el cambio en la calificación asignada por estos clientes en el NPS, en relación a los clientes neutros, se encuentra que un 21% de estos eran neutros y pasaron a ser detractores, un 3% eran neutros y siguen siendo neutros y un 10% eran neutros y pasaron a ser promotores. Esto muestra que en efecto, una parte de clientes neutros se convierte en detractores con el tiempo, si las causales de su insatisfacción son recurrentes y no son atendidas.

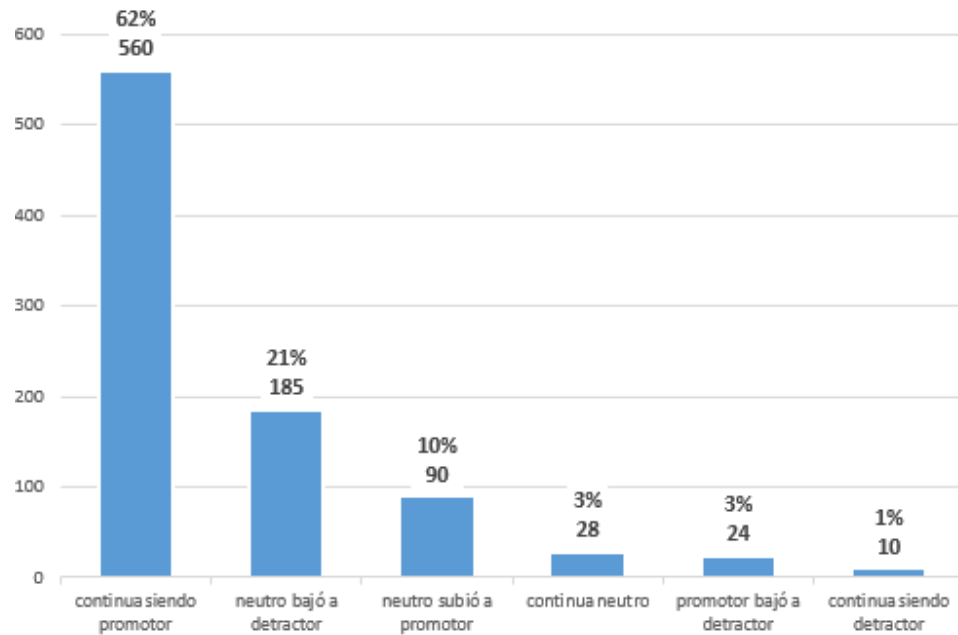


Figura 2: variación de la calificación en el NPS, para clientes en diferentes periodos

La solución que se propone para aumentar la intención de recomendación de los clientes es la identificación de factores a mejorar por medio del análisis de los verbatims de los clientes neutros, aplicando NLP ya que estos están expresados en lenguaje natural.

## **2. Objetivos**

### **2.1 Objetivo general**

Desarrollar una herramienta que permita priorizar estrategias para aumentar la intención de recomendación y la satisfacción de los clientes de una compañía del sector bancario, medidas por el NPS.

### **2.2 Objetivos específicos**

1. Identificar oportunidades de mejora en el servicio prestado por la empresa en los verbatims de la encuesta NPS, de los clientes neutros.
2. Encontrar los factores que más influyen en la satisfacción e insatisfacción del cliente.
3. Determinar cuáles son los clientes neutros con dolencias/inconformidades, ya que son clientes potenciales para convertirse en promotores.

## **3.Estado del arte**

### **3.1 Aprendizaje automático y procesamiento de lenguaje natural**

El procesamiento de lenguaje natural (NLP) trata de cómo interpretar el lenguaje que usan los humanos en su cotidianidad usando computadoras. Es una mezcla entre ciencias de la computación, inteligencia artificial y lingüística, el objetivo final es que las computadoras comprendan el contenido, los matices y el sentimiento en un documento escrito en lenguaje natural. Usando NLP se pueden extraer las características más relevantes en un texto, y organizarlas por categorías de sentimientos o percepciones, que pueden ser útiles para clasificación de clientes, identificación de factores a mejorar en los productos y/o servicios prestados por cualquier empresa.

El aprendizaje automático con un enfoque en NLP ha sido aplicado a múltiples contextos, y muestra ser muy útil para generar valor para el negocio, a partir de datos no estructurados, como lo son respuestas a preguntas abiertas en encuestas, comentarios en redes sociales, conversaciones telefónicas transcritas o reseñas en internet.

En un estudio publicado en la revista International Journal of Computational Intelligence Systems en el 2021 [27], se propuso una metodología para analizar las reseñas online y obtener oportunidades de mejora sobre sus productos, específicamente para una empresa de comercio electrónico en China. Se realiza un análisis de sentimientos para identificar los sentimientos y su intensidad en las reseñas para teléfonos móviles, teniendo en cuenta las propiedades dinámicas y diversificadas de los requerimientos de los clientes. Los resultados fueron útiles para los fabricantes identificar oportunidades de mejora en sus productos y poder dirigir algún producto a un grupo especial de clientes de acuerdo con los

requerimientos identificados en el análisis y así mejorar de manera eficiente la satisfacción del cliente.

Muktafin, E. H. en [28] utilizaron aprendizaje automático y NLP para clasificar clientes mediante conversaciones telefónicas transcritas en satisfechos e insatisfechos. Evaluar la satisfacción del cliente resulta muy importante para cualquier empresa, esta satisfacción o insatisfacción la puede dar a conocer el cliente en una encuesta o en una crítica escrita, que la empresa puede utilizar para identificar posibles mejoras, pero cuando solo unos cuantos clientes están dispuestos a responder la encuesta no se puede decir que esto representa el sentimiento general de los clientes, en estos casos se hace necesario buscar una fuente con más información, para este estudio fueron las conversaciones telefónicas. El objetivo era construir un sistema que no solo transcribiera las conversaciones telefónicas, sino que clasificara a los clientes en satisfechos e insatisfechos y permitiera obtener información para la evaluación y mejora de los servicios.

En este estudio utilizaron la técnica de los vecinos más cercanos (KNN) con un enfoque de NLP y el mejor resultado fue un modelo con 74% de exactitud, un 76% de precisión, y un 73,08% de recall, el modelo tiene un buen rendimiento ya que acierta el 74% de las veces entre las dos clases, de los clientes que clasifica como satisfechos el 76% realmente están satisfechos, se equivoca 24% de las veces que predice que el cliente está satisfecho y es capaz de identificar un 73,08% de los clientes que están satisfechos. El hallazgo principal fue que en las conversaciones con etiqueta "satisfecho" se identifica un sentimiento de satisfacción con el servicio y el cumplimiento de los deseos del cliente y en las conversaciones con la etiqueta "no satisfecho" los clientes no están satisfechos con el tiempo de espera, lo que revela un aspecto a mejorar para la empresa.

En una investigación [13], del año 2020 publicada en la revista *Computers in Biology and Medicine*, se muestra una aplicación para la detección del uso indebido de opioides, analizando comentarios de twitter. El uso indebido de opioides es un problema de salud importante en Estados Unidos, el cual puede provocar adicción, por eso es importante caracterizar el uso indebido de opioides, más allá de saber quién los usa indebidamente es encontrar los motivos del uso indebido.

Los investigadores, primero recopilaron datos de Twitter utilizando palabras clave relacionadas con los opioides, luego etiquetaron manualmente 6.988 tweets en las clases de consumo indebido de opioides por dolor, uso indebido recreativo y



no uso. Entrenaron y evaluaron clasificadores multiclase con un conjunto de prueba de 1.677 tweets. El modelo que logró el mejor rendimiento fue un XLNet con un F1-Score de 71% para la clase de uso indebido para el dolor y 79% para la clase de uso indebido recreativo. Lograron encontrar un modelo con un rendimiento prometedor, incluso detectando las diferencias sutiles entre el uso indebido relacionado con el dolor y el relacionado con la recreación. El aprendizaje automático y el NLP los ayudó a identificar palabras clave asociadas a las tres categorías, para permitir potencialmente el estudio de las características de las personas que exhiben tal comportamiento.

Estos estudios demuestran que los datos no estructurados como los de reseñas, encuestas con preguntas abiertas o redes sociales, tienen un gran potencial, que con las técnicas adecuadas de NLP, análisis de sentimientos y aprendizaje automático se puede extraer información relevante para las empresas, como factores a mejorar, factores que influyen en la satisfacción y la insatisfacción, las características que son más importantes para los clientes en sus productos y construir herramientas útiles y eficaces para proponer estrategias o productos adaptados a necesidades puntuales de los consumidores, que les permite aumentar la satisfacción del cliente de una manera rentable.

## 3.2 Procesamiento de la información

El tipo de información que se tiene en este trabajo es no estructurada, la cual corresponde a los verbatims del NPS. Para trabajar con bases de datos no estructuradas es necesario un procesamiento de esta información, en esta se debe convertir la información a entidades de longitud fija antes de usarlas como entrada en los algoritmos de aprendizaje automático. Para esto se lleva toda la información contenida en los verbatims a un corpus general, el cual es un conjunto estructurado de ejemplos de los términos usados en toda la colección de textos, este conjunto puede contener palabras o n-gramas.

La metodología que se recomienda seguir para la construcción del corpus es, eliminación de caracteres extraños y stopwords, búsqueda de n-gramas basados en asociaciones entre palabras, aplicación de lematización o derivación de los términos, construcción de un diccionario y finalmente del corpus que representa toda la información contenida en los documentos.

### 3.2.1 Normalización del texto

Usualmente los documentos contienen caracteres extraños que no aportan valor al contenido del texto y que por el contrario introducen ruido y hacen mucho más difícil la identificación de posibles tópicos. Por esto es necesario eliminar estos caracteres, los cuales pueden contener signos de puntuación extras, correos electrónicos, espacios extras entre palabras, saltos de línea o palabras de un solo carácter.

La idea principal es normalizar el texto, por eso también se suele convertir todo el texto a minúsculas, para que dos palabras no se identifiquen como diferentes solo por una tener mayúscula inicial y la otra no.

Ejemplo:

El **banco** tiene tasas altas para los créditos.

El **Banco** tiene tasas altas para los créditos.

Las dos frases anteriores tienen igual significado, sin embargo, la segunda tiene doble espacio luego de El y eso las hace diferentes, además la palabra banco en la segunda oración tiene mayúscula inicial y en la primera no, esto también hace diferentes estas dos palabras. Luego de la normalización estas dos frases serán consideradas iguales.

### 3.2.2 Tokenización

En [5] los autores definen la tokenización como el proceso que consiste en dividir todo el texto en una lista de tokens, que pueden ser palabras, oraciones, caracteres, entre otros. Gracias a la tokenización se puede reducir la tarea de búsqueda y reducir el uso del espacio de almacenamiento. Hay varias técnicas de tokenización, entre las cuales está el algoritmo de Porter

Ejemplo:

El banco tiene tasas muy altas para los créditos.

Luego de la tokenización por palabras la frase anterior queda como:

['El', 'banco', 'tiene', 'tasas', 'altas', 'para', 'los', 'créditos']

### 3.2.3 Stopwords

Las stopwords o palabras vacías, son palabras de alta frecuencia en todos los documentos, por lo tanto, no aportan un diferencial entre un documento y otro. De acuerdo con [5], las palabras vacías suelen tener poco contenido léxico y su presencia en un texto no logra distinguirlo de otros textos.

Las librerías Spacy y Nltk de python proporcionan una lista de stopwords entre sus funcionalidades para el idioma español, para este caso se decide usar Nltk, estas son algunas stopwords que contiene:

['la', 'el', 'una', 'has', 'hubiéramos', 'habido', 'habrían', 'le', 'tuviéramos', 'estén', 'estuviera', 'sus', 'serán', 'tendré', 'estando', 'mí', 'estuvieran']

Sin embargo, es necesario complementar la lista de stopwords, ya que la proporcionada para español por Nltk es corta.

### 3.2.4 n-gramas

Los n-gramas ayudan a incluir co-ocurrencias de palabras dentro del análisis de los documentos. En [46] se definen los n-gramas como una cadena de  $n$ , generalmente adyacentes, caracteres extraídos de una sección de texto continuo. De acuerdo a la longitud de los documentos, puede ser interesante trabajar con bigramas, trigramas, ...,etc.

Continuando con el mismo ejemplo de subsección 3.2.2, un bigrama en esa frase podría ser, `tasas_altas`.

### 3.2.5 Lematización y derivación

La derivación es el proceso de reducir las palabras a su raíz o quitar los sufijos.

La lematización se refiere a hacer uso adecuado del vocabulario y al análisis morfológico de las palabras. La lematización devuelve a las

palabras la forma base, o lo que se conoce como lema, después de comprender el contexto de la palabra en el documento.

La derivación y la lematización son métodos utilizados por los motores de búsqueda y los chatbots para analizar el significado de una palabra.

### **3.2.6 Bolsa de palabras**

Esta técnica se utiliza para preprocesar texto y extraer todas las características de un documento para usarlas en el modelado. También proporciona una representación del documento que explica la ocurrencia de las palabras, sin importarle su ubicación dentro de una oración en particular sino solamente si las palabras aparecen en el texto.

### **3.2.7 Corpus**

En lingüística y NLP, corpus (cuerpo en latín) se refiere a una colección de textos. Un corpus lingüístico es un conjunto amplio y estructurado de ejemplos reales de uso de la lengua. Estos ejemplos pueden obtenerse de textos escritos (los más comunes), o muestras orales (generalmente transcritas) [47].

## **3.3 Análisis de sentimientos**

Pang y Lee, definen en [61], el análisis de sentimiento como “el tratamiento computacional de opiniones, sentimientos y subjetividad en textos”, mientras en [62] lo definen como una técnica de investigación que analiza los sentimientos de una muestra determinada de textos que, normalmente, proviene de entornos digitales, como plataformas online o redes sociales. El análisis de sentimiento hace parte del campo del NLP y es el proceso que busca evaluar y determinar el sentimiento capturado en una selección de texto, con el sentimiento definido como sentimiento o emoción. Este sentimiento puede ser simplemente positivo, negativo o neutral. Para realizar análisis de sentimiento pueden aplicarse métodos con diccionarios de emociones preestablecidos, también conocidos como lexicón, o métodos como aprendizaje automático.

En una investigación publicada en el 2021, en la revista Journal of Business Management [26], realizaron análisis de sentimientos a las reseñas de

usuarios en línea con el objetivo de identificar factores importantes que la propuesta de valor debía incluir. Los datos usados para el caso de estudio fueron 3.865 comentarios de los compradores de Smart tv del sitio web de la tienda Digikala. En el estudio pretendían responder 2 preguntas: 1) ¿cuáles son los componentes de la propuesta de valor? y 2) ¿Cuál es el papel de cada factor de valor en el sentimiento de satisfacción o insatisfacción del cliente?

En el análisis de sentimientos se identificaron 30 factores importantes para la propuesta de valor en relación con el producto Smart tv. Quince elementos tuvieron un impacto positivo en la satisfacción del cliente, como tamaño de la pantalla, marca del producto, precio razonable, características y opciones, calidad del sonido, entre otras características del producto, 15 elementos tuvieron un impacto negativo en la propuesta de valor produciendo insatisfacción del cliente, como baja eficiencia del producto, empaque inadecuado, problemas de garantía, entre otros. En este caso el análisis de sentimientos permitió hallar los componentes más relevantes de la propuesta de valor del producto Smart tv, e identificar cuáles de estos componentes influyen en la satisfacción y cuáles en la insatisfacción del cliente con este producto.

Una de las principales fuentes de información no estructurada son las redes sociales, por eso en la bibliografía se encuentran muchas aplicaciones de NLP a datos extraídos de estas. En [63], realizaron análisis de sentimientos usando aprendizaje automático, sobre 72.564 datos extraídos de twitter acerca del aislamiento social obligatorio, como medida de prevención del Covid-19 adoptada por el gobierno. El objetivo era analizar los sentimientos subyacentes de los comentarios de Twitter relacionados con el aislamiento, identificando los temas y palabras más frecuentes utilizados en este contexto. Se probaron algoritmos como Naive Bayes, Naive Bayes kernel y Deep Learning, en las pruebas el algoritmo con mejores resultados fue el de Deep Learning. Clasificaron las opiniones en los sentimientos alegría, ira, miedo y tristeza, el sentimiento predominante de los usuarios de Twitter con respecto a la cuarentena fue el miedo con 49% de los comentarios, luego aparece la tristeza con 22%; en tercer lugar, la ira con 16% y, finalmente, la alegría con un 13%. En este contexto el análisis de sentimientos se usó para explotar la información que se encuentra en la red social twitter y descubrir los

sentimientos de los usuarios relacionados con el aislamiento social obligatorio.

Actualmente en Python, hay varias implementaciones disponibles para realizar análisis de sentimientos para diferentes idiomas, las librerías NLTK, TextBlob, y sentiment-analysis-spanish hacen un buen trabajo, pero para texto en español una buena opción es sentiment-analysis-spanish. La solución implementada en sentiment-analysis-spanish es una red neuronal convolucional entrenada con 800.000 reviews de las siguientes páginas: eltededor, decathlon, tripadvisor, filmaffinity and ebay [60].

### 3.4 Modelado de tópicos

El modelado de tópicos es una técnica de procesamiento de lenguaje natural y aprendizaje automático que se usa para determinar los temas presentes en un documento. Es capaz de determinar la probabilidad de que una palabra o frase pertenezca a un determinado tema y agrupar documentos en función de su similitud o cercanía. Lo hace analizando la frecuencia de palabras y frases en los documentos. Algunas aplicaciones de modelado de temas también incluyen resúmenes de texto, sistemas de recomendación, filtros de spam y similares.

Asignación latente de Dirichlet (LDA, por sus siglas en inglés) es una técnica de aprendizaje automático no supervisada que se usa comúnmente para el análisis de texto. Es un tipo de modelado de temas en el que las palabras se representan como temas y los documentos se representan como una colección de estos temas de palabras.

Risch en [48], define LDA como un modelo de tópicos generativo. Este modelo asume que cada documento de una colección contiene varios tópicos y la distribución de estos tópicos para cada documento es generada a partir de una distribución de Dirichlet, cada documento tendrá un tópico dominante y esto es determinado por lo pesos que tengan estos. Blei, Ng, & Jordan en [49] definen que en un modelo LDA cada documento es una mezcla de varios tópicos donde cada documento puede tener varios tópicos asignados mediante LDA.

En [50], Hammoe realiza un interesante trabajo con datos extraídos de twitter a cerca de compañías de telecomunicaciones de Argentina, con el objetivo de aplicar aprendizaje automático e identificar los tópicos de los que se hablan

acerca de dichas compañías. En ese trabajo el autor decide aplicar el algoritmo LDA, a pesar de que estudió otros, porque tiene un gran número de aplicaciones y es el más probado en este tipo de problemas, además desarrolla la aplicación en lenguaje Python con la librería Gensim.

Para entrenar los modelos, primero se procesaron 31.478 tweets del mes de Julio de 2018 vinculados a importantes compañías telefónicas. Para elegir el mejor modelo entre los probados, se utilizó la coherencia. Se probaron ocho modelos con diferentes números de tópicos y finalmente se eligió el modelo que proporcionó la mayor coherencia, justo antes de que el crecimiento de la coherencia dejará de ser significativo, este modelo fue uno con 38 tópicos, en los cuales se lograron identificar tópicos como reclamos por premios o promociones, respuesta a un problema inconcluso, problemas técnicos o clientes insatisfechos con voluntad de cambiar de empresa.

La Indexación semántica latente (LSI, por sus siglas en inglés) fue propuesta por Thomas K. Landauer (Pearson Knowledge Technologies) y Susan Dumais (Microsoft) entre otros, quienes presentaron la LSI como una nueva teoría general de adquisición y representación del conocimiento [51]. LSI consiste en un modelo estadístico que compara las similitudes semánticas entre segmentos de información textual compuesta por lexemas. Con LSI se espera determinar la relevancia de una palabra en relación con la totalidad del documento, otras partes o términos de este, basados en su ocurrencia o distancia entre ellos. El principio de LSI se basa en que muchas palabras utilizadas en varios textos pueden tener significados similares. La idea principal es emparejar lexemas considerando sus conceptos asociados y no solo por términos [52].

En [52], el objetivo es construir un modelo que permita clasificar textos informáticos considerando sus conceptos asociados y no solo sus términos. Usando LSI lograron construir un prototipo de documento para cada clase creada, y así al ingresar un documento nuevo se realiza la comparación con el prototipo de cada clase y se asignará a la clase con la cual tenga mayor similitud, y cuando no existe similitud con ninguna clase existente se crea una nueva clase.

Otra técnica para el modelado de tema es la factorización de matriz no negativa (NMF, por sus siglas en inglés), es una técnica de reducción de dimensionalidad que fue propuesta por Lee en [53] la cual consiste en

descomponer una matriz de números no negativos en dos matrices, una matriz base  $W$  y una matriz de expansión  $H$ .

Las matrices  $W$  y  $H$  se obtienen de la factorización de la matriz de entrada. Estos valores se comparan con la matriz de entrada mediante la divergencia generalizada de Kullback-Leiber, la cual es una métrica para medir la discrepancia entre los datos de la matriz original y la aproximación que se obtiene con la factorización. Para la optimización de valores se aplica el descenso del gradiente proyectado para evitar que los números se conviertan en negativos, pero también es lo que permite que los valores se aproximen al mínimo error. El hecho de que esta técnica use el descenso del gradiente es lo que permite que sea un modelo no supervisado, este algoritmo itera en los datos hasta conseguir la convergencia [53].

Los autores en [54], realizaron modelamiento de tópicos aplicados al estudio de ataques de Ingeniería social, en el cual aplicaron varios algoritmos de modelado de tópicos, LDA, LSI, HDP y NMF. Los resultados obtenidos para la coherencia en los cuatro modelos elegidos fueron cercanos entre sí, se seleccionaron los modelos LDA y HDP como los modelos que ofrecen una mejor adaptación a la naturaleza de los datos de mensajería instantánea, porque tienen una mejor precisión en la predicción del conjunto de datos real de ataques de ingeniería social.

En el trabajo de Yan, Guo, Lan, & Cheng [55], se plantea el modelo bitérmino (BTM, por sus siglas en inglés) para manejar el problema de los textos cortos ya que en estos textos se presenta baja co-ocurrencia de términos de un documento o esta es más esporádica, lo que dificulta el hallazgo de los temas más relevantes. El modelo de bitérminos puede lidiar con este problema porque aprende los tópicos modelando la generación de todos los bitérminos que contenta el corpus. Esta idea está sustentada en que los tópicos son grupos de palabras correlacionadas y que esta correlación es dada por patrones de ocurrencia de las palabras en los documentos. Estos patrones de ocurrencia también permitirán mejorar el problema dado en los textos cortos por la dispersión de los términos.

En textos cortos, se toma a cada documento como una unidad individual de contexto. Luego se extraen todos los bitérminos de cada documento y estos componen los datos de entrenamiento de BTM.



En la mayoría de las aplicaciones de modelado de tópicos, la librería usada de python es Gensim [64], que es una librería especializada en modelado de tópicos para humanos. Gensim permite tareas de preprocesamiento de información, entrenar modelos semánticos de NLP a gran escala y encontrar similitudes entre documentos. Otra librería muy útil para realizar clasificación de textos con modelado de temas es scikit-learn [65], sin embargo, Gensim trae implementados muchos más algoritmos para esta tarea puntual de NLP.

### 3.5 Técnicas de evaluación de modelos de tópicos

La evaluación del modelo de tópicos es el proceso de evaluar qué tan bien el modelo hace aquello para lo que fue diseñado y puede ayudar a decidir si el modelo ha capturado la estructura interna de un corpus. De manera general, la evaluación del modelo de tópicos puede ayudar a responder preguntas como:

- ¿Son comprensibles los tópicos identificados?
- ¿Son los tópicos coherentes?
- ¿El modelo de tópicos cumple el propósito para el que se está utilizando?

Sin embargo, evaluar un modelo de tópicos no siempre es fácil. Si un modelo de tópicos se utiliza para una tarea medible, como la clasificación, entonces su efectividad es relativamente fácil de calcular, por ejemplo, midiendo la proporción de clasificaciones exitosas, en caso de que se tenga al menos una muestra de los datos etiquetada. Pero si el modelo se usa para una tarea cualitativa, como explorar los temas semánticos en un corpus no estructurado, entonces la evaluación es más difícil.

La evaluación de los modelos de tópicos busca comprender qué tan fácil es para los humanos interpretar los tópicos producidos por el modelo. Dicho de otra manera, la evaluación del modelo de tópicos se trata de la "interpretabilidad humana" o "interpretabilidad semántica" de los tópicos.

Hay varias formas de evaluar los modelos de tópicos. Éstas incluyen:

1. **Juicio humano:** basado en observación, p. Ej. observar las primeras  $N$  palabras en un tema o basado en interpretación, p. Ej.

'intrusión de palabras' e 'intrusión de temas' para identificar las palabras o temas que "no pertenecen" en un tema o documento.

**2. Enfoques mixtos:** combinaciones de enfoques cuantitativos y basados en juicios.

### 3.5.1 Evaluación por juicio humano

#### Enfoque basado en la observación

La forma más fácil de evaluar un tema es mirar las palabras más probables del tema. Esto se puede hacer, por ejemplo, enumerando las diez palabras principales de cada tema o en otros formatos. Una forma visualmente atractiva de observar las palabras probables en un tema es a través de nubes de palabras.

El modelado de tópicos no proporciona orientación sobre el significado de ningún tema, por lo que etiquetar un tema requiere una interpretación humana. Más allá de observar las palabras más probables en un tema, los investigadores de la Universidad de Stanford han desarrollado un enfoque más completo basado en la observación llamado "Termitas".

La termita se describe como "una visualización de las distribuciones término-tema producidas por modelos de tema" [44]. En esta descripción, "término" se refiere a una "palabra", por lo que las "distribuciones término-tema" son "distribuciones palabra-tema".

Termita produce visualizaciones significativas al introducir dos cálculos:

- Una medida de 'relevancia', que identifica palabras que son más relevantes para los temas en los que aparecen (más allá de las meras frecuencias de sus conteos)
- Un método de 'seriación', para clasificar las palabras en agrupaciones más coherentes según el grado de similitud semántica entre ellas

### **Enfoque basado en la interpretación**

Los enfoques basados en la interpretación requieren más esfuerzo que los enfoques basados en la observación, pero producen mejores resultados. Estos enfoques se consideran un "estándar de oro" para evaluar modelos de tópicos, ya que utilizan el juicio humano para lograr el máximo efecto.

Una buena ilustración de estos se describe en un artículo de investigación de Jonathan Chang y otros (2009) [45] que desarrolló la "intrusión de palabras" y la "intrusión de temas" para ayudar a evaluar la coherencia semántica.

#### **Intrusión de palabras**

En la intrusión de palabras, los sujetos se presentan con grupos de seis palabras, cinco de las cuales pertenecen a un tema determinado y una que no: la palabra "intruso". Se pide a los sujetos que identifiquen la palabra intruso.

Para entender cómo funciona esto, considere el grupo de palabras:

[ perro, gato, caballo, manzana, cerdo, vaca]

En este caso el intruso es la palabra manzana, ya que el resto de las palabras están relacionadas con animales, por lo tanto, un posible tema para este conjunto de palabras es animales.

Si considera:

[ coche, profesor, ornitorrinco, ágil, azul, Zaire]

¿Cuál es el intruso en este grupo de palabras?

Es mucho más difícil de identificar, por lo que la mayoría elige el intruso al azar. Esto implica poca coherencia en el tema.

#### **Intrusión de temas**

De manera similar a la intrusión de palabras, en la intrusión de temas se pide a los sujetos que identifiquen el tema "intruso" de los grupos de temas que componen los documentos.

En esta tarea, a los temas se les muestra un título y un fragmento de un documento junto con cuatro temas. Tres de los temas tienen una alta probabilidad de pertenecer al documento, mientras que el tema restante tiene una probabilidad baja: el tema "intruso".

En cuanto a la intrusión de palabras, el tema del intruso a veces es fácil de identificar y en otras ocasiones no. El éxito con el que los sujetos pueden elegir correctamente al intruso ayuda a determinar el nivel de coherencia.

### **3.5.2 Evaluación por métricas cuantitativas**

Si bien los métodos de evaluación basados en el juicio humano pueden producir buenos resultados, su ejecución resulta costosa y requiere mucho tiempo. Además, el juicio humano no está claramente definido y los humanos no siempre están de acuerdo sobre cuál es un buen tópico y cuál no. Por el contrario, el atractivo de las métricas cuantitativas es la capacidad de estandarizar, automatizar y escalar la evaluación de modelos temáticos.

#### **Perplejidad**

Esta métrica evalúa la capacidad de un modelo de tópicos para predecir un conjunto de prueba después de haber sido entrenado en un conjunto de entrenamiento.

Se calcula dividiendo un conjunto de datos en dos partes: un conjunto de entrenamiento y un conjunto de prueba. La idea es entrenar un modelo de tema utilizando el conjunto de entrenamiento y luego probar el modelo en un conjunto de prueba que contiene documentos no vistos anteriormente.

Aunque la métrica de perplejidad es una elección natural para los modelos de temas desde un punto de vista técnico, no proporciona buenos resultados para la interpretación humana. Esto fue demostrado por la investigación, nuevamente por Jonathan Chang

y otros (2009) en [45] que encontró que la perplejidad no logró transmitir bien si los temas son coherentes o no.

Al comparar la perplejidad con los enfoques de juicio humano como la intrusión de palabras y la intrusión de temas, la investigación mostró una correlación negativa. Esto significa que a medida que mejora la puntuación de perplejidad, la interpretabilidad humana de los temas y las mezclas de temas empeora, en lugar de mejorar. Por lo tanto, la métrica de la perplejidad parece ser engañosa cuando se trata de la comprensión humana de tópicos.

### **Coherencia**

Se llama a un tema coherente, si las palabras principales tienen sentido juntas. Un ejemplo de un tema coherente es {juego, deporte, pelota, equipo}, mientras que un tema incoherente podría ser {juego, deporte, pelota, pingüino} [56].

Una de las deficiencias de la perplejidad es que no capta el contexto, es decir, la perplejidad no captura la relación entre las palabras de un tema o los temas de un documento.

Para superar esto, se han desarrollado enfoques que intentan capturar el contexto entre palabras en un tema. Utilizan medidas como la probabilidad condicional de la co-ocurrencia de palabras en un tópico. Estos enfoques se denominan colectivamente "coherencia".

La puntuación de coherencia se usa en el modelado de tópicos para medir qué tan interpretables son los tópicos para los humanos. En este caso, los tópicos se representan como las primeras  $N$  palabras con mayor probabilidad de pertenecer a ese tópico en particular. Brevemente, el puntaje de coherencia mide qué tan similares son estas palabras entre sí y se dice que un tema tiene alta coherencia si las palabras que lo definen tienen una alta probabilidad de aparecer juntas (co-ocurrir) en todos los documentos.

### **Puntuación de coherencia CV:**

Una de las métricas de coherencia más populares se llama CV. Crea vectores de contenido de palabras utilizando sus co-ocurrencias y, después de eso, calcula la puntuación utilizando información mutua normalizada por puntos y la similitud del coseno. Esta métrica es popular porque es la métrica predeterminada en el módulo de pipeline de coherencia de temas en la librería Gensim de Python. Sin embargo, actualmente presenta problemas, y no se obtiene respuesta cuando se ejecuta el comando.

### **Puntuación de coherencia UMass:**

UMASS fue propuesto por Mimno [57]. UMASS es muy similar a UCI excepto que usa la probabilidad condicional entre los pares de palabras principales.

La coherencia UMass es la más rápida de calcular en la librería Gensim, por lo que se recomienda usar esta en lugar de CV. Esta puntuación calcula la frecuencia con la que dos palabras aparecen juntas en el corpus y se define como:

$$C_{UMASS}(w_i, w_j) = \log \log \frac{D(w_i, w_j) + 1}{w_i}$$

donde  $D(w_i, w_j)$  indica cuantas veces las palabras  $w_i$  y  $w_j$  aparecen juntas en los documentos, y  $D(w_i)$  es cuantas veces la palabra  $w_i$  aparece sola. Cuanto mayor sea el número, mejor será la puntuación de coherencia. Además, esta medida no es simétrica, lo que significa que  $C_{UMASS}(w_i, w_j)$  no es igual a  $C_{UMASS}(w_j, w_i)$ . Se calcula la coherencia global del tema como los puntajes promedio de coherencia por pares en las N palabras principales que describen el tema.

### **Puntuación de coherencia UCI:**

La UCI fue propuesta por Newman [58]. Esta puntuación de coherencia se basa en las ventanas deslizantes y la información mutua puntual de todos los pares de palabras usando las mejores

$N$  palabras por ocurrencia. En lugar de calcular la frecuencia con la que aparecen dos palabras en el documento, se calcula la ocurrencia de palabras mediante una ventana deslizante. Es decir, que si dos palabras aparecen en un documento, pero no están juntas en la ventana deslizante, no se cuentan como co-ocurrencia.

Se define la puntuación UCI entre las palabras  $w_i$  y  $w_j$ , como:

$$C_{UCI}(w_i, w_j) = \log \log \frac{P(w_i, w_j) + 1}{P(w_i) \cdot P(w_j)}$$

Donde  $P(w_i)$  es la probabilidad de ver la palabra  $w_i$  en la ventana deslizante y  $P(w_i, w_j)$ , es la probabilidad de que las palabras  $w_i$  y  $w_j$ , aparezcan juntas en la ventana deslizante.

Evaluar modelos de tópicos realmente sigue siendo una tarea difícil. Hay varios enfoques disponibles, pero los mejores resultados provienen de la interpretación humana. Pero este es un ejercicio costoso y requiere mucho tiempo.

Los métodos de evaluación cuantitativa ofrecen los beneficios de la automatización y el escalado. La coherencia es el más popular de estos y es fácil de implementar en lenguajes de codificación ampliamente utilizados, como con Gensim en Python.

En la práctica, el mejor enfoque para evaluar los modelos de temas dependerá de las circunstancias. El conocimiento del dominio, la comprensión del propósito del modelo y el juicio ayudarán a decidir el mejor enfoque de evaluación.

## 4. Encuesta Net Promoter Score y muestra

Cuando una empresa quiere medir el porcentaje de lealtad de sus clientes puede utilizar el Net Promoter Score (NPS). Frederick F.Reichheld (2003), fue el primero en hablar sobre NPS en su artículo llamado “The One Number You Need to Know to Grow”, publicado en la revista Harvard Business Review [34], Frederick explica cómo mediante un estudio a 14 empresas de 6 industrias diferentes llegaron a la conclusión de que la pregunta: “en una escala de 0 a 10, ¿Qué tan probable es que recomiende la empresa X a un amigo o colega?”, podría ser mucho más efectiva para medir la lealtad de los clientes, que encuestas con hasta 20 preguntas, ya que el porcentaje de clientes que eran lo suficientemente entusiastas como para referir a un amigo se correlacionaba con la tasa media de crecimiento de varias de las empresas. Esta pregunta permite clasificar a los clientes en promotores, neutros y detractores, como se muestra en la figura 1.

El NPS se calcula restando el porcentaje de clientes que son detractores del porcentaje de clientes que son promotores y variará de -100 (peor) a 100 (mejor), es decir que para aumentar el NPS se necesita aumentar los promotores o disminuir los detractores.

$$\text{NPS} = \% \text{Promotores} - \% \text{Detractores}$$

De acuerdo con Customer Guru [43], una herramienta que ofrece el servicio de aplicar la encuesta de NPS para empresas, el NPS se debe interpretar por industrias, es decir, que para una empresa saber qué tan bien está el nivel de



satisfacción de sus clientes debe comparar su NPS con el de las demás empresas de la industria a la que pertenece, un buen NPS debe estar por encima del promedio del NPS de la industria.

Uno de los retos que puede tener una organización es aumentar la intención de recomendación por parte del cliente, porque gracias al voz a voz podrá crecer de manera costoeficiente. Para aumentar la recomendación se debe indagar sobre, qué servicios o factores debe mejorar. El NPS por sí solo no permite identificar las oportunidades de mejora que tiene la empresa, por eso a la pregunta inicial se le ha añadido una justificación por parte del cliente a su respuesta, esta justificación se almacena en lenguaje natural, tal cual lo dice el cliente y estos textos son llamados verbatims. Los verbatims son textos cortos, que en la mayoría de los casos mencionan varios temas o puntos importantes para el cliente. Un adecuado análisis de estos textos puede proporcionar oportunidades de mejora y ayudar a identificar problemas que pueden causar clientes con opiniones negativas y a su vez aumentar la preferencia e intención de recomendación.

Algunas empresas se encargan ellas mismas de realizar esta encuesta, pero cuando la empresa es muy grande optan por contratar este servicio con terceros expertos en mediciones de este tipo. La compañía financiera de este estudio contrata la empresa Ipsos [59], quienes se encargan de hacer la medición de NPS de manera trimestral a una muestra aleatoria de clientes. Esta encuesta se hace vía telefónica y el verbatim es transcrito, lo que implica que pueda tener errores de redacción desde su generación.

## **4.1 Población considerada**

Para este estudio se están considerando los clientes neutros de las encuestas realizadas en los años 2019, 2020 y 2021, es decir, hay disponible información de 12 encuestas porque cada año se realizan 4. Los clientes encuestados neutros en esos años en total son 7.000 y están distribuidos como se muestra en la Tabla 1:

<b>Año</b>	<b>Cientes neutros</b>
2019	2,683
2020	2,771
2021	1,546
<b>Total</b>	<b>7,000</b>

Tabla 1: distribución de los datos

Los siguientes son ejemplos del tipo de información con la que se va a trabajar, sin ninguna modificación desde la fuente, como se observa es información no estructurada sin ninguna etiqueta relacionada.

- “Préstamos más flexibles”
- “Servicio personalizado; cuando uno es buen cliente, hay asesores que lo atienden a uno en otras entidades.”
- “que sean más baratos en los costos en general por ejemplo si tiene una cuenta de ahorro es caro, si tiene una tarjeta de crédito cobran una cuota de manejo alta básicamente en esos dos productos”
- “los costos son muy altos la cuota de manejo en la tarjeta de crédito y débito. No da mas información”
- “Todo está bien”

## 5. Metodología

Para definir la metodología a seguir, se investigó sobre las herramientas disponibles, los algoritmos más aplicados, y se hicieron pruebas sobre estos realizando optimización de sus hyperparametros teniendo como criterio para evaluar la calidad de los tópicos la coherencia y un enfoque basado en la observación de los términos más relevantes de cada tópico.

Se decidió trabajar en el lenguaje Python, porque es un lenguaje accesible en la compañía dueña de los datos y es open source, lo cual es muy importante para este proyecto, porque se está realizando sin recursos económicos. En la tabla 2 se muestran las librerías más usadas actualmente para NLP de Python, con sus principales funciones y las limitaciones que se encontraron.

Identificación de factores a mejorar para aumentar la recomendación, análisis de NPS para clientes neutros a partir de procesamiento de lenguaje natural

Herramienta	Funcionalidades	Limitaciones
	Preprocesamiento de documentos	
	Análisis de sentimientos	
<b>Textblob</b>	Clasificadores	
	Traductor	
	Corrección de texto	Funcionalidades para texto en español
	Preprocesamiento de documentos	
<b>Nltk</b>	Análisis de sentimientos	
	Clasificadores	Funcionalidades y precisión para texto en español
<b>sentiment-analysis-spanish</b>	Análisis de sentimiento para documentos en español	
	Modelado de temas	
<b>Gensim</b>	Cálculo de coherencia	
<b>Spacy</b>	Preprocesamiento de datos disponible para texto en español	
<b>bitermplus</b>	Implementación del modelo de temas bitermino para textos cortos	
<b>PyLDAvis</b>	Visualización de temas para modelos LDA	No ofrece la representación gráfica para otro tipo de modelos

Tabla 2: librerías disponibles para NLP en Python

De la tabla 2 las mejores librerías para trabajar con documentos en español son sentiment-analysis-spanish para realizar el análisis de sentimientos, Nltk y Spacy para preprocesamiento, y Gensim para el modelado de temas. PyLDAvis es una buena opción para representar los tópicos gráficamente cuando se trabaja con modelos LDA.

Basado en el respaldo teórico y práctico, es decir en la documentación, los modelos con trabajos en aplicación en dónde se obtuvieron buenos resultados y las implementaciones disponibles en python, se definen los algoritmos que se considerarán en la búsqueda de un modelo óptimo, y sobre todo útil para el caso. Los modelos fueron LDA, LSI y NMF.

A continuación, se describen los pasos de la metodología desarrollada:

## 5.1 Extracción de información y análisis de sentimientos

La figura 3 muestra los pasos considerados en esta etapa de la metodología

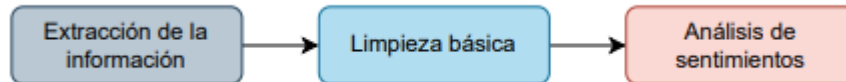


Figura 3: extracción de información y análisis de sentimientos

**Extracción de información:** en esta etapa se realiza la extracción de información de los verbatims de los clientes neutros para las encuestas realizadas en los años 2019, 2020 y 2021. Estos verbatims son almacenados en lenguaje natural y tienen la forma vista en la sección 4.1.

**Limpieza básica:** en esta limpieza inicial se remueven caracteres extraños, espacios o signos de puntuación extras de los documentos, para continuar con el análisis de sentimientos.

**Análisis de sentimientos:** se realiza el análisis de sentimientos de cada verbatim con la librería sentiment-analysis-spanish de Python. El resultado que proporciona el análisis de sentimientos es un número entre 0 y 1, que

para valores cercanos a 0 se interpreta como un sentimiento negativo, para valores cercanos a 1 se interpreta como un sentimiento positivo y valores cercanos a 0.5 se refieren a un sentimiento neutro.

Para este caso se definió la variable categórica *sentimiento\_identificado* como:

*Negativo si la puntuación es menor o igual a 0.4*

*Neutro, si la puntuación está entre 0.4 y 0.6*

*Positivo, si la puntuación es mayor o igual a 0.6*

La tabla 3 muestra los resultados obtenidos para cada año :

<b>Año</b>	<b>Sentimiento</b>	<b>Cientes</b>	<b>Porc_clientes</b>
2019	Negativo	2,374	88%
2019	Neutro	133	5%
2019	Positivo	180	7%
2020	Negativo	2,403	87%
2020	Neutro	111	4%
2020	Positivo	260	9%
2021	Negativo	1,349	88%
2021	Neutro	56	4%
2021	Positivo	134	9%
<b>Total</b>		<b>7,000</b>	

Tabla 3: sentimientos identificados en los verbatims

Se encuentra que la mayoría de los clientes, a pesar de ser clasificados como neutros en la encuesta del NPS, realmente tienen un sentimiento negativo en su verbatim.

Por ejemplo, el análisis de sentimientos identificó un sentimiento negativo en este verbatim:

- “Facilitarme mis créditos, teniendo en cuenta que nunca quede mal con dicho banco, con ninguno”.

En este verbatim identificó un sentimiento positivo:

- “Yo pienso que hace un buen trabajo, pequeños detalles pero mínimos”

Y en este un sentimiento neutro:

- “Es una decisión personal”

## 5.2 Procesamiento de la información

La figura 4 muestra los pasos considerados en esta etapa de la metodología

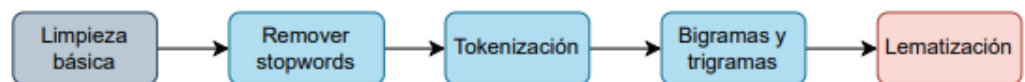


Figura 4: procesamiento de información

**Entrada:** texto con la primera etapa de limpieza aplicada, esta entrada es obtenida en el paso de limpieza básica descrito en la sección anterior.

**Remover stopwords:** luego, se eliminan las stopwords que contiene la librería Nltk y se crea una lista de palabras complementaria, ya que la de la librería nltk para español se queda corta en este sentido.

**Tokenización:** con los documentos libres de stopwords, se realiza la tokenización por palabras y se explora la tokenización por oraciones, pero por tratarse de textos cortos la mayoría de los documentos constituyen en sí una sola frase, entonces se decide trabajar únicamente con los tokens por palabras.

**Bigramas y trigramas:** con los documentos tokenizados, se crearon bigramas y trigramas para encontrar asociaciones entre palabras, que hacen más sentido juntas. La decisión de trabajar con bigramas y trigramas se debe a que los verbatim son textos cortos, lo que dificulta la formación de grupos de más de tres palabras con sentido.

**Lematización:** después de creados los bigramas y trigramas, se realiza la lematización para reducir los términos a su lema, para esto se usa la librería Spacy, la cual tiene esta implementación para el idioma español, aunque sus resultados revelan que es necesario un ajuste. Cuando se realiza la lematización se puede obtener información acerca de qué tipo de palabra es, basado en esta etiqueta que asigna Spacy, es posible reducir los lemas únicamente a sustantivos, verbos, adverbios y adjetivos.

## 5.3 Modelado de tópicos

La figura 5 muestra los pasos considerados en esta etapa de la metodología:

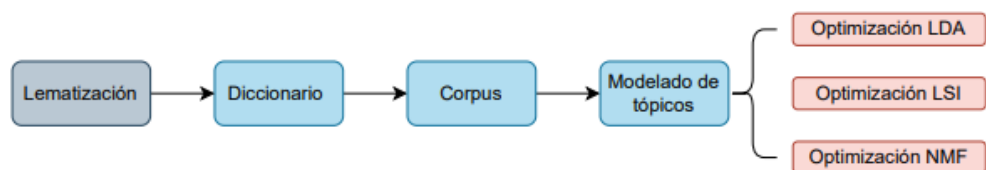


Figura 5: modelado de tópicos

**Entrada:** texto lematizado, creado en la sección anterior.

**Diccionario:** los algoritmos que se van a implementar en pasos posteriores requieren como argumento en sus funciones un diccionario.



El diccionario construido en este caso, contiene 2.515 tokens únicos que reflejan el vocabulario usado por los clientes en todos los documentos. Para la construcción del diccionario se tuvieron en cuenta únicamente aquellos tokens que se encontraban en al menos 3 documentos y los que estaban contenidos en no más del 80% de documentos.

**Corpus:** el corpus también es un argumento necesario para entrenar los modelos. Este se construye a partir del diccionario y consta de una lista de tuplas por documento, es decir cada lista representa un documento, y cada una va a contener una o más tuplas, que tendrán como primer elemento el término y como segundo elemento la cantidad de veces que sale ese término en ese documento.

Abajo se presentan 2 ejemplos, en la primera lista hay un trigram, el cual aparece una vez en ese documento y en la segunda lista hay nueve términos y cada uno aparece una vez en ese documento.

**Lista 1:** [('tasa\_interes\_alta', 1)]

**Lista 2:** [('buen', 1), ('banco', 1), ('productos\_costoso', 1), ('disponibilidad\_cajero', 1), ('lento', 1), ('corresponsales\_bancario', 1), ('positivo', 1), ('soporte', 1), ('amplia\_cobertura', 1)]

El hecho de que la mayoría de los términos salen una vez en el documento, se debe a que los textos son cortos.

### 5.3.1 Optimización de hiperparámetros:

#### Optimización LDA:

El algoritmo para el modelo LDA implementado en la librería gensim, permite los siguientes hiperparámetros:

- Beta: Distribución del número de palabras por tema
- Alpha: Distribución del número de temas por documento
- chunksize: Número de documentos que se usarán en cada fragmento de entrenamiento
- passes: Número de recorridos a través del corpus durante el entrenamiento.
- num\_topics: Número de tópicos

Para encontrar un modelo LDA óptimo, se entrenaron 588 modelos con el siguiente diccionario de hiperparámetros:

```
{'Validation_Set': [75% corpus, 100% corpus],  
'Topics': [2,4,6,8,10,12,14],  
'Alpha': [0.01, 0.21, 0.41, 0.61, 0.81, symmetric, asymmetric],  
'Beta': [0.01, 0.21, 0.41, 0.61, 0.81, symmetric]  
}
```

Tabla 4: diccionario de hiperparámetros para LDA

En la tabla 5 se presenta una muestra de los resultados obtenidos para la optimización y el resultado en azul fue el mejor obtenido.

<b>Validation_Set</b>	<b>Num_topics</b>	<b>Alpha</b>	<b>Beta</b>	<b>Coherencia</b>
<b>100% Corpus</b>	<b>2</b>	<b>0.01</b>	<b>0.81</b>	<b>-3.6</b>
100% Corpus	4	0.01	0.81	-5.4
100% Corpus	6	0.01	0.81	-5.9
100% Corpus	8	0.01	0.81	-7.9
100% Corpus	10	0.01	0.81	-8.7
100% Corpus	12	0.01	0.81	-9.3
100% Corpus	14	0.01	0.81	-10.8

Tabla 5: resultado obtenido aplicando LDA para el modelado de tópicos. El mejor resultado se resalta en azul

El mejor modelo entre los 588 probados fue el proporcionado por los siguientes hiperparámetros, con el cual se obtuvo una coherencia de -3.6:

- Chunksize= 1000,
- Passes= 10,
- Validation\_Set = 100% Corpus,
- Topics= 2,
- Alpha= 0.01,
- Beta= 0.81

### **Optimización LSI:**

El algoritmo para el modelo LSI implementado en la librería gensim, permite los siguientes hiperparámetros:

- chunksize: número de documentos que se usarán en cada lote de entrenamiento
- onepass: indica si se debe usar el algoritmo de un solo paso para el entrenamiento. False para un algoritmo de varias pasadas.
- power\_iters: número de pasos de iteración de potencia que se utilizarán. Aumentar el número de iteraciones de potencia mejora la precisión, pero reduce el rendimiento.
- extra\_samples: Muestras adicionales para usar. Puede mejorar la precisión.
- decay: Peso de las observaciones existentes en relación con las nuevas (tasa de aprendizaje)
- num\_topics: Número de tópicos

Para encontrar un modelo LSI óptimo, se entrenaron 56 modelos con el siguiente diccionario de hiperparámetros:

```
{'Validation_Set': [75% corpus, 100% corpus],  
'Topics': [2,4,6,8,10,12,14],  
Decay: [0.01, 0.31, 0.61, 0.91]  
}
```

Tabla 6: diccionario de hiperparametros para LSI

En la tabla 7 se presenta una muestra de los resultados obtenidos para la optimización y el resultado en azul fue el mejor obtenido.

<b>Validation_Set</b>	<b>Num_topics</b>	<b>Decay</b>	<b>Coherencia</b>
<b>100% Corpus</b>	<b>2</b>	<b>0.91</b>	<b>-4.3</b>
100% Corpus	4	0.91	-4.5
100% Corpus	6	0.91	-6.0
100% Corpus	8	0.91	-7.4
100% Corpus	10	0.91	-7.4
100% Corpus	12	0.91	-5.3
100% Corpus	14	0.91	-5.5

Tabla 7: resultado obtenido aplicando LSI para el modelado de tópicos. El mejor resultado se resalta en azul

El mejor modelo entre los 56 probados fue el proporcionado por los siguientes hiperparametros, con el cual se obtuvo una coherencia de -4.3:

- Chunksize= 1000,
- Onepass = 10,
- extra\_samples = 5,
- Validation\_Set = 100% Corpus,
- Topics= 2,
- Decay= 0.91

### Optimización NMF:

El algoritmo para el modelo NMF implementado en la librería gensim, permite los siguientes hiperparametros:

- chunksize: número de documentos que se usarán en cada lote de entrenamiento
- passes: Número de recorridos a través del corpus durante el entrenamiento.
- normalize: si se normaliza el resultado. Permite estimar la perplejidad, la coherencia, etc.num\_topics:
- num\_topics: Número de tópicos

Para encontrar un modelo NMF óptimo, se entrenaron 14 modelos con el siguiente diccionario de hiperparametros:

```
{'Validation_Set': [75% corpus, 100% corpus],  
'Topics': [2,4,6,8,10,12,14]  
}
```

Tabla 8: diccionario de hiperparametros para NMF

En la tabla 9 se presenta una muestra de los resultados obtenidos para la optimización y el resultado en azul fue el mejor obtenido.

<b>Validation_Set</b>	<b>Num_topics</b>	<b>Coherencia</b>
<b>100% Corpus</b>	<b>2</b>	<b>-3.8</b>
100% Corpus	4	-4.5
100% Corpus	6	-5.3
100% Corpus	8	-6.8
100% Corpus	10	-7.2
100% Corpus	12	-7.5
100% Corpus	14	-8.0

Tabla 9: resultado obtenido aplicando MNF para el modelado de tópicos. El mejor resultado se resalta en azul

- El mejor modelo entre los 14 probados fue el proporcionado por los siguientes hiperparámetros, con el cual se obtuvo una coherencia de -3.8:
- Chunksize= 1000,
- Onepass = 10,
- normalize =True,
- Validation\_Set = 100% Corpus,
- Topics= 2

## 5.4 Cantidad de tópicos óptimos y modelo seleccionado

De acuerdo con los enfoques descritos en la sección 3.5, en este trabajo se va a usar la métrica de la coherencia para elegir el número de tópicos óptimo, ya que permite realizar un trabajo automático de la selección del modelo, y posteriormente se va a usar el enfoque de observación para validar que el modelo elegido proporciona resultados coherentes.

Las siguientes gráficas muestran la evolución de la coherencia en función de la cantidad de tópicos para cada modelo, dado que para cada cantidad

de tópicos variaron los demás hyperparametros, la medida representada en las gráficas es un promedio de la coherencia agrupado por la cantidad de tópicos.

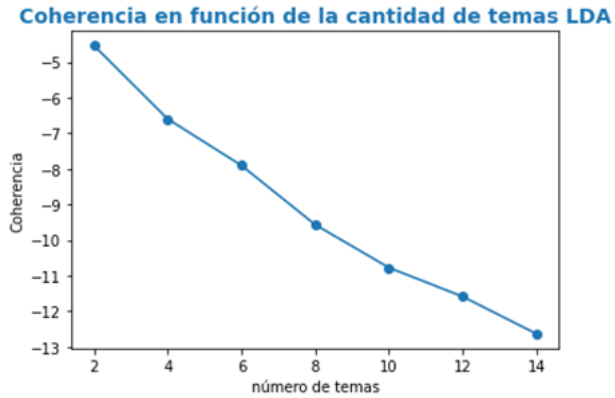


Figura 6: coherencia vs número de tópicos para LDA

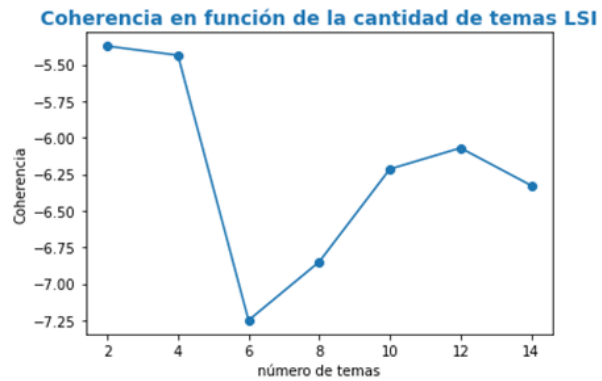


Figura 7: coherencia vs número de tópicos para LSI

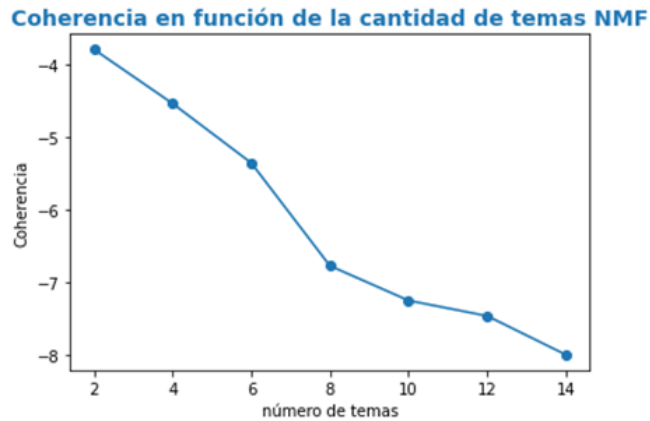


Figura 8: coherencia vs número de tópicos para NMF

En las figuras 6,7, y 8 se observa que la coherencia disminuye cuando se aumenta la cantidad de tópicos para los tres modelos y la coherencia máxima se alcanza con dos tópicos. Esto significa que los Verbatim de la encuesta NPS se encuentran muy concentrados en dos temas importantes.

Los algoritmos LDA, LSI y NMF sugieren dos tópicos, con diferencias en la coherencia muy pequeñas, LDA es el algoritmo que alcanza una mayor coherencia en sus tópicos como se muestra en la tabla 7, por lo tanto, se decide desarrollar la implementación con este modelo.

<b>Algoritmo</b>	<b>Coherencia</b>
LDA	-3.6
LSI	-4.3
NMF	-3.8

Tabla 10: mejores coherencias obtenidas



## 6. Resultados y Análisis

De acuerdo con los resultados obtenidos con la metodología aplicada, el mejor modelo entre los probados es el LDA con dos tópicos, en esta sección se van a interpretar sus resultados en el contexto del negocio, que en este caso se trata del sector financiero.

### 6.1 Caracterización de los tópicos

Una forma de evaluar y caracterizar el comportamiento de los tópicos es a través de la observación, pyLDAvis [66] ofrece una forma sencilla para obtener la representación gráfica de los tópicos y las palabras más relevantes para modelos LDA, como se muestra en las figuras 9 y 10.

En la figura 9 se observan los dos temas definidos con el modelo LDA, representado cada uno por una burbuja, en ella se ven los dos temas a una distancia considerable, esto es señal de un buen resultado, ya que se espera que las burbujas que representan los temas estén lo suficiente separadas y no se superpongan, esto significa que los tópicos son homogéneos en su interior y heterogéneos entre ellos.

A partir de los términos más relevantes en el corpus, que salen en la figura 10, se puede concluir que los clientes neutros, en general, hablan sobre los costos altos, las tasas de interés altas, cuotas de manejo y tiempo de espera en la atención.

Identificación de factores a mejorar para aumentar la recomendación, análisis de NPS para clientes neutros a partir de procesamiento de lenguaje natural

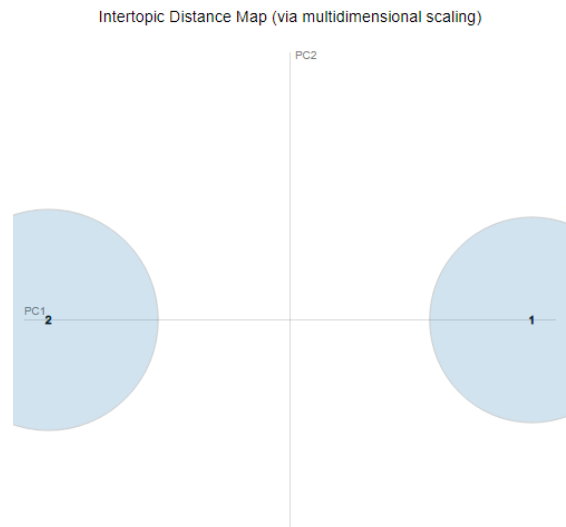


Figura 9: mapa de distancia entre tópicos

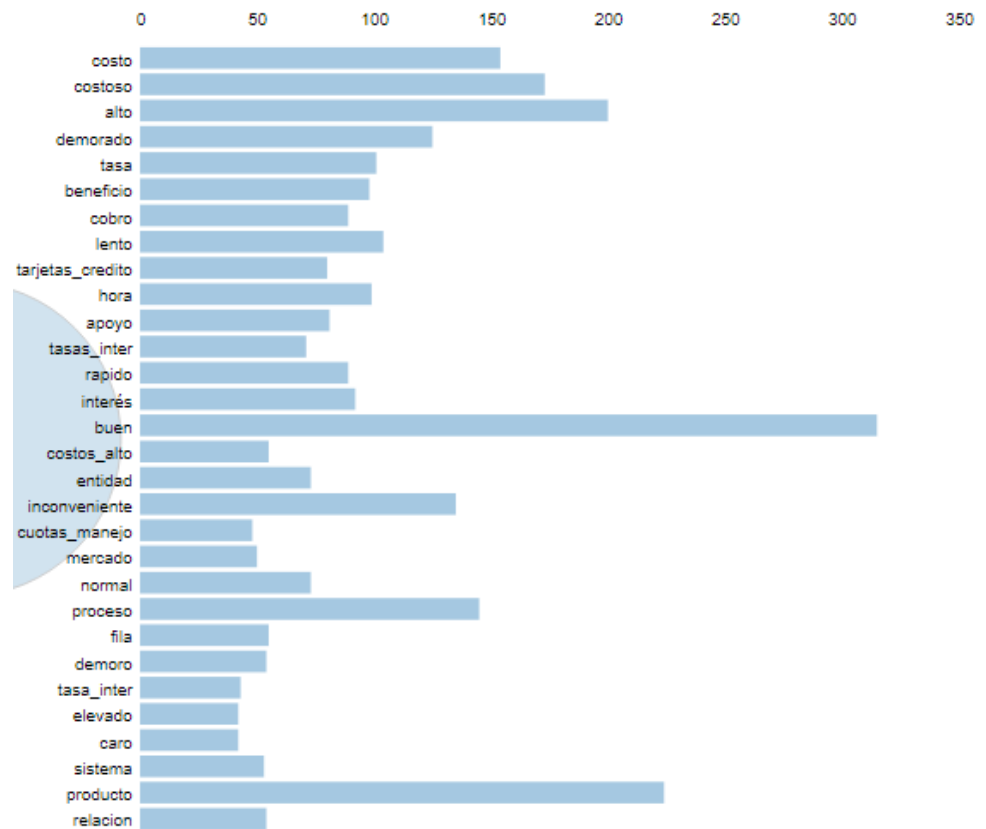


Figura 10: términos más frecuentes en el corpus

**Tópico: costos**

En la figura 11 se está representando uno de los tópicos, este tópico se puede identificar en general como costos, los clientes que tienen este tópico como dominante en su verbatim, hablan sobre costos altos, cuotas de manejo de tarjetas de crédito, tasa de interés en los créditos, y de la oferta de productos.

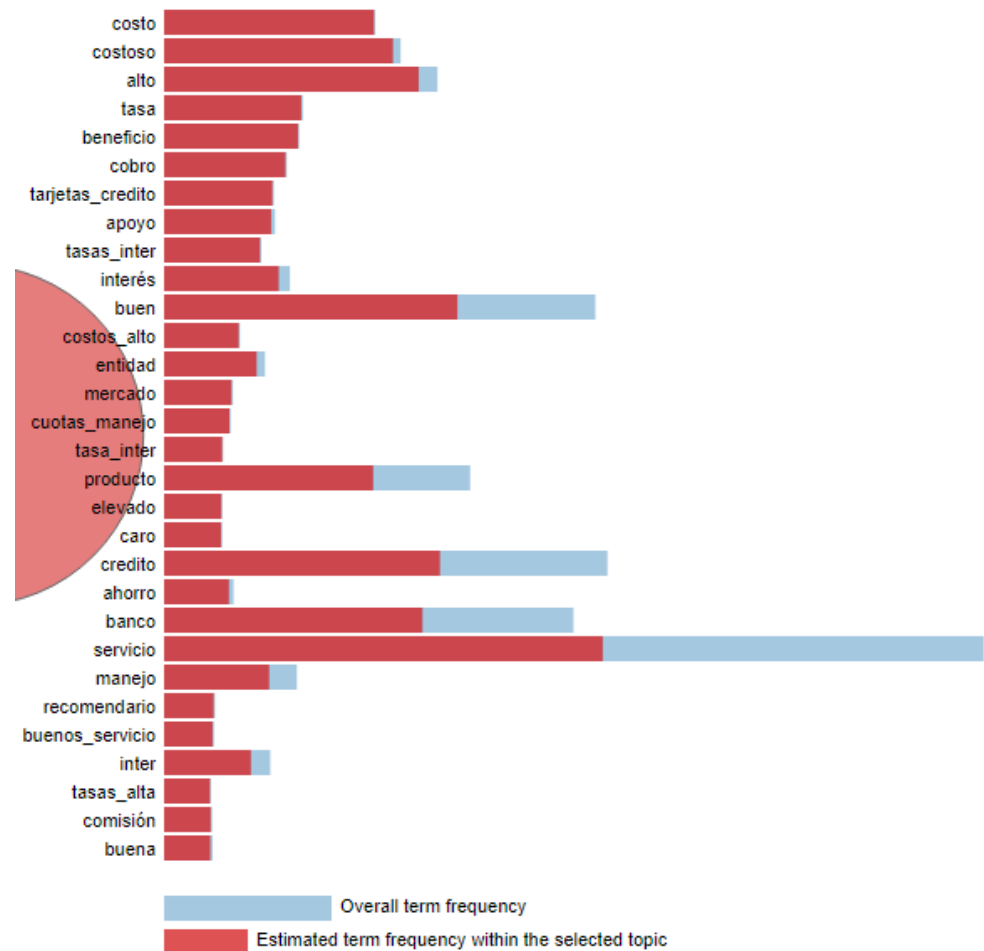


Figura 11: términos más frecuentes en el tópico costos

**Tópico: Servicio en los canales de atención**

En la figura 12 se observa la representación del tópicos dos, que por sus términos más relevantes se concluye que este se trata del servicio en los canales de atención, los clientes hablan de que el servicio es demorado, de que hay mucha fila en las sucursales, del servicio congestionado, y los inconvenientes que presenta el servicio.

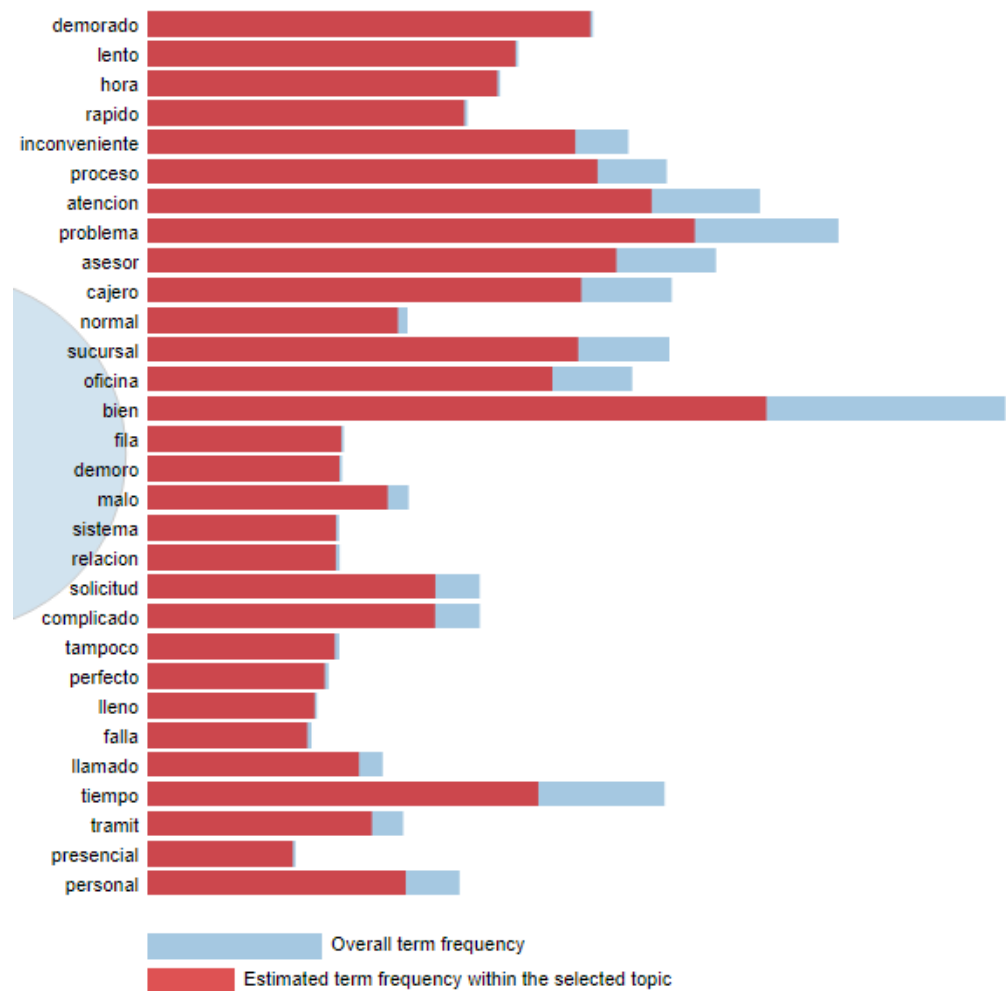


Figura 12: términos más frecuentes del tópico servicio en canales de atención

La tabla 11 muestra la distribución de los clientes entre los dos tópicos hallados:

<b>Tópico</b>	<b>Cantidad de clientes</b>	<b>Porcentaje</b>
Costos	3,548	50.69%
Servicio en los canales de atención	3,452	49.31%
<b>Total</b>	<b>7,000</b>	<b>100%</b>

Tabla 11: distribución de los clientes en los tópicos

**Documentos representativos:**

Ya que en un documento puede haber varios tópicos, el modelo asigna para cada tema un porcentaje de contribución en cada documento y para la interpretación de los resultados se debe elegir el tema más relevante para cada documento.

Dos opiniones representativas del tópico de costos son las siguientes,

- “que bajen las tasas, las tasas de intereses, tasas de las tarjetas y las cuotas de manejo.”
- “bajar o quitar los costos en las operaciones. son demasiado altas y hay entidades que dan ese tipo de facilidades, haciendo que el usuario se plantee trasladarse para las mismas.”

Dos opiniones representativas del tópico de servicio en los canales de atención son las siguientes:

- “mejor atención en bancos, más cajeros, más amplitud de los locales, hay ocasiones en que las filas son muy largas, hay una falla en su sistema por internet. (no da más información).”
- “yo creo que la atención un poquito, de pronto la rapidez hay veces se demora uno mucho allí.”

Algunos clientes también hablan sobre los dos temas, en su opinión, como el siguiente:

- “debido a los altos costos por usar sus servicios. Entiéndase los costos por retiros, cuotas de manejo etc. Aunque la atención es muy buena y tienen muchos canales de atención.”

En ese caso el modelo asigna el tema que encuentre más relevante en la opinión, por ejemplo, a esta opinión le asignó el tema de costos con un 73% de contribución.

## 6.2 Sentimiento dominante en cada tópico

Una vez que se establecen los tópicos encontrados, es interesante saber cuál es el sentimiento atado a cada tópico, porque no siempre las palabras claves ayudan a caracterizar el tópico en ese sentido. Para esto se propone usar el análisis de sentimiento, en donde ya se evidenció que la mayoría de los clientes neutros tienen un sentimiento negativo en el verbatim.

La tabla 12 presenta la distribución de los clientes para cada año considerado y para cada tópico en los sentimientos. Se observa que, durante los años 2019, 2020, y 2021 los temas de costos y de servicio en los canales de atención son constantes entre los clientes neutros, además de que en estos predomina el sentimiento negativo.

Estos resultados permiten concluir que no solo los clientes neutros hablan acerca de los costos y del servicio en canales de atención en sus verbatim, sino que, además, las opiniones acerca de estos dos temas en su mayoría son negativas, es decir, los clientes neutros encuestados en los años 2019, 2020 y 2021 están inconformes con los costos de las cuotas de manejo de las tarjetas de crédito, con las tasas de interés de los créditos, en general con los costos de los servicios prestados por la compañía y con la atención en los canales, además se puede deducir, por las palabras claves “fila” y “demorado” en la figura 11 que el canal del cual tienen más quejas es el de las sucursales.

Año	Tópico dominante	Sentimiento	Porc_clientes
2019	Costos	Negativo	40%
2019	Costos	Neutro	2%
2019	Costos	Positivo	3%
2019	Servicio en los canales de atención	Negativo	49%
2019	Servicio en los canales de atención	Neutro	3%
2019	Servicio en los canales de atención	Positivo	4%
2020	Costos	Negativo	46%
2020	Costos	Neutro	2%
2020	Costos	Positivo	5%
2020	Servicio en los canales de atención	Negativo	41%
2020	Servicio en los canales de atención	Neutro	2%
2020	Servicio en los canales de atención	Positivo	4%
2021	Costos	Negativo	37%
2021	Costos	Neutro	1%
2021	Costos	Positivo	3%
2021	Servicio en los canales de atención	Negativo	39%
2021	Servicio en los canales de atención	Neutro	1%
2021	Servicio en los canales de atención	Positivo	3%

Tabla 12: sentimiento dominante por tópico

### 6.3 Vinculación de los resultados con la realidad de los clientes

El siguiente análisis de información, se hace con el objetivo de evaluar la coherencia entre la opinión de los clientes en los verbatims, y su comportamiento real. El análisis se hace teniendo en cuenta los tópicos hallados para los clientes neutros y algunas variables que podrían reflejar un comportamiento relacionado con el tópico. En las secciones 6.3.1 y 6.3.2 se desea comparar el comportamiento de los clientes neutros con el de los clientes promotores, para indagar sobre la forma de materializar esa insatisfacción manifestada por los clientes neutros. Este análisis considera 7.000 clientes neutros y 20.837 promotores de las encuestas realizadas en los años 2019, 2020 y 2021.

### 6.3.1 Apertura y cancelación de productos

Las figuras 12 y 13 muestran la evolución del número de productos activos que tienen los clientes en la compañía, en la figura 12 cada etiqueta está relacionada con el tópico correspondiente y la comparación corresponde a un año después con respecto al momento de la realización de la encuesta .

Para clientes promotores y neutros se observa que alrededor del 80% disminuyó sus productos activos un año después de la encuesta. Entonces no se puede concluir que los clientes neutros disminuyan sus productos por su insatisfacción con los costos y servicio en los canales de atención, más bien puede que se deba al ciclo esperado de cancelación de los productos del portafolio de créditos.

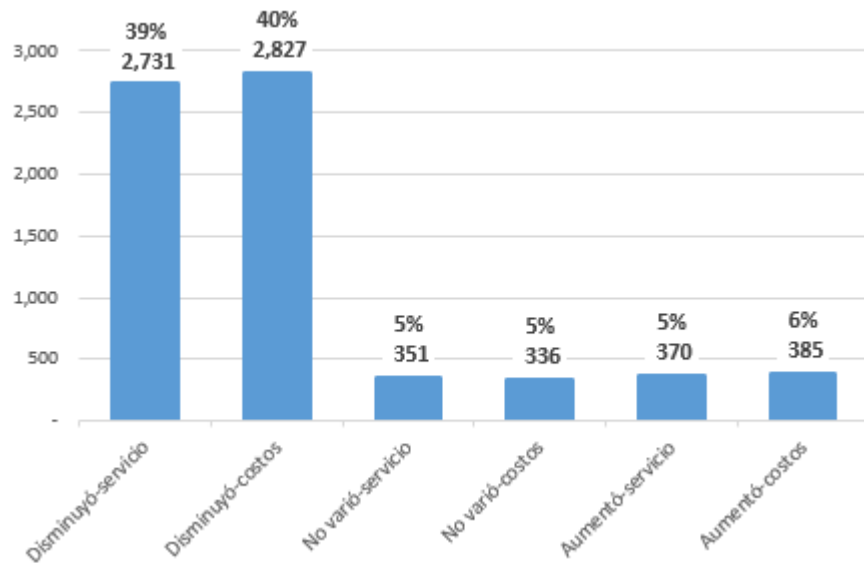


Figura 13: Comportamiento del número de productos en la compañía, un año después de la encuesta, para clientes neutros



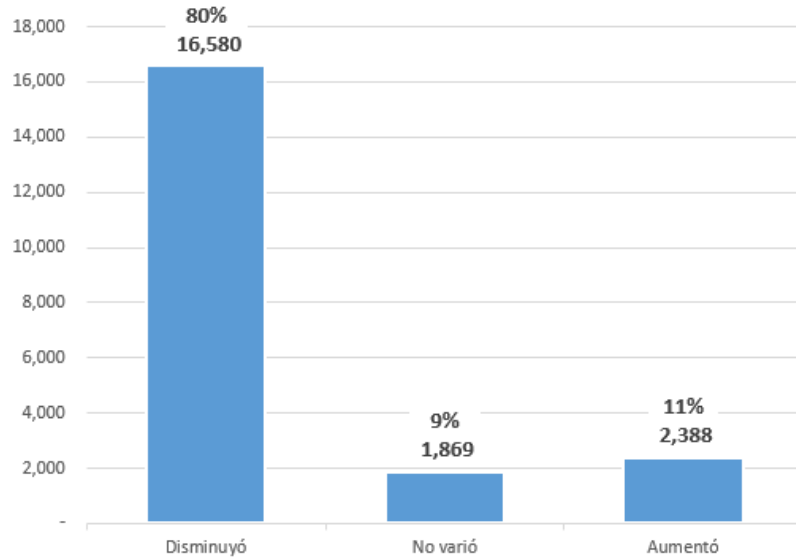


Figura 14: Comportamiento del número de productos en la compañía, un año después de la encuesta, para clientes promotores

La disminución en la cantidad de productos activos tanto para clientes neutros como promotores, puede traer consecuencias como la apertura de productos en otras entidades financieras. En la figura 15 se muestra que el 14% de los clientes, habla acerca de los costos, y aumentó el número de productos en otras entidades, mientras que el 13% de los clientes, hablan acerca del servicio prestado por la compañía y aumentó su número de productos con otras entidades financieras. Respecto a los clientes promotores en la figura 16 se observa que el 25% ha abierto productos en otras entidades en el último año.

La cancelación de productos en la compañía y apertura de estos en otras entidades financieras, no se debe necesariamente a que tan satisfechos están los clientes con la compañía actual.

Identificación de factores a mejorar para aumentar la recomendación, análisis de NPS para clientes neutros a partir de procesamiento de lenguaje natural

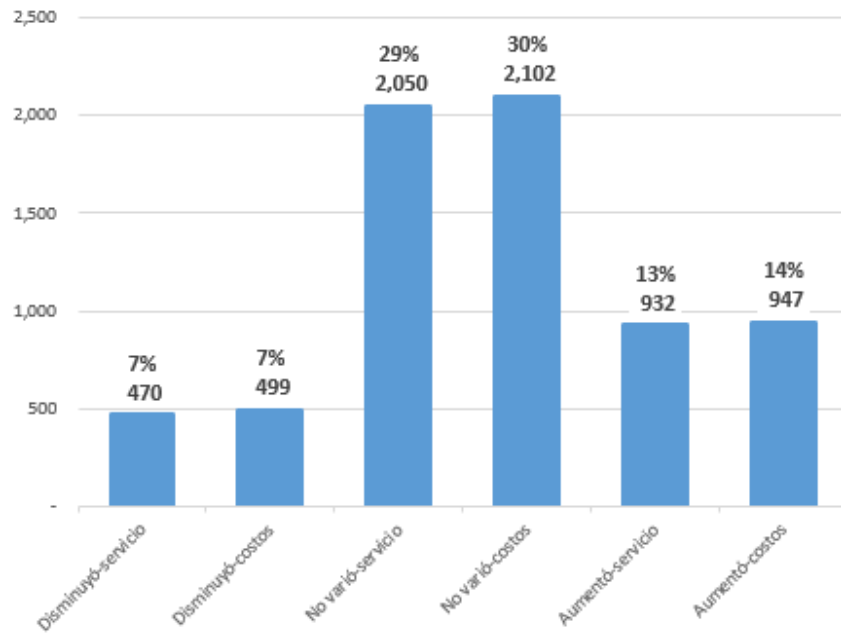


Figura 15: Comportamiento en el número de productos en otras entidades financieras, un año después de la encuesta, para clientes neutros

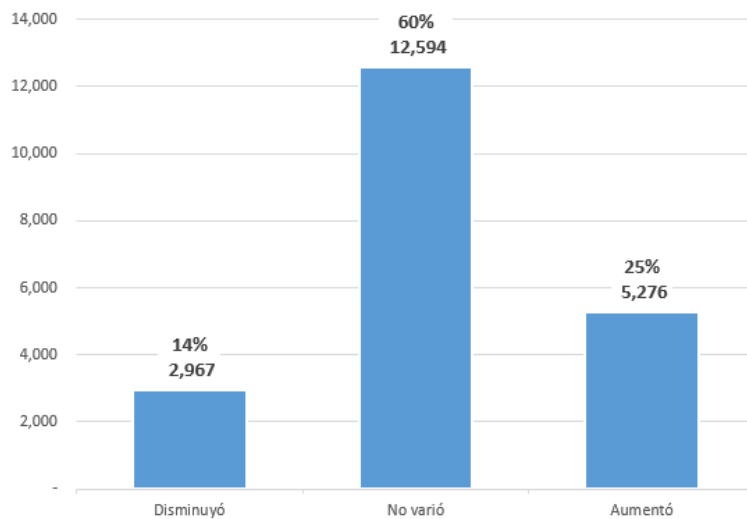


Figura 16: Comportamiento en el número de productos en otras entidades financieras, un año después de la encuesta, para clientes promotores

### 6.3.2 Deserción de clientes

La deserción es uno de los temas que más preocupa respecto a clientes insatisfechos, pero ¿cuántos clientes neutros o promotores en esta compañía, han materializado una deserción? y ¿será la insatisfacción motivo de deserción?, en las figuras 17 y 18 se muestra que la mayoría de los encuestados clasificados como neutros o promotores durante los años 2019, 2020 y 2021, un año después de la encuesta, continúan siendo clientes (para los clientes encuestados en el 2021, se considera la información más reciente, febrero 2022). Entonces no hay una tendencia que indique que los clientes insatisfechos tiendan a desertar.

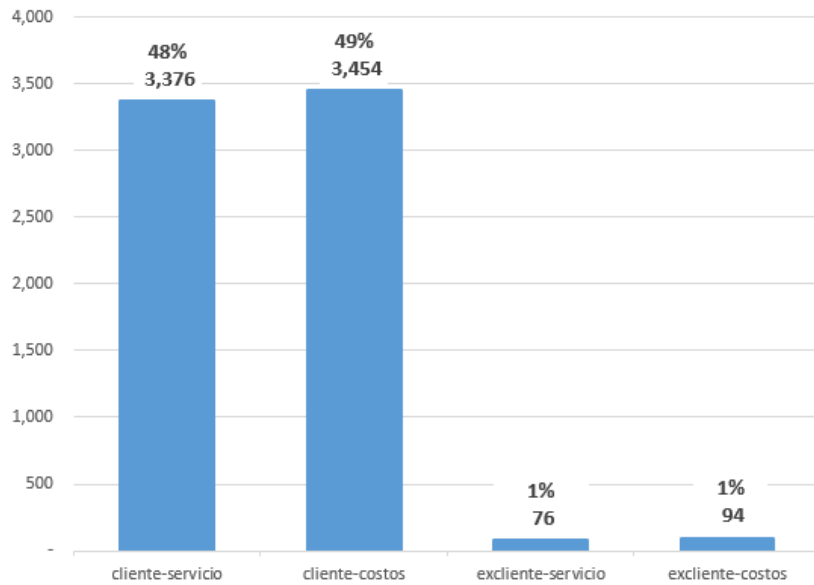


Figura 17: Clientes neutros desertores un año después de la encuesta

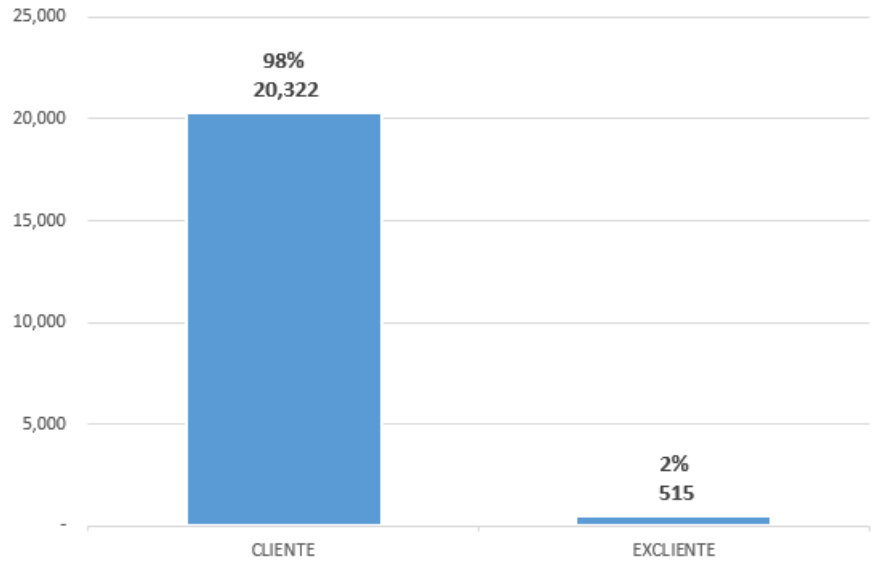


Figura 18: Clientes promotores desertores un año después de la encuesta

### 6.3.3 Cumplimiento del tiempo de promesa para la atención

El tiempo de promesa para la atención, es definido por la compañía como el número máximo de minutos que debe esperar el cliente antes de ser atendido en la sucursal física, esto depende del segmento del cliente y tipo de servicio, que puede ser asesoría o caja. El cumplimiento de este tiempo de promesa, es un factor importante para la satisfacción de los clientes, ya que la insatisfacción se ve reflejada en el tópico de servicio en los canales de atención.

En la figura 19 y 10, se observan los clientes neutros y promotores, respectivamente, a los cuales se les incumplió el tiempo de promesa en al menos un 10% de las veces que visitaron la sucursal durante el año anterior a la aplicación de la encuesta.

Respecto a los clientes neutros que visitaron la sucursal para el servicio de caja, un 9% de los clientes habla acerca del servicio y

se les incumplió el tiempo de atención promesa al menos un 10% de veces y un 9% de los clientes habla acerca de los costos y se les incumplió el tiempo de atención promesa al menos un 10% de las visitas.

Por otro lado, los clientes neutros que visitaron la sucursal física para el servicio de asesoría, un 9% de los clientes habla acerca del servicio y se les incumplió el tiempo de atención promesa al menos un 10% de veces y un 7% de los clientes habla acerca de los costos y se les incumplió el tiempo de atención promesa al menos un 10% de las visitas.

Estos resultados sugieren que el cumplimiento del tiempo de promesa en la atención no se ve reflejado necesariamente en la satisfacción del cliente, ya que probablemente el tiempo de promesa que la compañía se plantea para cada segmento y servicio, sigue siendo alto desde la percepción de los clientes.

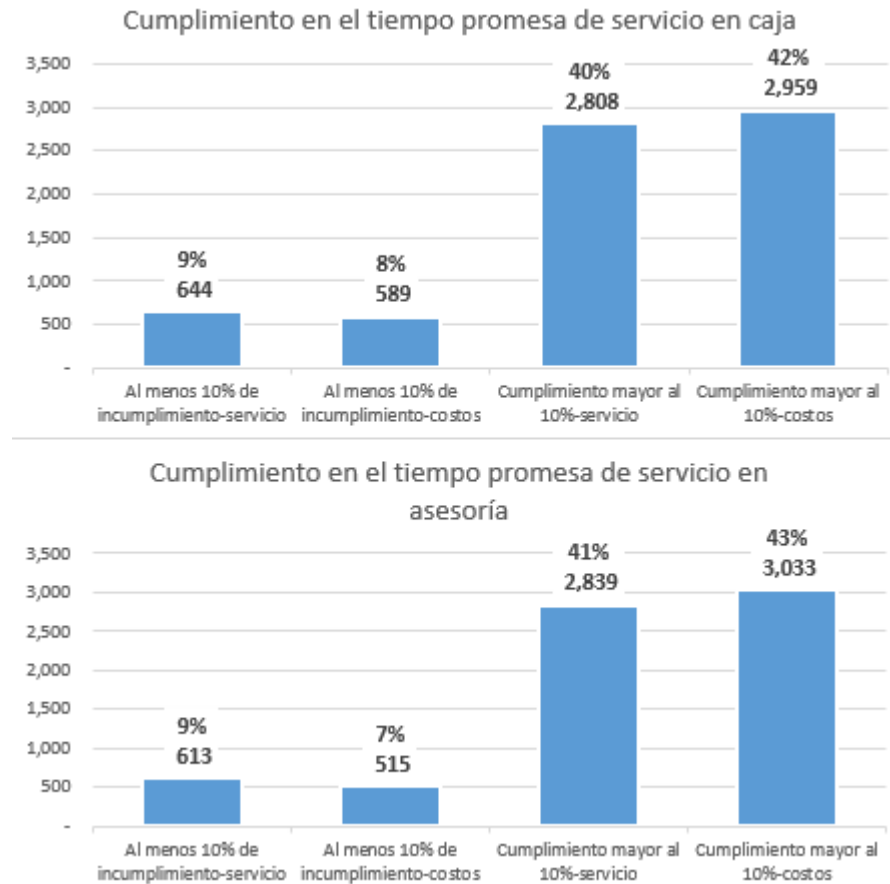


Figura 19: Cumplimiento en el tiempo de promesa, en el año anterior a la encuesta

En conclusión, la disminución de productos en la compañía y la apertura de estos en otras entidades financieras, no es un comportamiento propio de los clientes insatisfechos, ya que tanto clientes promotores como neutros lo tienen, además la insatisfacción revelada por los clientes neutros en el análisis de sentimientos la mayoría de las veces no se materializa en una deserción, y a pesar de que la compañía esté cumpliendo el tiempo promesa de atención en una proporción importante de las visitas a las sucursales, esto no se traduce en satisfacción de los clientes con el servicio prestado.

## 7. Conclusiones y recomendaciones

### 7.1 Conclusiones

Respecto al idioma español, todavía hay mucho por trabajar en temas relacionados a NLP. Sin embargo, con las librerías Nltk, Spacy, Gensim, se lograron resultados aceptables para los objetivos planteados.

El análisis de sentimientos permite revelar los sentimientos de los clientes atados a los tópicos encontrados, lo cual le da mucho más valor a los resultados del modelo de tópicos.

La optimización de los algoritmos LDA, LSI y NMF para este trabajo, dio resultados muy similares en la maximización de la coherencia, pero con la observación de los términos más relevantes se concluyó que el mejor modelo fue uno con la arquitectura LDA.

La tarea de modelado de temas sigue siendo una de las tareas más desafiantes en NLP, a pesar de toda la investigación en torno a este concepto.

Los resultados obtenidos, permiten descubrir que los clientes neutros la mayoría de las veces no tienen una opinión realmente neutra, sino que si tienen críticas sobre el servicio o productos ofrecidos por la organización y de allí se pueden obtener factores y sentimientos relevantes que permiten al negocio mejorar y llegar a fidelizar estos clientes.

A pesar de que los clientes neutros no tiendan a desertar de la empresa, esto no significa que no sean motivo de preocupación, ya que un voz a voz con comentarios negativos de la empresa puede representar una amenaza.

De acuerdo con los tópicos encontrados, los costos altos y el servicio al cliente son temas que en esta compañía han permanecido vigentes entre los clientes neutros a través de los tres años analizados, entonces, si se emprenden acciones de mejora en este sentido, es probable que se logre mejorar el NPS y posicionar mejor a la compañía en la industria financiera, respecto a esta encuesta de percepción por parte de los clientes.

## 7.2 Recomendaciones

En este trabajo se encontró necesario inspeccionar manualmente los tópicos para complementar y guiar la selección automática. Este tipo de inspecciones parecen recomendables para el modelado de tópicos.

Una extensión de este trabajo podría incluir el entrenamiento de modelos con un mayor volumen de datos, esto podría mejorar los resultados, encontrando más clientes que hablen acerca del mismo tema y aumentando las frecuencias de los términos.

Asimismo, se podrían entrenar otros modelos para encontrar sub tópicos, por ejemplo en el tópico costos, podría hallarse un tópico más específico como costos de las tasas de interés, costos de la cuota de manejo de la tarjeta de crédito. entre otros, o en el tópico de servicio en los canales de atención, evaluar si hay sub-tópicos como servicio en las sucursales, servicio en la línea telefónica, entre otros.



## Bibliografía

- [1] Athanasiou, V., & Maragoudakis, M. (2017). A novel, gradient boosting framework for sentiment analysis in languages where NLP resources are not plentiful: a case study for modern Greek. *Algorithms*, 10(1), 34.
- [2] Aziz, A. A., & Starkey, A. (2019). Predicting supervised machine learning performances for sentiment analysis using contextual-based approaches. *IEEE Access*, 8, 17722-17733.
- [3] Anderson, E. W., Fornell, C., & Lehmann, D. R. (1994). Customer satisfaction, market share, and profitability: Findings from Sweden. *Journal of marketing*, 58(3), 53-66.
- [4] Baviera, T. (2017). Técnicas para el análisis del sentimiento en Twitter: Aprendizaje Automático Supervisado y SentiStrength. *Dígitos*, 1(3), 33-50.
- [5] Bird S, Klein E, y Loper E, (2009), *Natural Language Processing with Python*, Estados Unidos de América, O'Reilly Media.
- [6] Cheng, M., & Jin, X. (2019). What do Airbnb users care about? An analysis of online review comments. *International Journal of Hospitality Management*, 76, 58-70.
- [7] Chiruzzo, L., Rosá, A., & Etcheverry, M. (2020). Análisis de Sentimiento para Tweets en Español: Algunos Experimentos con Foco en los Tweets Neutros. *Procesamiento del lenguaje natural*, (64), 109-116.
- [8] De la Torre, J., & del Consuelo, M. (2017). Nuevas técnicas de minería de textos: Aplicaciones (Doctoral dissertation, Universidad de Granada).
- [9] Feng, L., & Wei, W. (2020). A Combined Method of r-NPS and t-NPS Evaluations for Identification of Negative Triggers of Detractors' Experience. *Sustainability*, 12(4), 1652.

- [10] Fernández, J., Boldrini, E., Gómez, J. M., & Martínez-Barco, P. (2011). Análisis de Sentimientos y Minería de Opiniones: el corpus EmotiBlog. *Procesamiento del lenguaje natural*, 47, 179-187.
- [11] Fernández, J., Gutiérrez, Y., Gómez, J. M., Martínez-Barco, P., Montoyo, A., Muñoz, R. (2013). Sentiment analysis of spanish tweets using a ranking algorithm and skipgrams.
- [12] Fisher, N. I., & Kordupleski, R. E. (2018). What is Wrong with Net Promoter Score. *arXiv preprint arXiv:1806.10452*.
- [13] Fodeh, S. J., Al-Garadi, M., Elsankary, O., Perrone, J., Becker, W., & Sarker, A. (2020). Utilizing a multi-class classification approach to detect therapeutic and recreational misuse of opioids on Twitter. *Computers in Biology and Medicine*, 104132.
- [14] Gelbukh, A. (2010). Procesamiento de lenguaje natural y sus aplicaciones. *Komputer Sapiens*, 1, 6-11.
- [15] Henríquez, C., Pla, F., Hurtado, L. F., & Guzmán, J. (2017). Análisis de sentimientos a nivel de aspecto usando ontologías y aprendizaje automático. *Procesamiento del Lenguaje Natural*, (59), 49-56.
- [16] Hernández, M. B., & Gómez, J. M. (2013). Aplicaciones de procesamiento de lenguaje natural. *Revista Politécnica*, 32.
- [17] Korenius, T., Laurikkala, J., Järvelin, K., & Juhola, M. (2004, November). Stemming and lemmatization in the clustering of finnish text documents. In *Proceedings of the thirteenth ACM international conference on Information and knowledge management* (pp. 625-633).
- [18] Kumar, S., & Zymbler, M. (2019). A machine learning approach to analyze customer satisfaction from airline tweets. *Journal of Big Data*, 6(1), 1-16.
- [19] Mandal, P. C. (2014). Net promoter score: a conceptual analysis. *International Journal of Management Concepts and Philosophy*, 8(4), 209-219.
- [20] Mariñelarena-Dondena, L., Errecalde, M. L., & Solano, A. C. (2017). Extracción de conocimiento con técnicas de minería de textos aplicadas a la psicología. *Revista Argentina de Ciencias del Comportamiento*, 9(2), 65-76.
- [21] Mariñelarena-Dondena, L., Ferretti, E., Maragoudakis, M., Sapino, M., & Errecalde, M. L. (2017). Predicting depression: a comparative study of machine

learning approaches based on language usage. Cuadernos de Neuropsicología/Panamerican Journal of Neuropsychology, 11(3).

[22] Medina-Merodio, J. A., Pablos-Heredero, C. D., Jiménez-Rodríguez, M. L., Marcos-Ortega, L. D., Barchino-Plata, R., Rodríguez-García, D., & Gómez-Aguado, D. (2014). Analysis of customer satisfaction using surveys with open questions. *Dyna*, 81(188), 92-99.

[23] Melnic, E. L. (2016). Techniques for measuring customers' satisfaction in Banks. *Bulletin of the Transilvania University of Brasov. Economic Sciences. Series V*, 9(1), 23.

[24] Mesa Reyes, J. G. (2016). Procesamiento de lenguaje natural y su aplicación en servicios de hostelería.

[25] Moorthi, Y. L. R., & Mohan, B. C. (2017). Brand value proposition for bank customers in India. *International Journal of Bank Marketing*.

[26] Mousavi, S. M., & Amiri Aghdaie, S. F. (2021). Identifying the Constructive Elements of "Value Proposition" and their Impact on Customers' Satisfaction using Sentiment Analysis based on Text Mining. *Journal of Business Management*, 12(4), 1092-1116.

[27] Mu, R., Zheng, Y., Zhang, K., & Zhang, Y. (2021). Research on Customer Satisfaction Based on Multidimensional Analysis. *International Journal of Computational Intelligence Systems*, 14(1), 605-616.

[28] Muktafin, E. H. (2021). Sentiments analysis of customer satisfaction in public services using K-nearest neighbors algorithm and natural language processing approach. *Telkomnika*, 19(1), 146-154.

[29] Nazari, M., Shah Hosseini, M. A., & Tabatabaie Kalejahi, S. V. (2014). Impact of price perception factors on customer and price acceptance (Case study: MTN Irancell Company). *Journal of Business Management*, 6(3), 647-664.

[30] Paredes-Valverde, M. A., Colomo-Palacios, R., Salas-Zárate, M. D. P., & Valencia-García, R. (2017). Sentiment analysis in Spanish for improvement of products and services: a deep learning approach. *Scientific Programming*, 2017.

[31] Park, J. (2020). Framework for sentiment-driven evaluation of customer satisfaction with cosmetics brands. *IEEE Access*, 8, 98526-98538.

- [32] Park, M., McDonald, D., & Cha, M. (2013, June). Perception differences between the depressed and non-depressed users in twitter. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 7, No. 1).
- [33] Pérez, D., Alfonseca, E., Rodriguez, P., Gliozzo, A., Strapparava, C., & Magnini, B. (2005). About the effects of combining latent semantic analysis with natural language processing techniques for free-text assessment. *Revista signos*, 38(59), 325-343.
- [34] Reichheld, Frederick F. (diciembre 2003). "One Number You Need to Grow". *Harvard Business Review*.
- [35] Riaz, S., Fatima, M., Kamran, M., & Nisar, M. W. (2019). Opinion mining on large scale data using sentiment analysis and k-means clustering. *Cluster Computing*, 22(3), 7149-7164.
- [36] Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. *Information fusion*, 36, 10-25.
- [37] Tao, Y., Zhang, F., Shi, C., & Chen, Y. (2019). Social media data-based sentiment analysis of tourists' air quality perceptions. *Sustainability*, 11(18), 5070.
- [38] Tarnowska, K. A., & Ras, Z. (2021). NLP-Based Customer Loyalty Improvement Recommender System (CLIRS2). *Big Data and Cognitive Computing*, 5(1), 4.
- [39] Vallez, M., & Pedraza, R. (2007). *El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines*. Hipertext. net.
- [40] Villagra, A., Guzmán, A., Pandolfi, D., & Leguizamón, G. (2009). Análisis de medidas no-supervisadas de calidad en clusters obtenidos por K-means y Particle Swarm Optimization. *Ciencia y Tecnología*.
- [41] Wijngaards, I., Burger, M., & van Exel, J. (2020). Correction: The promise of open survey questions—The validation of text-based job satisfaction measures. *Plos one*, 15(7), e0236900.
- [42] Zainuddin, N., & Selamat, A. (2014, September). Sentiment analysis using support vector machine. In 2014 international conference on computer, communications, and control technology (I4CT) (pp. 333-337). IEEE.

- [43] Customer.guru - Net Promoter Score for E-Shops and SaaS [En línea] <https://customer.guru/> [último acceso: 22 Febrero 2022]
- [44] J. Chuang, CD Manning y J. Heer, Termite: Técnicas de visualización para evaluar modelos de temas textuales (2012), Departamento de Ciencias de la Computación de la Universidad de Stanford
- [45] J. Chang et al, Lectura de hojas de té: cómo los humanos interpretan los modelos temáticos (2009), Sistemas de procesamiento de información neuronal
- [46] Robertson, A. M., & Willett, P. (1998). Applications of n-grams in textual information systems. *Journal of Documentation*.
- [47] Welte, W. (1985). *Lingüística moderna, terminología y bibliografía* (Vol. 5). Gredos.
- [48] Risch, J. (2016). Detecting Twitter topics using latent Dirichlet allocation.
- [49] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- [50] Hammoe, L. (2018). Detección de tópicos: utilizando el modelo LDA.
- [51] Venegas, V. (2003). Análisis Semántico Latente: una panorámica de su desarrollo. *Revista signos*, 36(53), 121-138.
- [52] Vásquez, A. C. (2017). Clasificación de textos informáticos mediante indexación semántica latente. *Perfiles de Ingeniería*, 13(13), 109-122.
- [53] Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755), 788-791.
- [54] Velásquez García, L. E. (2021). Modelamiento de tópicos aplicados al estudio de ataques de Ingeniería social (Bachelor's thesis, Quito, 2021).
- [55] Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013, May). A biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 1445-1456).

- [56] Hamzeian, D. (2021). Using Machine Learning Algorithms for Finding the Topics of COVID-19 Open Research Dataset Automatically (Master's thesis, University of Waterloo).
- [57] Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011, July). Optimizing semantic coherence in topic models. In Proceedings of the 2011 conference on empirical methods in natural language processing (pp. 262-272).
- [58] Newman, D., Chemudugunta, C., & Smyth, P. (2006, August). Statistical entity-topic models. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 680-686).
- [59] Ipsos [En línea] <https://www.ipsos.com/es-es/nosotros> [último acceso: 22 Febrero 2022]
- [60] *Sentiment-analysis-spanish*. (2021, February 4). [En línea] <https://pypi.org/project/sentiment-analysis-spanish/> [último acceso: 22 Febrero 2022]
- [61] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, 2(1–2), 1-135.
- [62] Saura, J. R., Reyes-Menendez, A., & Palos-Sanchez, P. (2019). Are black Friday deals worth it? Mining Twitter users' sentiment and behavior response. *Journal of Open Innovation: Technology, Market, and Complexity*, 5(3), 58.
- [63] Arango Pastrana, C. A., & Osorio Andrade, C. F. (2021). Aislamiento social obligatorio: un análisis de sentimientos mediante machine learning. *Suma de Negocios*, 12(26), 1-13.
- [64] Gensim: topic modelling for humans [En línea] <https://radimrehurek.com/gensim/> [último acceso: 22 Febrero 2022]
- [65] scikit-learn: machine learning in Python — scikit-learn 1.0.2 documentation [En línea] <https://scikit-learn.org/stable/> [último acceso: 22 Febrero 2022]
- [66] pyLDAvis — pyLDAvis 2.1.2 documentation [En línea] <https://pyldavis.readthedocs.io/en/latest/readme.html> [último acceso: 22 Febrero 2022]