UNIVERSIDAD DE CONCEPCIÓN
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

SEGMENTACIÓN DE CONVERSACIÓN ENTRE CLIENTE-EJECUTIVO

TESIS PARA OPTAR AL TÍTULO PROFESIONAL DE INGENIERO CIVIL MATEMÁTICO

JOSÉ IRRIBARRA BASTÍAS

PROFESOR GUÍA:
Gabriel Villalobos Pedreros.

MIEMBROS DE LA COMISIÓN:
John Atkinson.
Julio Aracena.

CONCEPCIÓN, CHILE
2023

Con amor para toda mi familia,
Mis Padres, mi Hermana y mis Abuelos.

Saludos

Agradecimientos

Quiero agradecer en primera instancia a toda mi familia, que son los que han estado apoyándome toda la vida, mi padre que se ha esforzado infinitamente para que pueda sacar mi carrera y darme todo lo que necesite para estudiar, mi madre que con todo su cariño siempre me apoya y me da palabras de aliento, mi hermana que me ve estresado muchas veces y me termina retando. Siempre me apoyan aunque muchas veces no saben cómo ayudar, se los agradezco, también agradecer a mis abuelos que han estado aquí para mí, aunque mi abuelita ya no está conmigo. Esta tesis se la dedico a todos ellos, porque así como el mío, ha sido su sueño verme titulado, son ellos mi núcleo familiar que me ha dado la fuerza en todo momento para salir adelante y lograr todo lo que me propongo.

Tampoco quiero desmerecer el apoyo de mis amigos que he conocido aquí en la universidad, las salidas, los grupos de estudio, todas las experiencias en conjunto, todo has sido un lindo recuerdo digno de añorar en un futuro.

Finalmente a mi grupo de trabajo en esta tesis, a mi jefe Gabriel que ha tenido la paciencia de guiarme en este proyecto, y agradecerle a él y al equipo de trabajo todos los consejos que me ha dado para crecer tanto de manera profesional, como personal y me han mostrado cómo es este mundo laboral en el que próximamente me adentraré.

SEGMENTACIÓN DE CONVERSACIÓN ENTRE CLIENTE-EJECUTIVO

En la industria bancaria es fundamental ir modernizando la atención hacia los clientes. Cada vez se demanda más inmediatez en resolución de problemáticas, y con el aumento de clientes se es cada vez más difícil dar solución a los problemas de cada uno de ellos. Es por ello que es fundamental la auto atención digital. Si se es capaz de entender cuáles son las principales trabas que tienen los clientes se podrá dar soluciones a lo que ellos necesitan, con respuestas contextuales, procesos simples y ágiles, pero por sobre todo se podrá dar una experiencia al usuario que hará que cada uno de ellos pueda ser promotor del Banco Santander.

Por eso, para lograr estos objetivos es fundamental entender las problemáticas de nuestros clientes por los distintos canales, uno de los principales que se pueden encontrar más Insghits es el Chat con Ejecutivo, un canal resolutor de problemas que entrega una riquísima información de qué le pasa al cliente (para entender qué se debe mejorar) pero por sobre todo por qué llegó hasta ahí y poder ser proactivos y generar más confianza con ellos.

Para esto se necesita segmentar estas conversaciones, entender los principales tópicos y saber qué pasa dentro de ellas. Debido a esto que en colaboración con Banco Santander Chile, se define esta memoria de título. En ella se utilizarán modelos de deep learning, en particular técnicas asociadas a Natural Language Processing (NLP) en el software Python, para posteriormente poder desplegar en tiempo real los modelos matemáticos y poder tener un pulso en el instante de lo que pasa en el día a día en el Chat.

Tabla de Contenido

Índice de Tablas

Índice de Ilustraciones

Capítulo 1

Introducción

La interacción entre clientes y ejecutivos en diferentes plataformas de comunicación, como chats en línea, correos electrónicos y llamadas telefónicas, desempeña un papel crucial en la satisfacción del cliente, la retención y la adopción de productos o servicios. Para comprender y mejorar la relación y experiencia del cliente, es necesario analizar estas conversaciones y extraer información relevante que permita a las empresas tomar decisiones informadas y ajustar sus estrategias de atención al cliente.

En este contexto, la segmentación y análisis de las conversaciones entre clientes y ejecutivos adquiere una gran importancia, ya que proporciona una visión estructurada y organizada de las interacciones, permitiendo identificar áreas de oportunidad, detectar posibles problemas y mejorar la calidad del servicio brindado. Al segmentar las conversaciones y analizarlas en profundidad, es posible también evaluar la efectividad de las respuestas de los ejecutivos y adaptar las tácticas de comunicación para ofrecer un soporte más personalizado y eficiente.

En esta tesis, se aborda el problema de la segmentación de conversaciones entre clientes y ejecutivos mediante el uso de un modelo BERT y técnicas de clustering. El enfoque consiste en generar síntesis preliminar de alta calidad de las conversaciones utilizando el modelo BERT, lo que permite identificar palabras clave y temas relevantes en las interacciones de manera más efectiva. Estas síntesis de las conversaciones sirven como base para extraer información útil y categorizar las conversaciones de acuerdo a sus características, lo que facilita el análisis y la interpretación de los datos.

Además, la segmentación de las conversaciones permite obtener una perspectiva más precisa sobre la dinámica de la comunicación entre clientes y ejecutivos, identificando patrones, tendencias y áreas de mejora. Este enfoque tiene el potencial de transformar la forma en que las empresas entienden y gestionan sus interacciones con los clientes, impulsando la satisfacción del cliente y, en última instancia, contribuyendo al éxito empresarial.

Capítulo 2

Marco Teórico.

2.1. Definición y conceptos básicos

2.1.1. Conversaciones entre clientes y ejecutivos

Las conversaciones entre clientes y ejecutivos son interacciones que ocurren entre los representantes de una empresa (ejecutivos) y los consumidores o usuarios de sus productos o servicios (clientes). Estas conversaciones pueden llevarse a cabo a través de diversos medios de comunicación, como chats en línea, correos electrónicos, llamadas telefónicas, redes sociales y otros canales de atención al cliente.

En estas conversaciones, los clientes pueden plantear preguntas, solicitar información, expresar preocupaciones, hacer reclamos o proporcionar retroalimentación sobre productos y servicios. Por otro lado, los ejecutivos tienen la responsabilidad de responder a estas consultas, resolver problemas y brindar la información necesaria para satisfacer las necesidades y expectativas de los clientes.

Estas interacciones son fundamentales para mantener una relación sólida y duradera entre la empresa y sus clientes, ya que pueden afectar directamente la satisfacción del cliente, la percepción de la marca, la retención y la lealtad de los consumidores.

2.1.2. Segmentación de conversaciones y su importancia en la atención al cliente

La segmentación de conversaciones se refiere al proceso de dividir y organizar las conversaciones entre clientes y ejecutivos en grupos o categorías basados en su contenido, temas, palabras claves u otros atributos relevantes. Este proceso es esencial para analizar y comprender la dinámica de las interacciones y extraer información útil que permita mejorar la calidad y eficiencia de la atención al cliente.

La segmentación de conversaciones es importante en la atención al cliente por varias razones:

- Identificación de temas de interés: Al segmentar las conversaciones, es posible identificar los temas más recurrentes y las preocupaciones de los clientes, lo que facilita la priori-

zación de áreas de mejora y el desarrollo de estrategias para abordar dichos temas de manera eficaz.

- Evaluación del desempeño de los ejecutivos: La segmentación de conversaciones permite analizar el desempeño de los ejecutivos en términos de la calidad y efectividad de sus respuestas, lo que puede ser útil para la capacitación y el desarrollo del personal de atención al cliente.

- Personalización de la atención al cliente: La segmentación de conversaciones puede ayudar a identificar las necesidades y preferencias de los clientes en función de sus interacciones con los ejecutivos, lo que permite a las empresas ofrecer un servicio más personalizado y adaptado a las expectativas de cada cliente.

- Optimización de recursos: Al identificar los temas y preocupaciones más comunes en las conversaciones, las empresas pueden optimizar la asignación de recursos y esfuerzos para abordar de manera eficiente las necesidades de los clientes.

- Mejora de la experiencia del cliente: Al comprender las interacciones entre clientes y ejecutivos y abordar de manera proactiva las preocupaciones y necesidades de los clientes, las empresas pueden mejorar la experiencia del cliente, lo que puede resultar en una mayor satisfacción, retención y lealtad de los consumidores.

Así la segmentación de conversaciones es un proceso clave en la atención al cliente, ya que permite analizar y comprender las interacciones entre clientes y ejecutivos, identificar oportunidades de mejora y desarrollar estrategias efectivas para ofrecer un servicio de atención al cliente de alta calidad.

2.2. Algunas técnicas de procesamiento de lenguaje natural (NLP)

El procesamiento del lenguaje natural (NLP, por sus siglas en inglés) es una disciplina que se concentra en el entendimiento del lenguaje humano por parte de las computadoras. Esta área engloba aspectos de la ciencia de datos, inteligencia artificial y la lingüística. En el contexto de la segmentación de conversaciones, el NLP proporciona diversas técnicas y herramientas que permiten analizar y extraer información valiosa de las interacciones entre clientes y ejecutivos. A continuación, se presentan algunas de las técnicas de NLP más relevantes.

- Análisis de sentimientos [1]: El análisis de sentimientos se refiere al proceso de identificar y extraer las opiniones, emociones o actitudes expresadas en una secuencia de texto.

- Clasificación de texto [2][3]: La clasificación de texto es una tarea de NLP que implica asignar categorías o etiquetas a un texto en función de su contenido. En la segmentación de conversaciones, la clasificación de texto puede utilizarse para agrupar las interacciones en función de los temas discutidos o las preocupaciones planteadas por los clientes.

- Extracción de palabras clave [4]: La extracción de palabras clave es una técnica de NLP que identifica las palabras o frases más relevantes en un texto en función de su frecuencia, importancia o relación con otros términos. Al extraer palabras clave, se

simplifica el análisis y la comprensión de la información contenida en textos o frases, facilitando la obtención de información valiosa y la toma de decisiones basada en ellas. En la segmentación de conversaciones, la extracción de palabras claves puede utilizarse para identificar los temas más importantes en las interacciones y agrupar las conversaciones en función de estas palabras claves.

- Análisis de similitud y distancia entre textos [4]: El análisis de similitud y distancia entre textos implica medir cuán similares o diferentes son dos textos en función de su contenido, estructura o palabras clave. En la segmentación de conversaciones, esta técnica puede utilizarse para agrupar interacciones similares y detectar patrones en las conversaciones. Algunas medidas de similitud y distancia entre textos incluyen la similitud coseno [4][2] y la distancia de Levenshtein [5][6] , por nombrar algunas.

El objetivo principal de la segmentación de conversaciones es dividir las interacciones entre clientes y ejecutivos en segmentos o agrupaciones significativas que facilitan su análisis y comprensión. Para lograr este objetivo, es esencial considerar tres aspectos claves que afectan directamente a la calidad y la utilidad de los resultados obtenidos: la eficiencia del algoritmo de clustering [2][3], el impacto de la función de distancia o similitud [4] y la representación vectorial de los datos [2].

En primer lugar, el algoritmo de clustering desempeña un papel central en la segmentación de conversaciones. Dado que es común lidiar con grandes volúmenes de datos en aplicaciones de NLP, es vital seleccionar un algoritmo que pueda manejar eficientemente estas cantidades de datos y recursos computacionales. Los algoritmos de clustering varían en términos de complejidad, enfoque y capacidad para adaptarse a diferentes estructuras de datos. Por lo tanto, es fundamental evaluar y seleccionar el algoritmo de clustering adecuado según las características específicas del conjunto de datos y los requisitos de eficiencia.

El segundo aspecto a considerar es el impacto de la función de distancia o similitud en los resultados de la segmentación. La calidad y coherencia de las agrupaciones generadas depende en gran medida de la elección de la función de similitud. En el caso de datos de texto, la similitud del coseno [4] es una medida popular que evalúa la similitud entre dos vectores de palabras según el ángulo entre ellos. La selección de una función de similitud apropiada es crucial para garantizar que las agrupaciones sean relevantes y útiles en el análisis posterior. Por lo tanto, es importante elegir una función de similitud que capture de manera efectiva las relaciones entre los datos en el contexto específico de la tarea y el dominio.

Finalmente, una buena representación vectorial permite que los algoritmos de clustering identifiquen agrupaciones significativas y coherentes. En caso contrario, si las representaciones vectoriales no capturan adecuadamente las relaciones entre las palabras o frases, los resultados de la segmentación pueden ser de baja calidad y no representar la estructura subyacente de los datos.

2.3. Representación de palabras en un vector numérico

El avance en las técnicas de procesamiento del lenguaje natural ha llevado a la aparición de diversos enfoques para representar palabras y frases en forma de vectores. Estas representaciones vectoriales juegan un papel crucial en la realización de tareas de NLP, como la segmentación de conversaciones, ya que permiten capturar información semántica y sintáctica relevante del texto. En este capítulo, se compararán y analizarán dos enfoques principales para generar estas representaciones vectoriales: los métodos tradicionales que no capturan la semántica y la sintaxis de las palabras y los métodos neuronales más avanzados, como BERT [7], que sí lo hacen.

Los métodos tradicionales para generar representaciones vectoriales han sido utilizados durante mucho tiempo en el campo de NLP. Estos enfoques se basan en la idea de representar palabras como vectores dispersos de alta dimensionalidad [2]. Aunque estas representaciones pueden ser útiles para algunas tareas, carecen de la capacidad de capturar información semántica y sintáctica más profunda, lo que limita su utilidad en aplicaciones más avanzadas de NLP.

Por otro lado, los métodos neuronales que generan esta representaciones vectoriales han revolucionado el campo del procesamiento del lenguaje natural. Estos enfoques se basan en redes neuronales profundas [8] y arquitecturas de autoatención [9] para aprender representaciones vectoriales densas y de menor dimensionalidad que capturan información semántica y sintáctica contextualizada, las cuales a partir de ahora llamaremos Embeddings. A diferencia de los métodos tradicionales, los embeddings neuronales permiten capturar relaciones más complejas entre palabras y frases dentro de un contexto específico, lo que mejora significativamente el rendimiento en diversas tareas de NLP.

2.3.1. Modelos tradicionales

La representación de palabras en forma numérica es una técnica fundamental en el procesamiento del lenguaje natural (NLP) y en la construcción de modelos de aprendizaje automático que requieren datos de texto. En los enfoques tradicionales, se utilizan métodos simples pero eficientes para convertir palabras en números. Uno de ellos es asignar una etiqueta numérica única a cada palabra en un vocabulario predefinido. Por ejemplo, podemos asignar el número 1 a la palabra “casa”, el número 2 a la palabra “perro”, el número 3 a la palabra “árbol” y así sucesivamente.

Sin embargo, esta forma de representación puede tener algunas limitaciones y desventajas. Una de ellas es que no captura la relación semántica entre las palabras. Si utilizamos esta representación para entrenar algún modelo de Machine Learning, estaríamos entregando información al modelo de que la palabra “árbol” es tres veces más importante que la palabra “casa”, lo que no es cierto. Esto podría llevar a resultados inexactos en el modelo presentado.

Para evitar esto último, y hacer que todas las palabras tengan la misma importancia en el modelo se han desarrollado otros métodos de representación de palabras. *One-hot encoding*

[2] es una técnica en la que cada palabra se representa como un vector disperso de alta dimensionalidad con un 1 en la posición correspondiente a la palabra y 0 en todas las posiciones demás, basicamente se trata de un vector canónico para cada palabra, de ese modo se tiene que cada palabra es ortogonal y equidistante a cualquier otra. Aunque es una mejora con respecto de las etiquetas numéricas, sigue sin capturar las relaciones semánticas y sintácticas entre las palabras.

El método Bag of Words (BoW) [2][10] es otro enfoque que representa el texto describiendo las ocurrencias de palabras dentro de un documento, ignorando el orden y la gramática, pero manteniendo un registro de la frecuencia de cada palabra. Aunque esto permite capturar cierta información sobre la importancia de las palabras en un documento, aún no aborda adecuadamente las relaciones semánticas entre ellas.

Se trata de un modelo de representación de texto que se utiliza para transformar documentos de texto (que pueden ser frases, párrafos o libros, etc) en vectores numéricos que pueden ser procesados por algoritmos de Machine Learning. La idea es que cada documento de texto se represente como un conjunto desordenado de palabras (sin tener en cuenta el orden en que aparecen), lo que se conoce como "bolsa de palabras". Cada palabra se convierte en una característica o atributo en el modelo, y se cuenta el número de veces que aparece en el documento.

Por ejemplo, si tenemos tres frases:

- La ciencia es interesante.
- La ciencia es útil.
- La ciencia no es fácil.

La bolsa de palabras de estos documentos podría ser:

Tabla 2.1: Bolsa de palabras

Palabra	Documento 1	Documento 2	Documento 3
la	1	1	1
ciencia	1	1	1
es	1	1	1
interesante	1	0	0
útil	0	1	0
no	0	0	1
fácil	0	0	1

Por otro lado, existe otro método llamado *Term Frequency-Inverse Document Frequency (TF-IDF)* [11], que es una técnica de procesamiento de lenguaje natural (NLP) y minería de textos que se utiliza para evaluar la importancia de una palabra en un conjunto de documentos. TF-IDF consta de dos componentes principales:

- *Term Frequency* (TF): Es una medida de la frecuencia de una palabra en un documento. Se calcula dividiendo el número de veces que una palabra aparece en un documento por el número total de palabras en ese documento. Así, la frecuencia de una palabra en un documento específico es proporcional a su importancia en ese documento.

$$\text{TF}(t, d) = \left(\frac{\text{Número de veces que el término } t \text{ aparece en el documento } d}{\text{Número total de términos en el documento } d} \right)$$

- *Inverse Document Frequency* (IDF): Es una medida de la importancia de una palabra en todo el conjunto de documentos. Se calcula tomando el logaritmo del cociente entre el número total de documentos y el número de documentos que contienen la palabra en cuestión. Esto ayuda a disminuir el peso de las palabras comunes que aparecen en muchos documentos y aumentar el peso de las palabras menos comunes pero más específicas.

$$\text{IDF}(t, D) = \log \left(\frac{N}{n_t} \right)$$

  Donde:

  - N es el número total de documentos en el conjunto de documentos D.
  - n_t es el número de documentos en el conjunto D que contienen el término t.

Finalmente, el TF-IDF de una palabra se calcula multiplicando su TF y IDF asociado:

$$\text{TF-IDF}(t, d, D) = \text{TF}(t, d) * \text{IDF}(t, D)$$

Cuanto mayor sea el valor de TF-IDF para una palabra en un documento, más relevante e importante será esa palabra en ese documento en comparación con otros documentos del conjunto. TF-IDF es ampliamente utilizado en la recuperación de información, la clasificación de texto y la agrupación de documentos para identificar palabras clave y temas en grandes conjuntos de datos de texto.

Usando el ejemplo anterior, se tiene para TF e IDF los siguientes valores.

Tabla 2.2: Term Frequency (TF)

Palabra	Documento 1	Documento 2	Documento 3
la	1/4	1/4	1/5
ciencia	1/4	1/4	1/5
es	1/4	1/4	1/5
interesante	1/4	0	0
útil	0	1/4	0
no	0	0	1/5
fácil	0	0	1/5

Tabla 2.3: Inverse Document Frequency (IDF)

Palabra	IDF
la	$\log(3/3) = 0$
ciencia	$\log(3/3) = 0$
es	$\log(3/3) = 0$
interesante	$\log(3/1) = \log(3)$
útil	$\log(3/1) = \log(3)$
no	$\log(3/1) = \log(3)$
fácil	$\log(3/1) = \log(3)$

2.3.2. Modelos Neuronales

Dado que los métodos tradicionales tienen ciertas limitaciones al no capturar de manera efectiva las relaciones semánticas y sintácticas entre palabras, surgió la necesidad de desarrollar enfoques más avanzados y precisos para la representación de palabras en el ámbito del procesamiento del lenguaje natural. Los métodos basados en redes neuronales, aparecieron como una solución a estas limitaciones. Estos enfoques han revolucionado la representación de palabras y han demostrado ser más efectivos en una amplia gama de tareas de NLP. A diferencia de los métodos tradicionales, los métodos neuronales aprenden representaciones de palabras que capturan información semántica, sintáctica y contextual. En las siguientes secciones, se explorará en detalle algunos de los métodos neuronales más populares y destacados.

2.3.2.1. Word2Vec

Word2Vec [8] es un modelo que utiliza redes neuronales sencillas para aprender representaciones vectoriales de palabras basadas en su contexto. Estas representaciones vectoriales, también llamadas embeddings, capturan el significado y las relaciones entre las palabras en un espacio de menor dimensión. Word2Vec se basa en dos arquitecturas principales:

CBOW (Bolsa continua de palabras)[8][2]: Este enfoque intenta predecir una palabra específica según las palabras que la rodean en una oración. Para hacer esto, define una ventana de tamaño N alrededor de la palabra de interés, considerando N palabras tanto a la izquierda como a la derecha de la palabra en la entrada. Al aprender a predecir palabras en función de su contexto, CBOW captura las relaciones semánticas y sintácticas entre las palabras.

Skip-Gram[8][2]: Este enfoque es la idea opuesta a CBOW. Dada una palabra de entrada, Skip-Gram intenta predecir las palabras que la rodean en una oración, también utilizando una ventana de tamaño N. Al aprender a predecir el contexto en función de una palabra específica, Skip-Gram también captura las relaciones semánticas y sintácticas entre las palabras.

Ambas arquitecturas, CBOW y Skip-Gram, se entrenan en grandes conjuntos de texto para aprender las relaciones entre las palabras. Como resultado, se obtienen vectores que representan palabras de manera que palabras similares o relacionadas estén cercanas en el espacio vectorial. Estos embeddings de palabras son útiles en una variedad de tareas de procesamiento del lenguaje natural, como clasificación de texto, traducción automática y análisis de sentimientos.

2.3.2.2. BERT (Bidirectional Encoder Representations from Transformers)

BERT (Bidirectional Encoder Representations from Transformers) [7] es un modelo de lenguaje revolucionario que ha transformado el campo del procesamiento del lenguaje natural (NLP). Utiliza una arquitectura de Transformer [9] para aprender representaciones de palabras en función de su contexto, lo que permite una comprensión más rica y precisa del lenguaje humano.

A diferencia de otros modelos de lenguaje anteriores, BERT es bidireccional, esto es debido al uso de *mecanismos de atención* [12] que implementa. Estos mecanismos permiten que

BERT asigne diferentes pesos o "importancias" a distintas palabras en una oración, dependiendo de su relevancia en el contexto general. En lugar de tratar todas las palabras con igual importancia, los mecanismos de atención otorgan a BERT la habilidad de focalizar selectivamente en palabras que son esenciales para entender un contexto dado, independientemente de la distancia entre ellas en una oración.

Profundizando en los mecanismos de atención de BERT, el modelo no se limita a un mecanismo de atención simple, sino que emplea una técnica más compleja conocida como *atención de múltiples cabezas* [9]. Esto permite que el modelo puede generar varios conjuntos de pesos de atención simultáneamente para una misma entrada. Cada "cabeza" de atención puede centrarse en distintas relaciones o patrones dentro de la oración, permitiendo al modelo capturar múltiples tipos de información contextual. Gracias a esta capacidad, el modelo puede identificar y considerar relaciones tanto explícitas como implícitas entre palabras.

El entrenamiento de BERT se realiza mediante dos tareas principales: el Modelado del Lenguaje Enmascarado (MLM) y la Predicción de la Siguiente Oración (NSP). En el MLM, se enmascaran algunas palabras de una secuencia de entrada y se entrena al modelo para predecir las palabras enmascaradas en función de su contexto. Durante el entrenamiento, aproximadamente el 15 % de las palabras de cada secuencia se seleccionan al azar para ser enmascaradas. BERT intenta predecir la palabra original utilizando el contexto proporcionado por las palabras no enmascaradas que la rodean. El MLM permite a BERT aprender a entender el contexto y las relaciones semánticas entre palabras en ambas direcciones, es decir, tanto de izquierda a derecha como de derecha a izquierda.

En NSP, BERT recibe un par de oraciones como entrada y aprende a predecir si la segunda oración es una continuación lógica de la primera. Durante el entrenamiento, aproximadamente el 50 % de los pares de oraciones consisten en dos oraciones que se siguen en el corpus original, mientras que el otro 50 % contiene una segunda oración seleccionada al azar que no está relacionada con la primera. BERT aprende a clasificar estos pares como "IsNext" (la segunda oración sigue a la primera) o "NotNext" (la segunda oración no sigue a la primera). Esta tarea ayuda a BERT a aprender sobre la estructura y las relaciones entre las oraciones.

Una vez que BERT se ha preentrenado en un gran corpus de texto, puede ajustarse para adaptarse a tareas específicas de NLP, como análisis de sentimientos, traducción automática, respuesta a preguntas, etc. Durante el ajuste fino, se añade una capa adicional al modelo y se entrena con un conjunto de datos específicos de la tarea en cuestión. Esto permite a BERT personalizarse y mejorar su rendimiento en tareas específicas de NLP.

Capítulo 3

Metodología

En este capítulo se abordarán los métodos y técnicas específicas utilizadas en este estudio para enfrentar los desafíos en la segmentación de conversaciones entre clientes y ejecutivos en el contexto del procesamiento del lenguaje natural en español. Se discutirá cómo se llevaron a cabo los experimentos, desde la adaptación del modelo BERT del idioma inglés al español hasta la extracción de información y palabras clave mediante el uso de representaciones vectoriales y medidas de similitud, como la similitud del coseno.

En este capítulo, se describirá el proceso de ajuste del modelo BERT al idioma español, detallando las decisiones tomadas y los recursos utilizados para adaptar el modelo a las particularidades lingüísticas del español. A continuación, se explicará cómo se empleó este modelo adaptado para extraer información de las oraciones en el contexto de las conversaciones entre clientes y ejecutivos.

Posteriormente, se abordará cómo utilizando la información extraída del modelo BERT, se extrajeron palabras clave a través de representaciones vectoriales de las palabras y midiendo su similitud mediante la similitud del coseno. Este enfoque permitió identificar temas y patrones importantes dentro de las conversaciones segmentadas.

Además, basándose en la información extraída del modelo BERT, se adquirieron clusters asociados a las conversaciones, lo que permitió agrupar y analizar de manera más efectiva las interacciones entre clientes y ejecutivos.

Se mantendrán y analizarán los resultados obtenidos en cada etapa del proceso, poniendo de relieve las contribuciones y avances que este trabajo aporta al campo del NLP en la segmentación de conversaciones.

3.1. Adaptación de un modelo BERT en Inglés al Español

El proceso de ajuste o "fine-tuning" [13] del modelo BERT [7] preentrenado en inglés al idioma español implica la adaptación de un modelo de lenguaje profundo, específicamente el modelo `all-mpnet-base-v2` [14], para que comprenda y procese textos en español. Para ello, se replicará lo trabajado en [15], donde el modelo `all-mpnet-base-v2` se elige como modelo profesor debido a su sólido rendimiento en tareas de procesamiento de lenguaje natural en inglés. El modelo `bert-base-multilingual-cased` [7], un modelo multilingüe que ya comprende varios idiomas, incluido el español, se selecciona como modelo estudiante.

Este proceso de ajuste, se lleva a cabo en los siguientes pasos principalmente:

- Cada oración en el idioma español, debe tener su equivalente en inglés
- Se usa el Modelo Profesor (Inglés) para codificar la oración en inglés y obtener su vector de Embedding.
- Se usa el Modelo Estudiante (Multilingüe) para codificar tanto la oración en inglés como la oración en español y obtener sus vectores de Embeddings respectivos.
- Usando ambos embeddings del modelo estudiante se minimiza la función de perdida (Error cuadratico medio - MSE)

De este modo el modelo estudiante imita las representaciones vectoriales del modelo profesor, tanto en inglés como en español, mejorando significativamente las representaciones en el espacio vectorial de ambos idiomas.

El proceso de "fine-tuning" se lleva a cabo en dos fases principales:

- Ajuste del modelo de estudiante con datos paralelos en inglés y español: En esta etapa, se utilizan los datos del conjunto de datos TED2020 [16], el cual contiene oraciones en inglés de conversaciones TED del año 2020 y su respectiva traducción a diferentes idiomas, más particularmente en idioma español, donde en la columna "source" se encuentra la oración en inglés y en la columna "target" su traducción correspondiente en español, como en Figura 3.1. El objetivo es enseñar al modelo estudiante a producir representaciones vectoriales de oraciones en español similares a las del modelo profesor en inglés. El entrenamiento se realiza mediante la minimización de una función de pérdida basada en el error cuadrático medio (MSE) entre los embeddings de las oraciones en inglés y español generados por el modelo estudiante y los embeddings de las oraciones en inglés generada por el modelo profesor.

	source	target
0	Thank you so much , Chris.	Muchas gracias Chris.
1	And it's truly a great honor to have the oppor...	Y es en verdad un gran honor tener la oportuni...

Figura 3.1: Ejemplo del contenido TED2020 dataset

- Ajuste del modelo estudiante con terminología específica del Banco: Dado que el conjunto de datos TED2020 contiene oraciones de uso cotidiano, el modelo estudiante podría no comprender adecuadamente la terminología bancaria específica utilizada en Banco Santander. Para abordar este problema, se agrega al diccionario del modelo los nuevos conceptos claves del banco que se desean aprender y posteriormente se lleva a cabo una segunda fase de "fine-tuning", utilizando un conjunto de datos creados específicamente para este propósito. Este conjunto de datos incluye 34,000 pares de oraciones extraídas de conversaciones entre clientes y ejecutivos del banco en español. Se utilizan modelos de traducción automática [17][18][19], como los modelos Transformer, para traducir estas oraciones al inglés. Luego, se ajusta el modelo estudiante nuevamente utilizando estos datos paralelos en inglés y español con terminología bancaria, dejandola con un formato similar al dataset de TED2020, donde en la columna "source" se encuentran las oraciones traducidas al inglés y en la columna "target" la oración original en español, como en Figura 3.2.

	source	target
0	I need to see the possibility to hire a consum...	necesito ver la posibilidad para contratar un ...
1	Hello I want to perform the renewal of my card	hola quiero realizar la renovacion de mi tarjeta

Figura 3.2: Ejemplo del contenido de las traducciones de las conversaciones entre cliente-ejecutivo del banco

3.2. Extracción de Información usando BERT

El objetivo principal de este modelo es extraer información relevante de las conversaciones originales, que pueden ser las declaraciones del cliente o del ejecutivo. Para lograr este objetivo, se emplea un modelo basado en BERT, adaptado específicamente para la tarea de extracción de información.

Para esto, se hará uso de la arquitectura *Transformers* [9]. La cuál es una arquitectura de aprendizaje automático especialmente diseñada para entender secuencias de datos, como el lenguaje natural. Lo más distintivo de estos es su capacidad para prestar "atención" a diferentes partes de una oración, dándole mayor importancia a las secciones más relevantes de una oración gracias a sus mecanismos de atención [12].

Se implementa un modelo basado en BERT, al cual se le agrega una capa de Decodificación basada en Transformers [9][20], esta capa adicional es esencial para adaptar el modelo BERT a la tarea de extracción de información en conversaciones entre cliente y ejecutivo, ya que permite ajustar el modelo a esta tarea específica. Al agregar una capa decodificadora, se transforma el modelo en un modelo de secuencia a secuencia (seq2seq), lo que facilita la identificación y extracción de información relevante, puesto que la arquitectura Transformer en la capa decodificadora emplea mecanismos de atención para enfocar el modelo en segmentos específicos de la conversación que contienen dicha información relevante, lo que permite identificar y extraer información clave, como solicitudes del cliente o respuestas del ejecutivo, etc.

Esto es útil para analizar y procesar información de la conversación de manera más eficiente y adaptar el modelo a las necesidades específicas del proyecto. Al aprovechar las representaciones contextuales de alta calidad aprendidas por BERT y la capacidad de generar contenido de la capa decodificadora, se puede mejorar significativamente el rendimiento en la tarea de extracción de información en conversaciones entre cliente y ejecutivo.

Para el entrenamiento de este modelo seq2seq, se ha utilizado la base de conocimientos de las conversaciones proporcionadas por Banco Santander Chile, se constan de 1200 ejemplos de extracción de información, de los cuales 600 corresponden a las declaraciones del cliente y los 600 restantes a las declaraciones del ejecutivo. Se utilizan 100 frases adicionales de cada grupo como conjunto de prueba (200 en total) . El modelo preentrenado se ajusta con estos datos para producir predicciones de extracción de información que se acerquen lo más posible a los resultados deseados. En Figura 3.3, se puede ver un extracto de este conjunto de datos, donde "col1" contiene las oraciones del cliente o del ejecutivo de la que sea desea extraer información importante, mientras que "col2" contiene la extracción deseada de "col1".

	col1	col2
0	avisame si no puedes contactarme y te mando un...	no puedes contactarme y he podido recibir llam...
1	lo estoy buscando lo obtuve en una automotora ...	lo estoy buscando lo obtuve en julio del deme ...

Figura 3.3: Ejemplo del conjunto de datos utilizado para la Extracción de Información.

3.3. Clustering de la información extraida

Una vez que se ha extraido la información relevante en cada conversación, se procede a la identificación de los clusters asociados exclusivamente con la información proporcionada por el cliente. Es importante destacar que este modelo de clustering es de uso exclusivo y propiedad privada de Banco Santander Chile. Debido a su carácter confidencial, no se mostrarán los resultados concretos del proceso de obtención de los clusters, ni se detallará la arquitectura interna del modelo en el presente documento.

3.4. Extracción de palabras claves

Una vez que se ha obtenido la información relevante a través del proceso de extracción de información usando el modelo BERT, el siguiente paso es identificar las palabras clave más significativas en las frases extraídas. Para lograr esto, se emplea un enfoque basado en la similitud del coseno entre representaciones vectoriales de palabras y frases.

El proceso se lleva a cabo de la siguiente manera:

- Obtención de representaciones vectoriales de las frases extraídas: Utilizando el modelo BERT ajustado en la sección anterior, se obtienen las representaciones vectoriales para las frases resultantes de la extracción de información. Estos vectores capturan el significado semántico de cada frase.

- Obtención de representaciones vectoriales de las palabras en las frases extraídas: Se utiliza el mismo modelo BERT para obtener las representaciones vectoriales de cada palabra individual presente en las frases extraídas. Estos vectores representan el significado semántico de cada palabra en el contexto de la frase.

- Cálculo de la similitud del coseno entre las representaciones vectoriales: Para identificar las palabras clave en cada frase, se calcula la similitud del coseno entre el vector que representa la frase completa y los vectores de cada palabra individual en la frase. La similitud del coseno es una medida que compara la orientación de dos vectores en un espacio multidimensional, y su valor varía entre -1 y 1. Un valor cercano a 1 indica que los dos vectores tienen una orientación muy similar, lo que sugiere una relación semántica estrecha entre la palabra y la frase.

Capítulo 4

Resultados

En este capítulo, se presentarán y analizarán los resultados de los modelos ya entrenados en sus distintos casos de usos, entrenamiento que fue llevado a cabo de manera iterativa por lo que se mostrarán los resultados finales de la última iteración. El propósito de este capítulo es presentar la eficacia y los resultados obtenidos al adaptar el modelo BERT al español, así como los hallazgos conseguidos mediante la extracción de información y la extracción de palabras clave utilizando la representación vectorial y la similitud del coseno. Además, se mostrarán los resultados de la agrupación de la información extraída mediante técnicas de clustering.

4.1. Adaptación de BERT al español

Para medir el rendimiento de los modelos en esta tarea, se ha utilizado el conjunto de datos `STS2017` (STS: Semantic Textual Similarity) [21] . Este conjunto contiene pares de oraciones con una puntuación de similitud que varía entre 0 y 5, donde 5 indica que las frases tienen una completa equivalencia semántica, mientras que 0 indica que las frases no tienen relación alguna en su significado, en Figura 4.1 se puede ver un extracto del conjunto de datos utilizado, donde "frase_1" y "frase_2" son las columnas que contienen oraciones a las que se desea medir su similitud semántica y la columna "similitud" contiene la similitud de las frases anteriores en una escala entre 0 y 5.

	frase_1	frase_2	similitud
0	Un perro está con un juguete	Un perro tiene un juguete	4.8
1	La gente cocina en la parrilla	Los peatones miran con asombro mientras dos pe...	0.8

Figura 4.1: Ejemplo del contenido del dataset STS2017

Con el propósito de comparar las evaluaciones de similitud producidas por el modelo con las puntuaciones anteriores, es crucial transformar estos valores a una escala que oscile entre 0 y 1, esto permite que la similitud derivada de los embeddings, como la similitud del coseno, sea comparable con las evaluaciones del conjunto de datos anterior.

Para determinar cuán alineadas están las puntuaciones del modelo con las del conjunto de datos de prueba, empleamos la correlación de Spearman [22]. A diferencia de otras correlaciones, Spearman no asume distribución normal entre los datos, pues es una prueba no

paramétrica y es menos susceptible a valores atípicos, lo que la hace ideal para este tipo de evaluación. Así, se busca determinar si, cuando las evaluaciones humanas indican que un par de oraciones son similares, nuestro modelo produce evaluaciones consistentes con esas similitudes.

En la tabla siguiente, se muestran los resultados obtenidos en la tarea de adaptación de BERT al español:

Modelo	Spearman_corr $\rho \times 100$
`Modelo_profesor`	90.60
`Modelo_estudiante`	56.69
`Modelo_estudiante_es`	77.40
`Modelo_estudiante_es_FT`	76.85

En la prueba de correlación de Spearman, el modelo estudiante base, que no fue ajustado, presenta un rendimiento considerablemente inferior al del modelo profesor, con una diferencia de 34 puntos porcentuales en la misma tarea de similitud en la base a la data STS2017 pero en diferente idioma (STS2017_ingles para modelo profesor, STS2017_español para modelo estudiante). Este resultado era esperado ya que el modelo estudiante base no fue entrenado específicamente para el español.

La aplicación de "fine-tuning" al modelo estudiante base utilizando los datos TED2020 en inglés-español dio lugar al `modelo_estudiante_es`, el cual mostró una mejora considerable, llegando a una correlación de Spearman de 77.40.

Finalmente, se realizó una segunda etapa de "fine-tuning" utilizando los datos del Banco Santander, resultando en el `modelo_estudiante_es_FT`. Este modelo mostró un rendimiento muy similar al `modelo_estudiante_es` en las pruebas segun la correlación de Spearman. Aunque la adaptación al dominio bancario específico redujo muy ligeramente su rendimiento general en la comprensión del español, pero ahora este modelo es capaz de entender mejor las entidades bancarias y terminología específica, lo que lo hace más útil para tareas como la búsqueda semántica y la extracción de palabras clave en este contexto específico.

4.2. Extracción de Información usando BERT

El objetivo del modelo de extracción de información basado en BERT es identificar la o las frases que mejor representen la conversación. Para medir su rendimiento, se emplean dos métricas clave: Precisión y F1-score. Estas métricas se calculan a partir de la coincidencia de palabras entre las frases que el modelo identifica como relevantes y las frases etiquetadas manualmente como relevantes en el conjunto de datos de prueba.

La tabla a continuación muestra el rendimiento promedio del modelo en términos de estas métricas sobre el conjunto de datos de prueba:

Metrica	promedio ($\times 100$)
Precisión	76.7
f1-score	66.5

La precisión de 76.7 % indica que, en promedio, el modelo identifica correctamente el 76.7 % de las palabras relevantes por frase. Esta alta precisión implica que el modelo comete pocos errores de falso positivo, es decir, raramente identifica una palabra como relevante cuando no lo es.

Sin embargo, el F1-score de 66.5 % es ligeramente más bajo que la precisión, por lo que se puede inferir que el modelo podría no estar capturando algunas de las palabras en las frases que debiesen ser relevantes.

Es importante destacar que mientras la precisión se centra en la exactitud de las palabras que el modelo identifica como relevantes, el F1-score nos ofrece una perspectiva más completa, considerando tanto la precisión como el recall. Esto es fundamental porque, aunque el modelo tenga una alta precisión, esto no necesariamente se traduce en un rendimiento general óptimo. Un F1-score más bajo que la precisión sugiere que el modelo, aunque es preciso en sus identificaciones, puede estar fallando en recuperar todas las palabras relevantes que forman parte de las frases clave, indicando así la presencia de falsos negativos, i.e, que existen ciertas frases en toda la conversación que el modelo no ha sido capaz de extraer completamente, o sea que hay palabras en la extracciones que no fueron consideradas relevantes por el modelo, cuando en realidad si lo eran.

Notar que en Figura 4.2 se ilustra cuál es la cantidad (en %) de frases de la data de testeo tiene menos de una cierta precisión, por ejemplo, se ilustra que menos del 10 % de los datos poseen una precisión inferior al 52 %, o mejor aún, el 70 % de los datos tiene sobre un 71 % de precisión.

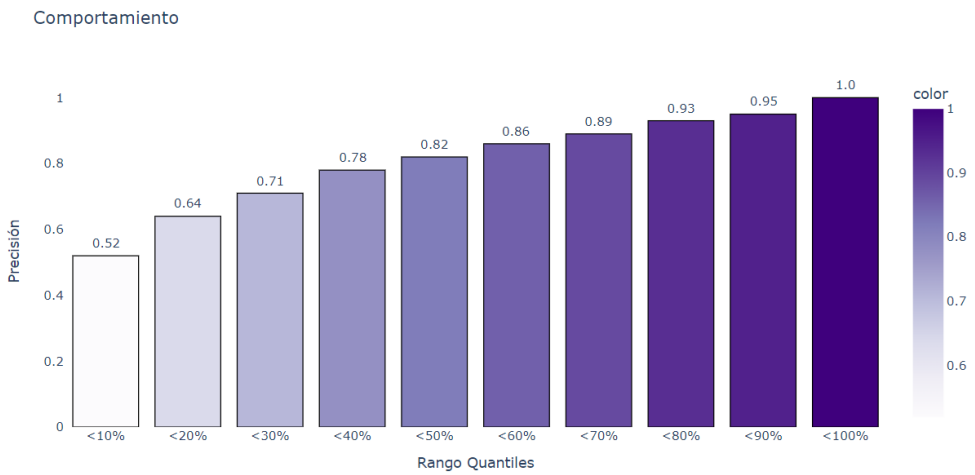


Figura 4.2: Quantiles/Precision

4.3. Clustering de la información extraída

Un primer análisis es obtener los temas o motivos de interacción con ejecutivo más concurridos, para ello se analizan las conversaciones entre clientes-ejecutivos (usando el modelo de Extracción de Información) entre los meses de Abril hasta Julio, para luego aplicar el modelo de clusters a la información extraída. En Figura 4.3 se ilustra cuáles son los clusters/tópicos más consultados por parte del cliente en el periodo mencionado.

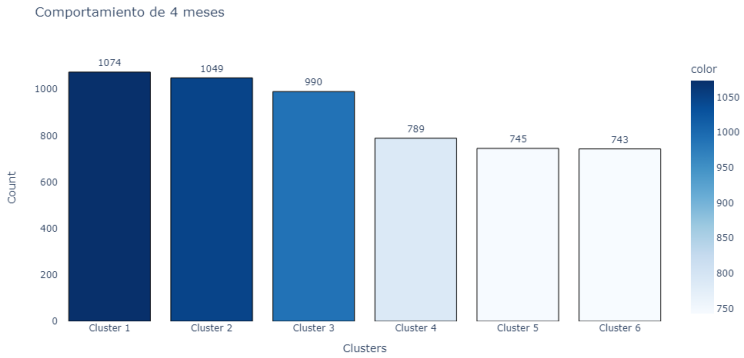
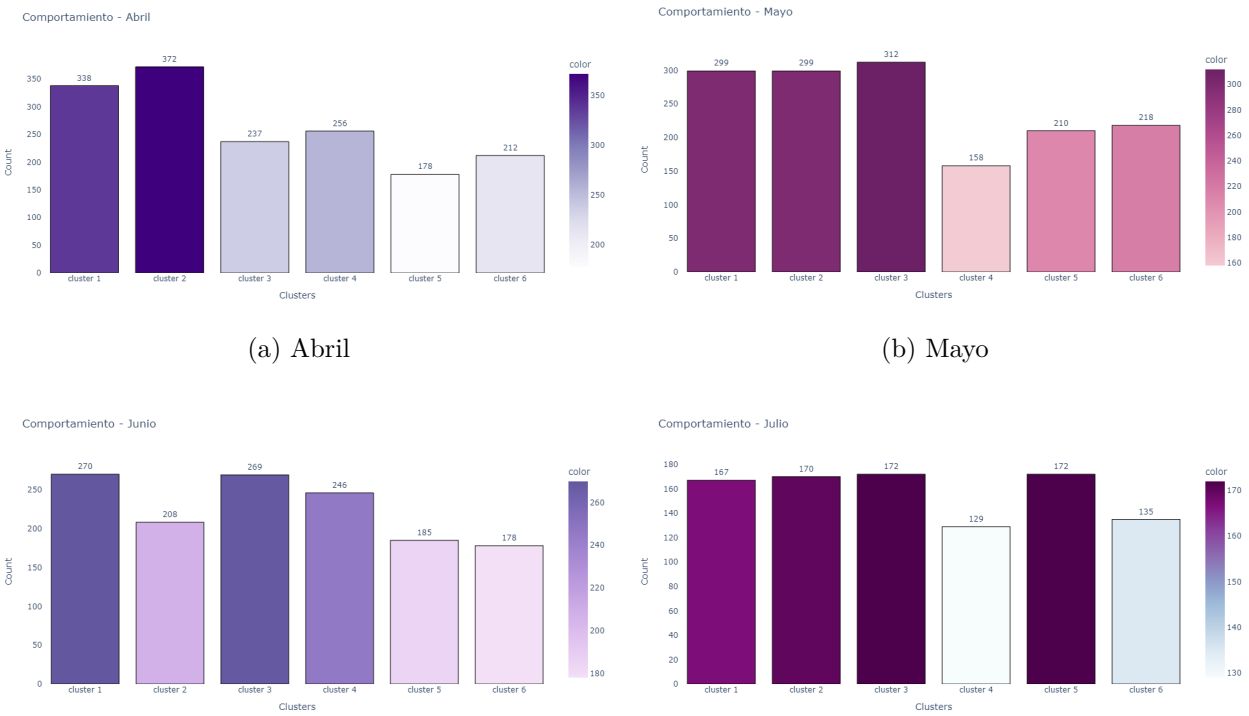


Figura 4.3: Clusters entre Abril y Julio

O bien, también es factible ver el comportamiento mensual de los clusters más frecuentes



(a) Abril

(b) Mayo

(c) Junio

(d) Julio

Figura 4.4: Clusters

4.4. Extracción de palabras clave

Haciendo uso del modelo de Palabras claves, nos damos cuenta por parte del cliente, que estas van en linea con lo encontrado en el modelo de cluster, por lo que se tiene una buena representación vectorial en el embedding ajustado.

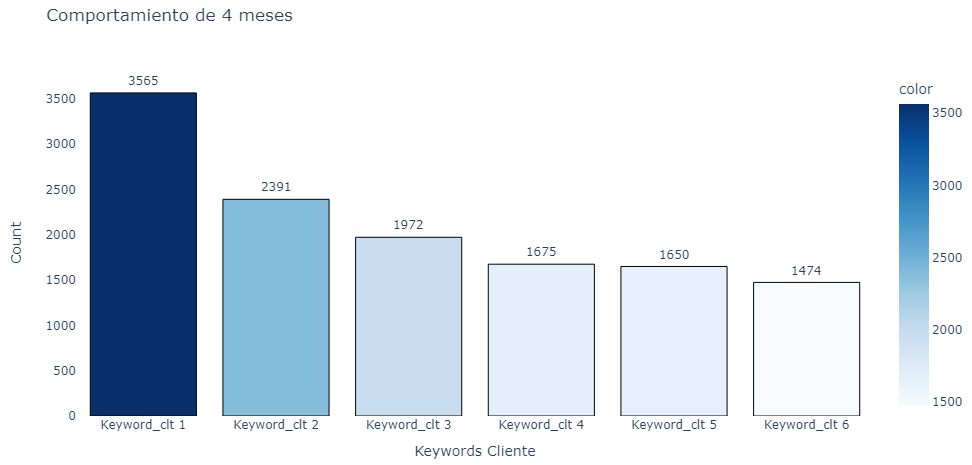
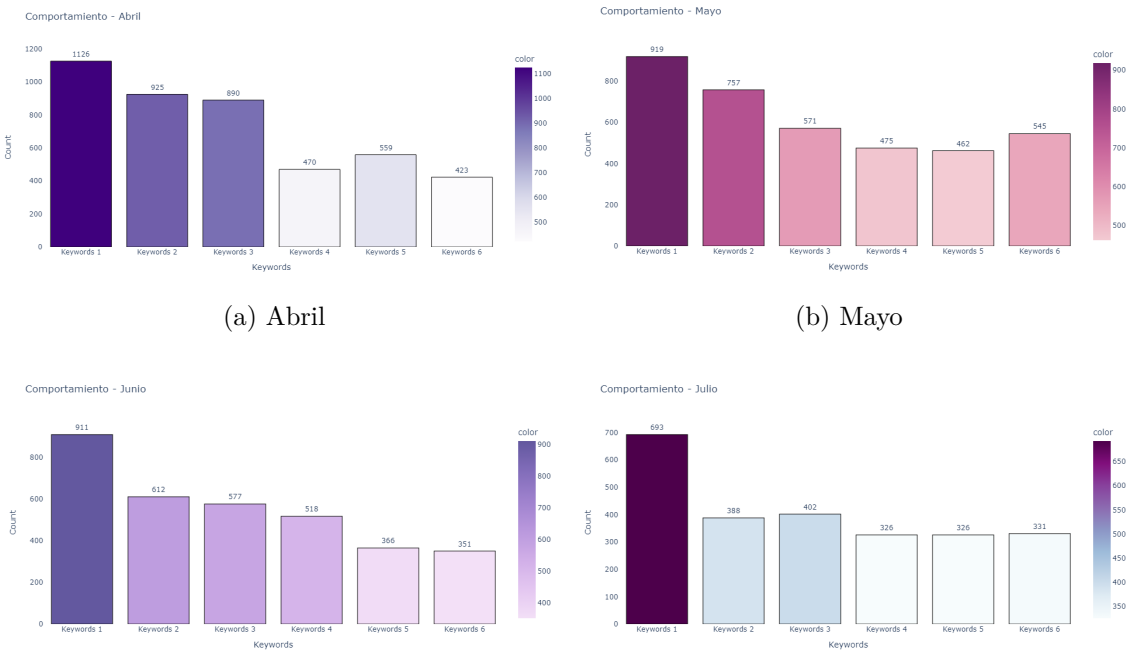


Figura 4.5: Palabras Clave del Cliente



(a) Abril

(b) Mayo

(c) Junio

(d) Julio

Figura 4.6: Palabras Clave del Cliente

Por otro lado, viendo las palabras claves del Ejecutivo, se ha podido tomar conciencia sobre las respuestas de estos, pudiendo saber en qué aspectos poder mejorar para la atención al cliente por parte del ejecutivo, y qué aspectos poder automatizar.

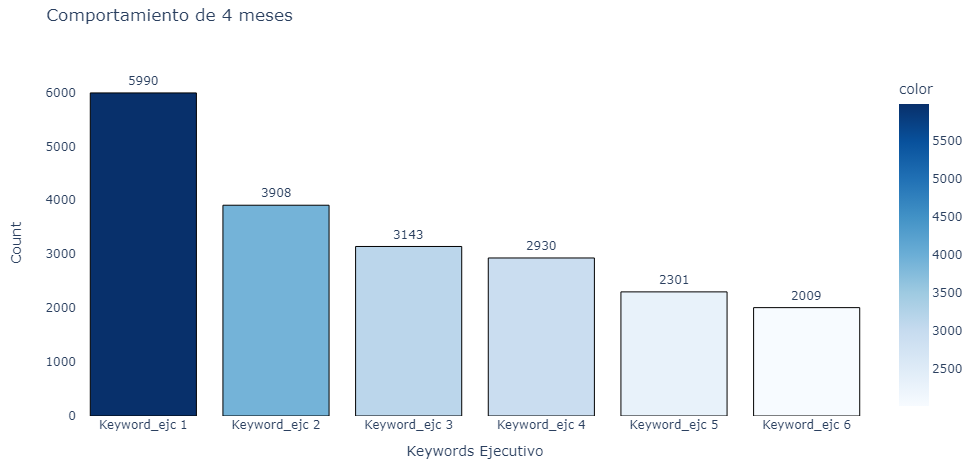
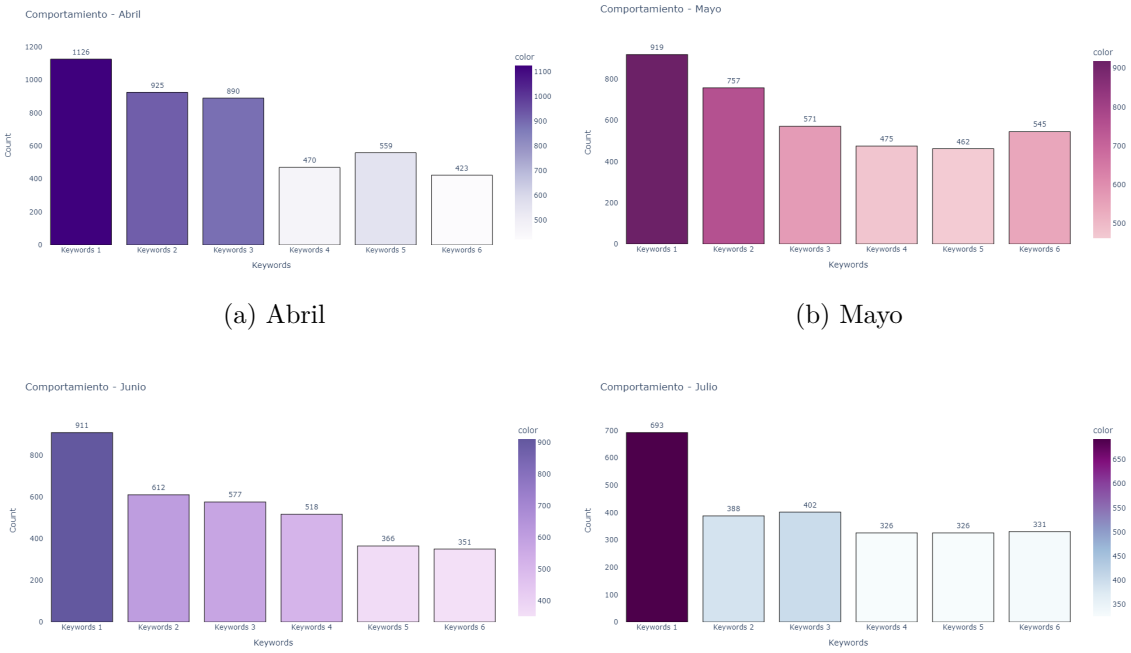


Figura 4.7: Palabras Clave del Ejecutivo



(a) Abril

(b) Mayo

(c) Junio

(d) Julio

Figura 4.8: Palabras Clave del Ejecutivo

Por otro lado, se ha de destacar que el comportamiento general de las palabras claves en el periodo de 4 meses, es estable y tiene el mismo comportamiento que en una vista mensual, esto ocurre tanto para el cliente como para el ejecutivo.

Capítulo 5

Conclusiones

En este trabajo, se ha adaptado y afinado un modelo de BERT para tareas de extracción de información específicas en el contexto bancario en español. El modelo BERT con capa decodificadora agregada, demostró ser eficaz para extraer información relevante de las interacciones entre clientes y ejecutivos luego del proceso de "fine-tuning".

Para medir la eficacia de este modelo en la tarea de extracción de información, se implementaron métricas de evaluación comunes, como la precisión y el F1-score. A pesar de que la precisión del modelo fue alta (76.7 %), lo que indica una alta tasa de verdaderos positivos, el F1-score fue relativamente más bajo (66.5 %), lo que sugiere que el modelo pueda estar omitiendo algunas palabras.

Además, para entender y visualizar mejor los temas más relevantes en las interacciones entre los clientes y los ejecutivos, se realizó un análisis de la clusterización de la información extraída. Esto proporcionó información útil sobre los temas más frecuentes y permitió al banco anticiparse a las necesidades de los clientes y ser consciente de las consultas más recurrentes.

Así, en este trabajo se han abordado una serie de técnicas y modelos, para poder ajustarlos y obtener los mejores resultados posibles para poder dar respuestas a incógnitas que se tenían por parte del banco, para que posteriormente la empresa pueda adaptarse tomando estos resultados en consideración y tener estos modelos desplegados en tiempo real para su posterior uso analítico.

5.1. Despliegue del modelo

Este Trabajo de tesis vino a resolver una problemática importantísima dentro del Banco, es por esto que es fundamental que este trabajo aplicado tenga un despliegue correcto y óptimo para que no quede solo en la teoría.

La evolución tecnológica nos permite hacer que estos modelos se utilicen en varias modalidades, por lo tanto implementaremos estos modelos de la siguiente forma:

1. Los modelos están desplegados en formato API post
2. Cada vez que el cliente termine el chat se ejecutará un proceso que llamará a esta API
3. Cuando la API entregue un resultado, este se guardará en tiempo real en los repositorios del banco para su posterior uso analítico. Esto es una ganancia relevante, ya que si posteriormente se quiere hacer analítica ya se tendrán los resultados del modelo guardado y no hay que procesar conversaciones posteriormente.
4. El resultado del Modelo además alimentará Dashboards para las unidades de Negocio que necesiten estar monitoreando el comportamiento de las conversaciones.

De esta manera los modelos están correctamente implementado y podrán ser correctamente utilizados por el Banco.

5.2. Trabajo futuro

Si bien, este trabajo ayuda a resolver una problemática importante dentro de la empresa, aún existe margen de mejora de los modelos, por lo que se puede hacer una revisión de estos para mejorar el rendimiento general, más particularmente se puede puede mejorar la puntuación de `Modelo_estudiante_es_FT` en español para que sea más cercano a la puntuación del `Modelo_profesor` en inglés y por ende tener mejores representaciones vectoriales en idioma español, y también mejorar aún más el modelo de Extracción de Información para disminuir la brecha entre la Precisión y el F1-score y aumentar ambas puntuaciones.

Por otro lado, este trabajo se enfocó solo en el chat entre cliente-ejecutivo, por ende es comprensible escalar estos modelos a las llamadas del call center, aplicando previamente un modelo Speech-to-Text (STT) para transcribir las llamadas y tener la interacción por escrito y poder proceder con el análisis de estos.

Bibliografía

[1] Ojamaa, B., Kristiina, P., y Muischnek, J. K., "Sentiment analysis on conversational texts.", www.filosoft.ee.

[2] Hirst, G., Goldberg, Y., Williams, P., Sennrich, R., Post, M., Koehn, P., Strötgen, J., Gertz, M., Gurevych, I., Eckle-Kohler, J., Matuschek, M., Cohen, S., Veale, T., Shutova, E., Klebanov, B. B., Heinz, J., Higuera, C. D. L., y Zaanen, M. V., "Synthesis lectures on human language technologies," 2015.

[3] Chowdhary, K. R., Fundamentals of artificial intelligence. Springer India, 2020, doi:10.1007/978-81-322-3972-7.

[4] Gunawan, D., Sembiring, C. A., y Budiman, M. A., "The implementation of cosine similarity to calculate text relevance between two documents," vol. 978, Institute of Physics Publishing, 2018, doi:10.1088/1742-6596/978/1/012120.

[5] Levenshtein, V. I. *et al.*, "Binary codes capable of correcting deletions, insertions, and reversals," en Soviet physics doklady, vol. 10, pp. 707–710, Soviet Union, 1966.

[6] Navarro, G., "A guided tour to approximate string matching.".

[7] Devlin, J., Chang, M.-W., Lee, K., y Toutanova, K., "Bert: Pre-training of deep bidirectional transformers for language understanding," 2019, https://arxiv.org/abs/1810.04805.

[8] Mikolov, T., Chen, K., Corrado, G., y Dean, J., "Efficient estimation of word representations in vector space," 2013, http://arxiv.org/abs/1301.3781.

[9] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., y Polosukhin, I., "Attention is all you need," 2017, http://arxiv.org/abs/1706.03762.

[10] Qader, W. A., Ameen, M. M., y Ahmed, B. I., "An overview of bag of words;importance, implementation, applications, and challenges," 2019, doi:10.1109/iec47844.2019.8950616.

[11] Ramos, J. *et al.*, "Using tf-idf to determine word relevance in document queries," en Proceedings of the first instructional conference on machine learning, vol. 242, pp. 29–48, Citeseer, 2003.

[12] Bahdanau, D., Cho, K., y Bengio, Y., "Neural machine translation by jointly learning to align and translate," 2014, https://arxiv.org/abs/1409.0473.

[13] Merchant, A., Rahimtoroghi, E., Pavlick, E., y Tenney, I., "What happens to bert embeddings during fine-tuning?," 2020, doi:10.48550/ARXIV.2004.14448.

[14] "all-mpnet-base-v2.", https://huggingface.co/sentence-transformers/all-mpnet-base-v2.

[15] Reimers, N. y Gurevych, I., "Making monolingual sentence embeddings multilingual using knowledge distillation," 2020, http://arxiv.org/abs/2004.09813.

[16] Reimers, N. y Gurevych, I., "Making monolingual sentence embeddings multilingual using knowledge distillation," en Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2020, https://arxiv.org/abs/2004.09813.

[17] Tiedemann, J. y Thottingal, S., "OPUS-MT — Building open translation services for the World," en Proceedings of the 22nd Annual Conferenec of the European Association for Machine Translation (EAMT), (Lisbon, Portugal), 2020.

[18] Tiedemann, J., "The Tatoeba Translation Challenge – Realistic data sets for low resource and multilingual MT," en Proceedings of the Fifth Conference on Machine Translation, (Online), pp. 1174–1182, Association for Computational Linguistics, 2020, https://www.aclweb.org/anthology/2020.wmt-1.139.

[19] Tiedemann, J. y Thottingal, S., "Helsinki-nlp/opus-mt-es-en, hugging face." https://huggingface.co/Helsinki-NLP/opus-mt-es-en, 2020.

[20] Rothe, S., Narayan, S., y Severyn, A., "Leveraging pre-trained checkpoints for sequence generation tasks," Transactions of the Association for Computational Linguistics, vol. 8, pp. 264–280, 2020, doi:10.1162/tacl_a_00313.

[21] Cer, D., Diab, M., Agirre, E., Lopez-Gazpio, I., y Specia, L., "SemEval-2017 task 1: Semantic textual similarity multilingual and crosslingual focused evaluation," en Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), Association for Computational Linguistics, 2017, doi:10.18653/v1/s17-2001.

[22] Zar, J. H., "Spearman rank correlation," 2005, doi:10.1002/0470011815.b2a15150.