



**Explorando la opinión de los usuarios de Twitter: Análisis de
sentimientos de marca mediante Deep Learning**

PROYECTO DE GRADO

Diego Fernando Soto Sarria

Julian Muñoz Bacca

Director

Javier Diaz Cely

FACULTAD DE INGENIERÍA

MAESTRÍA EN CIENCIA DE DATOS

SANTIAGO DE CALI

2023

Explorando la opinión de los usuarios de Twitter: Análisis de sentimientos de marca mediante Deep Learning

Diego Fernando Soto Sarria

Universidad ICESI, Cali

Julian Muñoz Bacca

Universidad ICESI, Cali

Trabajo de grado para optar por el título de Máster en Ciencia de Datos

Director

Javier Diaz Cely



FACULTAD DE INGENIERÍA

MAESTRÍA EN CIENCIA DE DATOS

SANTIAGO DE CALI

2023

RESUMEN

Identificar efectivamente las oportunidades de mejora es fundamental para toda organización; es por esto, que es de gran interés para las mismas tener conocimiento de la percepción de su marca en redes sociales como Twitter, donde sus clientes pueden expresarse pública y libremente.

Este estudio propone una solución teórico-práctica aplicando técnicas de minería de texto y Deep Learning sobre los tweets recopilados de los usuarios en 3 de las principales empresas prestadoras de servicios de telecomunicaciones como son Movistar, Claro y Tigo. Comparando las métricas de evaluación, en dos de las redes neuronales recurrentes mayormente usados en el análisis de sentimiento de texto, como son LSTM (Long Short-Term Memory) y GRU (Gated Recurrent Units).

Una vez realizada esta comparación, tanto GRU como LSTM obtuvieron muy buenos resultados en la métrica de evaluación y con poco sobre ajuste. Las pruebas ejecutadas con los modelos seleccionados mostraron una alta precisión en la clasificación de Tweets con sentimiento Negativo, con un porcentaje de Sensibilidad (Recall) en los datos de validación superiores al 94%. Sin embargo, en los Tweets con sentimientos No Negativos, la precisión fue más baja, con una Especificidad (Specificity) del 68%, 82.4% y 42.4% para Movistar, Claro y Tigo respectivamente, siendo significativamente bajo para este último.

La baja precisión para clasificar los Tweets no Negativos, se atribuyen a la gran variedad de temas para esta categoría, además de la baja cantidad de datos en comparación con los Negativos. Por lo tanto, para futuros estudios se recomienda el uso de un set de datos (Tweets) mucho más grande para mejorar la precisión en la clasificación de ambos grupos.

Gracias a esta clasificación y la identificación de aspectos negativos detectados en los diferentes comentarios en Twitter la solución propuesta permite gestionar de manera efectiva la experiencia de usuario mediante un tablero de control desarrollado en Power BI, este facilitará la supervisión de su presencia en esta red social, generando información clave que permita a la organización desarrollar estrategias de negocio basadas en datos que busquen abordar los problemas de manera efectiva y mejorar la calidad del servicio para satisfacer las necesidades del mercado.

TABLA DE CONTENIDO

1. INTRODUCCIÓN.....	9
1.1. Contexto, Antecedentes.....	9
1.2. Planteamiento del Problema.....	11
1.3. Objetivos del Proyecto.....	11
1.3.1. Objetivo General.....	11
1.3.2. Objetivos Específicos.....	11
2. MARCO TEÓRICO.....	13
2.1. Servicio de Atención al Cliente Social.....	13
2.2. Analítica de Negocio.....	15
2.2.1. Tipos de Analítica.....	16
2.2.1.1. Analítica Descriptiva.....	16
2.2.1.2. Analítica Predictiva.....	16
2.2.1.3. Analítica Prescriptiva.....	16
2.2.1.4. Analítica de Descubrimiento o Analítica de Sabiduría.....	16
2.3. Minería de Texto.....	17
2.3.1. KDD (Knowledge Discovery in Databases).....	17
2.3.2. KDT (Knowledge Discovery in Text).....	19
2.3.3. CRISP-DM.....	22
2.4. Análisis de Sentimiento.....	24
2.5. Deep Learning.....	25
2.5.1. Redes Neuronales Recurrentes (RNN).....	29
2.5.1.1. LSTM.....	31
2.5.1.2. GRU.....	34
3. ESTADO DEL ARTE.....	37
4. METODOLOGÍA.....	41
4.1. Fase 1: Comprensión del negocio.....	41
4.2. Fase 2: Comprensión de los datos.....	41
4.3. Fase 3: Preparación de los Datos.....	41
4.4. Fase 4: Modelamiento.....	42
4.5. Fase 5: Evaluación.....	44
4.6. Fase 6: Despliegue.....	45
5. RESULTADOS.....	46
5.1. Comprensión de los datos.....	46

5.1.1.	Negativos.....	49
5.1.2.	No Negativos	51
5.2.	Preparación de los Datos.....	53
5.2.1.	Vocabulario set de entrenamiento.....	53
5.2.2.	Representación secuencial.....	54
5.3.	Modelamiento	54
5.3.1.	Evaluación del Loss por época	55
5.4.	Evaluación	57
5.4.1.	Selección del punto de corte.....	57
5.4.2.	Métricas de desempeño.....	58
5.4.3.	Desempeño del modelo por Sentimiento	59
5.4.4.	Selección de modelo.....	60
5.5.	Despliegue del modelo	61
6.	CONCLUSIONES Y TRABAJOS FUTUROS	63
6.1.	Conclusiones	63
6.2.	Trabajos Futuros.....	64
7.	BIBLIOGRAFÍA.....	66

LISTA DE FIGURAS

Figura 1: Evoluciones técnicas de análisis de mercado. Fuente (Alfaro et al, 2012)	14
Figura 2: Tipos de Analítica. Adaptado (Raghupathi, W., & Raghupathi, V)	15
Figura 3: Fases Metodología KDD. Adaptada (Fayyad et al, 1996).....	18
Figura 4: Concepto de Jerarquía para conceptos tecnológicos. Adaptada (Feldman & Dagan,1995)	19
Figura 5: Metodología KDT. Fuente (Mansilla, 2018)	20
Figura 6: Metodología CRISP-DM. Fuente (IBM, 2021)	23
Figura 7: Relación IA, ML, Redes Neuronales y Deep Learning. Fuente (Kavlakoglu, 2020)	25
Figura 8: Relación IA, ML, Redes Neuronales y Deep Learning. Fuente (Kavlakoglu, 2020)	26
Figura 9: Modelo Red Neuronal Superficial. Fuente (Madhavan et al, 2020).....	27
Figura 10: Modelo Red Neuronal Profunda. Fuente (Madhavan et al, 2020)	28
Figura 11: Modelo Red Neuronal Recurrente. Fuente (Jones, 2017)	30
Figura 12: Modelo Red Neuronal Recurrente Simplificado. Fuente (Jones & Madhavan, 2021)33	
Figura 13: Modelo Base de una Red LSTM. Fuente (Jones & Madhavan, 2021)	33
Figura 14: Modelo Celda LSTM. Fuente (Olah, 2015).....	34
Figura 15: Modelo Celda GRU. Fuente (Olah, 2015)	35
Figura 16: Número de Tweets por día.....	47
Figura 17: Frecuencia de palabras Movistar	48
Figura 18: Frecuencia de palabras Claro	49
Figura 19: Frecuencia de palabras Tigo.....	49
Figura 20: Frecuencia de palabras Negativas Movistar.....	50
Figura 21: Frecuencia de palabras Negativas Claro.....	50
Figura 22: Frecuencia de palabras Negativas Tigo	50
Figura 23:Frecuencia de palabras No Negativas Movistar	52

Figura 24:Frecuencia de palabras No Negativas Claro	52
Figura 25:Frecuencia de palabras No Negativas Tigo.....	52
Figura 26: Evolución del Loss durante el entrenamiento por época Movistar	56
Figura 27: Evolución del Loss durante el entrenamiento por época Claro	56
Figura 28: Evolución del Loss durante el entrenamiento por época Tigo.....	56
Figura 29: Rendimiento de modelos Movistar	57
Figura 30: Rendimiento de modelos Claro	57
Figura 31: Rendimiento de modelos Tigo.....	57
Figura 32: Desempeño de modelos por Sentimiento Movistar	59
Figura 33: Desempeño de modelos por Sentimiento Claro	60
Figura 34 Desempeño de modelos por Sentimiento Tigo.....	60
Figura 35: Despliegue de modelo Movistar	61
Figura 36: Despliegue de modelo Claro	62
Figura 37: Despliegue de modelo Tigo.....	62

LISTA DE TABLAS

Tabla 1: Estado del Arte. Elaboración propia.	40
Tabla 2: Rango de hiperparámetros.....	43
Tabla 3: Métricas de evaluación.....	¡Error! Marcador no definido.
Tabla 4: Distribución por sentimiento	47
Tabla 5: Distribución por agrupación de sentimiento	48
Tabla 6: Vocabulario Movistar.....	53
Tabla 7: Vocabulario Claro.....	53
Tabla 8: Vocabulario Tigo	53
Tabla 9 Ejemplo Matriz de secuencia de Tokens	54
Tabla 10: Modelos Movistar	55
Tabla 11: Modelos Claro	55
Tabla 12: Modelos Tigo.....	55
Tabla 13: Umbral óptimo de los modelos	58
Tabla 14: Métricas de evaluación modelos	59

1. INTRODUCCIÓN

1.1. Contexto, Antecedentes

La era digital ha transformado la manera en que las organizaciones interactúan con sus usuarios, permitiendo una comunicación más ágil y eficiente; esto suma entonces mayores retos para las organizaciones en la gestión de sus canales de comunicación.

En este contexto, conceptos como la escucha digital o escucha en redes, son una estrategia clave para las organizaciones que desean conocer y mejorar la atención de sus usuarios, así como el interés de clientes potenciales; permitiéndoles identificar sus necesidades y preferencias, al igual que sus oportunidades y riesgos de manera inmediata y en cualquier momento.

La escucha digital consiste en el monitoreo y análisis de las interacciones sostenidas en canales digitales provistos por la organización, tales como redes sociales, siendo Twitter donde se concentra la mayor interacción entre usuarios y sus proveedores de productos o servicios, esto gracias a la libertad y la visibilidad pública que esta representa. Esta interacción digital administrada de manera adecuada puede tener un impacto significativo en la percepción de la marca y, como resultado, en los ingresos y utilidades de una organización.

De acuerdo con Kriss, P (2014). *The Value of Customer Experience, Quantified. Harvard Business Review*, los usuarios que tuvieron las mejores experiencias en relación con la atención recibida por sus proveedores gastan un 140% más en comparación con aquellos que tuvieron una mala experiencia.

Así mismo, menciona en su artículo, que un usuario que recibe una mala atención tiene solo un 43% de posibilidades de continuar con su servicio en el próximo año, en comparación a los usuarios que recibieron una atención superior, quienes tendrían un 74% de posibilidades de continuar con su servicio durante al menos otro año más, resumiendo así en una disminución de las tasas de abandono.

Resalta en su estudio que la principal justificación de las organizaciones para no invertir en mejorar sus procesos de calidad y efectividad en sus canales de atención es el alto costo de inversión que estos representan, sin embargo, en su proceso de investigación, ejecutivos de

grandes organizaciones concuerdan, que brindar excelentes experiencias en realidad reduce el costo de atender a los usuarios.

“los usuarios insatisfechos son costosos en términos de logística en la devolución de sus productos/servicios o en la asistencia de sus requerimientos si trasladamos esto en los costos operativos que conllevan.”

La gestión de esta interacción digital no es tarea fácil, es por esto, que, desde el gobierno de TI de cada organización, se deben definir lineamientos que incluyan la gobernabilidad de la experiencia del cliente (CX), *Customer Experience* por sus siglas en inglés, que faciliten el trabajo en conjunto y coordinado entre las distintas áreas de la organización y los proyectos de TI orientados a mejorar la experiencia del cliente.

Según el informe *“How IT Can Strengthen Customer Experience Governance”*, (Scheibenreif et al, 2023), el gobierno de CX asegura que los esfuerzos en el desarrollo de estrategias que mejoren la CX sean estratégicos, coordinados, coherentes y efectivos en toda la organización. Esto significa que las organizaciones que implementan un gobierno efectivo de CX pueden mejorar significativamente la satisfacción del cliente, su fidelidad, y por ende los resultados financieros.

Señala además que las empresas líderes en CX tienen más probabilidades de superar a sus competidores en ingresos, ganancias y retención de clientes. Entendiendo entonces los proyectos orientados a CX como iniciativas estratégicas dentro de una organización, por lo cual estima en su informe, que para el 2023 el 50% de los proyectos de TI se centrarán en la mejora de CX.

Según informe emitido por la Comisión de Regulación de Comunicaciones (CRC), en Colombia, los operadores con mayor participación del mercado son Claro, Movistar y Tigo-Une (Postdata,2023), donde sus métricas en relación a la satisfacción de cliente de los últimos años no han sido alentadoras.

El artículo publicado por el periódico la República, señala que en el 2021 se registraron cerca de 8,1 millones de reclamaciones; la mayoría de estas se presentaron para el servicio de internet fijo con 2,7 millones de reclamaciones registradas, seguido de telefonía móvil con 2,15 millones, televisión por suscripción con 1,59 millones, telefonía fija con 1,25 millones y finalmente de

internet móvil con 401.282. De este consolidado de reclamaciones, el 77,52% están agrupadas en cuatro categorías: no disponibilidad del servicio (29,64%); error en la factura, cobro o descuento injustificado (28,15%); intermitencia en el servicio (11,86%); y modificación de las condiciones acordadas (7,87%) (LaRepública, 2022).

Es por esta razón que las organizaciones necesitan una estrategia de interacción coherente en todos los canales digitales de atención, que sean basados en datos en tiempo real para satisfacer las expectativas de los consumidores, o corren el riesgo de quedarse rezagadas con respecto a la competencia. La clave es invertir en herramientas de interacción basadas en datos y personalizar las interacciones con cada cliente en sus términos.

1.2. Planteamiento del Problema

Monitorear y analizar de manera adecuada las interacciones de los clientes en los canales digitales mediante modelos de Deep Learning, ha demostrado tener grandes resultados para la analítica de negocios en la definición de planes estratégicos para mejorar la satisfacción de los clientes. El objetivo de este estudio es responder la siguiente pregunta:

¿Cuál es la percepción de los usuarios sobre la calidad de los servicios en Colombia de las empresas Movistar, Claro y Tigo en Twitter?

1.3. Objetivos del Proyecto

1.3.1. Objetivo General.

Evaluar la percepción de la calidad en la prestación de servicios de tres de las empresas de telecomunicaciones con mayor participación del mercado en Colombia, como son Movistar, Claro y Tigo en la red social Twitter.

1.3.2. Objetivos Específicos.

- Recopilar y construir la base de datos de los distintos comentarios de los usuarios en la red social Twitter con respecto a la marca de los operadores Movistar, Claro y Tigo, por medio de técnicas de minería de datos.
- Comparar el desempeño de los modelos LSTM y GRU de Deep Learning para Clasificar la polaridad (negativa y no negativa) de los últimos 10 mil comentarios recopilados de los usuarios en la red social Twitter, hasta el 29 de junio del 2022.

- Explorar y analizar el conjunto de datos recopilados mediante técnicas de análisis de sentimiento y analítica de negocios.
- Construir un tablero de visualización dinámica de las predicciones de los modelos desplegados para cada operador, que permita hacer seguimiento de la evolución de la percepción de marca en el tiempo, y encontrar patrones diferenciadores entre los Tweets por sentimiento.

2. MARCO TEÓRICO

2.1. Servicio de Atención al Cliente Social.

La tecnología y el acceso a internet ha revolucionado la forma en que interactuamos con el mundo. Desde el lanzamiento del primer smartphone y la explosión de las aplicaciones móviles de redes sociales como Twitter, Facebook, Instagram y YouTube la manera en que las empresas interactúan con sus clientes se ha impactado significativamente.

Las organizaciones vieron entonces en estos nuevos canales de difusión una oportunidad para incrementar sus ventas, Sin embargo, surgió un fenómeno el cual no dimensionaban, y es que esta comunicación funcionaba en ambos sentidos; donde cualquier persona expresaba sus ideas y experiencias de manera libre y sin restricciones.

Los clientes comenzaron a publicar sus inconformidades con el producto o servicio, o en búsqueda de referencias de terceros si tenían dudas o interés sobre los mismos; todos ellos, en su mayoría, tenían algo en común, habían tratado de resolver sus problemas o inquietudes a través de los canales tradicionales de Atención al Cliente, y muchos de ellos no habían obtenido una respuesta oportuna o satisfactoria. Es aquí donde las redes sociales brindaron esa nueva opción de contacto, convirtiéndose en un eje importante en la experiencia del cliente (CX) para las organizaciones.

Las organizaciones han comenzado a usar estos nuevos canales de comunicación con sus clientes, respondiendo sus preguntas y comentarios de manera directa y personalizada, buscando siempre la efectividad en la misma. Es esta gestión la que se define como *escucha digital* o *escucha en redes*, y surge entonces el concepto de Servicio de Atención al Cliente Social (Hill-Wilson & Blunt, 2013).

Pero ofrecer productos o servicios de excelente calidad ya no son suficientes para garantizar el éxito de una organización. En un entorno de mercado dinámico, es fundamental comprender cómo las percepciones son detonantes de emociones y sentimientos para los clientes de una organización, y cómo estas afectan directamente en sus ingresos y desempeño.

Ahora bien, debemos hacernos la pregunta, ¿Qué entendemos por el término experiencia del cliente? Si definimos el concepto desde el “uso o práctica”, la experiencia puede definirse como

todos los canales de interacción que tiene un cliente con la empresa, tales como internet, redes sociales, tienda física, los empleados, etc. Por otra parte, si abordamos el concepto desde el enfoque de la “vivencia”, su definición está directamente relacionada con aspectos emocionales, lo cual hace mucho más compleja su administración, involucrando otras ramas de investigación como el neuromarketing, análisis cuantitativo, análisis lingüístico y semántico, así como el uso de modelos estadísticos que permiten definir qué aspectos perceptivos impactan en determinados comportamientos de las personas.

Hasta hace unos años si una organización deseaba responder una pregunta de mercado, acudía a estudios de investigación tradicionales para obtener estos datos, y poder apoyarse en su toma de decisiones. Estos estudios se basaban en análisis cuantitativos soportados sobre encuestas o datos internos de la misma organización e investigaciones sustentadas en la observación de clientes en el entorno de la compra, recopilados mediante entrevistas personales o grupales, más conocidos como “Focus Group” (Alfaro et al, 2012). Esto ha evolucionado a la fecha para poder incluir en este análisis las nuevas ramas de investigación mencionadas previamente. (Ver Figura 1)

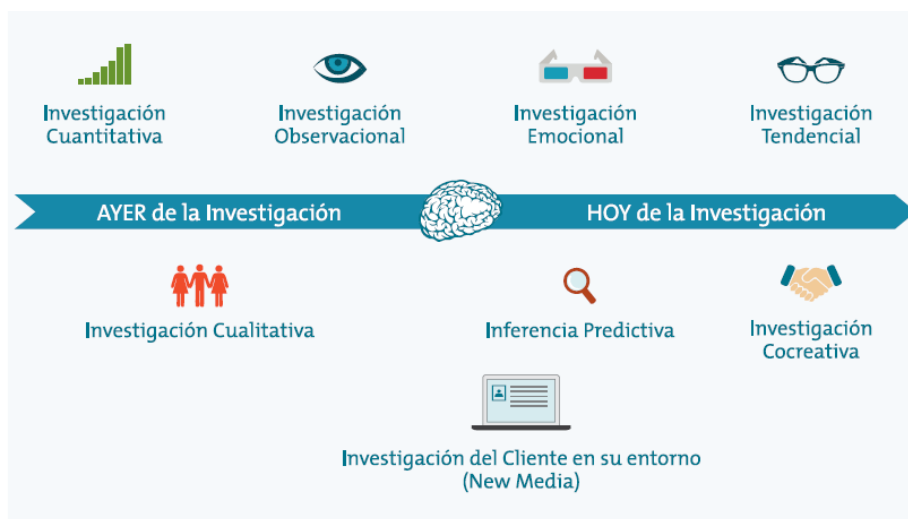


Figura 1: Evoluciones técnicas de análisis de mercado. Fuente (Alfaro et al, 2012)

La Gestión de Customer Experience, constituye entonces una propuesta estratégica para lograr una diferenciación frente a la competencia alrededor de la experiencia del cliente con los productos o servicios ofrecidos por una organización.

2.2. Analítica de Negocio

La analítica de negocio o BA (Business Analytics) por su nombre en inglés, es una disciplina que involucra todos los eslabones de una organización, desde la alta gerencia hasta el último contacto con el cliente, fomentando el intercambio de conocimiento basados en datos, de tal manera que se pueda medir, analizar y optimizar los mismos.

BA tiene como objetivo principal, proporcionar a las organizaciones información de valor que permita la toma de decisiones. Esta disciplina combina técnicas de minería de datos, modelos estadísticos, modelos matemáticos e inteligencia artificial para obtener una visión detallada y precisa del desempeño transversal de los procesos y métricas de una organización, permitiendo la identificación de patrones, tendencias y relaciones ocultas dentro de los datos.

El punto de partida para una organización es identificar la pregunta de negocio que quiere responder, de acuerdo con este enfoque, se puede categorizar cuatro tipos de BA como podemos observar en la Figura 2.

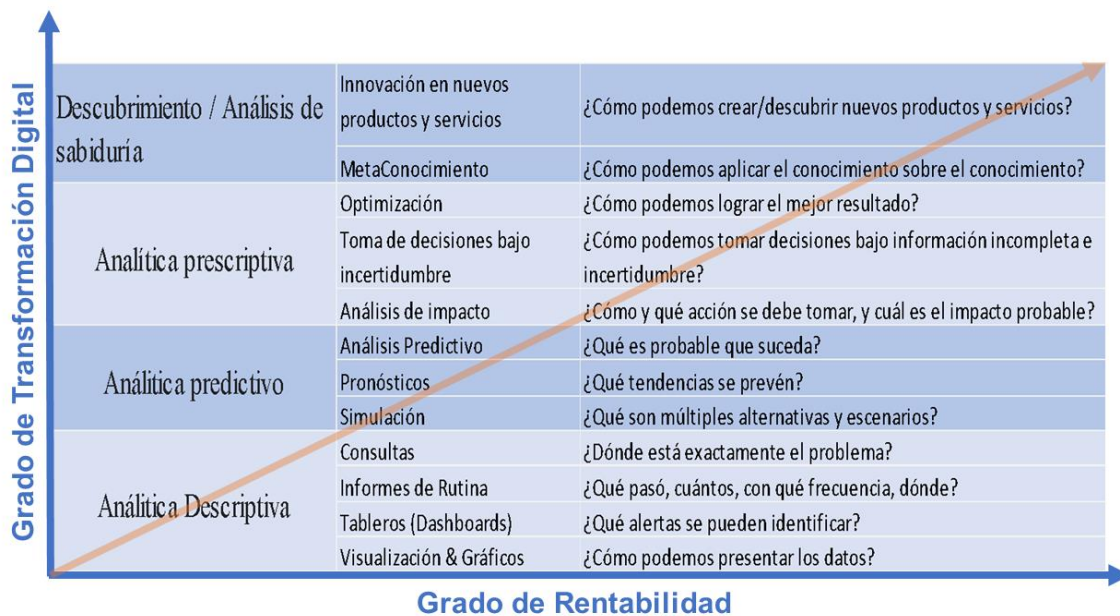


Figura 2: Tipos de Analítica. Adaptado (Raghupathi, W., & Raghupathi, V)

2.2.1. Tipos de Analítica

2.2.1.1. Analítica Descriptiva

Este tipo de analítica se enfoca en explorar lo que ha sucedido en el pasado a través del análisis de datos históricos. Aplicando técnicas de categorización, clasificación, caracterización y agregación de datos permite generar información útil a través de reportes y tableros (Dashboards) que ayuda a las organizaciones en la toma de decisiones basadas en hechos en el futuro.

2.2.1.2. Analítica Predictiva

Este tipo de analítica se enfoca en predecir lo que podría suceder en el futuro tomando como referencia lo conocido, es decir los datos históricos; aplicando a estos, modelos estadísticos, detectando tendencias, relaciones o patrones, extrapolando los mismos para generar un pronóstico informado en los datos. Permite de esta manera a las organizaciones identificar qué factores tienen mayor impacto en los resultados y tomar decisiones informadas sobre las consecuencias.

2.2.1.3. Analítica Prescriptiva

Este tipo de analítica se enfoca en determinar qué acciones se deben tomar para optimizar los resultados, utilizando modelos de optimización y simulación aplicados a los datos históricos, considerando los requisitos y limitaciones que se desean analizar. El principal objetivo consiste en brindar el mejor plan de acción que ayude a las organizaciones a entender cómo pueden lograr los mejores resultados posibles con los recursos disponibles.

2.2.1.4. Analítica de Descubrimiento o Analítica de Sabiduría

Este tipo de analítica se enfoca en descubrir patrones y relaciones en datos no estructurados. Esto implica el uso de técnicas avanzadas de procesamiento de lenguaje natural y minería de texto, con el objetivo de extraer información valiosa de grandes volúmenes de datos, como publicaciones en redes sociales, correos electrónicos y documentos.

2.3. Minería de Texto

La minería de texto es el proceso de analizar grandes volúmenes de fuentes de datos en formatos estructurados y no estructurados donde mediante el uso de diversas técnicas se busca identificar conceptos, temas clave, patrones y tendencias ocultas sin necesidad de conocer las palabras o los términos precisos que los autores han usado para expresar estos conceptos; aplicando técnicas de procesamiento lingüístico y procesamiento de lenguaje natural (PNL) (IBM, 2021).

Existen varias metodologías para abordar proyectos de minería de texto según el tipo de problema y el formato de los datos con los que se trabaja, entre estas las más relevantes encontramos:

2.3.1. KDD (Knowledge Discovery in Databases)

Se define KDD como “el proceso no trivial de identificar patrones válidos, novedosos y potencialmente útiles y, en última instancia, comprensibles a partir de los datos” (Fayyad et al, 1996).

Entendiendo estos conceptos como:

- **Válido:** Los patrones identificados deben ser precisos al procesar nuevos conjuntos de datos (aceptando cierto grado de incertidumbre).
- **Novedoso:** Los descubrimientos deben aportar algo desconocido para el usuario.
- **Potencialmente útil:** Los descubrimientos deben ser disparadores de planes de acción que reporten algún tipo de beneficio para el usuario.
- **Comprensible:** Los descubrimientos deben ser claros, que faciliten su fácil interpretación, revisión, validación y uso en la toma de decisiones.

KDD es un proceso interactivo e iterativo, en la cual el usuario debe estar involucrado en la toma de decisiones según la comprensión del dominio de la aplicación, es decir, la experticia y conocimiento adquirido sobre el problema a resolver. KDD usa entonces métodos de minería de datos para extraer información importante de acuerdo con una serie de especificaciones y restricciones preestablecidos sobre bases de datos estructuradas.

Esta metodología propone una serie de 5 fases, siendo la minería de datos una de ellas. cómo se detalla en la Figura 3 a continuación.



Figura 3: Fases Metodología KDD. Adaptada (Fayyad et al, 1996)

Selección: Esta etapa consiste en crear un conjunto de datos objetivo a partir de una base de datos más grande a las que se va a realizar la exploración.

Preprocesamiento: Esta etapa consiste en la limpieza y el preprocesamiento de los datos objetivo con el fin de contar con datos consistentes y en los formatos deseados para su revisión.

Transformación: Esta etapa consiste en encontrar las características más significativas para representar los datos, en este paso se aplican técnicas de transformación de los datos preprocesados para obtener una forma más adecuada para su análisis; reduciendo si es el caso, el número de variables a ser consideradas o agregar nuevas variables que aporten nuevas representaciones de los datos.

Minería de Datos: Esta etapa consiste en la búsqueda de patrones de interés. Es en esta etapa donde dependiendo del objetivo del caso de estudio se aplican modelos de regresión, clasificación o agrupamiento.

Interpretación / Evaluación: Esta etapa consiste en la interpretación y evaluación de los patrones según su relevancia y utilidad, así como la identificación de indicadores que sean disparadores de planes de acción.

2.3.2. KDT (Knowledge Discovery in Text)

La creciente necesidad de realizar trabajos de exploración ahora sobre bases de datos no estructuradas como correos electrónicos o publicaciones en redes sociales abrió el camino para metodologías como KDT, la cual es una adaptación de la ya existente KDD aplicada ahora al análisis de información textual (Feldman & Dagan,1995).

Como se menciona previamente, KDD requiere que los datos provengan de una base de datos estructurada, esta estructura debe reflejar la forma en la cual el usuario conceptualiza la naturaleza de los datos (Ver Figura 4). KDT tiene como foco, la exploración del texto proveniente de bases de datos no estructuradas, por lo cual hace uso de técnicas de etiquetado, generando palabras clave tomadas de un diccionario de datos controlado, el cual está organizado con una estructura jerárquica preestablecida.

De esta manera logra construir una estructura al texto que se está explorando. Conociendo esta jerarquía, se identifican las entidades de nivel superior, y se construye una indexación de las palabras contenidas en el texto, lo cual facilita encontrar la frecuencia de ocurrencia de palabras clave, pieza fundamental para los procesos de minería de datos.

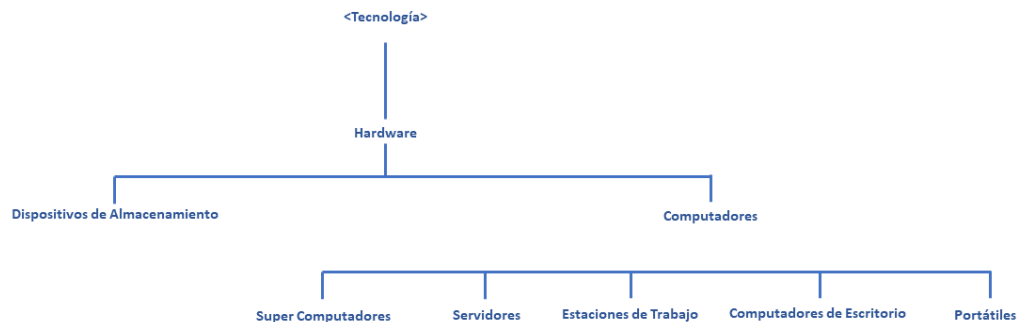


Figura 4: Concepto de Jerarquía para conceptos tecnológicos. Adaptada (Feldman & Dagan,1995)

KDT requiere entonces el dominio de técnicas adicionales para su procesamiento, tales como métodos de recuperación de información (Web Scraping) y procesamiento de lenguaje natural (PLN) para llevar a cabo esta tarea.

Esta metodología, propone una serie de 6 fases (Ver Figura 5), siendo nuevamente la minería de datos una de ellas.

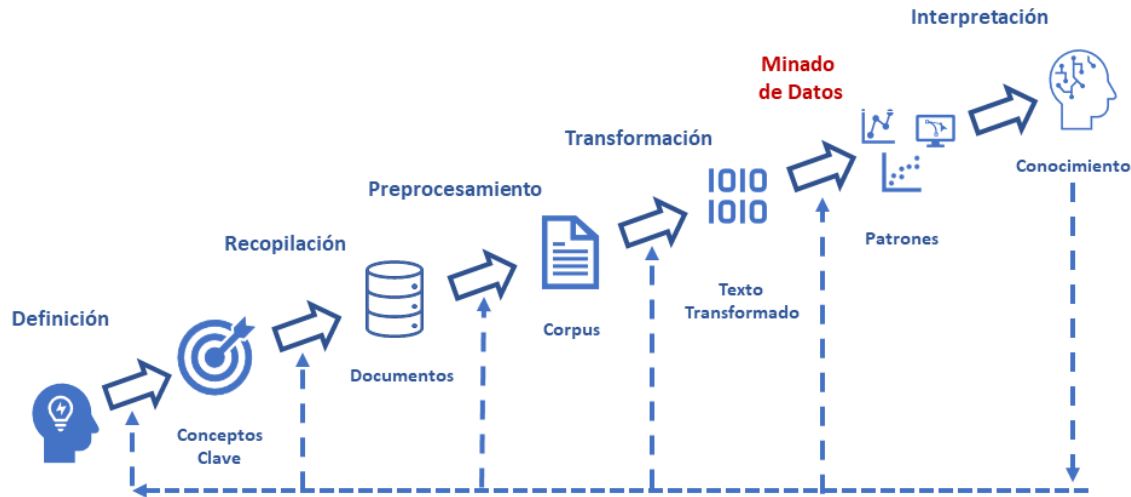


Figura 5: Metodología KDT. Fuente (Mansilla, 2018)

Definición: Esta etapa consiste en la definición de conceptos clave, los cuales deben estar alineados según el objetivo de la investigación, es decir deben ser consistentes en la búsqueda de una respuesta al planteamiento del problema propuesto en el caso de estudio.

Recopilación: Esta etapa se centra en la recolección de los datos que serán objeto de estudio, en particular los procesos automáticos soportados en técnicas de recuperación de texto como web scraping o mediante API's (Application Programming Interfaces) suministradas por los proveedores de los datos origen.

Pre-Procesamiento: Esta etapa se centra en limpieza de los datos, eliminando caracteres especiales, números, y palabras que no aportan información relevante para el análisis con el objetivo de garantizar la homogeneidad del texto.

Transformación: En esta etapa se concentra los procesos que marcan la diferencia con relación a KDD, es aquí donde se aplican técnicas de análisis de texto y análisis semántico como son:

- a) **Lematizado:** Es el proceso de separar los prefijos y sufijos de las palabras para derivar la forma y el significado de la raíz de la palabra, la cual se conoce como **lema**. Por

ejemplo, las palabras “llamando” y “contactando” tendrán como *lema* “llamar” y “contactar” respectivamente. Esta técnica mejora la recuperación de información al acotar el número de variables, que a su vez repercute en el número de niveles indexados, reduciendo así la complejidad en el análisis de texto.

- b) **Tokenización:** Este proceso consiste en desglosar el texto en oraciones, y estas a su vez en palabras, llevándolas a su unidad más significativa la cual se denominan tokens.

por ejemplo, la oración

“El servicio de internet de mi proveedor es muy rápido”.

Tras su proceso de tokenización, se obtiene la siguiente estructura

[“El”, “servicio”, “de”, “internet”, “de”, “mi”, “proveedor”, “es”, “muy”, “rápido”]

El objetivo es crear una representación estructurada del texto que haga más fácil el procesamiento de las técnicas de análisis de texto y minería de datos.

- c) **Etiquetado:** Este proceso consiste en aplicar técnicas de PLN, donde se asigna una etiqueta a cada token generado, de tal manera que sea posible identificar su función gramatical y su contexto.

De esta manera, continuando con el ejemplo de la oración

“El servicio de internet de mi proveedor es muy rápido”.

El resultado tras el proceso de etiquetado nos daría como resultado la siguiente clasificación

El: **Artículo determinante**

servicio: **Sustantivo**

de: **Preposición**

internet: **Sustantivo**

de: **Preposición**

mi: **Determinante posesivo**

proveedor: **Sustantivo**

es: **Verbo en tercera persona**

muy: **Adverbio**

rápido: **Adjetivo**

Este proceso permite a los algoritmos de PNL comprender mejor la estructura y significado de las oraciones, lo que facilita la precisión de tareas como el análisis de sentimiento.

Minado de Texto: Esta etapa consiste en la búsqueda de patrones de interés en una determinada forma de representación, dependiendo del objetivo de la exploración, es en esta etapa donde se pueden aplicar técnicas de análisis de sentimiento para clasificar los textos en positivos, negativos o neutros.

Interpretación: Esta etapa consiste en la interpretación y evaluación de los patrones según su relevancia y utilidad, así como la identificación de indicadores que sean disparadores de planes de acción.

2.3.3. CRISP-DM

CRISP-DM (Cross-Industry Standard Process for Data Mining), es una metodología que proporciona un marco estructurado y detallado para la planificación y ejecución de proyectos de minería de datos. Fue concebido en 1996 por un consorcio de empresas líderes en el campo de minería de datos como SPSS (entonces ISL), DaimlerChrysler (entonces Daimler-Benz), NCR y OHRA. Desde entonces se ha convertido en la metodología con mayor adopción por la industria y academia.

Gracias a su flexibilidad y su nivel de personalización, es fácilmente adaptable para abordar cualquier tipo de proyecto de minería de datos, desde bases de datos estructuradas como no estructuradas. En este orden las fases que componen la metodología pueden ser menos relevantes según el tipo de problema del caso de estudio.

Esta metodología, propone una serie de 6 fases (Ver Figura 6).

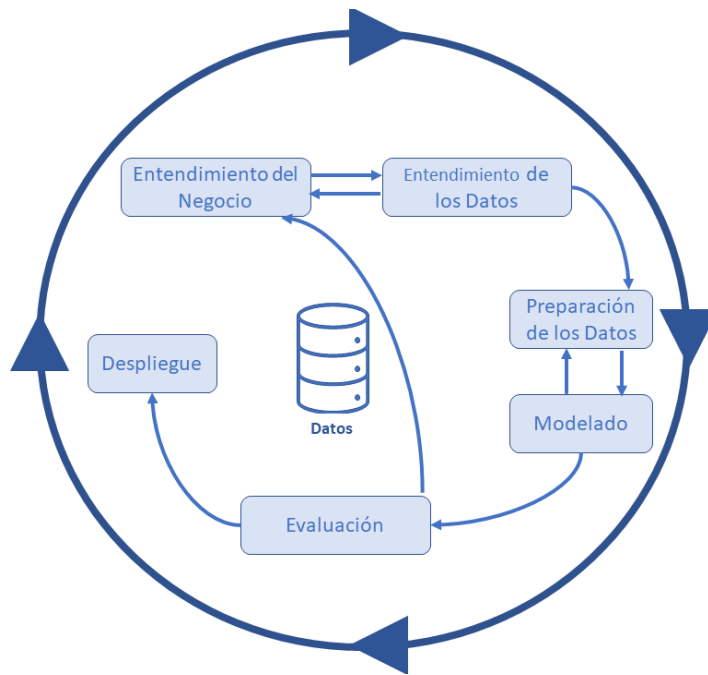


Figura 6: Metodología CRISP-DM. Fuente (IBM, 2021)

Entendimiento del negocio: Esta etapa inicial, se enfoca en comprender los requerimientos y objetivos del caso de estudio desde una perspectiva comercial; tras la comprensión de estos, se debe definir un plan de las técnicas de minería de datos a implementar con el fin del cumplimiento de los objetivos trazados.

Entendimiento de los datos: En esta etapa se ejecutan las tareas de recolección y exploración de los datos, lo cual permite la familiarización y detección de la calidad de los datos, así como la identificación preliminar de patrones o subconjuntos de datos interesantes que promuevan la formulación de las primeras hipótesis.

Preparación de los datos: Esta etapa genera como resultado la estructura de datos final, con la cual se aplicarán las técnicas de minería de datos que darán respuesta al objetivo del caso de estudio. En esta fase se aplicarán técnicas de limpieza y transformación de los datos, técnicas de agregación de variables adicionales y cambios de formato.

La estructura de los datos final, y el tratamiento de los datos aplicados para su obtención son cruciales para la técnica de modelado seleccionada, pues los datos requieren ser procesados según el modelo seleccionado. Es en esta fase donde CRISP-DM consolida las técnicas de KDD

y KDT, gracias a su flexibilidad para abordar cualquier tipo de caso de estudio de minería de datos.

Modelado: En esta etapa se construyen modelos de minería de datos utilizando técnicas como la regresión, clasificación, segmentación o asociación; sus parámetros son calibrados para obtener los valores óptimos.

Evaluación: En esta etapa los modelos seleccionados son evaluados según las métricas seleccionadas para determinar su precisión y calidad, alineando los resultados en búsqueda de responder adecuadamente los requerimientos y objetivos del caso de estudio.

Despliegue: Esta etapa debe garantizar que los resultados obtenidos generen conocimiento, y sean presentados de manera organizada y de fácil comprensión, de tal manera que sean disparadores de planes de acción.

2.4. Análisis de Sentimiento

Las opiniones, referencias y experiencia de terceros son determinantes para la mayoría de los seres humanos, puesto que son pieza clave como aspectos influenciadores para la toma de decisiones, comportamiento, o cómo percibimos el mundo.

El análisis de sentimiento en las redes sociales se enmarca en el campo de la investigación de mercados y la inteligencia de negocios en plataformas digitales, ya explorados en el apartado de CX.

En la actualidad, cada vez que queremos realizar una compra de un producto, o adquirir un servicio; nuestro primer instinto es conocer la experiencia de otros usuarios que previamente han adquirido alguno de estos; esto se ha facilitado gracias a la gran cantidad de información disponible en los comentarios o foros de discusión de plataformas digitales

Estos comentarios directos realizados por usuarios en relación con la experiencia vivida con productos, servicios u organizaciones se conocen como word-of-mouth (WOM) y es de gran importancia para medir y entender cómo se percibe una marca en línea. (Cheng, L. C., & Tsai, S. L., 2019)

Para llevar a cabo este tipo de análisis, se utilizan métodos de PLN y minería de texto para analizar los comentarios y extraer información relevante. Esto permite identificar patrones y

tendencias sobre el texto en cada publicación. Clasificando gracias al análisis de sentimientos, una opinión positiva, negativa o neutral sobre la marca. Esto puede hacerse mediante el uso de diccionarios de sentimientos pre-construidos, que asignan un puntaje de positivo o negativo a cada palabra en el diccionario, o mediante el entrenamiento de modelos de lenguaje natural utilizando un conjunto de datos etiquetados previamente con dicha clasificación, lo cual permite obtener una visión general del nivel de satisfacción o insatisfacción de los usuarios con la marca en cuestión, y puede ser muy útil para las empresas o personas públicas que desean mejorar su imagen en línea (Bing, 2022).

2.5. Deep Learning

Deep Learning, también conocido como aprendizaje profundo, es un subconjunto de la inteligencia artificial (IA), donde encontramos otras tecnologías como aprendizaje automático (ML) y redes neuronales. Una forma fácil de entender este concepto es pensar en ellos como un set de matrioshka como se ilustra en la Figura 7 a continuación.

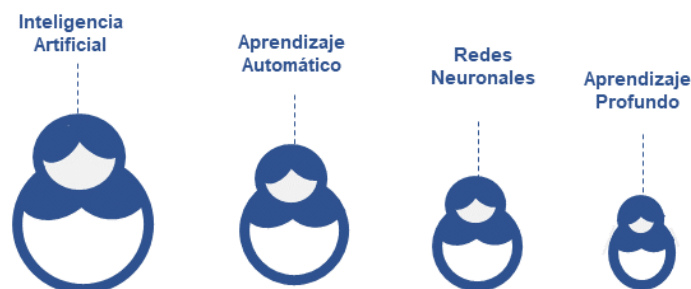


Figura 7: Relación IA, ML, Redes Neuronales y Deep Learning. Fuente (Kavlakoglu, 2020)

Observamos entonces que cada uno de estos es un complemento del componente anterior, es decir, aprendizaje automático es un subconjunto de inteligencia artificial. El aprendizaje profundo a su vez lo es del aprendizaje automático.

Siendo el número de capas compuestas por estas neuronas los determinantes en la definición entre un modelo de aprendizaje profundo o un modelo de aprendizaje automático (Kavlakoglu, 2020).

Una neurona artificial conocida también como perceptrón, es la unidad computacional base para la arquitectura de la inteligencia artificial. Este tiene como objetivo comprender la representación

de los datos que recibe como entrada, transformando los mismos en señales de salida. La Figura 8 ilustra el comportamiento de un perceptrón que recibe tres entradas:

- Conjunto de datos X_1 a X_n
- Pesos W_1 a W_n
- Constante, que corresponde a un término de sesgo b

En la primera fase dentro del perceptrón se calcula la suma ponderada del conjunto de los datos, sus pesos y el sesgo inducido.

$$Z = b + \sum_{i=1}^n W_i X_i$$

La segunda fase consiste en aplicar una función de activación al valor obtenido en Z , por lo general para modelos de predicción de probabilidades es comúnmente usada la función sigmoide, pues su salida asigna valores entre $[0, 1]$, lo cual facilita modelar límites de decisión no lineales complejos.

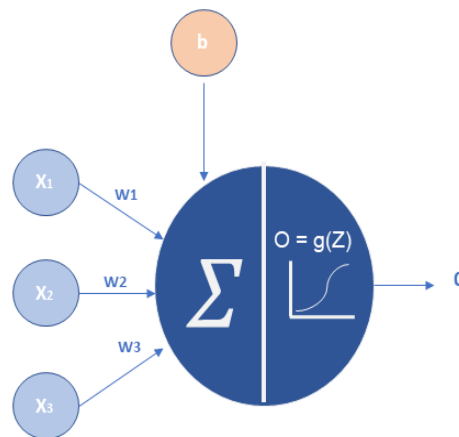


Figura 8: Relación IA, ML, Redes Neuronales y Deep Learning. Fuente (Kavlakoglu, 2020)

El perceptrón es entonces la unidad base de una red neuronal superficial. La cual está compuesta por tres capas: capa de entrada, capa oculta y capa de salida, cada una de estas capas cuenta con múltiples perceptrones como se observa en la Figura 9 a continuación.

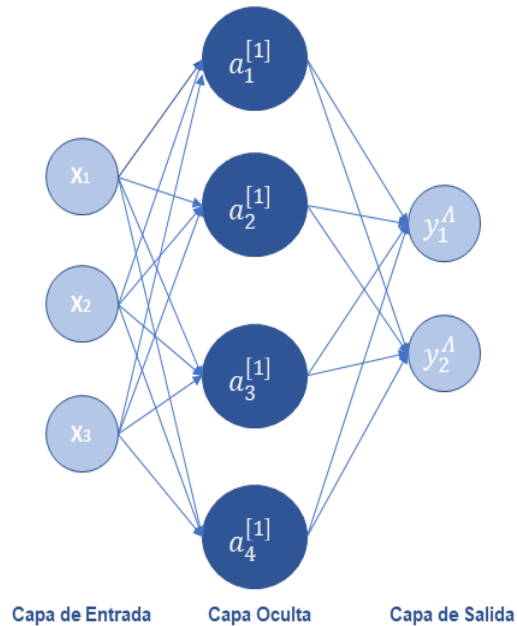


Figura 9: Modelo Red Neuronal Superficial. Fuente (Madhavan et al, 2020)

Cada perceptrón entonces realiza los cálculos mencionados previamente en cada una de las capas, incluyendo la capa de salida.

Como mencionábamos anteriormente, es justamente el número de capas ocultas el factor diferencial entre una red neuronal superficial y una red neuronal de aprendizaje profundo (Ver Figura 10). El concepto “profundo” en deep learning se refiere al número de capas ocultas en su arquitectura, agregar estas capas aumenta la complejidad de los problemas que pueden dar solución las redes neuronales. Cada perceptrón de estas capas ocultas, tienen múltiples conexiones con la capa que lo precede.

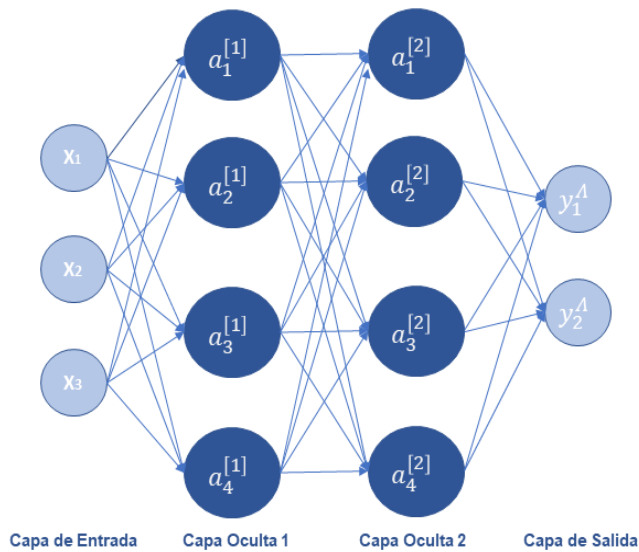


Figura 10: Modelo Red Neuronal Profunda. Fuente (Madhavan et al, 2020)

Gracias a las técnicas de propagación hacia adelante (forward propagation) y propagación inversa (back propagation) permiten que los modelos de aprendizaje profundo a medida que realiza estos cálculos y ajustes en cada iteración, con el tiempo sean más precisos.

Forward Propagation: En este proceso los datos se transmiten desde la capa de entrada a través de las capas ocultas hasta la capa de salida, generando una predicción o salida final ($O = g(Z)$).

Una vez se obtiene esta salida final, se calcula el error al comparar las salidas generadas, es decir las predicciones contra las salidas esperadas utilizando una función de pérdida, este error mide la discrepancia entre la salida de la red y los valores deseados.

Back Propagation: En este proceso, los cálculos y ajustes del error de predicción se ejecutan desde la capa de salida hacia la capa de entrada. Se calcula el error de la salida generada contra la salida esperada usando una función de pérdida. Adicionalmente se calcula la derivada de la función del error con respecto a los pesos, para obtener el gradiente de esta capa. Es este valor de gradiente lo que permite realizar los ajustes al modelo para optimizar la función objetivo de este.

Lo anterior describe entonces la arquitectura y procesamiento más simple de una red de aprendizaje profundo, lo cual es nuestro punto de partida para explorar uno de los tipos de redes neuronales de aprendizaje profundo para el estudio de PLN (Madhavan & Madan, 2020)

2.5.1. Redes Neuronales Recurrentes (RNN)

Las redes neuronales recurrentes, se utilizan comúnmente en el análisis PLN, debido a su capacidad de procesamiento de datos con una naturaleza secuencial. En el análisis de sentimiento, las palabras y frases de un texto tienen una estructura secuencial y temporal, por lo tanto, su significado depende del contexto en que se utilizan. Por ejemplo, analicemos la siguiente frase:

“El servicio de internet es bueno”

El contexto de esta frase se determina positivo, si consideramos el adjetivo “bueno”. Caso contrario, si tenemos la siguiente frase:

“El servicio de internet no es bueno”

En este caso, el contexto de la frase cambia al anteponer el adverbio “no” al adjetivo “bueno”.

Las RNN pueden modelar esta dependencia temporal al procesar secuencialmente las palabras en una oración y mantenerlas en memoria para evaluar así el contexto de la frase.

Cada palabra corresponde a la señal de entrada de una capa de la red, y la señal de salida de la neurona, se utiliza a su vez como señal de entrada para la siguiente neurona. Esto permite que la RNN considere el contexto previo al analizar cada palabra y hacer predicciones basadas en el contexto global de la frase.

Las redes RNN se basan en el principio de compartir parámetros a lo largo de las diferentes capas del modelo. Mientras las redes neuronales profundas asumen que las señales de entrada y de salida son independientes entre sí, la salida de las redes neuronales recurrentes depende de las señales de salida de las neuronas de las capas anteriores. Las RNN modelan este comportamiento al adicionar una capa de contexto (Ver Figura 11), a la arquitectura de las redes neuronales profundas, vista previamente.

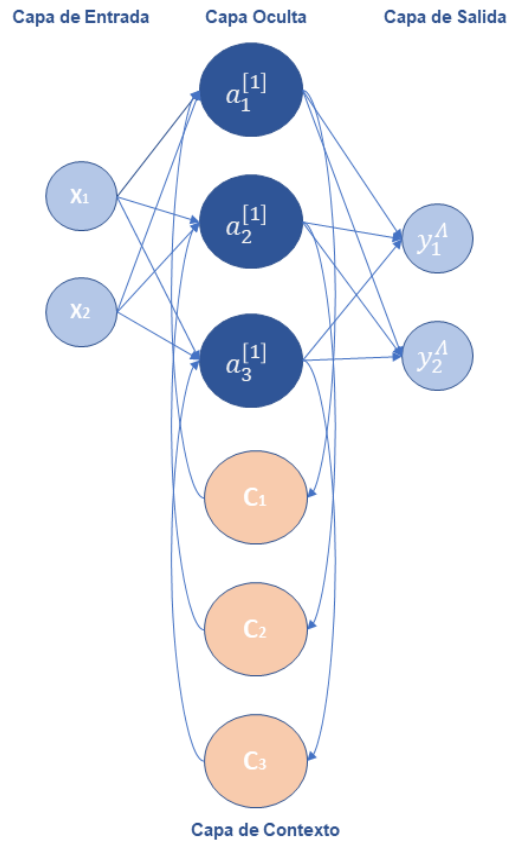


Figura 11: Modelo Red Neuronal Recurrente. Fuente (Jones, 2017)

En esta gráfica observamos como la capa oculta alimenta cada neurona de la capa de contexto, estos retienen en memoria las entradas pasadas, retornando este valor nuevamente como parámetro de entrada en la siguiente iteración.

Añadiendo de esta manera a nuestra función de suma ponderada este nuevo término.

$$Z = b + \sum_{i=1}^n W_i X_i + H_{i-1}$$

Donde,

- **Conjunto de datos** X_1 a X_n
- **Pesos** W_1 a W_n

- **Salidas** H_{i-1} de la capa anterior
- **Constante**, que corresponde a un término de sesgo b

Las RNN usan un algoritmo para determinar los gradientes en el proceso de back propagation conocido como BPTT (propagación inversa a través del tiempo), los principios de este algoritmo comparten el proceso de cálculo de gradiente tradicional, donde el modelo se entrena a sí mismo calculando los errores y ajustando los pesos desde la capa de salida hasta la capa de entrada, con la gran diferencia, que el BPTT considera la suma de los errores en cada paso temporal gracias a que los parámetros son compartidos en todo el modelo.

Esta propiedad enfrenta a las RNN a dos grandes problemas, conocidos como “explosión o desvanecimiento de gradiente”, los cuales están relacionados por el tamaño del gradiente en el tiempo, en otras palabras, la inclinación de la función de pérdida a lo largo de la curva de error. El desvanecimiento de gradiente se presenta, cuando el gradiente tiende a cero (0) a través de las iteraciones desde la capa de salida hasta la capa de entrada, esto quiere decir que el modelo ya no está aprendiendo. Por otra parte, la explosión de gradiente ocurre cuando su índice es demasiado grande, lo cual indica una inestabilidad en el modelo.

Como solución a estos problemas, se introducen algunas variantes a la arquitectura RNN como solución, siendo las más usadas LSTM y GRU.

2.5.1.1. LSTM

Las redes LSTM (Long Short-Term Memory) o Redes de Memoria a Largo Plazo, son una variación de las redes RNN diseñadas para procesar conjuntos de datos secuenciales y capturar sus dependencias a largo plazo.

Si consideramos la frase

“El servicio de internet es excelente, pero la facturación es pésima”

Una RNN tradicional, tomaría la primera palabra “El” como entrada y generaría su salida correspondiente. Luego, tomaría la siguiente palabra “servicio” y generaría una nueva salida basada en la información previa y la entrada actual; así sucesivamente hasta recorrer toda la frase.

El problema de desvanecimiento de gradiente se presenta a medida que ejecuta el proceso de propagación inversa a través de las capas para ajustar los pesos, este gradiente puede volverse cada vez más pequeño en cada iteración, debido a la multiplicación repetida de los gradientes en cada paso hacia la capa de entrada. Afectando la capacidad de la red RNN para capturar las relaciones a largo plazo. Es decir, RNN puede tener dificultad para establecer una conexión entre “servicio de internet” y “excelente” que se encuentran ubicados al comienzo de la frase, y “facturación” y “pésima” ubicadas en el otro extremo. Esto simplifica el concepto de “desvanecimiento de gradiente”, como vemos la información relevante puede volverse menos significativa a medida que los gradientes tienden a cero durante el procesamiento de largas oraciones de texto.

LSTM tiene una estructura más compleja que les permite recordar información relevante a lo largo del tiempo y evitar así el desvanecimiento de gradiente. La clave se encuentra en su unidad de memoria denominada “celda de memoria”, la cual se compone de dos componentes.

Estado de la celda (Cell State): Es la memoria principal de la LSTM y almacena información relevante de los datos de entrada a lo largo del tiempo.

Puertas (Gates): Las puertas en LSTM permiten controlar el flujo de información almacenada en la celda de memoria. Existen tres tipos principales de puertas para optimizar este flujo.

- **Puerta de Olvido (Forget Gate):** Encargada de decidir qué información almacenada previamente en la unidad de memoria debe ser olvidada, una vez se determina que la misma no es relevante para el contexto actual.
- **Puerta de Entrada (Input Gate):** Determina qué nueva información debe almacenarse en la unidad de memoria.
- **Puerta de Salida (Output Gate):** Controla qué información se transmite desde la unidad de memoria como dato de entrada de las capas ocultas del modelo.

La interacción entre estas puertas y la celda de memoria permite entonces que las redes LSTM puedan almacenar información relevante sobre las palabras anteriores y su significado en el contexto. A medida que el modelo procesa cada palabra, la unidad de memoria se actualiza y se ajusta para mantener información relevante (Jones & Madhavan, 2021).

Podemos reinterpretar la arquitectura de una red RNN como se ilustra en la Figura 12, de la siguiente manera.

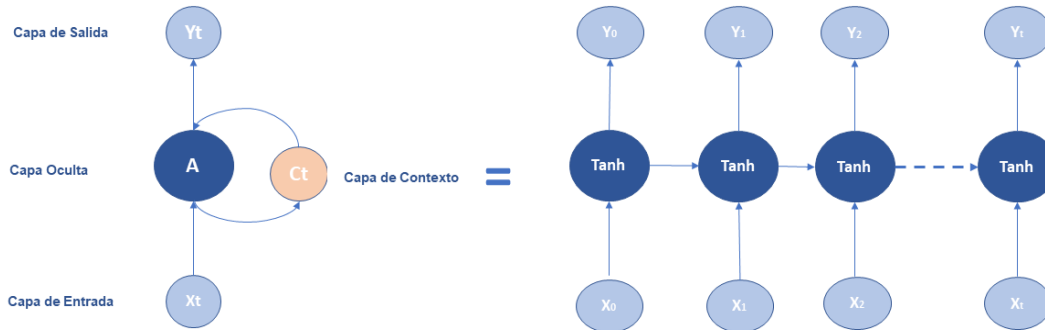


Figura 12: Modelo Red Neuronal Recurrente Simplificado. Fuente (Jones & Madhavan, 2021)

Adicionando los conceptos propuestos por LSTM, la arquitectura de la Red RNN en su modelo conceptual es el siguiente (Ver Figura 13).

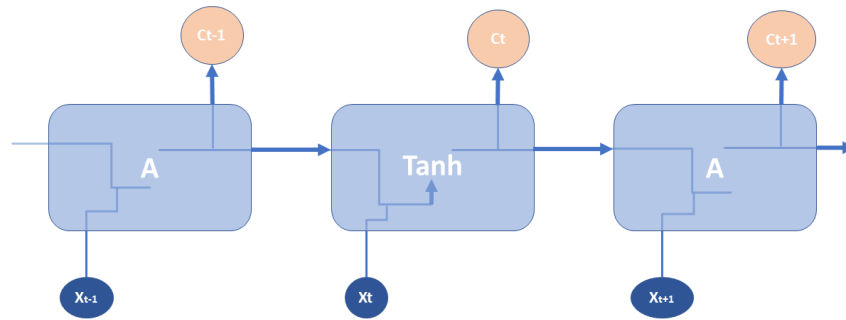


Figura 13: Modelo Base de una Red LSTM. Fuente (Jones & Madhavan, 2021)

Observamos entonces la forma básica de una neurona LSTM, internamente los componentes de puerta de olvido, puerta de entrada, puerta de salida interactúan entre ellas en función de las activaciones de la función sigmoide (Ver Figura 14).

En esta interacción se adiciona un nuevo componente conocido como **función Tanh**, la cual es la pieza clave para dar solución al problema de desvanecimiento de gradiente. Esta función tiene la tarea de comprimir los valores entre 1 y -1, de esta manera se garantiza que los valores de gradiente no sufran grandes o mínimas variaciones, lo cual como se expuso anteriormente dificultan el entrenamiento del modelo (Olah, 2015).

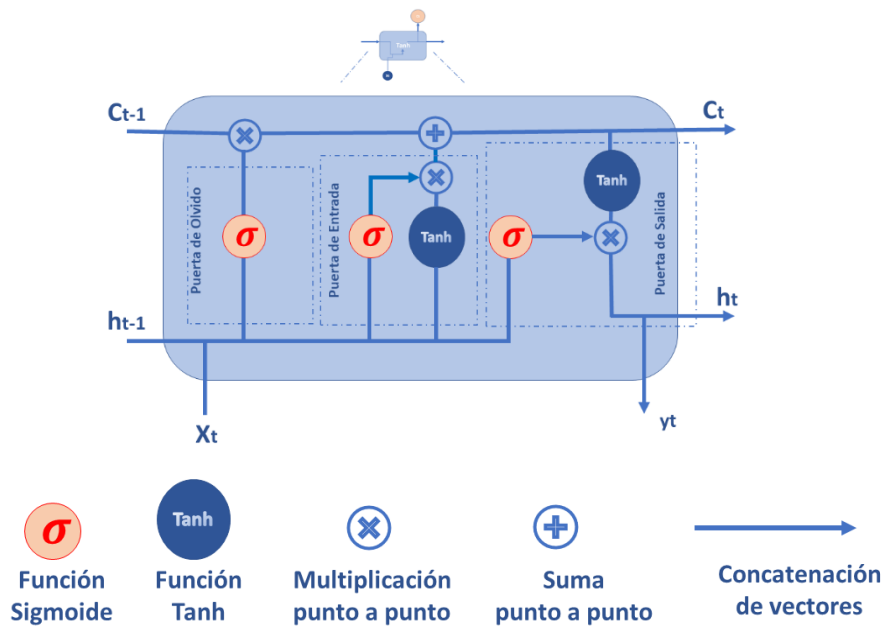


Figura 14: Modelo Celda LSTM. Fuente (Olah, 2015)

Donde,

- X son los datos de entrada a través del tiempo.
- h es el estado oculto
- y es la señal de salida
- C corresponde al estado de memoria interna.
- Constante, que corresponde a un término de sesgo b

Gracias a la interacción de estos componentes, LSTM puede evaluar la relevancia de los datos de entrada, y activar los componentes de entrada para determinar qué debe “olvidar” y que debe “conservar” en busca de tener mayor eficiencia durante el proceso de entrenamiento del modelo.

2.5.1.2. GRU

Las Redes GRU (Gated Recurrent Units) o Redes de Unidades Recurrentes Cerradas, son una variación más simple de las redes LSTM, estas conservan el concepto de las redes RNN, pero optimizan su estructura interna, en búsqueda de eficiencia y rendimiento computacional al

combinar algunos de sus componentes. Las redes GRU fusionan las puertas de entrada y de olvido en una sola que se denomina “Puerta de Actualización”. También fusiona el estado de la celda y el estado oculto como se detalla en la Figura 15.

Los componentes de la Red GRU son:

Puerta de Actualización (Update Gate): Encargada de mantener la información relevante y eliminar los datos que no son importantes.

Puerta de Reinicio (Reset Gate): Controla la cantidad de información que se debe “olvidar” durante el entrenamiento del modelo.

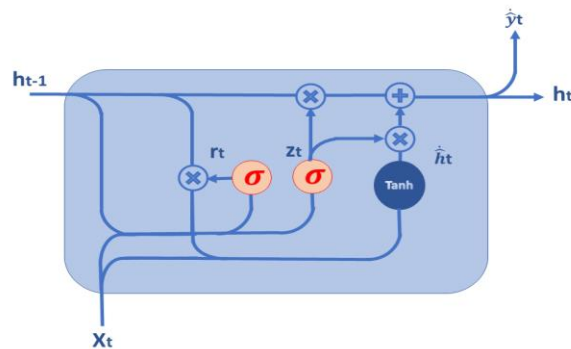


Figura 15: Modelo Celda GRU. Fuente (Olah, 2015)

Donde,

- **X** Es el vector de datos de entrada.
- **h** Es el vector de salida.
- **h[^]** Es el vector de activación candidato.
- **z_t** Es el vector de puerta de actualización.
- **r_t** Es el vector de puerta de reinicio.

Se observa entonces la simplicidad de la arquitectura interna de una celda GRU, gracias a contar con menos componentes y menos operaciones, el proceso de entrenamiento de modelos es más rápido en comparación a una red LSTM (Olah, 2015).

3. ESTADO DEL ARTE

El análisis de sentimiento de texto en entornos empresariales ha ganado relevancia en los últimos años tras la explosión de la era digital. Es por tal motivo que existen una variedad de investigaciones que abordan este tema desde diversos enfoques metodológicos, desde modelos estadísticos, machine learning hasta deep learning.

Los estudios y publicaciones recopilados para nuestra investigación comparten un punto común, y es la importancia de los datos para el análisis y definición de estrategias para una organización.

Sentiment Analysis and Opinion Mining (Liu, Bing, 2022): El autor de este documento realiza un extenso y detallado análisis sobre la estructura y complicaciones en la clasificación de análisis de sentimiento, así como los niveles existentes en este campo, tales como, nivel de documento y nivel de oración. En donde en el primer nivel, se realiza un análisis en conjunto de todo el documento y se realiza un acumulado de valoración de cada palabra, para finalmente hacer una estimación, y así clasificar si el documento en conjunto tiene una percepción negativa o positiva. Por otra parte, el nivel de oración realiza la valoración para realizar el mismo ejercicio. Este estudio realiza un análisis detallado en la detección de conceptos como sarcasmo, lo cual permite hacer una limpieza de los datos, lo cual resulta muy acertado para una clasificación de sentimiento con mayor asertividad.

BERT for Sentiment Analysis: Pre-trained and Fine-Tuned Alternatives (Souza, F. D., 2022, March): En esta publicación, se utilizó uno de los métodos más populares y recientes en la actualidad del Deep Learning, que es la transferencia de conocimiento con modelos preentrenados. En particular, para este estudio se consideró dos modelos: m-BERT and BERT_{imbau}, dos estrategias de diseño: preentrenado y ajustado, y como modelo de línea base incluyeron el TF-IDF, para el análisis de sentimientos en las reseñas de usuarios de portugués brasileño (clasificando como positivo o negativo), en donde hallaron que los modelos BERT_{imbau} preentrenados funcionaron mucho mejor que las relacionadas con el m-BERT multilingüe y que el enfoque de diseño propuesto, mejoró considerablemente los esquemas convencionales.

Contemporary Business Analytics: An Overview (Raghupathi, W. Raghupathi, V., 2021): Esta publicación confirma la importancia de la analítica de negocio dentro de una organización, permitiendo a esta optimizar diversas actividades comerciales como son el marketing, ventas,

operaciones y finanzas. Resaltando que la calidad de los indicadores y patrones identificados por esta analítica dependen directamente de la calidad de los datos en los que se basa, además de las habilidades y conocimientos de quienes la procesan. Discute además los diferentes tipos de analítica de datos, así como técnicas de análisis estadístico y el aprendizaje automático.

Sentiment Analysis of Movie Reviews based on Machine Learning (Yang, Z., 2020): Para realizar el estudio, obtuvo los datos de las reseñas de películas de diferentes personas del sitio web Kaggle y luego transformó los datos de texto en datos vectoriales utilizando el algoritmo TF-IDF, para posteriormente analizar el sentimiento de las reseñas al clasificarlos como negativo o positivo, utilizando modelos de regresión de Machine Learning como Lasso, Ridge y CatBoost, siendo este último el mejor con una precisión superior del 85%.

Deep Learning for Automated Sentiment Analysis of Social Media (Cheng & Tsai, 2019): En este artículo usaron las redes neuronales recurrentes LSTM, BiLS y GRU del Deep Learning para clasificar el sentimiento de los comentarios de un tráiler en YouTube y clasificar el sentimiento como positivo, negativo o neutral en donde destacaron que la complejidad y la naturaleza dinámica de este tipo de datos no estructurados dificultan la identificación precisa del sentimiento, pero aun así BiLS logró clasificar correctamente el 87% de estos.

A Deep Learning Architecture for Sentiment Analysis (Çano, E., & Morisio, M., 2018): Los autores realizan estudios comparando las métricas entre redes CNN y RNN con una capa LSTM, evaluando estos contra la arquitectura NgramCNN propuesta por los autores. Esta arquitectura de red neuronal para el análisis de sentimientos de documentos de texto extensos se evaluó utilizando diferentes tipos de conjuntos de datos de texto emocional y lograron una precisión del 91,2 % en las reseñas de películas populares de IMDB, superando las métricas generadas por los modelos tradicionales.

Sentiment Analysis of Twitter Data Using Machine Learning Techniques and Scikit-learn (Shihab & Yang, 2018): Este estudio probó diferentes modelos de Machine Learning, para analizar y clasificar los sentimientos de un conjunto de textos en Twitter como positivos y negativos, en donde lograron demostrar que BernoulliNB, LogisticRegression y el clasificador SGD alcanzaron una precisión de hasta el 75 %.

VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text (Hutto & Gilbert, 2014): Este estudio analiza en detalle el desempeño del modelo VADER

en tanto a la clasificación y evaluación del análisis de texto, descomponiendo este último en unidades de Lexicons (diccionarios que incluyen para cada palabra una valoración en alguna dimensión afectiva) en nuestro caso, positiva, negativa o neutra; y cotejando estos valores con resultados de otras técnicas con LIWC, ANEW y SentiWordNet, demostrando así la efectividad de VADER, una efectividad de clasificación del 96%.

Sentiment analysis of twitter data using machine learning approaches and semantic analysis (Gautam & Yadav,2014): Los autores de este estudio proponen un pre-procesamiento inicial de los datos, soportados en la estructura semántica del lenguaje. Donde realizan una descomposición de los tweets, extrayendo inicialmente los adjetivos, los cuales tienen un valor que determina si el comentario es positivo, negativo o neutro, y denominando esto último como un vector de característica (feature vector), con estos valores como dataset de entrada, se aplican modelos de clasificación de Machine Learning como Naive Bayes, Maximum Entropy, Support Vector Machine y WordNet y se comparan los índices de precisión de estos últimos en orden de desempeño de clasificación y precisión.

El material recopilado nos brinda un enfoque al planteamiento inicialmente propuesto y su importancia para gestionar la experiencia de usuario en una organización (Ver tabla 1).

#	Artículo	Año	Enfoque	Datos	Citación APA
1	Sentiment Analysis and Opinion Mining	2022	Minería de Datos y Procesamiento de lenguaje natural.	Documentos de Investigación Sin Implementación Puntual	Liu, Bing (2022)
2	BERT for Sentiment Analysis: Pre-trained and Fine-Tuned Alternatives	2022	BERT, BERT _{imabu} y TF-IDF	Open-Source Datasets Olist, Buscapé , B2W, UTLC Apps y UTLC Movies	Souza, F. D. (2022, March)
3	Contemporary Business Analytics: An Overview	2021	Analítica de Negocio y Minería de Datos.	Documentos de Investigación Sin Implementación Puntual	Raghupathi, W. & Raghupathi, V. (2021)
4	Sentiment Analysis of Movie Reviews based on Machine Learning	2020	TF-IDF, Machine Learning, Learning, Lasso, Ridge y CatBoost.	Open-Source Dataset recuperado de www.kaggle.com	Yang, Z. (2020, November)
5	Deep Learning for Automated	2019	Redes RNN LSTM, BiLS y GRU	Reseñas de películas obtenidas	Cheng, L. C., & Tsai, S. L.

	Sentiment Analysis of Social Media			desde www.IMDb.com	(2019, August)
6	A Deep Learning Architecture for Sentiment Analysis	2018	Redes CNN, Redes RNN LSTM y NgramCNN	Reseñas de películas obtenidas desde www.IMDb.com	Çano, E., & Morisio, M. (2018, April)
7	Sentiment Analysis of Twitter Data Using Machine Learning Techniques and Scikit-learn	2018	Machine Learning, BernoulliNB, LogisticRegression y Clasificador SGD.	Dataset de Tweets recopilados y suministrados por NLTK Corpora	Shihab & Yang (2018)
8	VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text (Hutto & Gilbert,2014)	2014	Modelo VADER, LIWC, ANEW y SentiWordNet.	Dataset de Tweets recopilados por los autores mediante la librería Pattern.en de Python	Hutto, C., & Gilbert, E. (2014)
9	Sentiment analysis of twitter data using machine learning approaches and semantic analysis	2014	Machine Learning, Naive Bayes, Maximun Entropy, SVM y WordNet	Dataset de Tweets previamente categorizado	Gautam, G., & Yadav, D. (2014, August)

Tabla 1: Estado del Arte. Elaboración propia.

El caso de estudio propuesto en esta investigación propone la aplicación de varias de las técnicas usadas en los objetos de investigación previos, como son las Redes Neuronales Recurrentes (RNN), de LSTM y GRU, así como procesos de extracción de minería de datos en un entorno lingüístico de la cultura colombiana, logrando categorizar el análisis de sentimiento bajo esta premisa, se logra hacer una comparación de ambos modelos acorde a sus métricas.

4. METODOLOGÍA

A continuación, se presentan las etapas del proyecto de nuestro caso de estudio, donde hemos adoptado la metodología CRISP-DM para implementar la solución al problema. Se consideraron diversas técnicas y herramientas mencionadas en las referencias encontradas en el estado del arte para el análisis de sentimiento de cadenas de texto.

4.1. Fase 1: Comprensión del negocio

Se eligieron como caso de estudio las cuentas de Twitter de tres de las principales empresas de telecomunicaciones en Colombia: Movistar, Claro y Tigo. El objetivo es medir la percepción de los usuarios en la red social en relación con el servicio que brindan estas empresas.

4.2. Fase 2: Comprensión de los datos

Se extrae una muestra de datos, a partir del API suministrado por Twitter mediante la librería `sntwitter` de Python, descargando los últimos 10.000 Tweets en español que mencionen la cuenta, incluyendo respuestas y tweets originados hasta el 30 de junio del 2022. Posterior a esto, se realizó la tarea de clasificar los Tweets de manera manual como Negativo, Neutro y Positivo de acuerdo con la percepción detectada en el comentario sobre el servicio que brinda la empresa.

4.3. Fase 3: Preparación de los Datos

Para la preparación de los dataos se realizaron los siguientes pasos:

- a. Se eliminan los Tweets con Spam o que no están relacionados con el servicio de telecomunicaciones, como publicidad, sorteos, eventos deportivos, conciertos, entre otros.
- b. Para el análisis exploratorio, se eliminaron los *stopwords*, que son palabras vacías que de manera individual no revelan patrones en los Tweets, las menciones a la cuenta de la empresa según corresponda (@Movistarco, @ClaroColombia, @Tigo_Colombia) pues aparece en cada Tweet por ser el parámetro de búsqueda en la descarga de datos. Además, se quitaron números, enlaces web, símbolos de puntuación a excepción de emoticones, y por último se convierte cada palabra a minúscula para disminuir el número de opciones en las que puede aparecer.

- c. Para la implementación de los modelos, se realiza la misma preparación, pero se conserva todas las palabras excluidas del diccionario de stopwords, pues dan contexto para la clasificación correcta del texto, y para generar un “vocabulario” más amplio.
- d. Se realizó una partición de los datos de, utilizando el 70% para el entrenamiento y el 30% restante para validación y comparación.
- e. Luego, a partir de los datos de entrenamiento procesados (limpios), generamos la tokenización, que consiste en definir una codificación para cada palabra única (token), para construir un **vocabulario**.

Para esto, se asigna el número 2 para la palabra más frecuente, 3 para la segunda más frecuente, y sucesivamente hasta la última palabra o carácter, es decir en orden descendente de la frecuencia de su uso en el set de entrenamiento, y adicionalmente, se asigna el número 0 al token “vacío” para tener secuencias de textos del mismo largo, y el número 1 al token ‘desconocido’ para palabras desconocidas, es decir palabras que no fueron consideradas para el entrenamiento del modelo, pero que pueden aparecer al clasificar nuevos Tweets como en la validación.

- f. Finalmente, a criterio del grupo investigador se genera una secuencia de tokens para cada Tweet con una longitud máxima de **40** tokens, considerando este valor como la media de tokens contenidos por Tweet. En donde los Tweets con menos palabras, se rellenan con un 0 según lo definido, mientras que los que superen la longitud establecida son recortados hasta 40 tokens. Con esto, pasamos de tener un vector con los textos originales, a una matriz numérica con la codificación de los tokens.

4.4. Fase 4: Modelamiento

A partir de la secuencia de tokens de los datos de entrenamiento, se construyen diferentes modelos de Redes Neuronales Recurrentes utilizando tanto LSTM y GRU. Se prueban diferentes combinaciones de hiperparámetros, para encontrar la configuración óptima para operador. Los modelos de Deep Learning considerados tienen la siguiente estructura:

- a) Se crea una instancia **Sequential**, que permite al modelo apilar capas una encima de la otra de forma lineal. Se añade una capa de **Embedding** al modelo, que representa la

secuencia de los tokens en una dimensionalidad reducida para encontrar relaciones semánticas entre las palabras.

- b) Se añade una capa de **Dropout** con probabilidad de 0.4 para regularizar el modelo y reducir el sobreajuste.
- c) Se añade una capa recurrente al modelo que puede ser **LSTM** o **GRU** para capturar información de secuencia en los datos.
- d) Se añade una capa densa con función de activación **Sigmoide** para obtener una salida binaria.
- e) Se compila el modelo utilizando la función de pérdida de **Binary Crossentropy**, el optimizador **RMSProp** o **Adam**, y se evalúa el rendimiento del ROC AUC del modelo durante el entrenamiento.

Para la selección de los hiperparámetros, se utiliza la optimización bayesiana para encontrar los valores óptimos del modelo para la clasificación de los Tweets. La tabla 2, presenta los hiperparámetros evaluados, una descripción corta y sus respectivos rangos de búsqueda

Hiperparámetro	Descripción	Rango
Learning Rate	Tasa de aprendizaje en el entrenamiento del modelo.	(0.0001, 0.1)
Neuronas de salida en capa Embedding	Dimensión de los vectores en la capa de embedding del modelo.	(32,64,128,256)
Capa recurrente	Define el tipo de modelo a utilizar.	(LSTM, GRU)
Neuronas de salida en la capa del modelo	El modelo recibe las neuronas de salida en la capa de Embedding y las usa para buscar patrones complejos en las secuencias de datos.	(32,64,128,256)
Optimizador	Ajusta de forma iterativa los pesos y sesgos del modelo, para minimizar la función de pérdida.	(RMSProp,Adam)
Número de épocas	Número de iteraciones máxima en el entrenamiento del modelo.	(6, 20)
Paciencia	Número máximo de épocas consecutivas sin encontrar reducción en el Loss de validación. Define la parada en el entrenamiento del modelo para evitar sobreajuste.	(3,5)

Tabla 2: Rango de hiperparámetros

Una vez definidos los hiperparámetros a usar en cada iteración, se realiza una validación cruzada con el método *K-Fold* con 3 divisiones para el set de entrenamiento, en donde en cada división, se realiza una nueva partición de entrenamiento y validación.

Después del entrenamiento, se calcula el área bajo la curva ROC (ROC AUC) para las predicciones en los conjuntos de entrenamiento y validación, para finalmente calcular el **Target**, que es el ROC AUC ajustado que penaliza el sobreajuste.

4.5. Fase 5: Evaluación

Para la selección del modelo, se escogen los 6 modelos con mayor *Target*, para los cuales comparamos las métricas para el problema de clasificación presentadas en la tabla 3, siendo estos medidos tanto en los datos de entrenamiento, como en los de validación.

Para esto, debido a que la clase mayoritaria (de mayor frecuencia) son los Tweet Negativos, definiremos un verdadero positivo como Tweets Negativo que fueron clasificados correctamente como negativos, y un falso negativo como un Tweet No Negativo clasificado incorrectamente como Negativo.

Métrica	Descripción
Precisión (Precision)	Proporción de Tweets Negativos clasificados correctamente con polaridad Negativa, en relación del total de predichos como Negativos.
Sensibilidad (Recall)	Proporción de Tweets Negativos clasificados correctamente con polaridad Negativa.
F1- Score	Medida que equilibra la bondad del modelo en términos de la precisión y la sensibilidad.
Especificidad (Specificity)	Proporción de Tweets No Negativos clasificados correctamente con polaridad No Negativa.
Exactitud (Accuracy)	Proporción de Tweets correctamente clasificados (tanto Negativos como No Negativos) en relación al total general.

Tabla 3: Métricas de evaluación

Después de seleccionar los hiperparámetros que mostraron el mejor ajuste del modelo, se entrena y se calcula el ROC AUC para los datos del conjunto de entrenamiento y de validación y se comparan.

Tras esta comparación, se elige el umbral óptimo para las predicciones del modelo que maximiza el F1-Score, para lograr un equilibrio entre el score de la precisión y la sensibilidad en la clasificación de los Tweets. De manera que se maximice la tasa de verdaderos positivos (TPR) y se minimice la tasa de falsos positivos (FRP), para que el modelo tenga una buena capacidad para discriminar.

Finalmente se evalúa la bondad de las métricas del modelo, por Clasificación de sentimiento.

4.6. Fase 6: Despliegue

Por último, el modelo seleccionado se utiliza para clasificar futuros Tweets, a los cuales se les realiza un seguimiento a partir de un tablero en Power BI, el cual permite interactuar dinámicamente la evolución de la percepción de la marca en el tiempo, logrando visualizar la frecuencia de Tweets, identificar palabras más utilizadas y generar una nube de textos por clasificación, que describen de manera global la experiencia de los usuarios en relación con su marca.

Logrando así, medir el impacto de las campañas de marketing, la percepción de la marca, de sus productos y servicios.

5. RESULTADOS

5.1. Comprensión de los datos

Como primer hallazgo, se encontró que un significativo porcentaje de los 10.000 Tweets descargados por cada operador pertenecen a diferentes usuarios corporativos de las mismas empresas de telecomunicaciones (@movistarco, @MovistarArenaCo, @clarocolombia, @claroteayuda, @Tigo_Colombia, @womcolombia). Los Tweets de estos usuarios no representan la percepción de marca sobre el servicio brindado y, por lo tanto, se consideraron como ruido y se excluyeron del estudio.

Además, se eliminaron muchos Tweets que no estaban relacionados con la percepción del servicio que presta la empresa. Para Movistar se depuraron menciones relacionadas con eventos de @MovistarArenaCo, en Claro menciones relacionadas con las carreras de la Formula 1, y en el caso de Tigo se eliminó más de la mitad de los Tweets, debido a las menciones del sorteo "AIEstadioConTigoVerde" de boletas para ir al estadio.

Después de la depuración, se obtuvieron 7.055 Tweets para Movistar, 8.000 para Claro y 2.658 Tweets para Tigo, siendo este último el más afectado, ya que tiene menos del 27% de datos iniciales para la construcción y validación de los modelos.

En la figura 16 se presenta la frecuencia de Tweets generados por día para cada operador, excluyendo aquí, los Tweets depurados en la fase de limpieza de datos.

En esta exploración, se puede observar que para las fechas entre el 13 al 29 de junio, se tienen frecuencias de publicación de entre 29 y 370 Tweets diarios, con un promedio de 136 Tweets, lo cual confirma la necesidad de poder automatizar la clasificación de estos, pues toma un tiempo considerable, realizar de manera manual el seguimiento, depuración y la clasificación de estos, para conocer en tiempo real la percepción del servicio por parte de los usuarios en Twitter.

En cuanto a la comparación de las 3 compañías, se observa un patrón similar entre ellas, destacando que Tigo tiene una mayor cantidad de menciones, seguido de Claro y por último Movistar, sin embargo, dichas diferencias no son muy grandes.

Algo a considerar a partir de este punto, es que, al realizar la depuración, el periodo de tiempo de los Tweets para Movistar es del 29 de abril al 29 de junio, para Claro del 14 de mayo al 29 de junio, y para Tigo tan solo del 14 al 29 de junio.

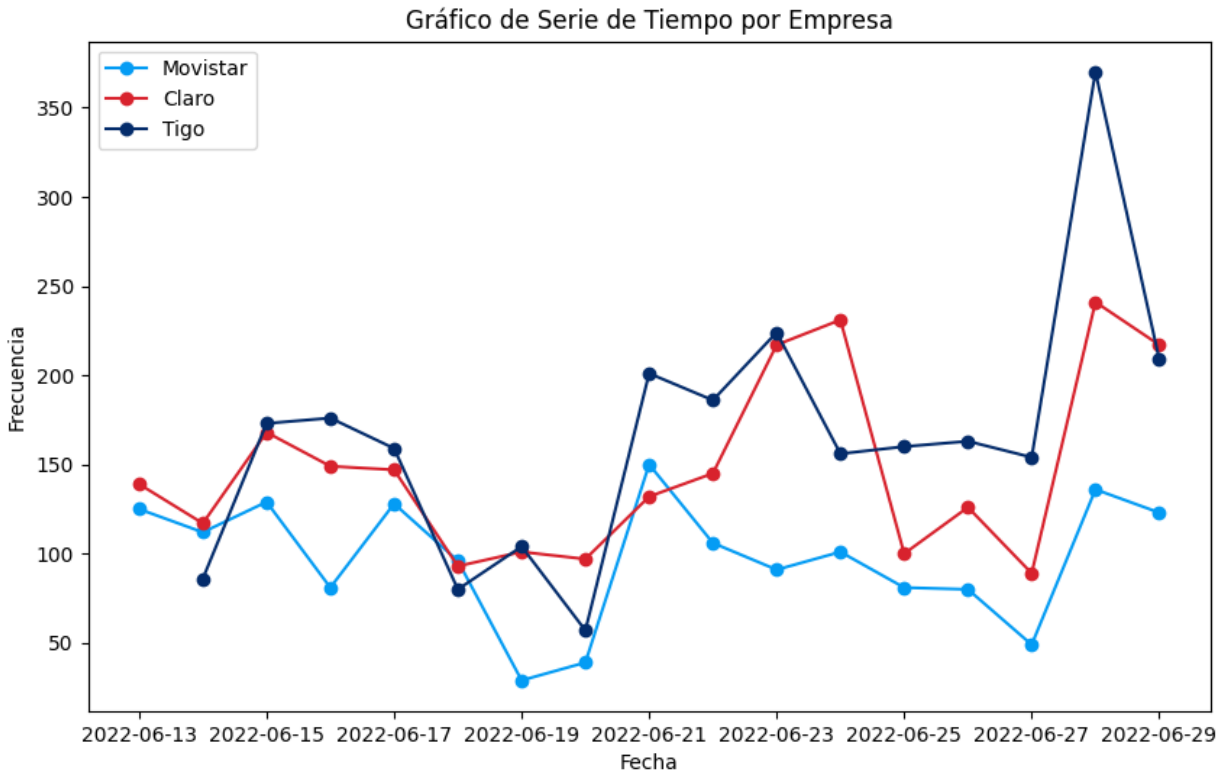


Figura 16: Número de Tweets por día

En, en la tabla 4, se observa que los operadores presentan una distribución similar en su clasificación en relación con el análisis de sentimiento percibido por sus usuarios, siendo en su mayoría comentarios Negativos, y encontrando un porcentaje muy bajo en los positivos, estando por debajo del 3%, lo cual dificulta el aprendizaje del modelo para clasificar esta etiqueta.

Clasificación	Movistar	Claro	Tigo
Negativo	5,107 (72.39%)	5,814 (72.68%)	2,224 (83.67%)
Neutro	1,769 (25.07%)	1,964 (24.55%)	397 (14.94%)
Positivo	179 (2.54%)	222 (2.78%)	37 (1.39%)
Total	7,055	8,000	2,658

Tabla 4: Distribución por sentimiento

Debido a la falta de datos para el entrenamiento y correcta predicción de las etiquetas de los Tweets Positivos, en la tabla 5, el grupo investigador decide agrupar los Tweets Neutros y los Positivos como “No Negativos” los cuales representan ahora alrededor del 27% para Movistar y Claro, pero siendo muy inferior aun para Tigo que es tan solo del 16.33% del total.

Clasificación	Movistar	Claro	Tigo
Negativo	5,107 (72.39%)	5,814 (72.68%)	2,224 (83.67%)
No Negativo	1,948 (27.61%)	2,186 (27.32%)	434 (16.33%)
Total	7,055	8,000	2,658

Tabla 5: Distribución por agrupación de sentimiento

Siguiendo el procedimiento establecido en la metodología, se realiza el proceso de limpieza de los textos. Esto nos facilita identificar patrones en los comentarios de los usuarios en la red social y calcular la frecuencia de ocurrencia.

En la siguientes figuras, se presenta el Top 20 de palabras más mencionadas y la nube de texto global para cada empresas. Se destaca que la palabra “Servicio” es la palabra mas utilizada, seguida de “@sicsuper”, “ineternet”, “días”, “línea”, “plan” y otras.

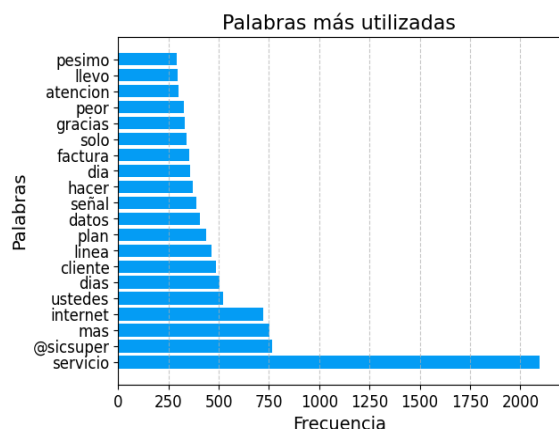


Figura 17: Frecuencia de palabras Movistar

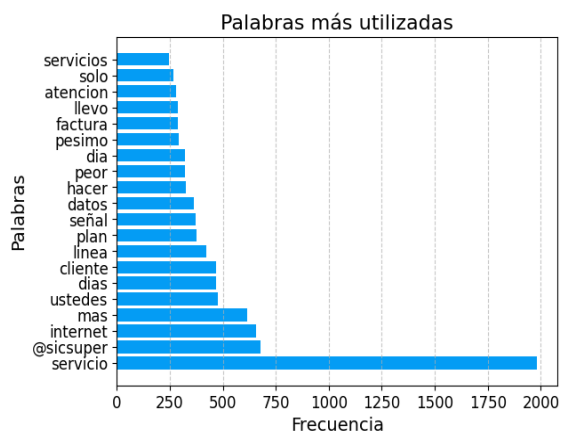


Figura 20: Frecuencia de palabras Negativas Movistar

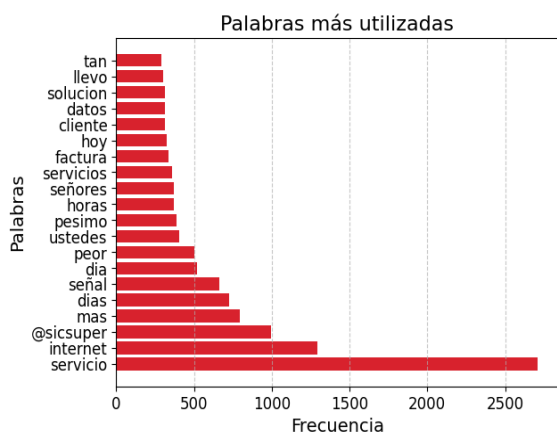


Figura 21: Frecuencia de palabras Negativas Claro

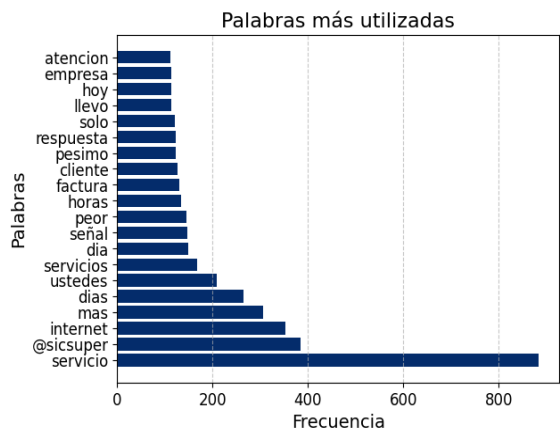


Figura 22: Frecuencia de palabras Negativas Tigo

5.1.2. No Negativos

Para los Tweets clasificados como “No Negativos”, se observan diferencias en la distribución de palabras. En general se destacan terminos como “gracias”, “más”, “mejor”.

Es importante resaltar que para Movistar, se encontraron muchas menciones relacionadas con “Muévete con ElegiApps TV” campaña que fue lanzada durante el periodo de evaluación y tuvo una gran aceptación por parte de los usuarios del servicio de Fibra optica de Movistar. En el caso de Claro, se mencionan palabras como “tecnologia” y “siempre”. Por otro lado, para Tigo, se observa que muchos usuarios mencionan querer cancelar su servicio actual de Claro, para cambiarse a Tigo.

Es relevante destacar que la muestra de Tweets clasificados como “No Negativos” es significativamente pequeña, como se mencionó en la depuración de los datos. Por lo tanto, se estima que estos Tweets no aportan un peso suficiente en la estimación y entrenamiento del modelo.

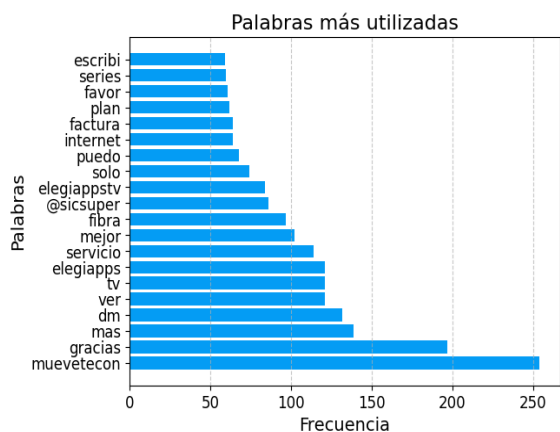


Figura 23:Frecuencia de palabras No Negativas Movistar

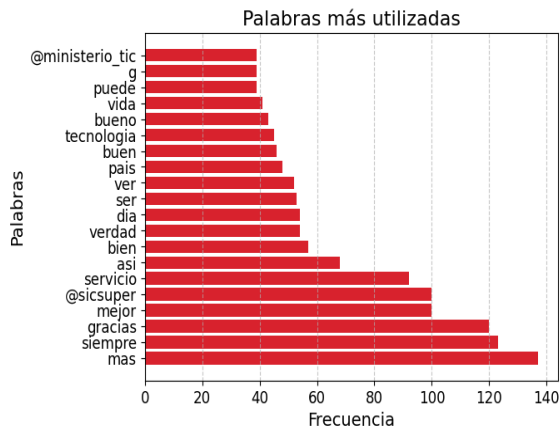


Figura 24:Frecuencia de palabras No Negativas Claro



Figura 25:Frecuencia de palabras No Negativas Tigo

5.2. Preparación de los Datos

5.2.1. Vocabulario set de entrenamiento

Después de realizar la limpieza de los datos, el proceso de tokenización y la construcción del vocabulario en los datos de entrenamiento, tal como se explicó en la sesión metodológica, se obtiene la secuencia de tokens para cada Tweet. El número de tokens para el conjunto de entrenamiento es de 9.604, 10.590 y 4.987 para Movistar, Claro y Tigo respectivamente. Las siguientes tablas, muestran el código asignado para los primeros tokens por operador.

Índice	Palabra	Índice	Palabra	Índice	Palabra
0	vacío	7	la	14	con
1	desconocido	8	el	15	es
2	@movistarco	9	me	16	un
3	de	10	a	17	para
4	que	11	en	18	los
5	y	12	por	19	lo
6	no	13	servicio	20	se

Tabla 6: Vocabulario Movistar

Índice	Palabra	Índice	Palabra	Índice	Palabra
0	Vacío	7	la	14	Con
1	Desconocido	8	no	15	por
2	'@clarocolombia'	9	a	16	@claroteayuda
3	De	10	en	17	un
4	Que	11	servicio	18	para
5	Y	12	me	19	los
6	El	13	es	20	se

Tabla 7: Vocabulario Claro

Índice	Palabra	Índice	Palabra	Índice	Palabra
0	vacío	7	el	14	con
1	desconocido	8	la	15	por
2	@tigo_colombia	9	a	16	un
3	de	10	me	17	para
4	que	11	servicio	18	los
5	y	12	en	19	se
6	no	13	es	20	lo

Tabla 8: Vocabulario Tigo

5.2.2. Representación secuencial

Después de definir los vocabularios de cada operador, se representa cada Tweet como un vector numérico, en donde se separa las palabras (tokens) de manera individual y se remplazan por su índice correspondiente en el vocabulario. De esta manera se convierte cada secuencia de tokens en una secuencia numérica como el ejemplo que se presenta en la tabla 9 a continuación.

2	36	37	...	0	0	0
2	201	38	...	0	0	0
2	201	38	...	111	27	1
...
1,480	1	34	...	0	0	0
1	1	1	...	0	0	0
1,480	1	3,284	...	0	0	0

Tabla 9 Ejemplo Matriz de secuencia de Tokens

5.3. Modelamiento

Las tablas de la 10 a 12 a continuación, muestran el Top 6 de los mejores hiperparámetros seleccionados, de 50 combinaciones en función del valor objetivo (Target) construidas en el set de entrenamiento para Movistar, Claro y Tigo respectivamente.

Los resultados de los 6 mejores modelos para las 3 empresas mostraron un desempeño superior utilizando el optimizador RMSProp. En el caso de Movistar, todos sus modelos utilizan GRU. En Tigo, la mayoría de los modelos son LSTM, mientras que en Claro se encontró un equilibrio entre ambos tipos de modelos.

No se observaron diferencias significativas en cuanto al Target de los modelos dentro de cada empresa, sin embargo, los de Claro tuvieron una ligera ventaja sobre los demás, mientras que los modelos de Tigo obtuvieron resultados inferiores en comparación.

Para cada operador, se selecciona los dos mejores modelos en cuanto al Target, para realizar la evaluación de las métricas de desempeño, las comparaciones y la selección del modelo final.

Modelo	Learning Rate	Neuronas Embedding	Modelo	Neuronas Modelo	Optimizador	Épocas	Paciencia	Target
Modelo 1	0.0849	32	GRU	128	RMSProp	15	3	0.8930
Modelo 2	0.0001	32	GRU	128	RMSProp	15	3	0.8871
Modelo 3	0.0001	64	GRU	256	RMSProp	15	3	0.8869
Modelo 4	0.0001	32	GRU	64	RMSProp	11	4	0.8864
Modelo 5	0.0761	32	GRU	256	RMSProp	16	3	0.8858
Modelo 6	0.0121	32	GRU	32	RMSProp	12	5	0.8857

Tabla 10: Modelos Movistar

Modelo	Learning Rate	Neuronas Embedding	Modelo	Neuronas Modelo	Optimizador	Épocas	Paciencia	Target
Modelo 1	0.1	32	GRU	64	RMSProp	16	3	0.9254
Modelo 2	0.0407	32	LSTM	64	RMSProp	6	3	0.9241
Modelo 3	0.0001	32	GRU	128	RMSProp	8	3	0.9241
Modelo 4	0.0656	32	GRU	32	RMSProp	6	4	0.9234
Modelo 5	0.0495	32	LSTM	64	RMSProp	12	3	0.9231
Modelo 6	0.0001	32	GRU	64	RMSProp	9	4	0.9227

Tabla 11: Modelos Claro

Modelo	Learning Rate	Neuronas Embedding	Modelo	Neuronas Modelo	Optimizador	Épocas	Paciencia	Target
Modelo 1	0.0363	32	LSTM	64	RMSProp	16	4	0.8134
Modelo 2	0.1	32	LSTM	32	RMSProp	16	4	0.8107
Modelo 3	0.0966	128	LSTM	128	RMSProp	14	4	0.8075
Modelo 4	0.0795	128	GRU	32	RMSProp	16	4	0.8041
Modelo 5	0.0330	32	GRU	64	RMSProp	16	5	0.8004
Modelo 6	0.0256	32	LSTM	64	RMSProp	16	5	0.7990

Tabla 12: Modelos Tigo

5.3.1. Evaluación del Loss por época

Las siguientes figuras, muestran la evolución de la pérdida (*Loss*) para los datos de entrenamiento y validación a lo largo de las épocas de entrenamiento de los dos modelos en cada operador. Estas graficas permiten evaluar el rendimiento, la convergencia y el sobreajuste. Se obtiene el *Loss* con índice más bajo en validación en la 6^a, 4^a y 8^a época para los dos modelos de Movistar, Claro y Tigo respectivamente. Entre estos operadores, Tigo presenta el mayor sobreajuste en sus modelos, ya que tiene una distancia más grande entre las dos líneas en los dos conjuntos de datos. Por otro lado, Claro es quien tiene una diferencia más pequeña de *Loss* entre ambos.

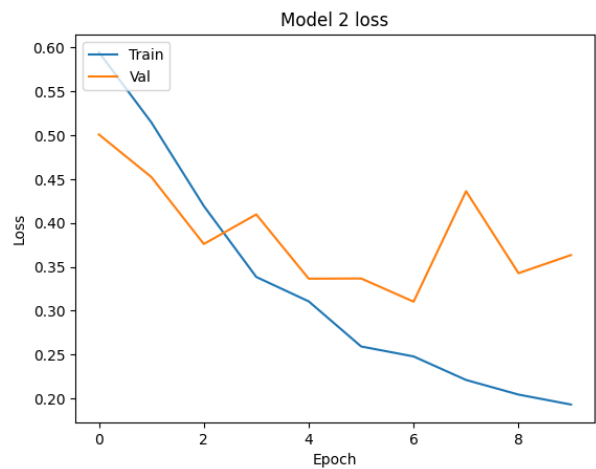
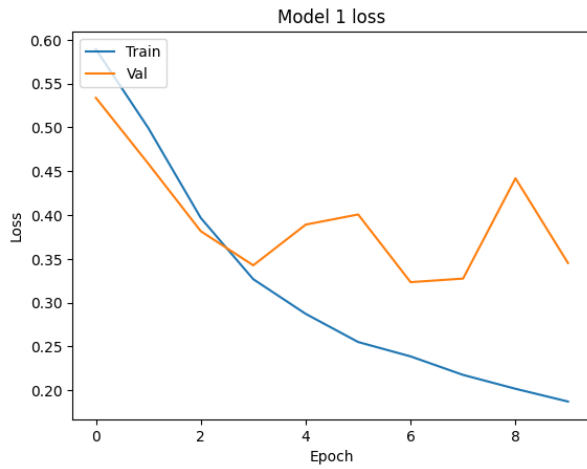


Figura 26: Evolución del Loss durante el entrenamiento por época Movistar

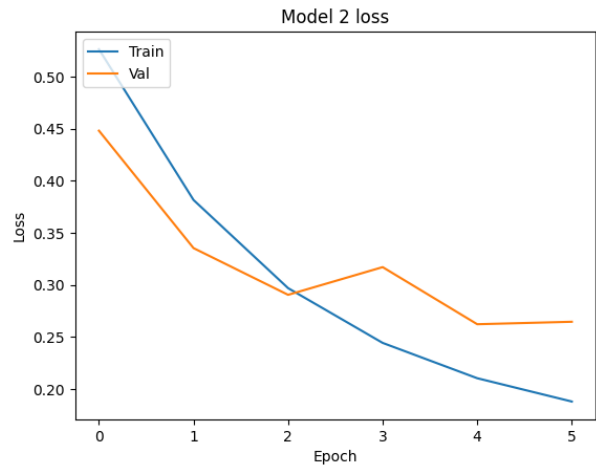
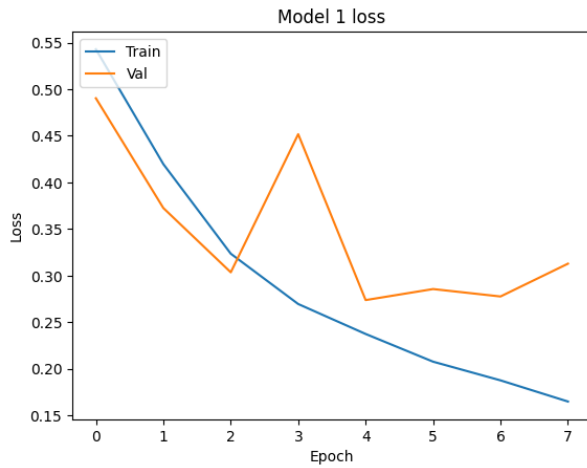


Figura 27: Evolución del Loss durante el entrenamiento por época Claro

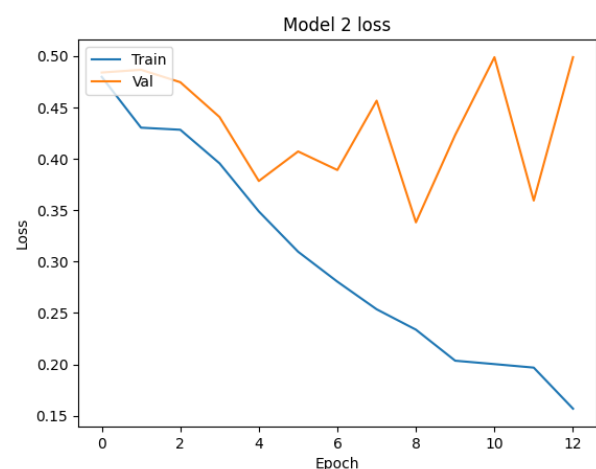
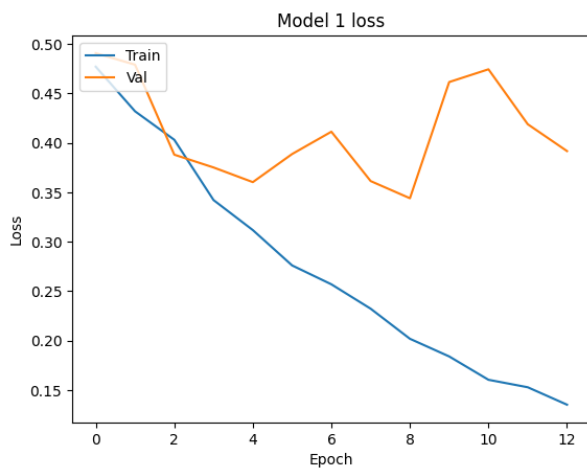


Figura 28: Evolución del Loss durante el entrenamiento por época Tigo

5.4. Evaluación

5.4.1. Selección del punto de corte

Ahora, se elige el punto de corte óptimo para clasificar los Tweets como “Negativos” o “No Negativos”. Para esto, se presentan las curvas con las métricas de evaluación del rendimiento en el conjunto de datos de validación, para los dos modelos seleccionados para cada operador.

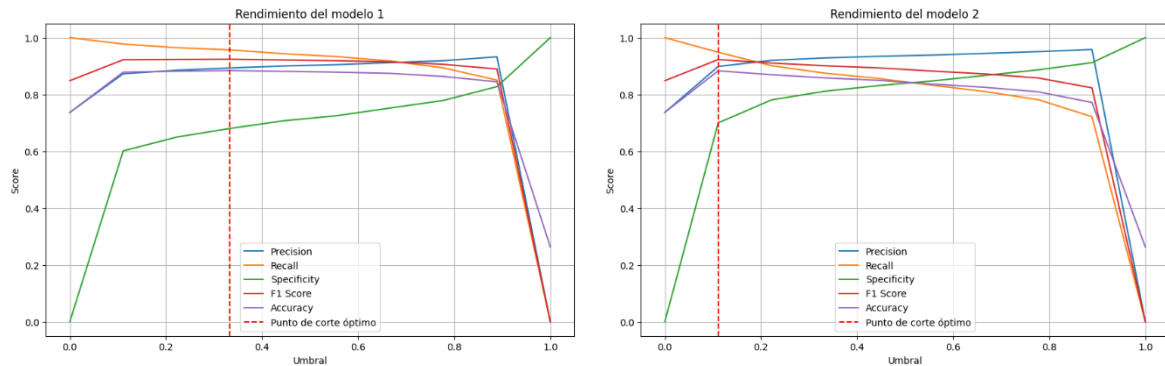


Figura 29: Rendimiento de modelos Movistar

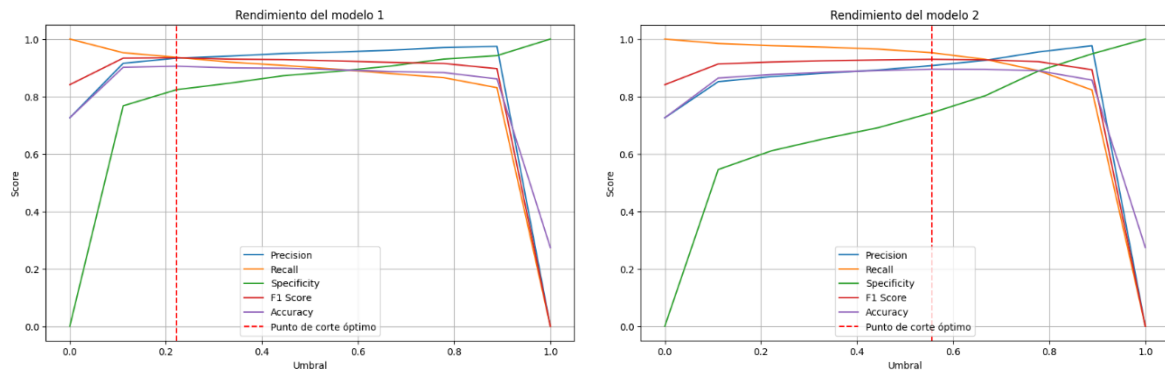


Figura 30: Rendimiento de modelos Claro

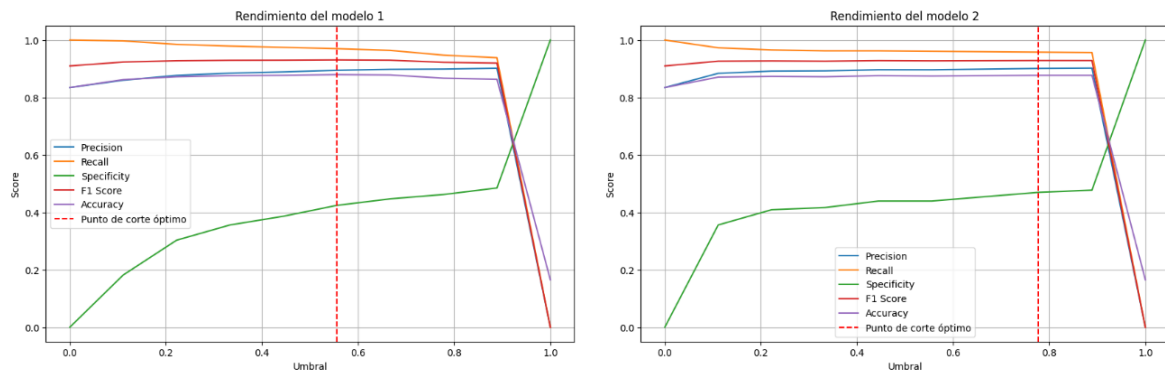


Figura 31: Rendimiento de modelos Tigo

De acuerdo con la sesión metodológica, para cada modelo se eligió como umbral óptimo el punto de corte que maximiza el **F1-Score**, con el objetivo de lograr un equilibrio entre el score de la **Precision** y el **Recall** en la clasificación de los Tweets. La tabla 13 a continuación presenta el punto de corte seleccionado en cada caso.

Operador	Punto de corte	
	Modelo 1	Modelo 2
Movistar	0.33	0.11
Claro	0.22	0.55
Tigo	0.55	0.77

Tabla 13: Umbral óptimo de los modelos

5.4.2. Métricas de desempeño

En la tabla 14, se muestran las métricas de evaluación de desempeño de los modelos, tanto en la partición de los datos de entrenamiento como en validación, para el umbral óptimo correspondiente de los modelos en cada operador.

Los modelos de los 3 operadores tienen resultados muy sobresalientes con scores superiores al 85% para todas las métricas, a excepción de la **Specificity** que se encuentra entre el 68% y 80% para Movistar, entre el 74% y 90% para Claro, y entre el 42% y 77% para Tigo, siendo este último muy inferior en comparación a los demás, especialmente en los datos de validación siendo del 42.4% para el primer modelo y del 47% para el segundo.

Al comparar los scores entre los datos de entrenamiento y validación, observamos diferencias muy pequeñas, siendo en su mayoría inferiores a un 5%. Esto nos indica que los modelos no sufren de sobreajuste. Sin embargo, nuevamente, para la **Specificity** estas diferencias son significativas, especialmente en Tigo, donde el score de validación es más de un 30% más bajo en comparación con el entrenamiento.

En cuanto a la comparación entre los dos mejores modelos por operador, los scores son muy similares. El primer modelo tiene un rendimiento ligeramente superior en algunas métricas, mientras que el segundo modelo tiene un rendimiento mejor para otras. Sin embargo, estas diferencias no son significativas.

Operador	Métrica	Modelo 1		Modelo 2	
		Entrenamiento	Validación	Entrenamiento	Validación
Movistar	Precisión	92.2%	89.3%	92.0%	89.9%
	Recall	97.5%	95.7%	97.5%	94.9%
	F1 Score	94.8%	92.4%	94.7%	92.3%
	Accuracy	92.3%	88.4%	92.1%	88.3%
	Specificity	79.1%	68.0%	78.4%	70.0%
Claro	Precisión	96.1%	93.4%	94.6%	90.8%
	Recall	95.2%	93.6%	96.2%	95.2%
	F1 Score	95.6%	93.5%	95.4%	92.9%
	Accuracy	93.7%	90.5%	93.2%	89.5%
	Specificity	89.7%	82.4%	85.3%	74.3%

Tigo	Precisión	95.6%	89.5%	94.9%	90.1%
	Recall	98.3%	97.0%	98.2%	95.8%
	F1 Score	96.9%	93.1%	96.5%	92.9%
	Acurracy	94.8%	88.0%	94.0%	87.7%
	Specificity	76.5%	42.4%	72.5%	47.0%

Tabla 14: Métricas de evaluación modelos

5.4.3. Desempeño del modelo por Sentimiento

En este apartado, se presentan las gráficas con el desempeño de las métricas en los datos de validación por Clasificación de sentimiento (Negativo y No Negativo).

En los 3 operadores observamos una distribución similar en las métricas de evaluación. Se destaca que los modelo tiene una mayor capacidad para clasificar los “Negativo” en comparación con los “No Negativos”.

Además, en los “No Negativos” la **Precision** es mayor que el score del **Recall** y **F1 – Score**, mientras que en los “Negativos” el score de la **Precision** es el más bajo.

Estas observaciones indican que los modelos tienen una mejor bondad para clasificar el sentimiento de un Tweet “Negativo” correctamente como “Negativo”, pero puede tener dificultades para clasificar correctamente los “No Negativos”, lo cual se atribuye a tener muestras más pequeñas para el entrenamiento de esta última clase.

Estas diferencias en la bondad de los modelos por sentimiento, puede deberse además a que se encontró que en los Tweets “Negativos” hay más homogeneidad en cuanto a su contenido, como problemas en el servicio de internet, televisión, datos o facturación, mientras que en los “No Negativos” se encuentran una variedad mucho más amplia de temas, que dificulta el aprendizaje del modelo para detectar patrones para una correcta clasificación.

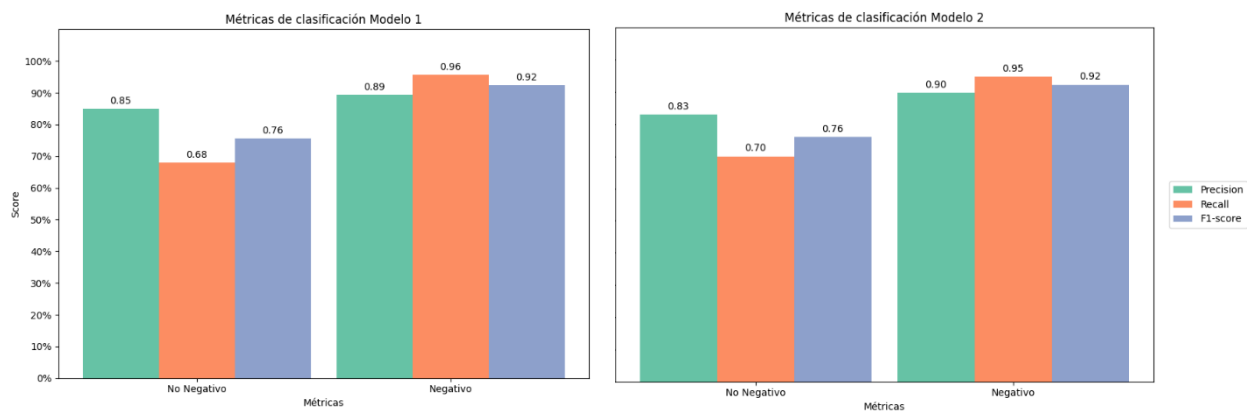


Figura 32: Desempeño de modelos por Sentimiento Movistar

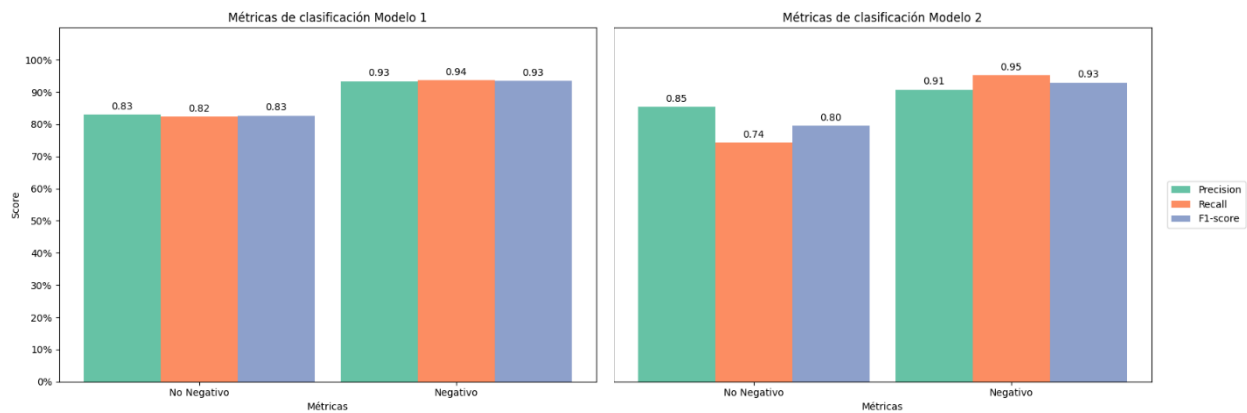


Figura 33: Desempeño de modelos por Sentimiento Claro

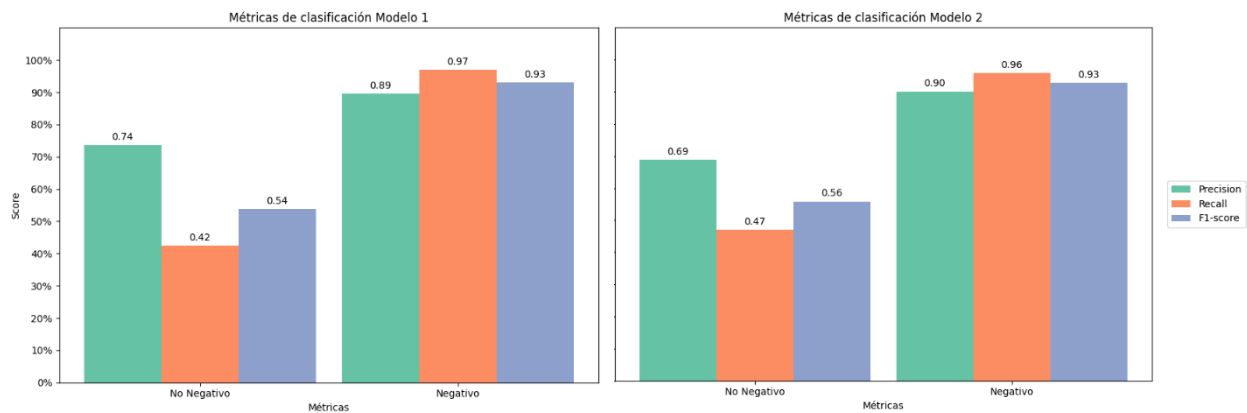


Figura 34 Desempeño de modelos por Sentimiento Tigo

5.4.4. Selección de modelo

Ahora, se decide el modelo final para cada operador, en función de los resultados, comparaciones y conclusiones presentadas de las subsecciones anteriores.

Para Movistar, no hay diferencias significativas en la evaluación de los dos modelos, debido a que tienen la misma estructura e hiperparámetros, a excepción de que el *Learning Rate* es más pequeño para el segundo modelo. A pesar de ello, se prefiere el modelo 1 que presentó una ligera ventaja en las métricas de desempeño en validación general y por sentimiento.

Para Claro, hay una mayor diferencia en la estructura de los dos modelos. El primer modelo es una red neuronal recurrente GRU, mientras que el segundo modelo es una red neuronal recurrente LSTM. Aunque ambos modelos tienen muy buen desempeño, se seleccionó el modelo

1, debido a su mayor rendimiento en las métricas para los datos de validación a nivel general como por sentimiento. Además, el modelo seleccionado se destaca por tener una mayor homogeneidad en los scores por sentimiento.

Finalmente, para el operador de Tigo, ambos modelos son redes neuronales recurrentes LSTM que comparten hiperparámetros muy similares, lo que resulta en que sus scores también sean cercanos. Sin embargo, al igual que en los demás operadores, se eligió el modelo 1, debido a que tiene un ligero rendimiento superior en las métricas para los datos de validación. Además, la **Precisión** en los Tweets “No Negativos”, son un 5% más altos en el modelo seleccionado en comparación con el segundo modelo.

5.5. Despliegue del modelo

Una vez seleccionados, entrenados y validados los modelos, se procede al despliegue para clasificar nuevos Tweets. Para esto, primero se debe realizar el procesamiento de depuración y limpieza de los datos descargados como se explicó en la sesión metodológica.

A continuación, se obtienen las predicciones sobre estos nuevos datos y los resultados se visualizan en un tablero dinámico, que permite hacer seguimiento periódicamente y comparaciones de la percepción de los usuarios en la red social Twitter.

En las siguientes figuras, se presentan los tableros dinámicos construidos para los 3 operadores de telecomunicaciones, en el periodo de tiempo de la semana del 2 al 9 de mayo de 2023.



Figura 35: Despliegue de modelo Movistar

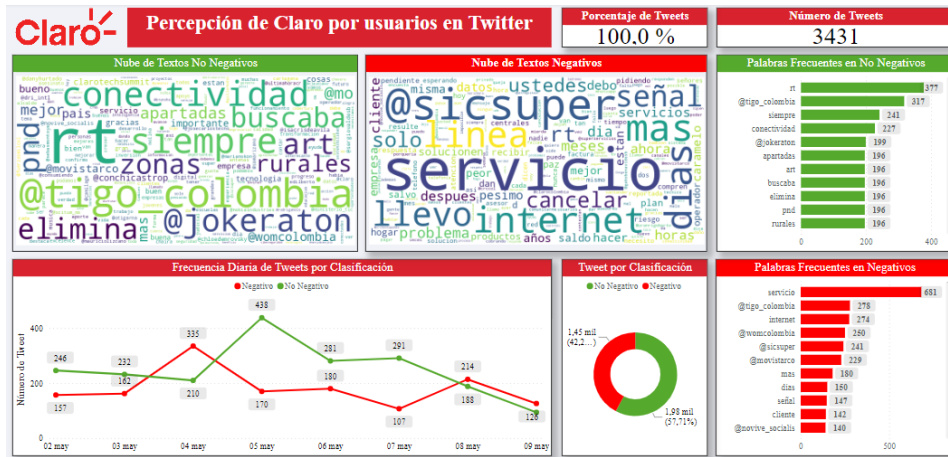


Figura 36: Despliegue de modelo Claro

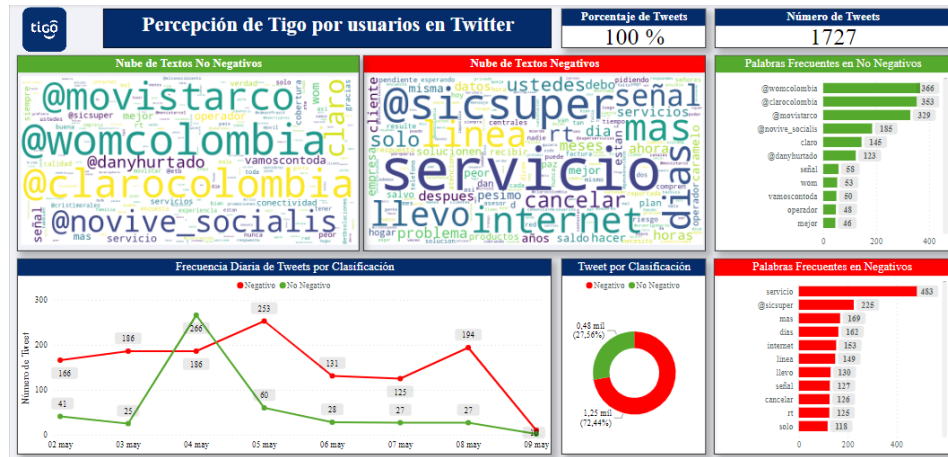


Figura 37: Despliegue de modelo Tigo

6. CONCLUSIONES Y TRABAJOS FUTUROS

6.1. Conclusiones

Durante la recopilación y construcción la base de datos de los comentarios de usuarios en Twitter que mencionaban a los operadores de telecomunicaciones se observó que un alto porcentaje de estos comentarios no estaban relacionados con el servicio que presta principalmente la empresa. Se encontró una gran cantidad de Tweets relacionados con eventos deportivos, sorteos, conciertos entre otros temas no relevantes para el objetivo de la investigación. Este hallazgo demuestra la gran importancia del proceso de limpieza y depuración de los datos, antes del entrenamiento de los modelos para no obtener resultados e interpretaciones incorrectas.

Los modelos desarrollados y desplegados en este estudio permiten clasificar el sentimiento de los Tweets de los usuarios como “Negativos” o “No Negativos” para los 3 operadores de telecomunicaciones. Estos modelos, tienen una bondad bastante alta para clasificar los Tweets Negativos correctamente con polaridad “Negativa”, con un porcentaje de **Sensibilidad (Recall)** en los datos de validación superiores al 94%. Mientras que, el porcentaje de los Tweets con sentimientos “No Negativos”, correctamente clasificados como “No Negativos” **Especificidad (Specificity)** es del 68%, 82.4% y 42.4% para Movistar, Claro y Tigo respectivamente, siendo significativamente bajo para este último.

La falta de precisión para los Tweets “No Negativos” se puede atribuir a varios factores. En primer lugar, la amplia variedad de temas presentes en los comentarios de esta categoría, lo que dificulta encontrar patrones distintivos. Además, de la baja cantidad de datos de esta categoría en comparación con los Negativos, que implica un vocabulario más reducido para el aprendizaje del modelo, y la correcta clasificación de estos Tweets. Lo anterior conlleva a que el modelo tienda a sobre ajustarse a los datos de entrenamiento y tenga dificultades para generalizar correctamente en casos fuera de este conjunto de datos. Esto se vuelve más evidente en la clasificación para el operador de Tigo, donde el número de Tweets “No Negativos”, es bastante reducida después de aplicar los procesos de limpieza y depuración de los datos.

Los modelos LSTM y GRU propuestos, revelaron resultados satisfactorios con diferencias mínimas en las métricas de evaluación. Sin embargo, al comparar por operador, se observó que en Movistar los 6 mejores modelos fueron redes neuronales recurrentes GRU. En el caso de la selección del modelo final para Claro se prefirió GRU en lugar de LSTM, ya que demostró una

ligera ventaja en las métricas. Por otro lado, para Tigo, se prefirió utilizar el modelo LSTM para la clasificación del sentimiento de los Tweets debido a que obtuvo porcentajes más altos en las métricas de evaluación.

El tablero de visualización dinámica desarrollado para el análisis de las predicciones de los modelos desplegados para cada operador permite realizar un análisis del negocio basado en datos. Mediante este panel de control, se facilita realizar consultas dinámicas del número de Tweets por día y la frecuencia de palabras según la clasificación de sentimiento, así como el análisis comparativo entre los 3 operadores. Estas herramientas proporcionan información actualizada de las opiniones de los usuarios en la red social de Twitter, logrando una ventaja competitiva.

6.2. Trabajos Futuros

Para futuras investigaciones se recomienda utilizar un conjunto de datos más amplio, para tener un número de Tweets suficientemente grande para ambas categorías. Se aconseja usar alrededor de 5.000 Tweets para cada clase de sentimiento, tanto "Negativos", como "No Negativos". Con esto, se espera que los modelos tengan una mejor capacidad para generalizar y clasificar de manera más precisa ambos grupos y disminuir el sobreajuste, especialmente en los Tweets en la categoría "No Negativo". De igual manera se propone definir la longitud máxima de tokens de la representación secuencial, como un hiperparámetro de búsqueda, de esta manera su exploración sería dinámica según el set de entrenamiento y no a criterio del grupo investigador.

De igual manera se propone evaluar el escenario y comprobar si se obtiene una bondad similar o superior, al utilizar un solo modelo que consolide la información de los 3 operadores, utilizando los Tweets y la compañía objetivo como datos de entrada para la clasificación del sentimiento, de manera que se simplifique el proceso de inferencia y reducir la carga computacional.

También se propone explorar y comparar otros modelos de Deep Learning como el *BERT (Bidirectional Encoder Representations from Transformers)*, para el análisis de sentimientos. De acuerdo con el estado del arte, *BERT* y en general los modelos basados en *Transformers*, han demostrado lograr mejores resultados en diversas tareas de procesamiento de lenguaje natural, incluido el análisis de sentimientos. Realizar esta comparación ayudaría a determinar si *BERT*

supera a los modelos recurrentes tradicionales propuesto en términos de la precisión de la clasificación de sentimientos.

En cuanto al tablero dinámico construido en el despliegue, se debe seguir actualizando con nuevos Tweets periódicamente y complementarlo con métricas adicionales relevantes para evaluar la percepción de marca de los operadores en Twitter. Esto proporcionará una visión más completa y actualizada de la reputación de la empresa en la red social, permitiendo identificar oportunidades de mejora y tomar decisiones estratégicas basadas en el análisis de datos con información reciente.

7. BIBLIOGRAFÍA

Kriss, P (2014). The Value of Customer Experience, Quantified. Harvard Business Review. The Harvard Business Review.

Recuperado de:

<https://hbr.org/2014/08/the-value-of-customer-experience-quantified>

Scheibenreif, D., Chiu, M., Nordlund, C., Thompson, E.(2023). How IT Can Strengthen Customer Experience Governance. Gartner.

Recuperado de:

<https://www.gartner.com/en/doc/728141-how-it-can-strengthen-customer-experience-governance>

Postdata (2023). Data Flash 2023-01 - Portabilidad Numérica Móvil. Comisión de Regulación de Comunicaciones CRC.

Recuperado de:

<https://www.postdata.gov.co/dataflash/data-flash-2023-001-portabilidad-numerica-movil>

LaRepública (2022). Conozca los operadores móviles con mayor número de quejas por parte de usuarios.

Recuperado de:

<https://www.larepublica.co/empresas/conozca-los-operadores-moviles-con-el-mayor-numero-de-quejas-por-parte-de-usuarios-3363706#:~:text=Para%20empezar%2C%20los%20datos%20de,por%20parte%20de%20sus%20usuarios>

Hill-Wilson, M., & Blunt, C. (2013). Delivering effective social customer service: How to redefine the way you manage customer experience and your corporate reputation. John Wiley & Sons.

Alfaro, E., Velilla, J., Brunetta, H., Navarro, B., & Molina, C. (2012). Customer experience: una visión multidimensional del marketing de experiencias.

Recuperado de

https://boscolg.files.wordpress.com/2013/09/ebook_customerexperience.pdf

Raghupathi, W., & Raghupathi, V. (2021). Contemporary business analytics: An Overview. Data, 6(8), 86.

Recuperado de:

<https://doi.org/10.3390/data6080086>

IBM Documentation (2021). About text mining.

Recuperado de:

<https://www.ibm.com/docs/en/spss-modeler/saas?topic=analytics-about-text-mining>

IBM Documentation (2021). Conceptos básicos de ayuda CRISP-DM.

Recuperado de:

<https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, 39(11), 27-34.

Recuperado de:

<https://dl.acm.org/doi/pdf/10.1145/240455.240464>

Feldman, R., & Dagan, I. (1995, August). Knowledge Discovery in Textual Databases (KDT). In KDD (Vol. 95, pp. 112-117).

Recuperado de:

<https://www.academia.edu/download/49469467/KDD95-012.pdf>

Mansilla, J. (2018). Knowledge Discovery in Text. LinkedIn.

Recuperado de: <https://www.linkedin.com/pulse/knowledge-discovery-text-javier-mansilla/?trackingId=tHu%2BJEDxQVSdTKswYBUc7g%3D%3D>

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

Recuperado de:

<https://www.cs.uic.edu/~liub/FBS/liub-SA-and-OM-book.pdf>

Souza, F. D., & Filho, J. B. D. O. E. S. (2022, March). BERT for Sentiment Analysis: Pre-trained and Fine-Tuned Alternatives. In Computational Processing of the Portuguese Language: 15th International Conference, PROPOR 2022, Fortaleza, Brazil, March 21–23, 2022, Proceedings (pp. 209-218). Cham: Springer International Publishing.

Cheng, L. C., & Tsai, S. L. (2019, August). Deep learning for automated sentiment analysis of social media. In Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining (pp. 1001-1004).

Kavlakoglu, E. (2020). AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the difference?. IBM Documentation

Recuperado de:

<https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>

Madhavan, S. & Madan, P. (2020). An introduction to deep Learning. IBM Documentation.

Recuperado de:

<https://developer.ibm.com/learningpaths/get-started-with-deep-learning/an-introduction-to-deep-learning/>

Jones, T. (2017). Recurrent Neural Networks Deep Dive. IBM Documentation

Recuperado de:

https://developer.ibm.com/articles/cc-cognitive-recurrent-neural-networks/?mhsr=ibmsearch_a&mhq=architecture%20for%20RNN

Madhavan, S & Jones, T. (2021). Deep Learning Architectures. IBM Documentation.

Recuperado de:

https://developer.ibm.com/articles/cc-machine-learning-deep-learning-architectures/?mhsr=ibmsearch_a&mhq=architecture%20for%20RNN

Olah, C. (2015). Understanding lstm networks. Personal Blog.

Recuperado de:

<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Provost, F., & Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking. " O'Reilly Media, Inc."

Yang, Z. (2020, November). Sentiment Analysis of Movie Reviews based on Machine Learning. In 2020 2nd International Workshop on Artificial Intelligence and Education (pp. 1-4).

Çano, E., & Morisio, M. (2018, April). A deep learning architecture for sentiment analysis. In Proceedings of the International Conference on Geoinformatics and Data Analysis (pp. 122-126).

Elbagir, S., & Yang, J. (2018, December). Sentiment analysis of twitter data using machine learning techniques and scikit-learn. In Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence (pp. 1-5).

Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the international AAAI conference on web and social media (Vol. 8, No. 1, pp. 216-225).

Recuperado de:

<https://ojs.aaai.org/index.php/icwsm/article/view/14550>

Gautam, G., & Yadav, D. (2014, August). Sentiment analysis of twitter data using machine learning approaches and semantic analysis. In 2014 Seventh international conference on contemporary computing (IC3) (pp. 437-442). IEEE.