



**ANÁLISIS DE SENTIMIENTO DE MENSAJES DE CORREO
ELECTRÓNICO BAJO UN ENFOQUE DE TRANSFER LEARNING
EMAIL SENTIMENT ANALYSIS UNDER A TRANSFER LEARNING
APPROACH**

Yohann Montaña Reyes, jmontanor@libertadores.edu.co

RESUMEN

La nueva dinámica introducida por la pandemia del covid-19 aceleró la velocidad en que las organizaciones de todas las industrias demandan la entrega de resultados a sus clientes internos y externos, generación de eficiencias operacionales de una manera ágil y capitalización al máximo de las inversiones. Según (Reichheld) incrementar la retención de los clientes en un 5% puede aumentar las utilidades en un 25% debido a que los clientes se sienten más cómodos trabajando con una compañía que conocen y refieren a otros clientes. El análisis de sentimiento por sus siglas en español (AS) también llamado minería de opinión, es un campo de estudio que analiza las opiniones de las personas, sentimientos, evaluaciones, apreciaciones, actitudes y emociones sobre diferentes entidades (Liu, 2012).

Utilizando técnicas de transfer learning para el análisis de sentimiento de correos electrónicos, se optimiza la cantidad y calidad de los datos utilizados, comparado con métodos tradicionales. El modelo basado BERT permitió llegar a un 79% de exhaustividad, reduciendo el esfuerzo en las tareas de preprocesamiento de datos.

Palabras clave: Análisis de sentimiento, aprendizaje automático, gerencia comercial, retención de clientes, BERT.

ABSTRACT

The new dynamics introduced by the covid-19 pandemic accelerate the speed at which organizations in all industries demand the delivery services to their internal and external customers, the generation of operational efficiencies and capitalization their investments to achieve its strategic objectives. According to (Reichheld) increasing customer retention by 5% can increase profits by 25% because customers feel more comfortable working with a company they know and refer other customers. Sentiment analysis in Spanish (AS) also called opinion mining, is a field of study that analyzes people's opinions, feelings, evaluations, appreciations, attitudes, and emotions about different entities (Liu, 2012).

Applying transfer learning techniques for sentiment analysis of emails, the quantity and quality of the data used is optimized, compared to traditional methods. The BERT-based model allowed reaching 79% completeness, reducing the effort in data pre-processing tasks.

Keywords: Sentiment analysis, machine learning, business management, customer retention, BERT.

INTRODUCCIÓN

Las compañías de servicios en su búsqueda de estar alineadas con los objetivos estratégicos de sus clientes mantienen una constante evolución para aumentar el valor entregado en cada una de sus soluciones. En la carrera de entrega de valor y como consecuencia aumento de sus ingresos, las *estrategias de retención de cliente* juegan un papel decisivo para la fidelización y posicionamiento, según (Emmett C. Murphy, 2002) al incrementar las estrategias de retención de clientes en un 5% se aportará un 10% en la disminución de costos operativos relacionados con el proceso de ventas. El desarrollo de *estrategias de retención de clientes* no es desconocido en el campo empresarial al igual que las herramientas que se utilizan para tal fin, lo que ha comenzado a cambiar es el uso de nuevas maneras sobre las cuales se apoyan dichas estrategias.

Como parte del procesamiento del lenguaje natural (NLP) el AS, permite identificar opiniones relacionadas con un objeto (Fernández, Gómez, Boldrini, & Martínez-Barco, 2011) debido a que la información subjetiva contiene un valor relevante, que permiten en gran medida generar campañas de marketing o mejorar productos en algunos casos. El AS

presenta aplicaciones en diferentes industrias como lo son, opinión política a nivel de las redes sociales como Twitter (Alva & Abraham, 2020), sector de entretenimiento como lo es el cine para evaluación de críticas en sitios como *muchocine* (Domínguez & Fabela) y en la revisión de resúmenes para la predicción de ventas (Yu, Liu, & Huang, 2012) por citar algunos.

(Liu, 2012) propone investigar el AS de tres maneras principales, *a nivel de documento*, en este nivel se busca identificar la opinión total del documento – Positivo o negativo. *A nivel de oración* en este nivel se busca identificar las oraciones de manera individual como – positivo, negativo o neutro -, donde neutro usualmente no tiene opinión. Y *a nivel de aspectos* donde el análisis a nivel de documento y a nivel de la oración no se puede detectar un sentimiento positivo o negativo (e.g. “aunque el servicio no es bueno, me continúa gustando este restaurante”). Dentro de los métodos NLP relacionados con el AS se encuentran (Ruijun, Yuqian, Changjiang, & Ming, 2019) *Tradicionales* donde se incluyen aproximaciones basada en léxico y sin aplicación de redes neuronales como clasificadores, basados en *Deep Learning* presentando el uso de Bag of Words (BoW) y Word Embedding, y por último en *Transfer Learning*, como un método que utiliza la similitud de los datos y distribuciones para aplicar un conocimiento de un dominio en otro. Bajo este último método (Luca, Irigoitia, Perez, & Pons) coinciden que al utilizar una base de conocimiento entrenado y transferirla para realizar otra tarea requiere menos cantidad y calidad de datos.

El presente trabajo entrega una aproximación para el uso de AS en el correo electrónico corporativo buscando validar la posibilidad de minimizar la cantidad y calidad de datos para llegar a una exhaustividad del 80% con métodos de *transfer learning*. Por otro lado, se compara con el rendimiento bajo un enfoque *tradicional* con algoritmo de optimización *Gradiente Descendente Estocástico* (SGD) (Pedregosa, y otros, 2011), el uso de este modelo presenta eficiencia en el costo computacional ya que se pueden obtener valores para lograr un rendimiento adecuado (Regalado, 2020).

Palabras clave: Análisis de sentimiento, transfer learning, procesamiento de lenguaje natural, redes neuronales, gradiente descendente.

METODOLOGÍA

La metodología utilizada para el presente análisis se basó en Cross Industry Standard Process for Data Mining- CRIS-PM, como guía para el desarrollo de la minería de los datos (IBM, 2021) a continuación las diferentes etapas:

Comprensión del negocio

Según la encuesta realizada por Invesp (Saleh, 2011) la probabilidad de cerrar una venta a un cliente existente se encuentra entre el 60% - 70%, mientras que cerrar una venta para un cliente nuevo la probabilidad esperada es del 5% - 20%. Por lo anterior, desarrollar las estrategias adecuadas para la retención de los clientes permitirá a las compañías de servicios, optimizar sus esfuerzos para incrementar utilidades y mantener relaciones a largo plazo con cada uno de sus clientes.

Comprensión de los Datos

El presente estudio se basó en datos anonimizados de 3.450 correos electrónicos y etiquetados de manera manual 340 de ellos en tres polaridades positivo, neutro y negativo. La tabla de datos fue organizada con únicamente dos atributos “review” que corresponde al cuerpo del mensaje y “label” que corresponde a la polaridad asignada al “review”. Se tiene etiquetados 151 mensajes como neutro, 131 mensajes como negativos y 58 mensajes como positivos.

La revisión preliminar sin aplicar ningún preprocesamiento de datos incluyendo toda la historia del mensaje, entendiendo como histórico toda la información desde el mensaje inicial, todas sus respuestas y reenvíos que incluye, presenta un total de 1.120.274 palabras y en promedio 298,66 palabras por mensaje. En este punto se presenta un reto importante frente a la clasificación del mensaje, debido a que al mantener el histórico afecta el sentimiento final o no tenemos certeza sobre qué mensaje está evaluando los modelos de clasificación finalmente.

La figura 1 presenta una muestra de 112 mensajes, los cuales fueron depurados de manera manual excluyendo información histórica (respuestas y reenvíos) contenida en el cuerpo del mensaje, lo cual permitirá una clasificación individual de los mensajes.

Histograma tamaño mensajes sin procesar

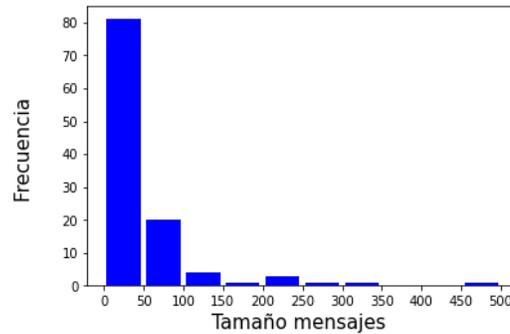


Figura 1 – Tamaño de mensjaes sin procesar

Se observa una reducción significativa en el promedio de palabras por mensaje $50,64$, lo que permitirá tener mejor eficiencia en la ejecución de los modelos manteniendo el objetivo planteado de clasificación del sentimiento del mensaje más reciente.

Como se observa el vocabulario en la nube de palabras *ver anexo b*, en el ambiente corporativo es en gran parte cordial, cuidadoso, repetitivo y practico debido a la naturaleza de los interlocutores, sumando a ello el evitar algún tipo de violación protocolaria o reglamentaria que podría resultar en un compliance interno o externo para la organización.

Nube de palabras para mensajes texto tokenizado



Figura 2

La figura 2, identifican palabras en su gran mayoría solicitando “respuesta” de manera “urgente” a alguna solicitud o servicio.

Nube de palabras para mensajes texto tokenizado



Figura 3

La figura 3, Se identifican palabras relacionadas con la ejecución de un buen trabajo y reconociendo a todo el equipo de trabajo de manera general, evitando dirigir a colaboradores individuales.

Preparación de datos

La preparación de los datos del primer modelo que corresponde a *transfer learning*, utiliza *Bidirectional Encoder Representations from Transformers* (BERT) (Devlin, Cahng, Lee, & Toutanova) el cual obtiene representaciones bidireccionales a partir de redes transformers, proponiendo en lugar crear un modelo que aprenda a interpretar el lenguaje natural de manera general y luego agregar capas que permita entrenarlo para tareas específicas, esto es BERT. Posteriormente se integra con redes neuronales de (Sotaquirá, 2020) y con un corpus pre-entrenado en español de *dccuchile/bert-base-spanish-wwm-uncased* (Cañete, y otros, 2020). Dentro de las principales ventajas que encontramos con BERT es la gran cantidad de datos con la cual fue pre-entrenado. Se entrega al tokenizador de BERT, el texto base del estudio para generar los tokens_Ids que corresponden a los códigos correspondientes del modelo pre-entrenado. Posteriormente se incluyen los tokens de BERT, que permiten al modelo entender cuando inicia el texto *token de clasificación* [CLS] y en el caso de estudio el *token de separación* [SEP] o donde finaliza el texto, lo anterior a través de funciones *enconde_plus*, ver anexo E.

Con lo anterior se tiene el texto codificado para llevarlo a una red neuronal, la cual realiza los procesos de clasificación para los datos de entrenamiento y validación.

La preparación de los mismos datos para el modelo *tradicional*, parte de aplicar el pipeline de la *tabla 1*:

Característica	Librería / modificación / descripción
Stopwords	Nltk.corpus – stopwords
Inclusión de palabras adicionales como stopwords	'buenos','buen','dia','favor','el','gracia','atento','dias','gracias','nuevo','siguiente'
Normalización del texto en minúsculas	Str.lower
Eliminación de acentos	Unicodedata
Corrección de palabras mal escritas dentro del texto	Spellchecker
Tokenización del texto	spaCy - es_core_news_lg
Tf-idf	sklearn.feature_extraction.text

Tabla 1 – Pipeline del preprocesamiento de los datos en el modelo tradicional

Se utiliza finalmente Term frequency – Inverse document frequency (TF-IDF) en este caso, para realizar los procesos de vectorización y después enviar el texto a los procesos de clasificación.

Modelos

Se cargaron 340 correos electrónicos etiquetados con polaridades 0,1 y 2 indicando “negativo”, “neutro” y “positivo” respectivamente, con una separación de datos del 80% para entrenamiento y 20% para validación. A pesar de que la cantidad de datos etiquetados es relativamente pequeño, BERT ha sido pre-entrenado con 31.002 palabras del idioma español.

Para ejecutar el entrenamiento del modelo de BERT y después su carga, se configura la red neuronal de tipo *feed forward* (Pytorch) la cual contine 768 neuronas de entrada que son las características del modelo pre-entrenado y 3 neuronas de salida que son la clasificación final o polaridad. Son ejecutadas 5 epochs durante la fase de entrenamiento con dropouts del 30%, 60% como mecanismo para evitar el *overfitting* (Castañares, 2020) debido a la cantidad de datos con la cual se está trabajando – *ver gráfico 3*.

Se ejecutaron los 2 modelos y analizaron los resultados de la clasificación en cada época para los datos de entrenamiento y validación, junto con la matriz de confusión para la época número 5 debido a que esta última resume los resultados finales para cada set de datos. La figura 6, presenta gráficamente la evolución de cada época.

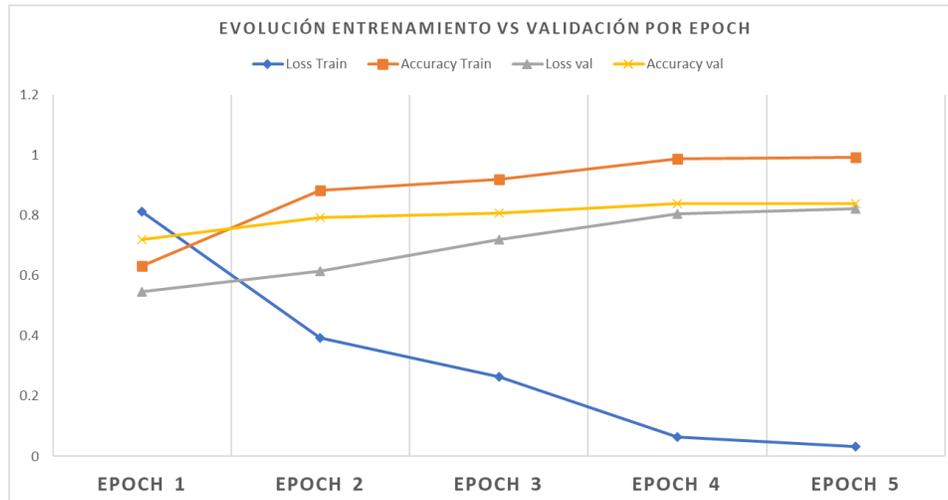


Figura 4 - Presenta la evolución de los **loss train** y **accuracy train** para cada epoch, identificando en la **epoch 5** presencia de overfitting, el cual es reducido con la inclusión de dropout al 90% Ver gráfico con corrección **Anexo C**. Igualmente Se presenta el contraste de **loss val** y **accuracy val** el cual identifica el aumento del error según avanza las epoch, donde para la **epoch 5** se puede observar que el loss y el accuracy para los **datos de validación** se estabilizan.

Por otro lado, el entrenamiento del modelo *Tradicional* incluyó vectorización haciendo uso del algoritmo Term Frequency Invert Document Frequency TF-IDF de (scikit-learn), excluyendo los términos que se encuentren en el 80% de los documentos y términos poco utilizados que se encuentren en menos de 2 documentos. La función de pérdida utilizada fue Support Vector Machines y la penalización L2 la cual permite trabajar contra el overfitting (neptuneblog) al reducir los pesos de características sin eliminarlas. También se trabajan el modo balanceado para ajustar los pesos. Finalmente se genera el reporte de clasificación y la matriz de confusión para evaluar los resultados.

Evaluación

En primera instancia se presentan los resultados obtenidos por los 2 modelos de transfer learning utilizando BERT y redes neuronales feed forward:

Polaridad	Precision	Recall	F1-score
2	0,75	0,67	0,71

1	0,78	0,85	0,82
0	0,83	0,76	0,79
Accuracy	0,79		

Tabla 2 – matrix de confusión – datos de validación – modelo transfer learning

Para el modelo de la tabla 2 presenta un recall sobre los datos de validación del 76% y el overfitting fue reducido sobre los datos de entrenamiento. Al calcular la matriz de confusión gráfico para este último modelo podemos visualizar que únicamente 8 mensajes no fueron clasificados de manera “negativa” sobre los datos de validación. La figura 7, presenta el resultado utilizando la matriz de confusión.

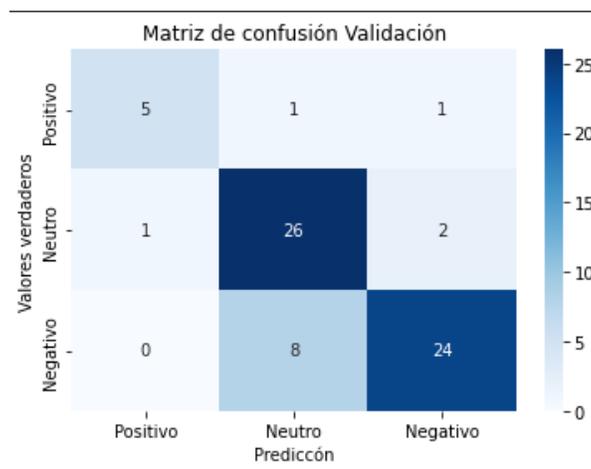


Figura 5 – matriz de confusión

En segunda instancia los resultados obtenidos por el modelo *Tradicional* utilizando descenso por el gradiente estocástico encontramos que el informe de clasificación presentado en el grafico 4, del 72%.

Polaridad	Precision	Recall	F1-score
2	0,69	0,62	0,65
1	0,71	0,76	0,73
0	0,68	0,72	0,70
Accuracy	0,70		

Tabla 3 – matrix de confusión – datos de validación – modelo tradicional

Aunque parece poca la diferencia entre los dos modelos, la matriz de confusión correspondiente en el figura 9 al presentar datos absolutos indica en totalidad 14 mensajes clasificados de polaridad “negativa” de manera errónea.



Figura 6 - Matriz de confusión - modelo tradicional (Ruijun, Yuqian, Changjiang, & Ming, 2019)

DISCUSIÓN DE RESULTADOS

El presente estudio entrega una aproximación para el uso de AS en el correo electrónico corporativo para el diseño de estrategias dirigidas a la retención de clientes y en consecuencia aumentando la utilidad en la entrega de sus soluciones. Los modelos que presentaron mejores resultados fueron bajo el contexto de *transfer learning* – BERT – y redes neuronales disminuyendo las actividades de preprocesamiento y cantidad de datos utilizados según (Luca, Irigoitia, Perez, & Pons) que con modelos *Tradicionales* (Ruijun, Yuqian, Changjiang, & Ming, 2019). Se obtuvo la meta de rendimiento esperado en los dos modelos, aunque en la búsqueda de sentimiento “negativo” para dirigir las estrategias de retención el modelo de transfer learning generó menos errores.

El AS a nivel empresarial presenta retos diferentes frente a su aplicación en el contexto de las redes sociales como Facebook, twitter, intragram, tiendas online etc. Lo anterior debido a que el vocabulario utilizado en redes sociales tiende a ser más agresivo cuando es negativo, muy expresivo cuando es positivo permitiendo realizar una clasificación de opinión más exacta.

En contraste su aplicación al contexto corporativo requiere mayores esfuerzos en el etiquetado de la fuente datos y preprocesamiento de los datos para garantizar el significado

semántico correcto, debido al vocabulario corporativo que debe ser manejado entre las organizaciones y que en algunos casos puede ser ambiguo. También se encuentran retos importantes frente a la complejidad del léxico utilizado como lo son las palabras positivas o negativas que pueden ser ambiguas, palabras positivas o negativas que no expresan ningún sentimiento, detección del sarcasmo y algunas oraciones pueden expresar un sentimiento sin que necesariamente incluyan palabras positivas o negativas.

CONCLUSIONES

El presente estudio logro validar que al utilizar métodos de *transfer learning* junto con redes neuronales *feed forward* reducen esfuerzos en términos de preprocesamiento de los datos para proyectos de NLP, permitiendo obtener niveles de rendimiento aceptables con baja cantidad y calidad de datos. Lo anterior permite optimizar de igual manera los costos y esfuerzos en términos de etiquetado de datos a nivel corporativo. Se alcanzan de una manera ágil niveles de exhaustividad de 79%.

Es importante tener en cuenta que los modelos *tradicionales* pueden presentar un mayor control frente a los parámetros de preprocesamiento de los datos debido a que el pipeline permite una personalización diferente e igualmente con un mayor afinamiento llegar a valores aceptables en términos de exactitud y exhasutividad.

		Validación	precision	recall	f1-score	support		
			2	1.00	0.64	0.78	11	
	3			1	0.92	0.67	0.77	33
				0	0.62	0.96	0.75	24
			accuracy			0.76	68	
			macro avg	0.85	0.75	0.77	68	
		weighted avg	0.83	0.76	0.77	68		
		Entrenamiento	precision	recall	f1-score	support		
3			2	0.87	0.85	0.86	47	
			1	0.92	0.92	0.92	118	
		0	0.88	0.89	0.88	107		
		accuracy			0.89	272		
		macro avg	0.89	0.88	0.89	272		
		weighted avg	0.89	0.89	0.89	272		
4		Validación	precision	recall	f1-score	support		
			2	0.89	0.73	0.80	11	
	4			1	0.74	0.94	0.83	33
				0	0.94	0.67	0.78	24
			accuracy			0.81	68	
			macro avg	0.86	0.78	0.80	68	
		weighted avg	0.83	0.81	0.81	68		
		Entrenamiento	precision	recall	f1-score	support		
4			2	0.98	0.98	0.98	47	
			1	0.96	0.97	0.96	118	
		0	0.96	0.95	0.96	107		
		accuracy			0.96	272		
		macro avg	0.97	0.97	0.97	272		
		weighted avg	0.96	0.96	0.96	272		
4		Validación	precision	recall	f1-score	support		
			2	0.62	0.73	0.67	11	
	4			1	0.75	0.91	0.82	33
				0	1.00	0.62	0.77	24
			accuracy			0.78	68	
			macro avg	0.79	0.75	0.75	68	
		weighted avg	0.82	0.78	0.78	68		

Dropout	Conjunto de datos	Resultado clasificador
---------	-------------------	------------------------

30%	Entrenamiento	<pre> precision recall f1-score support 2 1.00 1.00 1.00 47 1 0.99 1.00 1.00 122 0 1.00 0.99 1.00 103 accuracy macro avg 1.00 1.00 1.00 272 weighted avg 1.00 1.00 1.00 272 </pre>
	Validación	<pre> precision recall f1-score support 2 0.80 0.73 0.76 11 1 0.83 0.86 0.85 29 0 0.86 0.86 0.86 28 accuracy macro avg 0.83 0.82 0.82 68 weighted avg 0.84 0.84 0.84 68 </pre>

Tabla 4 – Resultados en la ejecución epoch 5. Modelo transfer learning

La tabla 4 presenta un overfitting para los datos de entrenamiento lo que debe ser ajustado para buscar que el modelo pueda generalizar de manera correcta.

Dropout	Conjunto de datos	Resultado clasificador
60%	Entrenamiento	<pre> Reporte de Clasificación - ENTRENAMIENTO: precision recall f1-score support 2 0.81 0.66 0.73 44 1 0.80 0.87 0.83 121 0 0.85 0.82 0.83 107 accuracy macro avg 0.82 0.78 0.80 272 weighted avg 0.82 0.82 0.81 272 </pre>
	Validación	<pre> Reporte de Clasificación - VALIDACIÓN: precision recall f1-score support 2 1.00 0.50 0.67 14 1 0.61 1.00 0.76 30 0 1.00 0.50 0.67 24 accuracy macro avg 0.87 0.67 0.70 68 weighted avg 0.83 0.72 0.71 68 </pre>

Tabla 5 - Resultados en la ejecución epoch 2. Modelo transfer learning.

La tabla 5 presenta la corrección de overfitting para los datos de entrenamiento y los datos en validación presentando un accuracy del 72% sobre los datos de validación. En este punto el recall para la detección de los mensajes con polaridad “negativa” nos presenta una detección únicamente del 50% lo que deberá ser ajustado para mejorar los procesos de identificación de clientes que requieren intervención a nivel de estrategias corporativas.

Dropout	Conjunto de datos	Resultado clasificador
---------	-------------------	------------------------

60%	Entrenamiento	<pre> precision recall f1-score support 2 1.00 0.92 0.96 49 1 0.91 0.97 0.94 117 0 0.97 0.93 0.95 106 accuracy 0.95 272 macro avg 0.96 0.94 0.95 272 weighted avg 0.95 0.95 0.95 272 </pre>
	Validación	<pre> precision recall f1-score support 2 0.75 0.67 0.71 9 1 0.78 0.85 0.82 34 0 0.83 0.76 0.79 25 accuracy 0.79 68 macro avg 0.79 0.76 0.77 68 weighted avg 0.79 0.79 0.79 68 </pre>

Tabla 6 – Resultados en la ejecución epoch 4. Modelo transfer learning.

D. Matriz de confusión modelo para dropout del 30%.

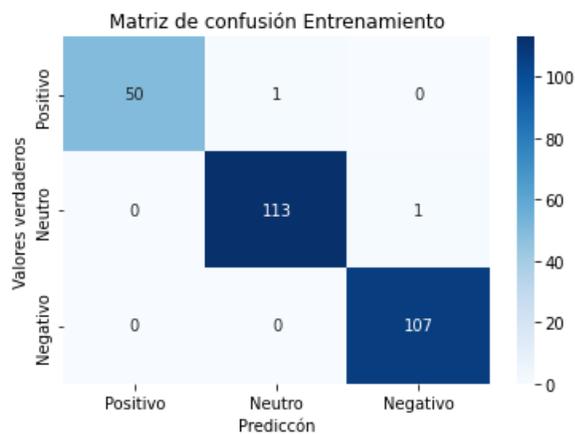


Figura 8 - matriz confusión - calculada con overfitting

E. Inclusión tokens_id y tokens de bert.

```

Frase: me encanta esa pelicula
Tokens: ['me', 'encanta', 'esa', 'pelicula']
Tokens numéricos: [1094, 5494, 1607, 29105]

```

Figura 9 - Ejemplo de tokens generados y tokens numéricos de un modelo pre-entrenado.

```

['[CLS]', 'me', 'encanta', 'esa', 'pelicula', '[SEP]']
tensor([ 4, 1094, 5494, 1607, 29105, 5])
tensor([1, 1, 1, 1, 1, 1])

```

Figura 10 - Muestra de la inclusión de los tokens especiales de BERT

Referencias

- Alva, S., & Abraham, D. (2020). *Universidad internacional de la Rioja - UNIR*. Obtenido de <https://reunir.unir.net/handle/123456789/11364>
- Cañete, Chaperon, J. a., Fuentes, G. a., Ho, R. a., Kang, J.-H. a., & Jorge, H. a. (2020). *Hugging Face*. Obtenido de Spanish Pre-Trained BERT Model and Evaluation Data: <https://huggingface.co/dccuchile/bert-base-spanish-wwm-uncased>
- Castañares, A. A. (Febrero de 2020). *Universidad Politecnica de Madrid*. Obtenido de https://oa.upm.es/57875/1/TFM_ALEJANDRO_ANTONA_CASTANARES.pdf
- ColombiaTech. (2021). *Valora AnalitiK*. Obtenido de https://www.valoraanalitik.com/wp-content/uploads/2022/02/VF_Distrito-Colombia-Report-2021-1.pdf
- Cuerva, L. L. (2020). *Universidad Politecnica de Valencia*. Obtenido de <https://riunet.upv.es/bitstream/handle/10251/171362/Lopez%20-%20Analisis%20de%20sentimiento%20a%20nivel%20aspectual%20en%20el%20dominio%20de%20la%20experiencia%20de%20usuario%20en%20a....pdf?sequence=1>
- Devlin, J., Cahng, M.-W., Lee, K., & Toutanova, K. (s.f.). *arxiv.org*. Obtenido de BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding: <https://arxiv.org/pdf/1810.04805.pdf>
- Domínguez, M. F., & Fabela, J. L. (s.f.). *Universidad Autonoma del Estado de Mexico*. Obtenido de https://rcs.cic.ipn.mx/2020_149_8/Analisis%20de%20Sentimiento%20a%20nivel%20de%20documento%20en%20criticas%20de%20cine%20en%20espanol.pdf
- Emmett C. Murphy, M. A. (2002). *The 10 Critical Elements for Success in Volatile Times*. Prentice Hall Press.
- Fernández, J., Gómez, J. M., Boldrini, E., & Martínez-Barco, P. (Septiembre de 2011). *Redalyc*. Obtenido de Sociedad Española para el Procesamiento del Lenguaje Natural: <https://www.redalyc.org/pdf/5157/515751747019.pdf>
- https://oa.upm.es/57875/1/TFM_ALEJANDRO_ANTONA_CASTANARES.pdf. (Febrero de 2020). *Universidad Politecnica de Madrid*. Obtenido de https://oa.upm.es/57875/1/TFM_ALEJANDRO_ANTONA_CASTANARES.pdf
- IBM. (17 de agosto de 2021). Obtenido de <https://www.ibm.com/docs/es/spss-modeler/SaaS?topic=dm-crisp-help-overview>
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Graeme Hirst.
- Luca, A. d., Irigoitia, M., Perez, G., & Pons, C. (s.f.). *Univesidad Nacional de la Plata*. Obtenido de Uso de la Técnica de Transfer Learning en Machine Learning:

http://sedici.unlp.edu.ar/bitstream/handle/10915/130474/Documento_completo.pdf?sequence=1

neptuneblog. (s.f.). Obtenido de <https://neptune.ai/blog/fighting-overfitting-with-l1-or-l2-regularization>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2011). Scikit-learn: Machine Learning in {P}ython.

Pytorch. (s.f.). Obtenido de <https://pytorch.org/docs/stable/nn.html#linear-layers>

Regalado, N. (2020). Obtenido de Redes neuronales, ¿ donde esta la magia ?:
<http://cms.dm.uba.ar/academico/carreras/licenciatura/tesis/2020/Regalado.pdf>

Reichheld, F. (s.f.). *Bain & Company*. Obtenido de
https://media.bain.com/Images/BB_Prescription_cutting_costs.pdf

Ruijun, L., Yuqian, S., Changjiang, J., & Ming, J. (2019). *IEEE.org*. Obtenido de A Survey of Sentiment Analysis Based on Transfer Learning:
<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8746210>

Saleh, k. (Noviembre de 2011). *Invesp*. Obtenido de www.invespro.com:
<https://www.invespro.com/blog/customer-acquisition-retention/>

scikit-learn. (s.f.). *scikit-learn.org*. Obtenido de https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

Sotaquirá, M. (2020). *Analisis de sentimiento con BERT*. Obtenido de Codificando Bits:
https://github.com/codificandobits/Analisis_de_sentimientos_con_BERT

spaCy. (s.f.). *Spanish pipeline optimized for CPU*. Obtenido de <https://spacy.io/models/es>

Yu, X., Liu, Y., & Huang, X. (2012). *Research Gate*. Obtenido de
https://www.researchgate.net/publication/220073572_Mining_Online_Reviews_for_Predicting_Sales_Performance_A_Case_Study_in_the_Movie_Domain