Comparación de modelos de redes neuronales para la clasificación de emociones en textos

Cesar Peñaranda Chaves[1], Stephanie Vega López[2], Joseph Rivera Noguera[3]

cesar.penaranda@ucr.ac.cr, stepanie.vegalopez@ucr.ac.cr, joseph.rivera@ucr.ac.cr

RESUMEN

El presente estudio tuvo como objetivo comparar el desempeño de tres modelos de redes neuronales —Perceptrón Multicapa (MLP), Red Neuronal Convolucional (CNN) y Red Neuronal Recurrente Bidireccional (BiRNN)— para la clasificación de emociones en textos, utilizando el conjunto de datos "Emotion Dataset". Este corpus contiene más de 400 mil registros categorizados en seis emociones: alegría, tristeza, enojo, miedo, amor y sorpresa.

Los modelos fueron entrenados y evaluados mediante métricas de precisión y pérdida en los conjuntos de entrenamiento, validación y prueba. Los resultados indicaron que la Red Neuronal Recurrente (RNN) fue la más efectiva, alcanzando una precisión del 94,27% en el conjunto de prueba, superando a la CNN (93,56%) y al MLP (89,81%). La RNN mostró un rendimiento especialmente robusto en emociones como alegría y tristeza, aunque presentó limitaciones en la clasificación de emociones menos representadas como amor y sorpresa, reflejado en menores valores de recall y F1-score para estas categorías.

Adicionalmente, se utilizó la técnica LIME para interpretar las predicciones del modelo, lo que permitió identificar las palabras clave que influenciaron cada clasificación, otorgando mayor transparencia al proceso de toma de decisiones del modelo.

El estudio reafirma la superioridad de las RNN para tareas de clasificación emocional en textos, dada su capacidad para capturar la secuencia y el contexto del lenguaje natural. Se sugiere que futuras investigaciones exploren otras técnicas de vectorización y modelos como CARER, además de utilizar conjuntos de datos balanceados para mejorar la clasificación en emociones menos frecuentes.

PALABRAS CLAVE: Embedding, Precisión, Validación, LIME

[1] Graduado, Bachillerato en Estadística, Universidad de Costa Rica.
[2] Graduada, Bachillerato en Estadística, Universidad de Costa Rica.
[3] Estudiante, Bachillerato en Estadística, Universidad de Costa Rica.

INTRODUCCIÓN

En los últimos años, la ciencia de datos ha cobrado gran relevancia en diversos campos, destacándose particularmente en el procesamiento de texto. Este auge ha sido impulsado por la necesidad de clasificar y analizar grandes volúmenes de datos textuales provenientes de diversas fuentes, como correos electrónicos, comentarios en plataformas en línea y hasta redes sociales. En este contexto, una de las áreas más importantes de investigación es la clasificación de emociones.

El clasificar las emociones impresas de manera textual juega un papel crucial en aplicaciones como en atención al cliente, donde el análisis de los sentimientos de los usuarios permite ofrecer respuestas más personalizadas y efectivas tal como lo menciona Figueroa Sacoto (2021) el cual indica que mediante el uso de *chatbot* entrenados mediante una red recurrente y el procesamiento de lenguaje natural estos sistemas son capaces de analizar y clasificar emociones en tiempo real a partir de las conversaciones.

Además, de acuerdo con Vázquez Arias (2021), el uso de herramientas inteligentes para el análisis de la perspectiva de los consumidores, especialmente en redes sociales, permite a las empresas comprender mejor las necesidades y expectativas de sus clientes. El análisis de opiniones sobre productos o servicios, mediante técnicas de procesamiento de lenguaje natural y clasificación de emociones, ayuda a identificar patrones en las valoraciones y comentarios, lo que facilita la toma de decisiones estratégicas para mejorar la oferta comercial y la atención al cliente, así mismo permite personalizar las sugerencias de contenido según el estado emocional percibido en sus interacciones.

En áreas como la salud, existen estudios como el realizado por Kabir et al. (2022) el cual demuestra cómo el procesamiento de publicaciones sobre salud mental en redes sociales puede utilizarse para detectar la severidad de la depresión en los usuarios, aplicando técnicas avanzadas de procesamiento de lenguaje natural. Este enfoque no solo ayuda a los profesionales de la salud a monitorear el bienestar emocional de los pacientes de manera remota, sino que también permite una intervención temprana sin la necesidad de contacto físico constante.

De esta manera, las redes neuronales artificiales se han posicionado como una herramienta eficaz para comprender, procesar y reaccionar ante las emociones humanas en el entorno digital, dado que pueden aprender patrones complejos de los datos textuales y clasificarlos en diversas categorías emocionales existentes.

En la bibliografía consultada han explorado distintas arquitecturas de redes neuronales para abordar tareas específicas relacionadas con el análisis de datos textuales. Por ejemplo, Figueroa implementa redes neuronales de memoria a largo plazo (LSTM), una variante de las redes neuronales recurrentes (RNN), en aplicaciones de procesamiento de lenguaje natural (PLN), como los *chatbots*. Estas redes son especialmente efectivas debido a su capacidad para "recordar" información relevante a largo plazo y ajustarse a nuevas entradas sin perder el contexto de las interacciones previas, lo que las convierte en una herramienta valiosa para mantener coherencia y relevancia en las conversaciones. Por su parte, Vázquez Arias utiliza redes RNN en el desarrollo de una herramienta inteligente para analizar las perspectivas de los consumidores, destacando su eficacia al captar patrones secuenciales en datos textuales y mejorar la comprensión de las emociones subyacentes.

En ambos casos, las RNN demostraron un rendimiento superior debido a su capacidad para manejar datos secuenciales y contextuales, lo que resalta su aplicabilidad en tareas relacionadas con la clasificación de emociones en texto. A partir de este antecedente, el presente trabajo tiene como objetivo comparar el desempeño de tres modelos: un perceptrón multicapa (MLP), una red neuronal convolucional (CNN) y una red neuronal recurrente bidireccional (BiRNN). Esta comparación busca determinar cuál de estas arquitecturas es más adecuada para la clasificación de emociones en texto, aportando una perspectiva clara sobre su efectividad y contribuyendo a la selección del modelo más adecuado para futuras aplicaciones prácticas en este ámbito.

METODOLOGÍA

El presente estudio utilizó el conjunto de datos “Emotion Dataset” (Saravia et al., 2018), una colección de textos preprocesados diseñada para la clasificación de emociones en el ámbito del procesamiento de lenguaje natural (PLN). Este conjunto consta de 416,809 registros distribuidos en seis categorías emocionales: alegría (141,067 registros), tristeza (121,187), enojo (57,317), miedo (47,712), amor (34,554) y sorpresa (14,972). Los datos fueron preprocesados mediante un algoritmo semi supervisado basado en grafos, enriquecido con *word embeddings,* que captura las sutilezas lingüísticas inherentes a las emociones expresadas en los textos.

Inicialmente, se dividió el conjunto en tres subconjuntos: entrenamiento (80%), validación (10%) y prueba (10%), garantizando un proceso de evaluación justo para los modelos. Luego, se analizó la frecuencia de palabras por categoría emocional, generando un conteo de términos más comunes y visualizaciones en gráficos de barras. Posteriormente, se aplicó la técnica de *tokenización* para convertir las etiquetas emocionales en formato numérico, ajustando el codificador sobre el conjunto de entrenamiento y aplicándolo a los conjuntos de

validación y prueba. Para garantizar la uniformidad, se realizó un proceso de *padding* con una longitud máxima de 100 palabras por secuencia. Con los datos preparados, se procedió a entrenar los modelos propuestos.

El primer modelo consiste en una red MLP con una capa de *embedding* que convierte las palabras en vectores densos en un espacio de dimensión 128. Una capa de *pooling* promedia los vectores de características, reduciendo la dimensionalidad y generando un único vector representativo de toda la secuencia. La arquitectura incluye una capa densa intermedia con 56 neuronas y activación *ReLU*, seguida de una capa de salida con 6 neuronas y activación *softmax* para clasificar las emociones. El modelo se compiló con el optimizador Adam y la función de pérdida de entropía cruzada categórica. Para evitar el sobreajuste, se implementó la técnica de *Early Stopping,* configurando un máximo de 50 épocas y deteniendo el entrenamiento si la pérdida de validación no mejoraba en 5 épocas consecutivas.

El segundo modelo utiliza una arquitectura CNN con una capa de *embedding* similar al modelo MLP. Le siguen dos capas convolucionales, cada una con 128 filtros y un tamaño de núcleo de 5, utilizando activación *ReLU* para extraer patrones locales del texto. Cada capa convolucional está seguida por una capa de *max pooling* para reducir la dimensionalidad. Este modelo también incluye una capa densa con 128 neuronas y activación *ReLU,* culminando con una capa de salida idéntica al modelo anterior. Se mantuvieron las mismas configuraciones de optimización, función de pérdida y *Early Stopping.*

El tercer modelo emplea una arquitectura basada en redes neuronales recurrentes (RNN) bidireccionales con Long Short-Term Memory (LSTM). Comienza con una capa de embedding y una capa Bidirectional que envuelve una LSTM con 64 unidades, permitiendo que la información fluya en ambas direcciones para capturar dependencias a largo plazo en el texto. Se agregó una capa de normalización de *batch* para mejorar la estabilidad del entrenamiento. Luego, una segunda capa *Bidirectional* LSTM con 32 unidades genera la entrada para una capa densa de 32 neuronas con activación *ReLU.* La salida final consiste en 6 neuronas con activación *softmax,* configurada de forma similar a los modelos anteriores.

Los modelos fueron comparados utilizando métricas de pérdida y precisión en los conjuntos de entrenamiento y validación. Gráficos de la evolución de estas métricas a lo largo de los *epocs* facilitaron una comparación directa, permitiendo identificar posibles sobre ajustes o diferencias en el rendimiento. Adicionalmente, se generaron visualizaciones con subgráficos para comparar simultáneamente la pérdida y precisión de los tres modelos.

Finalmente, se evaluó el mejor modelo en el conjunto de prueba, transformando las predicciones numéricas a sus clases originales para una interpretación directa. Además, se utilizó la técnica LIME (Local Interpretable Model-agnostic Explanations) para interpretar las

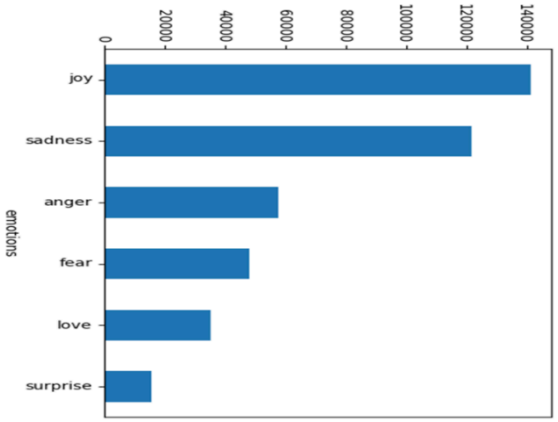
predicciones del modelo en textos seleccionados. LIME destacó las palabras clave que contribuyeron a la clasificación, proporcionando mayor transparencia en el comportamiento del modelo y facilitando su análisis.

RESULTADOS

El análisis descriptivo reveló cómo se presenta en la figura 1 que la emoción más frecuente en la base de datos es alegría, representando un 33,84% de los registros totales, mientras que la menos frecuente es la emoción de sorpresa, con un 3,59% de los registros.

Figura 1

Frecuencia de oraciones presentes por emoción



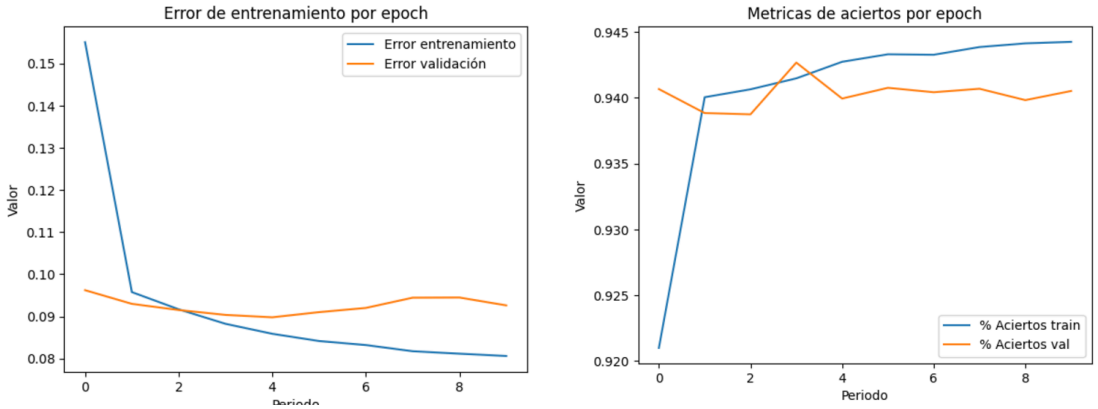
En relación con la estructura textual, el análisis exploratorio identificó patrones claros en la construcción de oraciones vinculadas a emociones. Las diez palabras más representativas para cada emoción destacaron consistencias importantes en el uso del lenguaje, independientemente de la emoción asociada. Las figuras 2 y 3 presentes en Anexos muestran estos patrones, mientras que la Figura 4 en anexos confirma la predominancia de ciertas palabras mediante nubes de palabras organizadas por emoción.

Asimismo, en cuanto al desempeño de los modelos, se encontró que el modelo Multi-Layer Perceptron (MLP) alcanzó una precisión del 90% y una pérdida de 0,1844 en el conjunto de entrenamiento, mientras que en el conjunto de prueba logró una precisión del 89,81%. Por su parte, la Red Neuronal Convolucional (CNN) presentó un rendimiento superior con una precisión del 93,52% y una pérdida de 0,1063 en el entrenamiento, logrando un 93,56% de precisión en el conjunto de prueba.

Finalmente, la Red Neuronal Recurrente (RNN) como se observa en la figura 5 demostró ser el modelo más eficaz, con un 94,30% de precisión y una pérdida de 0,0882 en el entrenamiento, alcanzando un 94,27% de precisión en los datos de prueba.

Figura 5

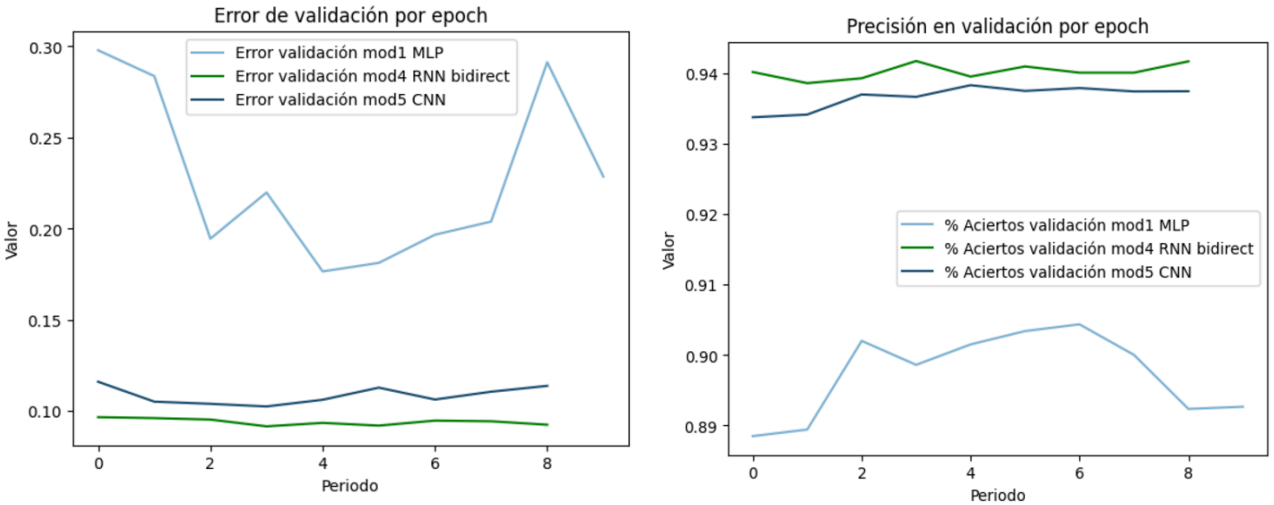
Métricas modelo RNN



Al realizar la comparación objetivo de los tres modelos ante la tarea de clasificar las emociones presentes en los textos utilizando las métricas de rendimiento tal como se observa en la figura 6 se evidencia que, de los tres modelos realizados, el mejor modelo para clasificar las emociones en los textos es la Red Neuronal Recurrente (RNN) tras tener un error de validación muy bajo menor a 0.10 y una precisión superior a 0.94.

Figura 6

Gráficos de comparación de las métricas de rendimiento para los modelos

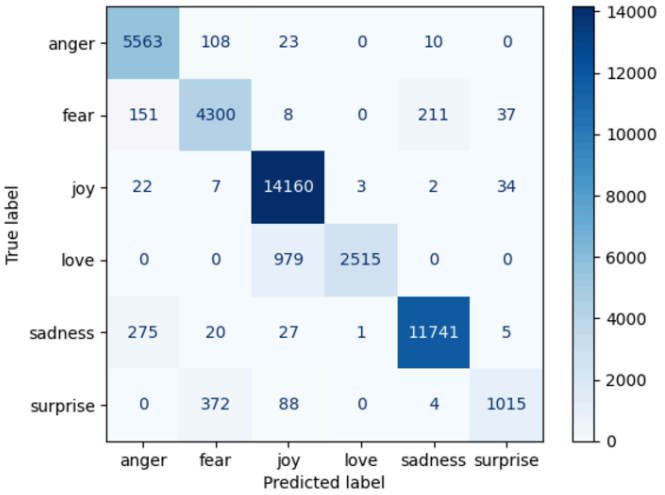


En cuanto al modelo seleccionado con mejor rendimiento (RNN), se puede observar que en la matriz de confusión presente en la figura 8 la mayoría de las emociones es clasificada correctamente, aunque existe cierta confusión a la hora de catalogar entre amor y alegría

donde el modelo no es capaz de clasificar adecuadamente los escritos de amor y clasifica un número alto (979) como textos de alegría.

Figura 8

Matriz de confusión del modelo RNN



De igual manera al observar las métricas de rendimiento según las emociones presentes en el cuadro 1, se destaca que donde se posee una muestra de menor tamaño, se presentaron también menores en Recall y F-Score, esto sucedió de igual forma para los demás modelos tal como se presenta en el cuadro 2 y 3 en anexos, no obstante, aunque estos valores disminuyen con el tamaño de la muestra, los mismos se encuentran en rangos de valores adecuados, es decir, las métricas muestran un buen desempeño y consistencia, las cuales en su mayoría se encuentran mayor a 0,9 a excepción de Amor y Sorpresa.

Cuadro 1

Métricas de rendimiento del modelo RNN según emociones

Emociones	Precisión	Recall	F1-score	Muestra
Enojo	0,93	0,98	0,95	5704
Miedo	0,89	0,91	0,90	4707
Alegría	0,93	1,00	0,96	14228
Amor	1,00	0,72	0,84	3494
Tristeza	0,98	0,97	0,98	12069
Sorpresa	0,93	0,69	0,79	1479

En cuanto a las clasificaciones de los textos, se puede observar en el cuadro 4 algunos ejemplos de clasificación realizados por el modelo RNN, en los casos que están mal catalogados los textos se puede adjudicar que el modelo no posee un análisis del contexto, por lo cual clasifica erróneamente, como es el caso de la observación 17 “hated” la cual es una palabra

característica de enojo, pero en el contexto se refiere a una sensación de autorrechazo. Con respecto a las observaciones 24 y 25, aunque presentan palabras ampliamente relacionadas con el valor real el modelo no es capaz de acertar adecuadamente, esto puede estar relacionado por ambigüedad de las oraciones.

Cuadro 4

Textos según observación por clasificación

Observación	Texto	Real	Predicha
38	i hope youll look at this often especially when were fighting or youre feeling insecure	Fear	Fear
39	i can bear any severe pain but when i am down with common cold i simply feel irritated and bugged down	Anger	Anger
41	ill save that accountable thing for another day when im feeling like exposing my tender bits	Love	Love
17	i cannot but feel dahl would have **hated**	Sadness	Anger
24	i feel like my **beloved** rpg s are falling under siege of these trading card games	Love	Joy
25	**i feel the love** a project for sweet olivia a href http alittlebitofdetail	Love	Joy

Finalmente, a modo de ejemplo mediante el uso de Local Interpretable Model-agnostic Explanations (LIME por sus siglas en inglés) como se observa en la figura 11 en el ejemplo que el modelo identifica que las palabras "feels" y "strong" son determinantes para clasificar la emoción principal del texto como "joy", con una probabilidad del 100%.

Figura 11

Ejemplo de clasificación mediante LIME

CONCLUSIONES

Los modelos planteados brindaron resultados realmente convenientes, además, se resalta lo que la teoría menciona, los modelos de Redes Neurales Recurrentes (RNN) se adecúa de mejor manera para el análisis de entrenamientos en textos, el cual se contrasta en lo mencionado en la introducción, las RNN demostraron un rendimiento superior debido a su capacidad para manejar datos secuenciales y contextuales, lo que resalta su aplicabilidad en

tareas relacionadas con la clasificación de emociones en texto, asimismo, quedó en evidencia que los modelos Multi-Layer Perceptron (MLP) son menos eficientes para este tipo de estudio, esto mediante las métricas de rendimiento obtenidas.

Por otra parte, al comparar con otras investigaciones con el mismo conjunto de datos, como el de Saravia, E., Liu, H.-C. T., Huang, Y.-H., Wu, J., & Chen, Y.-S. (2018) se encontraron discrepancias en las métricas de rendimiento presentadas, en mencionado proyecto el modelo con los mejores valores se dio mediante un modelo CARER *our enriched patterns*, el cual logró una precisión de un 0,81, seguido de un modelo CNN *word vectors* con un accuracy de 0,68. En dicho estudio RNN *word2ve*c alcanzó una precisión de 0,53. Sin embargo, con análisis más recientes como el presentado por Katoch99 (2021) a través de Github, logró mediante un modelo RNN una precisión de 98,50% y una pérdida de 0,0538 en entrenamiento, similar a lo encontrado en el presente estudio.

Dado a los problemas de clasificación, aunque son pocos, se recomienda para futuros estudios la implementación de otros tipos de vectorizaciones con el fin de ver si mejora la clasificación en aquellos textos donde la ambigüedad de los textos no permite una adecuada asignación en las emociones, además, de implementar modelos como el de CARER para comparar y ver el comportamiento respecto a los modelos presentados.

Por último, para mejorar las métricas de precisión en el modelo seleccionado en aquellas cuyo valor son menores se recomienda la utilización de un conjunto de emociones de forma balanceada con el fin que exista un entrenamiento más conciso entre todas las categorías emocionales.

REFERENCIAS BIBLIOGRÁFICAS

Figueroa Sacoto, S. S. (2021). *Diseño y desarrollo de un chatbot usando redes neuronales recurrentes y procesamiento de lenguaje natural para tiendas virtuales en comercio electrónico* (Bachelor's thesis).

Kabir, M. K., Islam, M., Kabir, A. N. B., Haque, A., & Rhaman, M. K. (2022). Detection of depression severity using Bengali social media posts on mental health: study using natural language processing techniques. *JMIR Formative Research*, *6*(9), e36118

Katoch99 (2021). Twitter-Emotion-Recognition. https://github.com/katoch99/Twitter-Emotion-Recognition

Saravia, E., Liu, H.-C. T., Huang, Y.-H., Wu, J., & Chen, Y.-S. (2018). *CARER: Contextualized affect representations for emotion recognition*. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 3687–3697). Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1404

Saravia, E., Liu, H.-C. T., Huang, Y.-H., Wu, J., & Chen, Y.-S. (2018). *CARER: Contextualized affect representations for emotion recognition*. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* https://www.researchgate.net/publication/334116313_CARER_Contextualized_Affect_Representations_for_Emotion_Recognition.

Vázquez Arias, E. J. (2021). *Diseño y desarrollo de una herramienta inteligente para el análisis de la perspectiva de los consumidores sobre productos específicos de empresas dentro de sus redes sociales* (Bachelor's thesis)

ANEXOS

Figura 2

Top 10 de palabras más frecuentes en oraciones de enojo y miedo

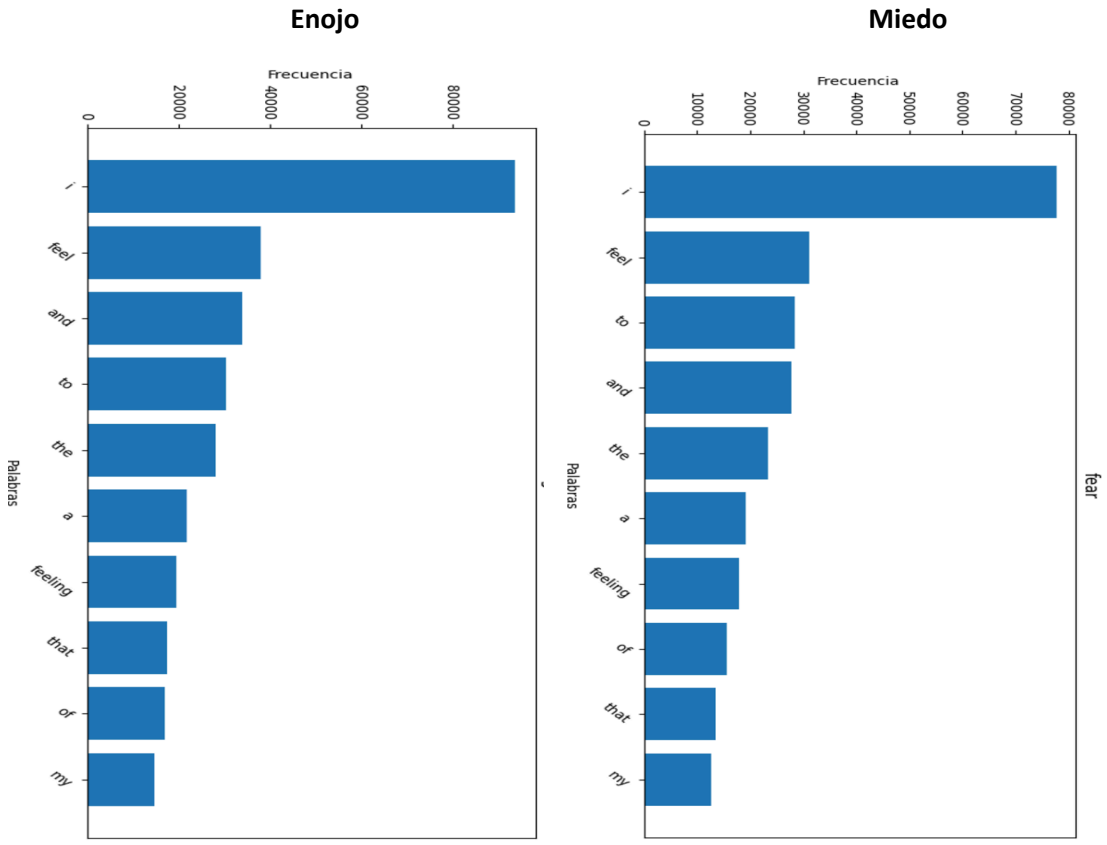


Figura 3

Top 10 de palabras más frecuentes en oraciones de alegría, amor, tristeza y sorpresa

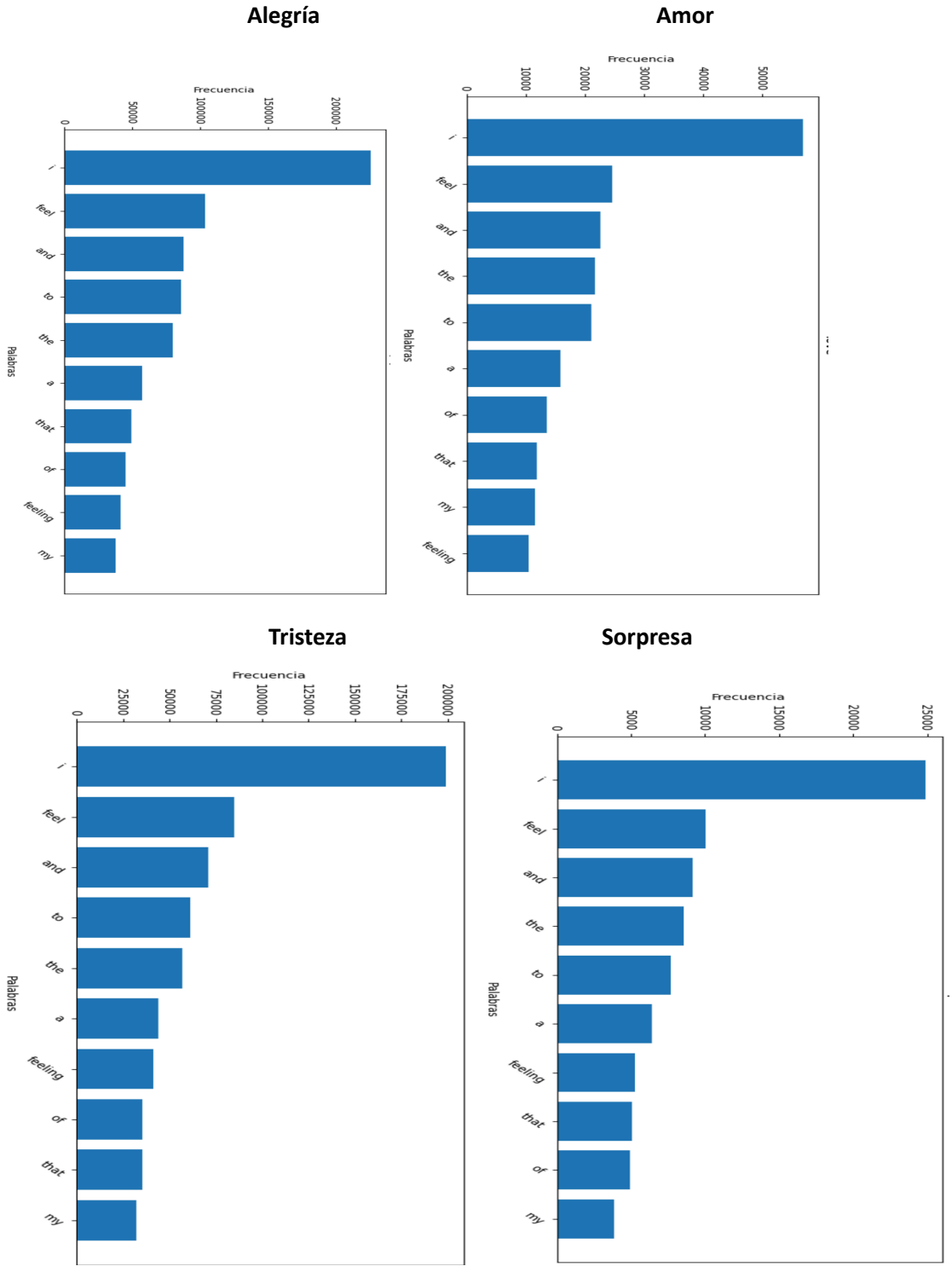
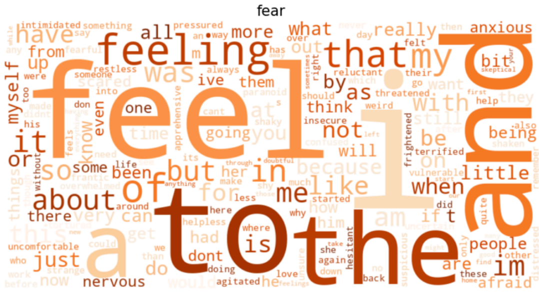


Figura 4

Matriz de nubes de palabras según emociones

fear

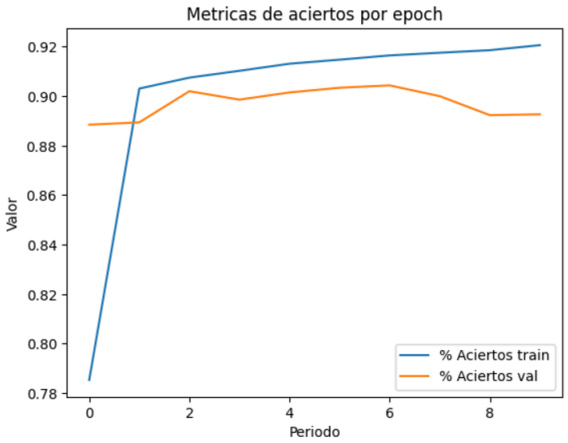


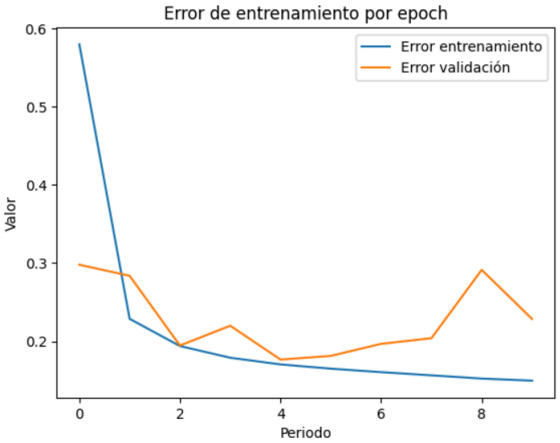
surprise

Figura 7

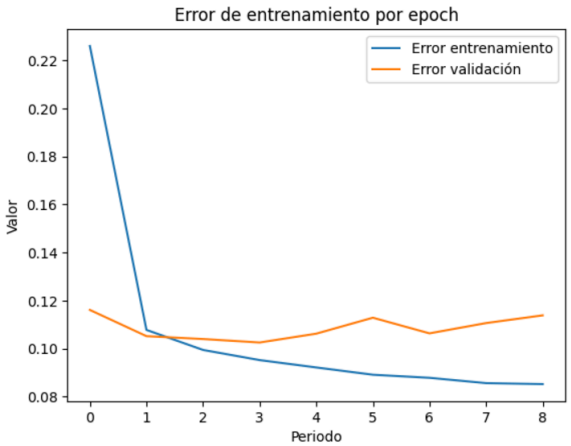
Métricas de rendimiento por modeles

MLP





CNN



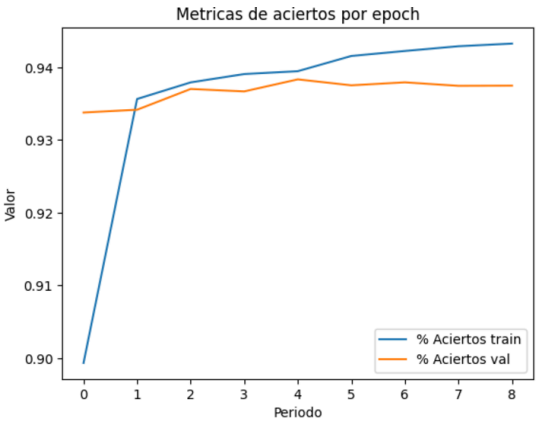


Figura 8

Matriz de métricas de rendimiento por modelos

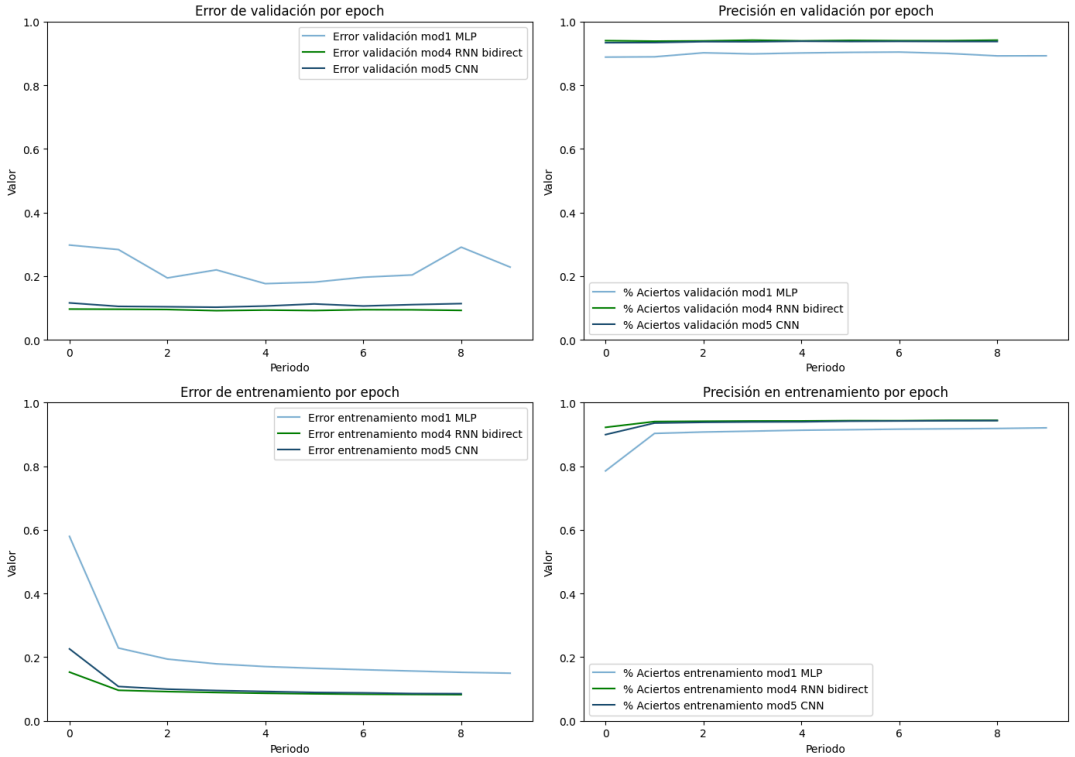


Figura 9

Matriz de confusión modelo MLP

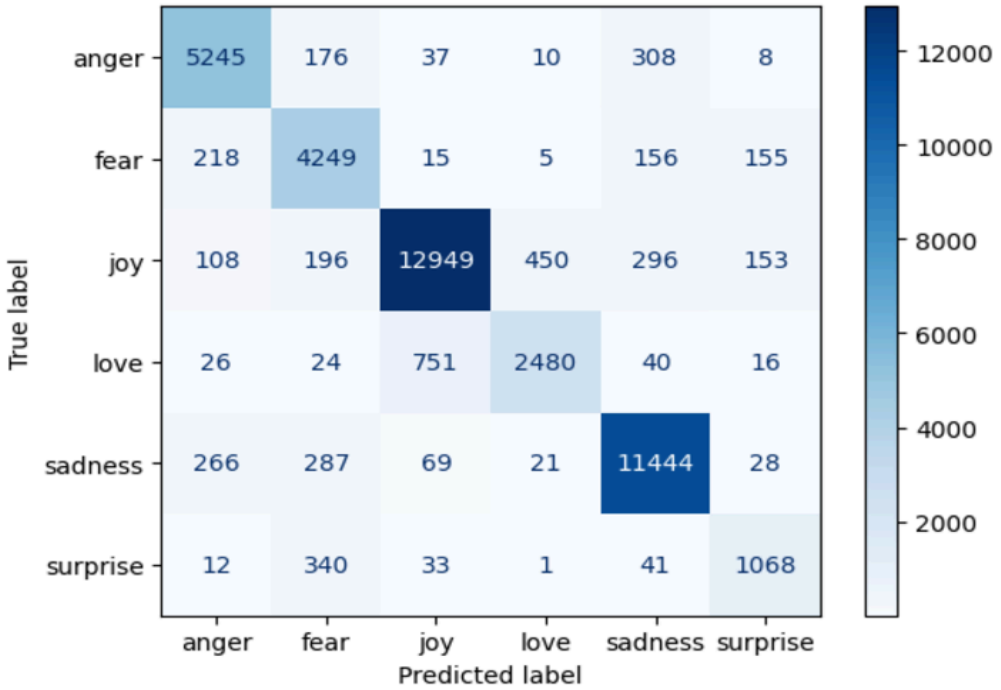
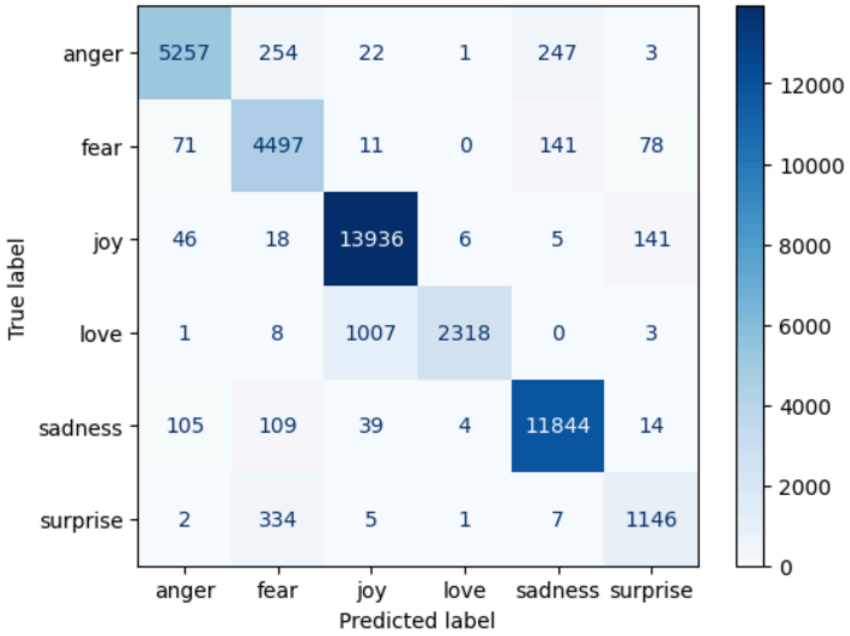


Figura 10

Matriz de confusión modelo CNN



Cuadro 2

Métricas de rendimiento del modelo MLP según emociones

Emociones	Precisión	Recall	F1-score	Muestra
Enojo	0,89	0.91	0,90	5784
Miedo	0,81	0,89	0,84	4798
Alegría	0,93	0,91	0,92	14152
Amor	0,84	0,74	0,79	3337
Tristeza	0,93	0,94	0,94	12115
Sorpresa	0,75	0,71	0,73	1495

Cuadro 3

Métricas de rendimiento del modelo CNN según emociones

Emociones	Precisión	Recall	F1-score	Muestra
Enojo	0,96	0,91	0,93	5784
Miedo	0,86	0,94	0,90	4798
Alegría	0,93	0,98	0,96	14152
Amor	0,99	0,69	0,82	3337
Tristeza	0,97	0,98	0,97	12115
Sorpresa	0,83	0,77	0,80	1495