

Universidad del Valle de Guatemala  
Data Science  
Departamento de Ciencias de la Computación  
Ciclo II, 2024  
Fecha de entrega: 28 oct 2024

Proyecto 2  
**Resultados Iniciales**

Guillermo Alfonso Furlán Estrada	Carné 20713
Roberto Francisco Ríos Morales	Carné 20979
Diego Andrés Alonzo Medinilla	Carné 20172
Diego Alejandro Perdomo Sagastume	Carné 20204

Guatemala, 28 de octubre de 2024

# Investigación de algoritmos

## RNN (Redes Neuronales Recurrentes)

Las RNN es un modelo de aprendizaje profundo diseñado para procesar y convertir una secuencia de datos en una salida secuencial específica. De manera que sus entradas pueden ser palabras, oraciones, datos de series de manera que se analizan cómo es que funcionan en virtud de reglas semánticas y sintácticas.

La unidad básica para esta arquitectura son las neuronas que son nodos de procesamiento, así mismo, está compuesta por una arquitectura de capas de entrada, salida y ocultas de manera que la capa de entrada recibe la información, las capas ocultas tratan los datos secuencialmente paso a paso aunque poseen un bucle (recurrencia) para poder recordar y utilizar entradas pasadas para hacer predicciones futuras de manera que modelan cierta *memoria a corto plazo*; finalmente la capa de salida entrega un resultado.

Es precisamente el hecho de que tengan esta memoria a corto plazo y la capacidad de recordar secuencias lo que les permite ser útiles en reconocimiento de voz, traducción automática, y en este caso particular analizar los títulos y los resúmenes de los artículos biomédicos para identificar entidades y predecir las relaciones entre las entidades como asociaciones, correlaciones o interacciones. Además, el hecho de que los artículos científicos poseen una estructura secuencial en las palabras y frases bajo un contexto dependiente las RNN pueden modelar estas dependencias con una mayor precisión que otros algoritmos.

## LSTM (Redes de memoria a corto y largo plazo)

Las LSTM es un derivado de las RNN ya que permite al modelo la capacidad de almacenar más memoria para adaptarse a una línea del tiempo más larga. A diferencia de una RNN esta puede utilizar varias secuencias anteriores para mejorar su predicción ya que agrega una celda de capa oculta que está controlada por una puerta de salida, olvido y de entrada a nueva información de manera que a pesar que olvidará información nueva mantendrá información de tandas pasadas.

En pocas palabras las LSTM presentan una solución a un problema común que poseen las RNN que es llamado *problema del gradiente que desaparece* que consiste en que en el largo plazo las RNN son incapaces de aprender y capturar dependencias a largo plazo, poseen una memoria a corto plazo y no a largo plazo. El hecho de que las LSTM tienen la capacidad de recordar a largo plazo permite que tengan un contexto global de una secuencia y no a corto plazo lo que permite que se capturen relaciones y dependencias más precisas concluyendo en predicciones más realistas y exactas. Sin embargo, un problema que poseen las LSTM es el hecho de que están más propensas a ser sobreentrenadas además de que poseen una mayor cantidad de variables para su entrenamiento lo que las hace sustancialmente más complejas.

Para este contexto, las LSTM gracias a sus relaciones de largo plazo permite identificar la posición y conexión entre las entidades en los diferentes artículos, asociándolas a partir de contextos relacionados o similares como puede ser química, biotecnología, biomédica, etcétera.

## BERT (Bidirectional Encoder Representation from Transformers)

El modelo de BERT utiliza como base la arquitectura de transformadores lo que lo hace más eficiente que un modelo de RNN o LSTM. La arquitectura de BERT al basarse en transformadores utiliza el

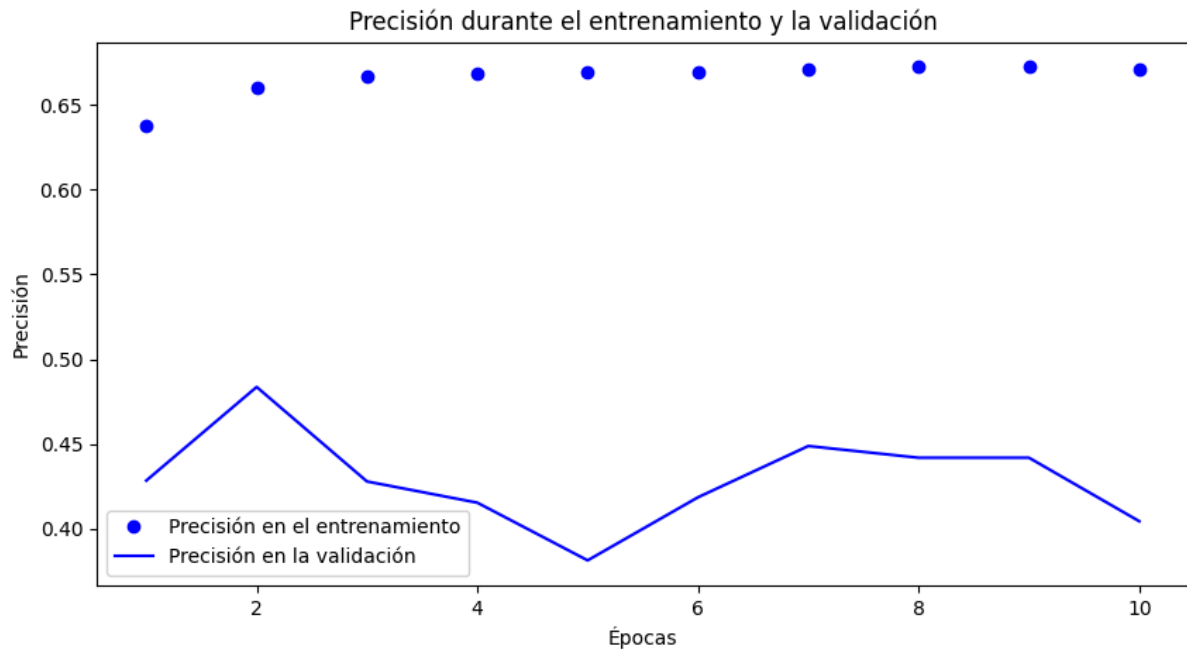
contexto total y no una memoria a corto plazo como las RNN, de manera que utiliza un codificador y estrategias de auto atención para poder extraer lo más importante del input esto es lo que permite establecer las relaciones contextuales existentes entre las distintas palabras de la entrada.

Este modelo puede modelar con precisión las relaciones de las entidades así como cualquier tipo de conexión basado en contexto, esto incluso se ha visto en un proyecto llamado BioBert donde se minaban artículos biomédicos para mejorar la precisión de tareas.

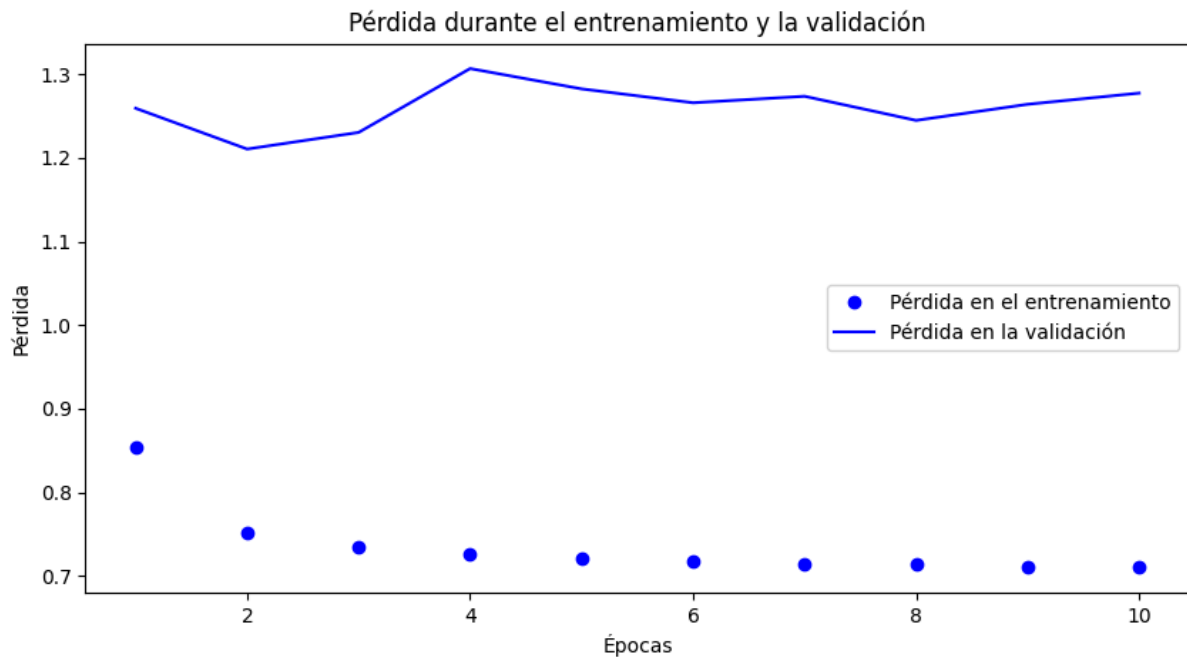
# Discusión y Resultados

## LSTM

Gráfica 1. Pérdida a través de las epochs del entrenamiento y la validación.

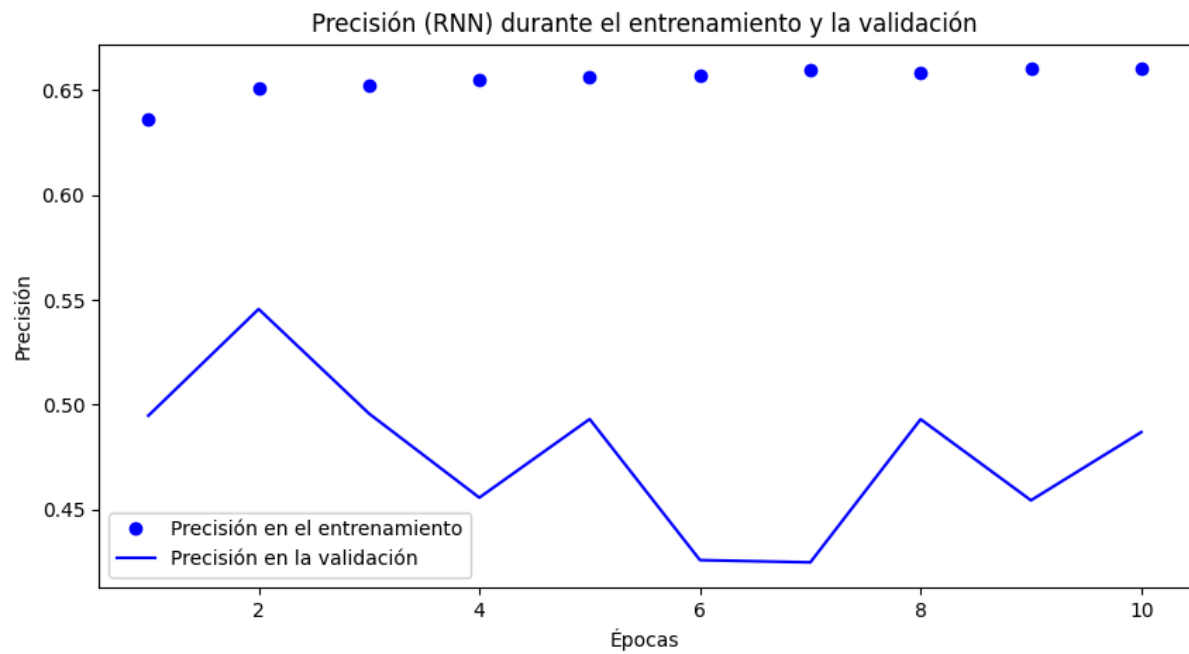


Gráfica 2. Precisión a través de las epochs del entrenamiento y la validación.

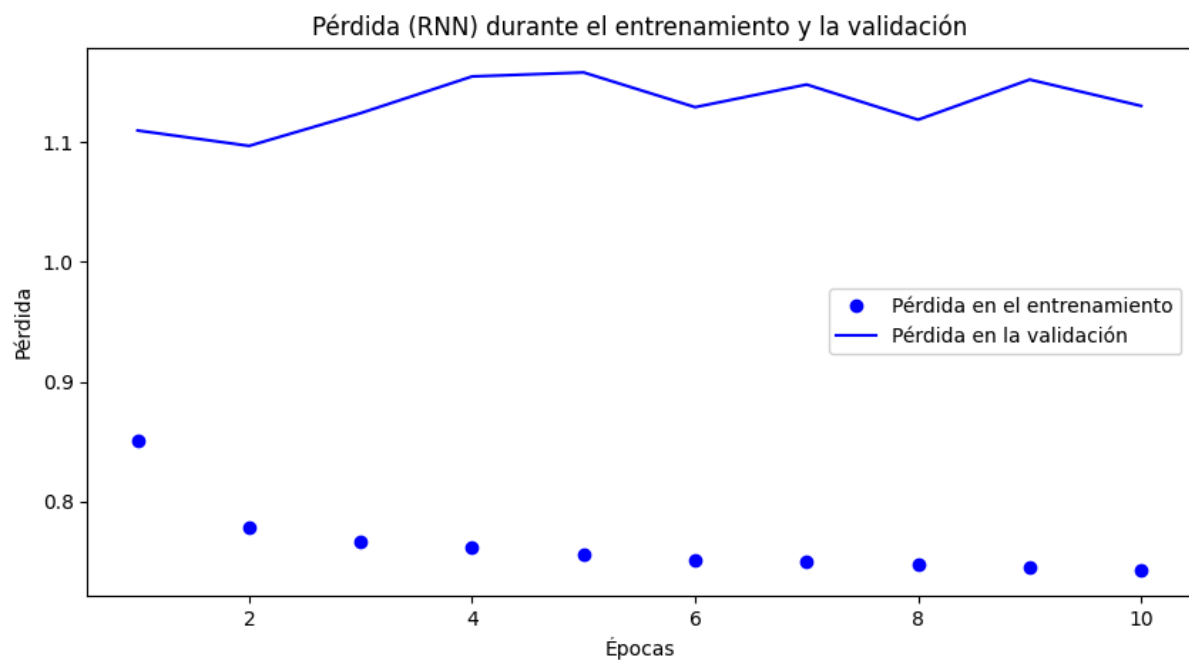


## RNN

Gráfica 3. Precisión a través de las epochs del entrenamiento y la validación.

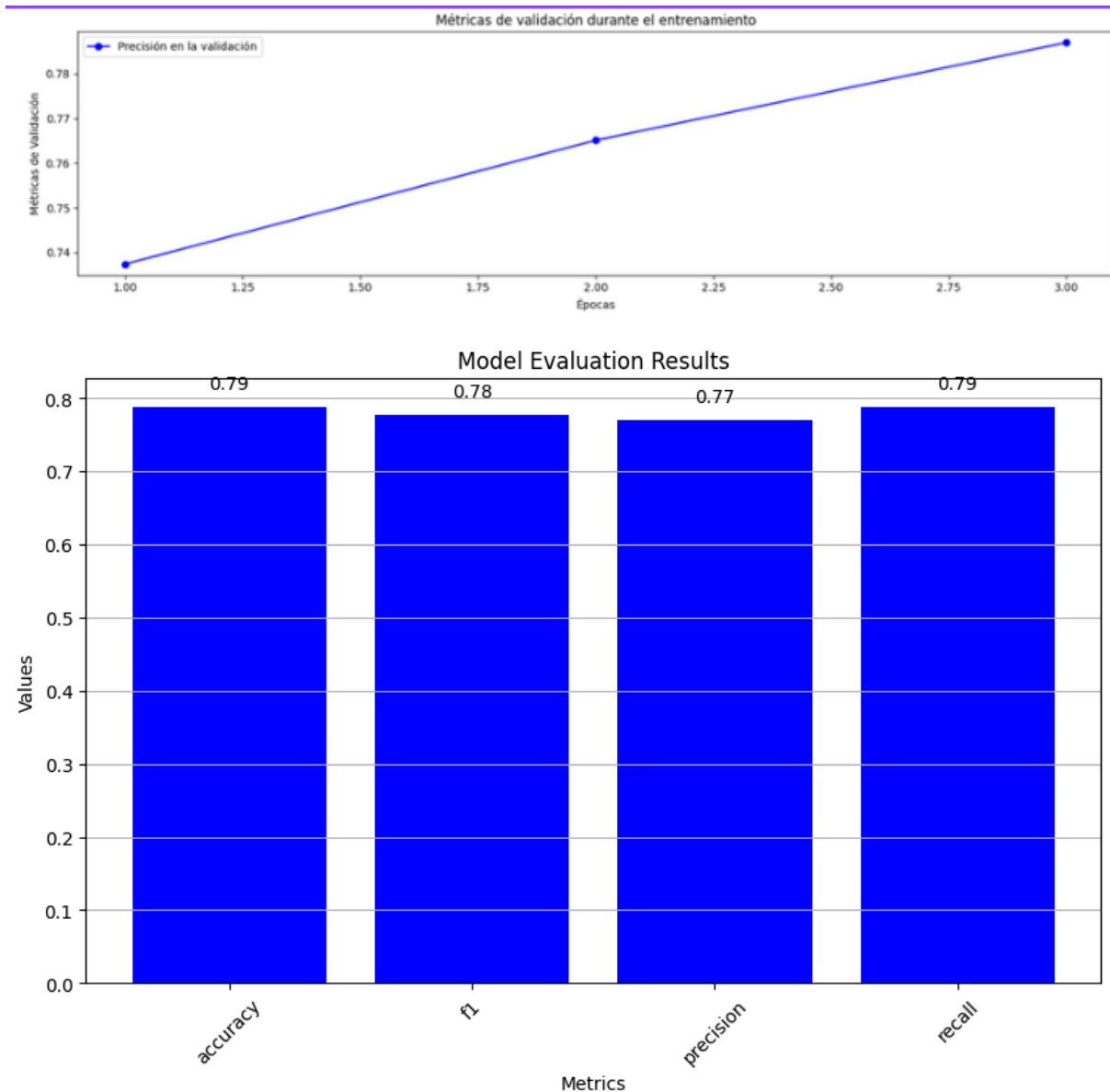


Gráfica 4. Pérdida a través de las epochs del entrenamiento y la validación.



## BERT

Gráfica 5. Resultados de la evaluación del modelo



En estos resultados preliminares se puede observar que para los LSTM el entrenamiento ha sido bastante decente dado que se observaba una precisión del 65 al 70% sin embargo, al validar este modelo se observó un desempeño pobre significando un 55% aunque si mejoraba a través de las épocas no es lo suficientemente bueno como para tomarlo en consideración como un modelo real para modelar el problema de los artículos.

Para el modelo de RNN se observa una precisión de entrenamiento del 66% y 61% para el de validación, esto tiene sentido dado que el hecho de que algunas entidades aparecen muy poco no requieren de tener una memoria muy grande como podría ser en un modelo de LSTM lo que en este caso son un poco más ideales que las LSTM.

Finalmente, el modelo BERT tuvo el mejor resultado con diferencia ya que a partir de su modelo de auto atención permite extraer lo más importante de los textos. Esto explica por qué su rendimiento fue del 78% con el entrenamiento y un 79% en la validación.

# Referencias

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2019). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4), 1234-1240. <https://doi.org/10.1093/bioinformatics/btz682>

Daniel. (2024, 20 mayo). *Memoria a largo plazo a corto plazo (LSTM): ¿Qué es?* Formación En Ciencia de Datos | DataScientest.com. <https://datascientest.com/es/memoria-a-largo-plazo-a-corto-plazo-lstm>

*¿Qué es una RNN?: Explicación sobre redes neuronales recurrentes: AWS.* (s. f.). Amazon Web Services, Inc. [https://aws.amazon.com/es/what-is/recurrent-neural-network/#:~:text=Una%20red%20neuronal%20recurrente%20\(RNN,salida%20de%20datos%20secuencial%20espec%C3%ADfica.](https://aws.amazon.com/es/what-is/recurrent-neural-network/#:~:text=Una%20red%20neuronal%20recurrente%20(RNN,salida%20de%20datos%20secuencial%20espec%C3%ADfica.)

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.1706.03762>