

Evaluación Parcial N°3

Redes Recurrentes/Transformers Estudiante

Sigla	Nombre Asignatura	Tiempo Asignado	% Ponderación
DLY0100	Deep Learning	2 semanas/ 10 minutos de presentación	40%

1. Situación Evaluativa:

x	Entrega de encargo
	encargo

×	Presentación
---	--------------

2. Instrucciones

Descripción general de la evaluación

En esta evaluación, los/las estudiantes implementarán arquitecturas basadas en RNN y Transformers para tareas de procesamiento de lenguaje natural (NLP). El propósito es la aplicación de modelos RNN y Transformers, ajustando sus componentes clave para resolver tareas como la traducción automática o generación de texto, utilizando frameworks avanzados como TensorFlow o PyTorch.

En esta evaluación, el/la docente entregará los sets de datos y dos códigos fuentes base, uno para RNN y otro para TRANSFORMERS. Los/as estudiantes deben modificar/completar los códigos, ejecutarlos, entrenar los modelos, calcular las métricas e implementar cambios a los hiperparámetros para mejorar dichas métricas.

Los/as estudiantes deben entregar dos códigos fuentes, uno para RNN y otro para TRANSFORMER.



- El propósito de esta evaluación es evaluar los siguientes Indicadores de Logro:
 - o IL 2.3 Propone arquitecturas RNN base y LSTM, ajustando hiperparámetros para lograr un mejor rendimiento y dar solución a una problemática de predicción, clasificación, según el caso de estudio.
 - o IL 2.4 Evalúa el desempeño de los modelos de clasificación y predicción de datos no estructurados, considerando distintas métricas y requerimientos, para resolver un problema planteado.
 - o IL 2.5 Implementa arquitecturas basadas en encoder-decoder y transformer, explicando sus componentes clave y demostrando su aplicación en problemas de procesamiento de lenguaje natural (NLP) como la traducción automática o generación de texto.
- Esta evaluación consiste en una entrega de encargo con presentación y tiene un 40% de ponderación sobre la nota final de la asignatura.
 - Tiempo asignado para esta evaluación es de 2 semanas, desde la fecha de la entrega de la instrucción, hasta su entrega y defensa, cuentan con 10 minutos para su presentación y se realiza en *equipos de dos personas* en laboratorio.
 - o La distribución de los porcentajes de las situaciones evaluativas que componen esta evaluación es la siguiente:

l Evaluación		Tipo de situación evaluativa	Distribución de porcentajes
5 '' D ' N 2	400/	A Encargo	30%
Evaluación Parcial N° 3	40%	B Presentación	70%

Instrucciones Específicas

Dimensión encargo:

Los/as estudiantes deben entregar dos códigos fuentes, uno para RNN y otro para TRANSFORMER, en base a los cuales realizarán su presentación.

Primer entregable: Un modelo RNN mejorado, documentado y analizado con sus respectivas métricas. Para ello, se debe considerar:

- Implementar una red neuronal recurrente (RNN, LSTM o GRU) en Python utilizando TensorFlow o PyTorch.
- Justifica la elección entre RNN y LSTM, explicando con precisión los ajustes de hiperparámetros y su impacto en el desempeño del modelo. Utiliza métricas como accuracy y F1-score de manera efectiva para respaldar su elección, mostrando un análisis profundo y comparativo.
- Ajustar los hiperparámetros para optimizar el rendimiento del modelo.



- Describe de manera clara y detallada el impacto de los hiperparámetros ajustados en la precisión y eficacia del modelo. Sustenta su análisis con evidencia técnica sólida, como gráficos, tablas comparativas o resultados experimentales. Explica de manera precisa cómo cada hiperparámetro influye en el rendimiento del modelo.
- Evaluar el desempeño del modelo con métricas como accuracy, precision, recall, y F1-score.
- Analiza el desempeño del modelo utilizando métricas como accuracy, F1-score, precisión y recall. Interpreta los resultados de forma clara y precisa, relacionándolos directamente con los ajustes de hiperparámetros realizados. Sustenta su análisis con evidencia técnica sólida (gráficos, tablas, etc.) y ofrece conclusiones bien fundamentadas.
- Justifica la selección del modelo final, explicando el impacto de las épocas utilizadas, la tasa de aprendizaje y el tamaño del lote en las métricas de evaluación. Sustenta su análisis con evidencia técnica sólida (gráficos, tablas comparativas, etc.) y ofrece conclusiones bien fundamentadas, proponiendo mejoras basadas en los resultados obtenidos.
- Presentar el desarrollo de la solución (en el COLAB) explicando las decisiones tomadas y justificando los ajustes realizados en la red recurrente.

Segundo entregable: un modelo TRANSFORMER mejorado, documentado y analizado con sus respectivas métricas. Para ello, se debe considerar:

- Implementar una arquitectura transformer básica (utilizando encoder-decoder y multi-head attention).
- Implementa de manera correcta y eficiente una arquitectura Transformer básica, describiendo de forma clara y detallada sus componentes clave (encoder, decoder y mecanismo de autoatención). Explica con precisión el funcionamiento de cada componente y su interacción dentro de la arquitectura. Además, proporciona ejemplos o visualizaciones que refuerzan la comprensión de la implementación.
- Explicar los componentes clave del modelo y cómo influyen en su desempeño.
- Aplicar el transformer a una tarea de NLP como traducción automática o generación de texto.
- Evaluar el desempeño del modelo utilizando métricas relevantes, como BLEU (Bilingual Evaluation Understudy) Score o ROUGE.
- Describe el funcionamiento de los elementos esenciales del modelo, explicando su rol y contribución al procesamiento de lenguaje natural (NLP).

 Analiza exhaustivamente la calidad de los resultados a través de ejemplos de salida aplicados a un problema de NLP, interpretando los aciertos y errores del modelo de manera crítica.
- Presentar el desarrollo de la solución (en el COLAB) explicando cómo implementaron, ajustaron y evaluaron el modelo transformer.



Dimensión presentación:

Se espera que, a partir del informe o encargo entregado, desarrollen una presentación, en la cual defenderán y evidenciarán los aspectos más importantes trabajados, debe utilizar un lenguaje acorde a la disciplina.

Para ello debe considerar evidencia relevante de su informe y realizar las acciones especificadas:

- Implementar una red neuronal recurrente (RNN, LSTM o GRU) en Python utilizando TensorFlow o PyTorch. / Ajustar los hiperparámetros para optimizar el rendimiento del modelo.
- Explica la selección de la arquitectura y los hiperparámetros ajustados. Compara diferentes configuraciones de manera detallada, justificando su impacto en el desempeño del modelo con métricas como accuracy y F1-score, respaldado con ejemplos concretos de su evolución. Demuestra un dominio completo del tema y utiliza recursos visuales o ejemplos prácticos para reforzar su explicación.
- Evaluar el desempeño del modelo con métricas como accuracy, precision, recall, y F1-score.
- Explica el significado de cada métrica utilizada en la evaluación del modelo, relacionándolas con el problema planteado. Compara de manera detallada el rendimiento de diferentes configuraciones, justificando la mejor opción con argumentos sólidos y ejemplos concretos. Demuestra un dominio completo del tema.
- Presentar el desarrollo de la solución (en el COLAB) explicando las decisiones tomadas y justificando los ajustes realizados en la red recurrente.

Segundo entregable (40% de la nota): un modelo TRANSFORMER mejorado, documentado y analizado con sus respectivas métricas. Para ello, se debe considerar:

- Implementar una arquitectura transformer básica (utilizando encoder-decoder y multi-head attention). / Explicar los componentes clave del modelo y cómo influyen en su desempeño. / Aplicar el transformer a una tarea de NLP como traducción automática o generación de texto. / Evaluar el desempeño del modelo utilizando métricas relevantes, como BLEU (Bilingual Evaluation Understudy) Score o ROUGE.
- Explica de manera clara, concisa y estructurada el funcionamiento del modelo, destacando de forma precisa las diferencias clave entre arquitecturas tradicionales (como RNN o LSTM) y transformadores (por ejemplo, mecanismo de autoatención, paralelización, etc.). Demuestra un dominio completo del tema. Además, emplea recursos visuales o gráficos para reforzar su explicación.

Orientaciones para la implementación de la evaluación (se aplica a los dos entregables).

Productos a Entregar



CUADERNO JUPYTER EN GOOGLE COLAB (considerado como informe)

Instrucciones generales

- Los/as estudiantes deben desarrollar un cuaderno Jupyter en Google Colab, documentando cada paso de la implementación.
- El cuaderno debe incluir explicaciones en texto (Markdown), código bien estructurado y visualización de resultados (gráficos y tablas).
- Se recomienda el uso de TensorFlow/Keras o PyTorch para la implementación del modelo.
- Se debe proporcionar una sección de conclusiones, donde se analicen los resultados obtenidos.

Estructura sugerida del cuaderno

- Introducción: Descripción del problema (ej. traducción automática, generación de texto).
- Carga y preprocesamiento de datos: Explicación del dataset utilizado, transformaciones y limpieza de datos necesarias, Conversión de datos en embeddings adecuados para el modelo.
- Definición del modelo
- Justificación de la selección del modelo y sus configuraciones: Número de capas, Tamaño del batch, Tasa de aprendizaje.
- Entrenamiento y ajuste de hiperparámetros.
- Evaluación del modelo: Uso de métricas.
- Análisis y mejora del modelo: Comparación de distintas configuraciones y ajustes realizados.
- Conclusiones: Reflexión sobre el desempeño del modelo y posibles mejoras.

Pautas de documentación

- Markdown debe utilizarse para explicar cada bloque de código.
- Comentarios en el código para facilitar la comprensión de la implementación.
- Uso de gráficos y tablas para visualizar resultados.
- Códigos bien estructurados y modularizados.

PRESENTACIÓN ORAL

Objetivo

• Los/as estudiantes deben exponer los principales hallazgos y aprendizajes de su implementación en un máximo de 10 minutos, utilizando su cuaderno Jupyter como base para la presentación.

Estructura sugerida

- Introducción al problema y objetivos.
- Explicación de la arquitectura del modelo.
- Decisiones clave en la implementación y configuración.
- Resultados obtenidos y análisis de métricas.
- Comparación de configuraciones y mejoras implementadas.



Conclusiones y reflexiones finales.

o Requisitos de la presentación

- Explicar de manera clara y concisa el desarrollo del modelo.
- Utilizar gráficos y tablas para ilustrar los resultados.
- Justificar decisiones tomadas en la implementación.
- Gestionar bien el tiempo de exposición. Máximo 10 minutos, dividirse equitativamente la defensa.

Se entiende como trabajo en pareja el desarrollo del encargo, la presentación y defensa son individuales y dependerá del desempeño particular de cada estudiante.



3. Pauta de Evaluación

Tipo de Pauta: Rúbrica

Categoría	% logro	Descripción niveles de logro
Muy buen desempeño	100%	Demuestra un desempeño destacado, evidenciando el logro de todos los aspectos evaluados en el indicador.
Buen desempeño	80%	Demuestra un alto desempeño del indicador, presentando pequeñas omisiones, dificultades y/o errores.
Desempeño aceptable	60%	Demuestra un desempeño competente, evidenciando el logro de los elementos básicos del indicador, pero con omisiones, dificultades o errores.
Desempeño incipiente	30%	Presenta importantes omisiones, dificultades o errores en el desempeño, que no permiten evidenciar los elementos básicos del logro del indicador, por lo que no puede ser considerado competente.
Desempeño no logrado	0%	Presenta ausencia o incorrecto desempeño.

	Categorías de Respuesta					Ponderación
Indicador de Evaluación	Muy buen desempeño 100%	Buen desempeño 80%	Desempeño aceptable 60%	Desempeño incipiente 30%	Desempeño no logrado 0%	Indicador de Evaluación
Dimensión Encargo (En p	arejas)					
IEE 2.3.1 Justifica la	Justifica la elección	Justifica la elección	Justifica la elección	Justifica	No justifica la	3%
elección entre RNN y	entre RNN y LSTM,	entre RNN y LSTM,	entre RNN y LSTM,	superficialmente la	selección entre	
LSTM, explicando los	explicando con	explicando los ajustes	mencionando	elección entre RNN y	modelos o no realiza	
ajustes de	precisión los ajustes	de hiperparámetros y	algunos ajustes de	LSTM, sin explicar	la elección.	
hiperparámetros y su	de hiperparámetros y	su efecto en el	hiperparámetros	claramente los ajustes		
efecto en el	su impacto en el	desempeño del	pero su impacto es	de hiperparámetros o		
desempeño del	desempeño del	modelo. Utiliza	limitado en el modelo.	su efecto en el		
modelo, con base en	modelo. Utiliza	métricas como	Utiliza métricas como	modelo. Utiliza		
métricas como	métricas como	accuracy y F1-score	accuracy y F1-score,	métricas como		
accuracy y F1-score.	accuracy y F1-score	para respaldar su	pero el análisis es	accuracy y F1-score		
	de manera efectiva	elección, aunque el	superficial y carece de	de manera limitada o		
	para respaldar su					

MALETA DUOCUC

	elección, mostrando un análisis profundo y comparativo.	análisis tiene algunas omisiones.	profundidad comparativa.	incorrecta. El análisis es confuso.		
IEE 2.3.2 Describe el impacto de los hiperparámetros ajustados (épocas, tasa de aprendizaje, tamaño del lote) en la precisión y eficacia del modelo, sustentándolo con evidencia técnica.	Describe de manera clara y detallada el impacto de los hiperparámetros ajustados en la precisión y eficacia del modelo. Sustenta su análisis con evidencia técnica sólida, como gráficos, tablas comparativas o resultados experimentales. Explica de manera precisa cómo cada hiperparámetro influye en el rendimiento del modelo.	Describe el impacto de los hiperparámetros ajustados en la precisión y eficacia del modelo, sustentando su análisis con evidencia técnica, aunque falta profundidad. Explica de manera general cómo los hiperparámetros afectan el rendimiento, pero puede omite detalles específicos.	Describe de manera básica el impacto de los hiperparámetros ajustados en la precisión y eficacia del modelo. Menciona alguna evidencia técnica, pero esta es poco detallada. El análisis es superficial y no profundiza en la relación entre los hiperparámetros y el rendimiento del modelo.	Describe superficialmente el impacto de los hiperparámetros ajustados, sin explicar claramente su relación con la precisión y eficacia del modelo. La evidencia técnica es escasa o poco relevante. El análisis es incompleto y carece de conclusiones claras. Explicación con poca claridad y sin comparaciones relevantes.	No describe el impacto de los hiperparámetros ajustados ni su relación con la precisión y eficacia del modelo. No presenta evidencia técnica o esta es incorrecta o irrelevante.	4%
IEE 2.4.1 Analiza el desempeño del modelo mediante métricas como accuracy, F1-score, precisión y recall, interpretando los resultados en función de los ajustes de hiperparámetros.	Analiza el desempeño del modelo utilizando métricas como accuracy, F1-score, precisión y recall. Interpreta los resultados de forma clara y precisa, relacionándolos directamente con los ajustes de hiperparámetros realizados. Sustenta su análisis con evidencia técnica	Analiza el desempeño del modelo utilizando métricas como accuracy, F1-score, precisión y recall, interpretando los resultados de manera general. Relaciona los resultados con los ajustes de hiperparámetros, aunque con menor profundidad que en el nivel superior. Presenta evidencia	Analiza de manera básica el desempeño del modelo utilizando algunas métricas (accuracy, F1-score, precisión o recall). La interpretación de los resultados es superficial y no profundiza en la relación con los ajustes de hiperparámetros. La evidencia técnica y las	Analiza superficialmente el desempeño del modelo utilizando alguna métrica, pero sin interpretar adecuadamente los resultados o relacionarlos con los ajustes de hiperparámetros. La evidencia técnica es escasa o poco relevante. El análisis es incompleto y	No analiza el desempeño del modelo ni utiliza métricas como accuracy, F1-score, precisión o recall. No interpreta los resultados ni los relaciona con los ajustes de hiperparámetros.	5%



	sólida (gráficos, tablas, etc.) y ofrece conclusiones bien fundamentadas.	técnica, pero puede carecer de detalles específicos o conclusiones más elaboradas.	conclusiones son poco detalladas.	carece de conclusiones claras.		
IEE 2.4.2 Justifica la selección del modelo final con base en el impacto de las épocas utilizadas, la tasa de aprendizaje y el tamaño del lote en las métricas de evaluación.	Justifica la selección del modelo final, explicando el impacto de las épocas utilizadas, la tasa de aprendizaje y el tamaño del lote en las métricas de evaluación. Sustenta su análisis con evidencia técnica sólida (gráficos, tablas comparativas, etc.) y ofrece conclusiones bien fundamentadas, proponiendo mejoras basadas en los resultados obtenidos.	Justifica la selección del modelo final, explicando el impacto de las épocas, la tasa de aprendizaje y el tamaño del lote en las métricas de evaluación. Presenta evidencia técnica, aunque con menor profundidad que en el nivel superior. Las conclusiones son claras, pero omite algunos aspectos para la selección.	Justifica la selección del modelo final, mencionando el impacto de las épocas, la tasa de aprendizaje y el tamaño del lote en las métricas de evaluación. La evidencia técnica es limitada y el análisis es superficial, sin profundizar en la relación entre los hiperparámetros y el rendimiento del modelo.	Justifica superficialmente la selección del modelo final, sin explicar claramente el impacto de las épocas, la tasa de aprendizaje o el tamaño del lote en las métricas de evaluación. La evidencia técnica es escasa, confusa o poco relevante. El análisis es incompleto y carece de conclusiones claras.	No justifica la selección del modelo final ni explica el impacto de las épocas, la tasa de aprendizaje o el tamaño del lote en las métricas de evaluación	5%
IEE 2.5.1 Implementa una arquitectura Transformer básica, describiendo sus componentes clave (como el encoder, decoder)y mecanismo de autoatención.	Implementa de manera correcta y eficiente una arquitectura Transformer básica, describiendo de forma clara y detallada sus componentes clave (encoder, decoder y mecanismo de autoatención). Explica con precisión el funcionamiento de	Implementa una arquitectura Transformer básica, describiendo sus componentes clave (encoder, decoder y mecanismo de autoatención). La explicación es clara, pero omite algunos detalles técnicos o ejemplos ilustrativos. La implementación es funcional	Implementa una arquitectura Transformer básica, mencionando de manera general sus componentes clave (encoder, decoder y mecanismo de autoatención). La descripción es superficial y carece de detalles técnicos profundos. La implementación	Implementa una arquitectura Transformer básica, mencionando algunos de sus componentes clave (encoder, decoder o mecanismo de autoatención) de forma incompleta. La descripción es confusa y la implementación presenta errores	No implementa una arquitectura Transformer básica ni describe sus componentes clave.	7%



IEE 2.5.2 Describe el funcionamiento de los elementos esenciales del modelo y analiza la calidad de los resultados a través de ejemplos de salida aplicados a un problema de NLP.	cada componente y su interacción dentro de la arquitectura. Además, proporciona ejemplos o visualizaciones que refuerzan la comprensión de la implementación. Describe el funcionamiento de los elementos esenciales del modelo, explicando su rol y contribución al procesamiento de lenguaje natural (NLP). Analiza exhaustivamente la calidad de los resultados a través de ejemplos de salida aplicados a un problema de NLP, interpretando los aciertos y errores del modelo de manera	demuestra comprensión de los conceptos principales. Describe el funcionamiento de los elementos esenciales del modelo, explicando su rol en el procesamiento de lenguaje natural (NLP). Analiza la calidad de los resultados mediante ejemplos de salida aplicados a un problema de NLP. La interpretación de los resultados es clara, pero omitió algunos elementos de este.	puede presentar algunos errores menores. Describe el funcionamiento de los elementos esenciales del modelo, mencionando su rol en el procesamiento de lenguaje natural (NLP). Analiza la calidad de los resultados a través de ejemplos de salida, pero el análisis es superficial y no profundiza en la interpretación de los aciertos o errores del modelo.	significativos que afectan su funcionalidad. Describe superficialmente el funcionamiento de algunos elementos esenciales del modelo, sin explicar claramente su rol en el procesamiento de lenguaje natural (NLP). El análisis de la calidad de los resultados carece de ejemplos de salida relevantes. La interpretación de los resultados es incompleta o poco clara.	No describe el funcionamiento de los elementos esenciales del modelo ni analiza la calidad de los resultados.	6%
	crítica.				Porcentaje encargo	30%
Dimensión presentación ((Individual)				, 0-	
IEP 2.3.1 Explica de manera clara la selección de la arquitectura y los hiperparámetros ajustados, comparando diferentes configuraciones y	Explica la selección de la arquitectura y los hiperparámetros ajustados. Compara diferentes configuraciones de manera detallada, justificando su	Explica la selección de la arquitectura y los hiperparámetros ajustados, comparando diferentes configuraciones y justificando su	Explica de manera general la selección de la arquitectura y los hiperparámetros ajustados, mencionando algunas comparaciones entre	Explica de manera superficial la selección de la arquitectura y los hiperparámetros ajustados, sin comparar configuraciones ni	No explica la selección de la arquitectura ni los hiperparámetros ajustados.	17%



justificando su impacto en el desempeño del modelo con las métricas de accuracy, F1-score y ejemplos concretos de su evolución.	impacto en el desempeño del modelo con métricas como accuracy y F1-score, respaldado con ejemplos concretos de su evolución. Demuestra un dominio completo del tema y utiliza recursos visuales o ejemplos prácticos para reforzar su explicación.	impacto en el desempeño del modelo con métricas como accuracy y F1-score. Incluye ejemplos, pero omite algunos aspectos en ellos. Demuestra un buen manejo del tema.	configuraciones y su impacto en el desempeño del modelo. Utiliza métricas como accuracy y F1-score, pero el análisis es superficial y los ejemplos son limitados. Demuestra un manejo básico del tema.	justificar adecuadamente su impacto en el desempeño del modelo. Las métricas como accuracy y F1- score se mencionan de forma incompleta, y los ejemplos son escasos o irrelevantes. Demuestra un manejo limitado del tema		
IEP 2.4.1 Explica con claridad el significado de cada métrica utilizada en la evaluación del modelo, comparando el rendimiento de diferentes configuraciones y justificando la mejor opción según el problema planteado.	explicación. Explica el significado de cada métrica utilizada en la evaluación del modelo, relacionándolas con el problema planteado. Compara de manera detallada el rendimiento de diferentes configuraciones, justificando la mejor opción con argumentos sólidos y ejemplos concretos. Demuestra un dominio completo del tema.	Explica con claridad el significado de las métricas utilizadas en la evaluación del modelo, relacionándolas con el problema planteado. Compara el rendimiento de diferentes configuraciones y justifica la mejor opción, aunque puede carecer de profundidad en algunos aspectos. Su maneja del tema es correcto, pero omite algunos aspectos de importancia.	Explica el significado de las métricas utilizadas en la evaluación del modelo, mencionando su relación con el problema planteado. Compara el rendimiento de algunas configuraciones, pero el análisis es superficial y la justificación de la mejor opción es básica. Demuestra un manejo básico del tema.	Explica superficialmente el significado de algunas métricas utilizadas en la evaluación del modelo, sin relacionarlas claramente con el problema planteado. La comparación de configuraciones es limitada o confusa, y la justificación de la mejor opción es incompleta o poco clara. Demuestra un manejo limitado del tema.	No explica el significado de las métricas utilizadas en la evaluación del modelo ni las relaciona con el problema planteado.	25%
IEP 2.5.1 Explica de manera clara y concisa el funcionamiento del modelo, destacando las	Explica de manera clara, concisa y estructurada el funcionamiento del	Explica el funcionamiento del modelo, destacando las diferencias clave	Explica el funcionamiento del modelo, mencionando	Explica superficialmente el funcionamiento del modelo, con errores	No explica el funcionamiento del modelo ni destaca las diferencias entre	28%



diferencias clave entre arquitecturas tradicionales y transformadores.	modelo, destacando de forma precisa las diferencias clave entre arquitecturas tradicionales (como RNN o LSTM) y transformadores (por ejemplo, mecanismo de autoatención, paralelización, etc.). Demuestra un dominio completo del tema. Además, emplea recursos visuales o gráficos para reforzar su explicación.	tradicionales y transformadores. La explicación es precisa, aunque omite algunos detalles técnicos o	entre arquitecturas tradicionales y transformadores. La explicación es	en las diferencias clave entre arquitecturas tradicionales y transformadores. La explicación es confusa o incompleta, y no se profundiza en los conceptos. Demuestra un manejo limitado del tema.	arquitecturas tradicionales y transformadores	
Porcentaje presentación						
					Total	100%