

## Informe Ejecutivo

### *Análisis y Evaluación de Modelos de Clasificación*

---

#### Presentado por:

Jiliar Antonio Silgado Cardona

*Ingeniero de Software - Tech Manager*

---

#### Universidad EAN

*Facultad de Ingeniería*

*El presente informe tiene como objetivo analizar y evaluar diferentes modelos de clasificación aplicados a datos de texto, con énfasis en el rendimiento de modelos de redes neuronales y arquitecturas avanzadas como BERT. Dada la creciente relevancia de la inteligencia artificial y el procesamiento de lenguaje natural (NLP) en la resolución de problemas complejos en diversas industrias, es fundamental contar con herramientas que permitan seleccionar los modelos más eficientes y efectivos para tareas de clasificación.*

*La temática de este informe es crucial debido a la importancia de entender cómo las diferentes arquitecturas de modelos impactan en los resultados de clasificación, lo que puede tener implicaciones significativas en áreas como la automatización, la analítica avanzada, la toma de decisiones en tiempo real, y la optimización de procesos. Este análisis aporta información valiosa para mejorar la precisión y la eficiencia de los modelos, lo que se traduce en un mejor aprovechamiento de los recursos y un avance hacia la implementación exitosa de soluciones basadas en IA.*

# **Resumen Ejecutivo del Proyecto de Procesamiento de Lenguaje Natural (NLP)**

## **Introducción**

En un mundo impulsado por la información, la capacidad de procesar y clasificar grandes volúmenes de texto es esencial para organizaciones y sistemas automatizados. Este proyecto se centra en el desarrollo de un sistema avanzado de clasificación de noticias mediante el uso de técnicas modernas de Procesamiento de Lenguaje Natural (NLP). Utilizando arquitecturas de vanguardia como BERT, redes neuronales recurrentes (RNN) y redes de memoria a largo plazo (LSTM), el objetivo es categorizar noticias en temas especializados como deportes, cultura, economía y justicia con alta precisión y eficiencia.

A través de una metodología bien estructurada que abarca desde la exploración de datos hasta la implementación de modelos, se busca no solo evaluar el rendimiento de estas arquitecturas, sino también optimizarlas para maximizar su aplicabilidad en entornos reales. Este trabajo resalta la importancia del NLP en la automatización y mejora de procesos informativos, contribuyendo a la toma de decisiones basada en datos estructurados y organizados.

## **Resumen del Proyecto**

El proyecto tiene como objetivo desarrollar un sistema avanzado de clasificación de noticias utilizando modelos de Procesamiento de Lenguaje Natural (NLP), incluyendo BERT, RNN y LSTM. Este sistema permite categorizar noticias en temas especializados como deportes, cultura, economía y justicia. Para ello, se evaluaron múltiples enfoques y arquitecturas, desde el preprocesamiento de texto hasta la evaluación de modelos, con el fin de seleccionar la metodología más eficiente y precisa.

El modelo BERT fue elegido como candidato principal debido a su capacidad para comprender el contexto y las relaciones entre palabras. Asimismo, se realizaron comparaciones con modelos de redes neuronales recurrentes (RNN) y redes de

memoria a largo plazo (LSTM) para identificar el enfoque óptimo. Se llevaron a cabo pruebas exhaustivas y análisis de resultados para garantizar la calidad y aplicabilidad del modelo en entornos reales.

## Objetivos del Proyecto

- Desarrollar un modelo avanzado de clasificación de noticias: Utilizar arquitecturas modernas como BERT, RNN y LSTM para categorizar noticias de forma precisa.
- Evaluar el rendimiento de múltiples modelos: Comparar métricas clave como precisión, recall y F1-score entre diferentes enfoques.
- Optimizar el rendimiento del modelo: Implementar técnicas de regularización y optimización de hiperparámetros para mejorar la generalización del modelo.
- Proveer recomendaciones para implementación en producción: Diseñar un flujo de trabajo reproducible y aplicable en sistemas automatizados.

## Metodología del Proyecto

### 1. Selección y Preparación de Modelos

- Modelos preentrenados evaluados:
  - ``dccuchile/bert-base-spanish-wwm-cased``
  - ``bertin-project/bertin-roberta-base-spanish``
  - ``mrm8488/bert-base-spanish-wwm-uncased``
- Preparación de datos: Tokenización, ajuste de secuencias y configuración de entradas para garantizar compatibilidad con BERT.

## **2. Carga y Exploración de Datos**

- Análisis de la distribución de categorías mediante visualizaciones gráficas.
- Filtrado y balanceo de datos para optimizar el rendimiento de los modelos.

## **3. División del Conjunto de Datos**

- División aleatoria en conjuntos de entrenamiento (70%), validación (15%) y prueba (15%) manteniendo la proporción de categorías.

## **4. Implementación y Entrenamiento de Modelos**

- RNN y LSTM: Construcción de modelos con capas de embedding y recurrentes, incluyendo regularización mediante dropout.
- BERT: Ajuste fino del modelo preentrenado con hiperparámetros optimizados, incluyendo tasa de aprendizaje y número de épocas.

## **5. Evaluación de Modelos**

- Métricas calculadas: Precisión, recall, F1-score y matriz de confusión.
- Análisis de errores para identificar áreas de mejora.

## **6. Optimización de Hiperparámetros**

- Aplicación de búsqueda en cuadrícula y aleatoria para encontrar configuraciones óptimas.
- Ajuste de hiperparámetros como tamaño de lote, tasa de aprendizaje y número de épocas.

## **7. Pruebas y Validación Final**

- Evaluación en el conjunto de prueba para medir la capacidad de generalización.
- Análisis de las categorías con mayor y menor precisión.

## 8. Visualización de Resultados

- Generación de curvas de aprendizaje, matrices de confusión y gráficos de palabras clave.

## 9. Conclusiones y Recomendaciones

- Resumen del rendimiento de los modelos comparados:
  - RNN: Adecuado para patrones locales.
  - LSTM: Mejor captura de dependencias temporales.
  - BERT: Sobresaliente en contexto y semántica.

## Hallazgos

### Modelo RNN

Métrica	Valor
Pérdida	1.3412
Precisión	0.3904
Recall	0.3904
F1-Score	0.2192

### Matriz de Confusión:

	Predicción: Cultura	Predicción: Deportes	Predicción: Economía	Predicción: Justicia
Cultura	0	86	0	0
Deportes	0	146	0	0
Economía	0	73	0	0
Justicia	0	69	0	0

Reporte de Clasificación:

Clase	Precisión	Recall	F1-Score	Soporte
Cultura	0.00	0.00	0.00	86
Deportes	0.39	1.00	0.56	146
Economía	0.00	0.00	0.00	73
Justicia	0.00	0.00	0.00	69

- El modelo RNN mostró una precisión general baja, destacándose solo en la categoría "deportes", con un recall perfecto. Sin embargo, las otras categorías no lograron un buen rendimiento.

Modelo RNN Optimizado

Métrica	Valor
Pérdida	1.6826
Precisión	0.2861
Recall	0.2861
F1-Score	0.2741

Matriz de Confusión:

	Predicción: Cultura	Predicción: Deportes	Predicción: Economía	Predicción: Justicia
Cultura	39	23	14	10
Deportes	70	49	16	11
Economía	27	21	16	9
Justicia	26	28	12	3

Reporte de Clasificación:

Clase	Precisión	Recall	F1-Score	Soporte
-------	-----------	--------	----------	---------

Cultura	0.24	0.45	0.31	86
Deportes	0.40	0.34	0.37	146
Economía	0.28	0.22	0.24	73
Justicia	0.09	0.04	0.06	69

- A pesar de las optimizaciones, este modelo aún presenta un rendimiento limitado, especialmente en las categorías de "cultura" y "justicia". La precisión es baja en general, con mejores resultados en "deportes".

---

### Modelo LSTM

Métrica	Valor
Pérdida	1.6826
Precisión	0.2861
Recall	0.3904
F1-Score	0.2192

### Matriz de Confusión:

	Predicción: Cultura	Predicción: Deportes	Predicción: Economía	Predicción: Justicia
Cultura	0	86	0	0
Deportes	0	146	0	0
Economía	0	73	0	0
Justicia	0	69	0	0

### Reporte de Clasificación:

Clase	Precisión	Recall	F1-Score	Soporte
Cultura	0.00	0.00	0.00	86
Deportes	0.39	1.00	0.56	146
Economía	0.00	0.00	0.00	73

Justicia	0.00	0.00	0.00	69
----------	------	------	------	----

- El modelo LSTM mostró una precisión baja y un recall perfecto solo en "deportes", sin un rendimiento destacable en las otras categorías.

#### Modelo LSTM Optimizado

Métrica	Valor
Pérdida	1.3360
Precisión	0.3904
Recall	0.3904
F1-Score	0.2192

#### Matriz de Confusión:

	Predicción: Cultura	Predicción: Deportes	Predicción: Economía	Predicción: Justicia
Cultura	0	86	0	0
Deportes	0	146	0	0
Economía	0	73	0	0
Justicia	0	69	0	0

#### Reporte de Clasificación:

Clase	Precisión	Recall	F1-Score	Soporte
Cultura	0.00	0.00	0.00	86
Deportes	0.39	1.00	0.56	146
Economía	0.00	0.00	0.00	73
Justicia	0.00	0.00	0.00	69



- El modelo LSTM optimizado mostró ligeras mejoras en precisión y recall comparado con el modelo LSTM no optimizado, aunque aún con bajos resultados en la mayoría de las categorías.

---

**Modelo BERT: "mrm8488/bert-base-spanish-wwm-uncased"**

Métrica	Valor
Pérdida	27.0157
Precisión	0.9390
Recall	0.9390
F1-Score	0.9393

**Matriz de Confusión:**

	Predicción: Clase 0	Predicción: Clase 1	Predicción: Clase 2	Predicción: Clase 3
Clase 0	319	0	16	5
Clase 1	15	571	1	2
Clase 2	9	1	265	18
Clase 3	11	4	9	247

**Reporte de Clasificación:**

Clase	Precisión	Recall	F1-Score	Soporte
Clase 0	0.90	0.94	0.92	340
Clase 1	0.99	0.97	0.98	589
Clase 2	0.91	0.90	0.91	293
Clase 3	0.91	0.91	0.91	271

- El modelo BERT mostró un excelente desempeño con altas métricas en precisión, recall y F1-score, destacando particularmente en la categoría "deportes", que tuvo una excelente precisión y recall.

**Modelo BERT Optimizado: "mrm8488/bert-base-spanish-wwm-uncased"**

Métrica	Valor
Pérdida	0.1190
Precisión	0.9658
Pérdida de validación	0.4355
Precisión de validación	0.8850

**Matriz de Confusión:**

	Predicción 0	Predicción 1	Predicción 2	Predicción 3
Clase 0	1466	152	97	60
Clase 1	90	2674	45	46
Clase 2	70	106	1178	101
Clase 3	73	86	167	1054

Reporte de Clasificación:

Clase	Precisión	Recall	F1-Score	Soporte
Clase 0	0.86	0.83	0.84	1775
Clase 1	0.89	0.94	0.91	2855
Clase 2	0.79	0.81	0.80	1455
Clase 3	0.84	0.76	0.80	1380
Total	0.85	0.85	0.85	7465

**Comentario:** El modelo BERT optimizado presenta una ligera mejora en comparación con su versión anterior, con un rendimiento equilibrado en todas las clases y métricas de evaluación muy buenas en términos generales.

---

Resultados

Comparativa de Modelos

A continuación se presenta una comparativa de los modelos entrenados: RNN, RNN Optimizado, LSTM, LSTM Optimizado, BERT y BERT Optimizado, con sus respectivos resultados en términos de precisión, recall, F1-Score y matriz de confusión.

Modelo	Pérdida	Precisión	Recall	F1-Score
RNN	1.3412	0.3904	0.3904	0.2192

<b>RNN Optimizado</b>	1.6826	0.2861	0.2861	0.2741
<b>LSTM</b>	1.6826	0.2861	0.3904	0.2192
<b>LSTM Optimizado</b>	1.3360	0.3904	0.3904	0.2192
<b>BERT</b>	27.0157	0.9390	0.9390	0.9393
<b>BERT Optimizado</b>	0.4274	0.8536	0.8536	0.8526

## Análisis y Resultados:

### Modelos Recurrentes (RNN y LSTM):

- Los **modelos RNN y LSTM** muestran un desempeño relativamente bajo, con precisión y recall similares (alrededor de 0.39).
- A pesar de tener una precisión de 0.39, los modelos no logran un rendimiento notable en cuanto a la clasificación de todas las clases, con F1-scores bajos, especialmente para las clases "cultura", "economía" y "justicia", que tienen un valor de 0.00 en los reportes de clasificación.
- La optimización de estos modelos no mejora significativamente los resultados, especialmente en el caso del **RNN Optimizado**, que incluso muestra un empeoramiento en la precisión (0.2861).

### Modelos BERT:

- Los modelos **BERT y BERT Optimizado** presentan resultados impresionantes, especialmente el modelo **BERT** con una precisión, recall y F1-score de 0.9390, 0.9390 y 0.9393, respectivamente.
- El modelo **BERT Optimizado** muestra una ligera disminución en el rendimiento, con precisión, recall y F1-score en torno a 0.85, lo cual sigue siendo excelente pero inferior al modelo original de BERT.

- A pesar de la disminución, **BERT Optimizado** es claramente superior a todos los modelos recurrentes en cuanto a los principales indicadores de evaluación.

### **Conclusión:**

El **modelo BERT** ("mrm8488/bert-base-spanish-wwm-uncased") presenta el mejor rendimiento de todos los modelos evaluados. Con una alta precisión (0.9390), recall (0.9390) y F1-score (0.9393), se destaca de manera significativa en comparación con los modelos RNN, LSTM y sus versiones optimizadas. Aunque la optimización de BERT (BERT Optimizado) tiene un rendimiento ligeramente inferior, sigue siendo considerablemente más efectivo que los modelos basados en redes neuronales recurrentes (RNN y LSTM), que muestran un desempeño más modesto en términos de las métricas de evaluación.

Por lo tanto, **BERT** es el modelo con mejor rendimiento en esta evaluación.

## Bibliografía

- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Recuperado de <https://arxiv.org/abs/1810.04805>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. Recuperado de <https://www.deeplearningbook.org>
- González, R. J., & Hernández, L. A. (2020). Introducción a las redes neuronales recurrentes (RNN) y redes de memoria a largo plazo (LSTM). *Revista de Computación e Inteligencia Artificial*, 5(2), 45–60. Recuperado de <https://revistacomputacion.universidad.edu>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. Recuperado de <https://arxiv.org/abs/1301.3781>
- Brownlee, J. (2021). *Deep Learning for Natural Language Processing*. Machine Learning Mastery. Recuperado de <https://machinelearningmastery.com>