

Optimización de modelos de inteligencia artificial para la clasificación de información empresarial: un enfoque basado en particle swarm optimization ¹⁴

Optimization of artificial intelligence models for business information classification: a particle swarm optimization approach

Resumen

El crecimiento de los datos no estructurados en entornos organizacionales ha incrementado la necesidad de desarrollar herramientas analíticas capaces de transformar información textual en conocimiento útil para la toma de decisiones. En este contexto, el estudio analiza el impacto de la optimización de hiperparámetros en el desempeño de modelos de inteligencia artificial aplicados a la clasificación de información empresarial, particularmente noticias económicas. La investigación se fundamenta en fuentes secundarias de literatura científica especializada y en el uso del corpus Reuters Newswire Topics, compuesto por noticias financieras reales. Metodológicamente, se adopta un enfoque experimental en tres fases: construcción de un modelo base de red neuronal profunda (MLP), optimización de hiperparámetros mediante Particle Swarm Optimization (PSO) (Clerc, 2006) y evaluación comparativa del desempeño. Se emplean métricas como exactitud y F1 macro. Los resultados evidencian que, aunque el modelo base alcanza niveles aceptables de exactitud, presenta limitaciones en la clasificación de categorías minoritarias. La optimización mediante PSO permitió mejorar significativamente el desempeño, destacándose un incremento cercano al 18 % en la medida F1 macro. Se concluye que la optimización de hiperparámetros es un componente crítico en el diseño de modelos de aprendizaje profundo, ya que mejora su capacidad de generalización y su utilidad en contextos de analítica empresarial y toma de decisiones.

¹⁴Edgardo Andrés Gordillo Moreno. Maestrante en Análisis de Datos y Sistemas Inteligentes de la Universidad Santo Tomás. Magíster en Administración de la Universidad del Tolima, Especialista en Gerencia de Proyectos y Administrador de Empresas de la Universidad del Tolima. E-mail: eagordillom@ut.edu.co. Docente catedrático del IDEAD de la Universidad del Tolima.

Palabras clave: Inteligencia artificial (C45), Métodos computacionales (C63), Analítica empresarial (M15), Toma de decisiones (D83), Innovación tecnológica (O33).

Abstract

The growing volume of unstructured data in organizational environments has increased the need for analytical tools capable of transforming textual information into useful knowledge for decision-making. In this context, this study analyzes the impact of hyperparameter optimization on the performance of artificial intelligence models applied to the classification of business information, particularly economic news. The research is based on secondary sources, including specialized scientific literature and the Reuters Newswire Topics dataset, which consists of real financial news widely used in natural language processing studies. Methodologically, an experimental approach is adopted in three phases: the development of a baseline deep learning model (Multilayer Perceptron), hyperparameter optimization using Particle Swarm Optimization (PSO) (Clerc, 2006), and a comparative performance evaluation. Metrics such as accuracy and macro F1-score are used. The results show that, although the baseline model achieves acceptable accuracy, it presents limitations in classifying minority categories. The application of PSO led to a significant improvement in performance, with an increase of approximately 18% in the macro F1-score. It is concluded that hyperparameter optimization is a critical component in the design of deep learning models, as it enhances their generalization capacity and practical usefulness in business analytics and decision-making contexts.

Keywords: Artificial intelligence (C45), Computational optimization (C63), Business analytics (M15), Decision-making (D83), Technological innovation (O33).

Introducción

En entornos organizacionales caracterizados por el creciente volumen de información, la capacidad de transformar datos en conocimiento útil se ha convertido en un elemento clave para la toma de decisiones. En particular, el análisis de información textual proveniente de noticias económicas, informes financieros y documentos corporativos exige herramientas que superen las limitaciones de la interpretación manual frente a la velocidad y complejidad

de los datos (Provost y Fawcett, 2013). En este contexto, la analítica avanzada y las técnicas de aprendizaje profundo han emergido como soluciones relevantes, al permitir la identificación de patrones complejos en datos estructurados y no estructurados, especialmente en tareas de clasificación de texto dentro del procesamiento del lenguaje natural (Domingos, 2012; LeCun et al. 2015). Estas técnicas facilitan la modelación de relaciones no lineales y la extracción automática de representaciones semánticas a partir de grandes volúmenes de información (Goodfellow et al. 2016; Raiaan et al. 2024). Sin embargo, su desempeño depende en gran medida de la adecuada configuración de los hiperparámetros, los cuales inciden directamente en la capacidad de generalización y la estabilidad del entrenamiento.

La selección de hiperparámetros en redes neuronales profundas constituye un problema complejo debido a la alta dimensionalidad del espacio de búsqueda y a la interacción no lineal entre los parámetros (Krizhevsky et al. 2012). Tradicionalmente, este proceso se ha abordado mediante estrategias empíricas basadas en prueba y error, lo que implica altos costos computacionales y una fuerte dependencia de la experiencia del investigador. En respuesta a estas limitaciones, la literatura reciente ha explorado el uso de algoritmos metaheurísticos como alternativa para automatizar la optimización, permitiendo explorar eficientemente espacios de búsqueda complejos sin requerir información explícita sobre la función objetivo (Alpaydin, 2021; Darwish et al. 2020). Entre estos, el algoritmo *Particle Swarm Optimization (PSO)* ha demostrado un desempeño destacado al equilibrar los procesos de exploración global y explotación local mediante la interacción de múltiples soluciones candidatas (Clerc, 2006).

Desde una perspectiva aplicada, la integración de técnicas de optimización evolutiva con modelos de aprendizaje profundo representa una oportunidad para mejorar los sistemas de clasificación automática en contextos empresariales (LeCun et al. 2015). En escenarios de inteligencia competitiva, la correcta categorización de noticias económicas facilita la identificación de tendencias, la detección temprana de riesgos y el análisis de dinámicas de mercado (Murphy, 2012). En este sentido, el presente estudio analiza el impacto de la optimización de hiperparámetros mediante PSO en el desempeño de un modelo de red neuronal profunda aplicado a la clasificación multiclase de noticias financieras. Para ello, se plantea un diseño experimental en tres etapas: (i) construcción de un modelo base, (ii)

optimización mediante PSO y (iii) comparación de resultados antes y después del proceso (Clerc, 2006). Metodológicamente, el análisis se basa en el corpus Reuters y emplea métricas como la exactitud y la F1 macro para evaluar tanto el desempeño global como la capacidad de clasificación en categorías minoritarias (Sebastiani, 2002).

1. Marco Teórico

1.1. Analítica de datos y aprendizaje profundo en contextos empresariales

El crecimiento exponencial de los datos en entornos organizacionales ha transformado la forma en que las empresas gestionan la información y toman decisiones. En particular, la disponibilidad de grandes volúmenes de datos no estructurados, como noticias económicas, informes financieros y documentos corporativos, ha generado la necesidad de desarrollar herramientas analíticas capaces de extraer conocimiento relevante de manera eficiente (Pedregosa, et al. 2011). En este contexto, la analítica avanzada y el uso de técnicas de inteligencia artificial se han consolidado como componentes fundamentales dentro de los sistemas de inteligencia empresarial (Chen et al. 2012; Davenport & Ronanki, 2018). Dentro de estas tecnologías, el aprendizaje profundo ha demostrado un alto potencial para el procesamiento de información textual, debido a su capacidad para modelar relaciones complejas mediante arquitecturas multicapa que permiten aprender representaciones jerárquicas de los datos (Goodfellow et al. 2016; Young et al. 2017). A diferencia de los enfoques tradicionales de aprendizaje automático, estos modelos reducen la dependencia de la ingeniería manual de características y logran capturar patrones semánticos de forma automática, lo que ha favorecido su aplicación en tareas como clasificación de documentos, análisis de sentimientos y minería de datos (Minaee et al. 2021; Liu, 2020; Devlin et al. 2019). En el ámbito empresarial, estas capacidades permiten transformar información textual en insumos estratégicos para la toma de decisiones. La clasificación automática de noticias económicas, por ejemplo, facilita la identificación de tendencias, el monitoreo de sectores productivos y la detección de eventos relevantes que pueden afectar el desempeño organizacional (Wamba et al. 2017; McAfee & Brynjolfsson, 2012). En este sentido, modelos como el *Multilayer Perceptron (MLP)*, cuando se aplican sobre representaciones vectoriales del texto, continúan siendo una alternativa eficiente para abordar problemas de clasificación

multiclase en contextos de analítica empresarial (Zhang et al. 2010; Jordan & Mitchell, 2015).

1.2.Optimización de hiperparámetros y algoritmos metaheurísticos

A pesar de los avances en modelos de aprendizaje profundo, su desempeño depende en gran medida de la adecuada configuración de hiperparámetros, los cuales determinan la capacidad de generalización, la estabilidad del entrenamiento y la calidad de las predicciones (Bengio, 2019; Chollet, 2021). Elementos como la tasa de aprendizaje, el número de neuronas, los mecanismos de regularización y el tamaño de los lotes de entrenamiento interactúan de manera compleja, lo que dificulta su ajuste mediante métodos tradicionales (Liu, 2020). Históricamente, la selección de estos parámetros se ha realizado mediante técnicas como búsqueda en cuadrícula o búsqueda aleatoria. Sin embargo, estos enfoques presentan limitaciones cuando el espacio de búsqueda es amplio o altamente no lineal, lo que ha motivado la exploración de alternativas más eficientes (Bergstra & Bengio, 2012; Snoek et al. 2012). En este contexto, los algoritmos metaheurísticos han emergido como una solución viable, al permitir la exploración simultánea de múltiples regiones del espacio de soluciones sin requerir supuestos estrictos sobre la función objetivo (Blum & Roli, 2003; Yang, 2021). Dentro de este grupo, el algoritmo *Particle Swarm Optimization (PSO)* (Clerc, 2006) ha sido ampliamente utilizado en la optimización de modelos de aprendizaje automático. Este método, inspirado en el comportamiento colectivo de sistemas naturales, permite equilibrar la exploración y la explotación del espacio de búsqueda mediante la interacción de múltiples soluciones candidatas (Kennedy & Eberhart, 1995; Poli et al. 2007). Su aplicación en redes neuronales profundas ha mostrado mejoras consistentes en el desempeño de los modelos, particularmente en escenarios donde la configuración de hiperparámetros resulta determinante (Krizhevsky, et al. 2012; Mirjalili et al. 2019; Raiaan et al. 2024). Desde una perspectiva empresarial, la optimización eficiente de modelos analíticos no solo mejora la precisión de las predicciones, sino que también contribuye a la construcción de sistemas más robustos y confiables, capaces de adaptarse a la complejidad de los entornos organizacionales (Talbi, 2009).

1.3. Evaluación del desempeño en modelos de clasificación multiclase

La evaluación del desempeño de modelos de clasificación multiclase constituye un aspecto fundamental en el análisis de su utilidad práctica. Aunque la exactitud es una métrica ampliamente utilizada, su capacidad para reflejar el comportamiento real del modelo es limitada en escenarios donde existe desbalance entre las clases (Sokolova & Lapalme, 2009). En este sentido, métricas como la precisión, el recall y la medida F1, particularmente en su versión macro, permiten evaluar el desempeño del modelo considerando de manera equilibrada todas las categorías (Powers, 2020; Chicco & Jurman, 2020). Este enfoque resulta especialmente relevante en contextos empresariales, donde la identificación de eventos poco frecuentes puede tener un impacto significativo en la toma de decisiones estratégicas (Pedregosa, et al. 2011). Adicionalmente, herramientas como la matriz de confusión permiten analizar los patrones de error del modelo, facilitando la identificación de categorías que presentan mayores dificultades de clasificación y orientando posibles mejoras en el diseño del sistema analítico (Aggarwal, 2018; Hastie, et al. (2009).

2. Metodología

2.1. Enfoque metodológico y diseño experimental

El estudio se desarrolla bajo un enfoque experimental, orientado a analizar cómo la optimización de hiperparámetros incide en el desempeño de modelos de aprendizaje profundo aplicados a la clasificación de información textual en contextos empresariales. Este tipo de aproximación permite examinar, en condiciones controladas, la relación entre la configuración del modelo y su capacidad para generalizar patrones en datos no estructurados, aspecto particularmente relevante en escenarios donde la información presenta alta dimensionalidad y distribución heterogénea entre clases. El diseño metodológico se organiza en tres momentos articulados. En una primera fase, se construye un modelo base de red neuronal profunda con una configuración inicial de hiperparámetros, lo que permite establecer una línea de referencia para el análisis. En una segunda fase, se implementa un proceso de optimización mediante el algoritmo *Particle Swarm Optimization (PSO)*, con el propósito de explorar de manera sistemática el espacio de soluciones. Finalmente, en una tercera fase, se comparan los resultados obtenidos antes y después del proceso de

optimización, lo que permite evaluar su impacto sobre el desempeño del modelo (Raiaan et al. 2024; Zeng et al. 2025). Para reducir posibles sesgos en la estimación de resultados, se adopta una estrategia de partición de datos en tres subconjuntos: entrenamiento, validación y prueba. Esta separación permite diferenciar claramente entre el proceso de ajuste de parámetros internos del modelo, la evaluación durante la optimización y la medición final del desempeño, siguiendo prácticas ampliamente aceptadas en el campo del aprendizaje automático (Goodfellow et al. 2016; Hutter, et al. 2019).

2.2. Datos y preparación de la información

El análisis se realiza a partir del conjunto de datos *Reuters Newswire Topics*, uno de los corpus más utilizados en investigaciones de procesamiento del lenguaje natural para tareas de clasificación automática de textos. Este conjunto de datos tiene su origen en la agencia internacional de noticias Reuters, reconocida por la producción de información financiera, económica y corporativa a escala global. El corpus está compuesto por noticias reales clasificadas en múltiples categorías temáticas, lo que lo convierte en una fuente pertinente para el estudio de problemas asociados a la organización y análisis de información empresarial.

La relevancia de este dataset radica no solo en su uso extendido dentro de la literatura científica, sino también en la naturaleza de la información que contiene. Al tratarse de noticias económicas y corporativas, permite aproximar escenarios reales en los que las organizaciones deben procesar grandes volúmenes de información textual para identificar tendencias, monitorear sectores productivos y apoyar la toma de decisiones estratégicas (Minaee et al. 2021; Chen et al. 2012). En este sentido, su utilización no se limita a un ejercicio técnico, sino que se vincula directamente con aplicaciones de inteligencia competitiva y analítica empresarial (Wamba et al. 2017). Adicionalmente, el uso de este corpus facilita la comparabilidad de resultados con investigaciones previas, dado que ha sido ampliamente documentado en estudios sobre clasificación de texto y aprendizaje profundo (Aggarwal, 2018; Zhang et al. 2010). Esta característica fortalece la validez externa del estudio y contribuye a la reproducibilidad del experimento. Dado que los documentos presentan longitudes variables, se requiere un proceso de transformación que permita convertirlos en representaciones numéricas de dimensión fija. Para ello, se emplea una

codificación multi-hot, mediante la cual cada documento se representa como un vector binario que indica la presencia o ausencia de términos dentro de un vocabulario previamente definido. Con el fin de mantener un equilibrio entre riqueza semántica y eficiencia computacional, el vocabulario se limita a las 10.000 palabras más frecuentes del corpus. De manera complementaria, las etiquetas de clase son transformadas mediante codificación one-hot, lo que permite representar cada categoría dentro de un espacio vectorial adecuado para el uso de funciones de pérdida en problemas de clasificación multiclase. Posteriormente, los datos son divididos en conjuntos de entrenamiento, validación y prueba, utilizados de manera diferenciada a lo largo del experimento.

2.3. Modelo base y estrategia de optimización

Como punto de partida, se implementa un modelo de red neuronal profunda de tipo Multilayer Perceptron (MLP), seleccionado por su capacidad para trabajar con representaciones vectoriales de alta dimensionalidad en problemas de clasificación de texto (Aggarwal, 2018; Minaee et al. 2021). Este tipo de arquitectura permite capturar relaciones no lineales entre las características del texto y las categorías objetivo, lo que resulta especialmente útil en escenarios donde los patrones semánticos son complejos. La arquitectura del modelo incluye dos capas ocultas con funciones de activación ReLU (Jordan & Mitchell, 2015), acompañadas de mecanismos de regularización mediante dropout, los cuales contribuyen a reducir el riesgo de sobreajuste durante el entrenamiento (Chollet, 2021). La capa de salida utiliza una función de activación softmax, permitiendo estimar la probabilidad de pertenencia de cada documento a las diferentes categorías del problema.

El proceso de entrenamiento se realiza mediante el optimizador Adam, ampliamente utilizado en aprendizaje profundo debido a su capacidad para ajustar dinámicamente la tasa de aprendizaje y favorecer una convergencia estable (Kingma & Ba, 2015). Como función de pérdida se emplea la entropía cruzada categórica, adecuada para problemas de clasificación multiclase. Con el fin de mejorar el desempeño del modelo, se incorpora un proceso de optimización de hiperparámetros basado en el algoritmo Particle Swarm Optimization. En este contexto, cada partícula representa una configuración específica del modelo, definida por parámetros como el número de neuronas en las capas ocultas, la tasa de aprendizaje, la proporción de dropout, la regularización y el tamaño del lote de entrenamiento (Kennedy &

Eberhart, 1995; Poli et al. 2007). El algoritmo opera mediante un proceso iterativo en el cual las partículas ajustan su posición en función de su mejor desempeño individual y del mejor desempeño global observado en el enjambre. Este mecanismo permite explorar el espacio de búsqueda de manera eficiente, identificando configuraciones que mejoran el desempeño del modelo en comparación con aquellas obtenidas mediante métodos tradicionales (Imran, et al. 2013; Mirjalili et al. 2019). Desde una perspectiva aplicada, este tipo de enfoque permite automatizar el ajuste de modelos analíticos, reduciendo la dependencia de procesos manuales y mejorando la calidad de los resultados.

2.4. Evaluación del desempeño del modelo

La evaluación del desempeño del modelo se realiza mediante un conjunto de métricas que permiten analizar tanto el rendimiento global como el comportamiento del sistema en cada categoría. En este sentido, se utilizan la exactitud (accuracy) y la medida F1 macro como indicadores principales. Mientras que la exactitud permite medir la proporción de predicciones correctas, la F1 macro resulta particularmente útil en escenarios donde existe desbalance entre clases, ya que evalúa el desempeño promedio considerando todas las categorías por igual (Sokolova & Lapalme, 2009; Powers, 2020). Esta característica es especialmente relevante en contextos empresariales, donde la correcta identificación de eventos poco frecuentes puede tener implicaciones significativas en la toma de decisiones. De manera complementaria, se emplea la matriz de confusión como herramienta de análisis, lo que permite identificar patrones de error y comprender con mayor detalle el comportamiento del modelo en cada categoría. Este tipo de análisis facilita la identificación de oportunidades de mejora en la configuración del modelo y en el proceso de entrenamiento (Aggarwal, 2018; Chicco & Jurman, 2020).

3. Hallazgos y discusión

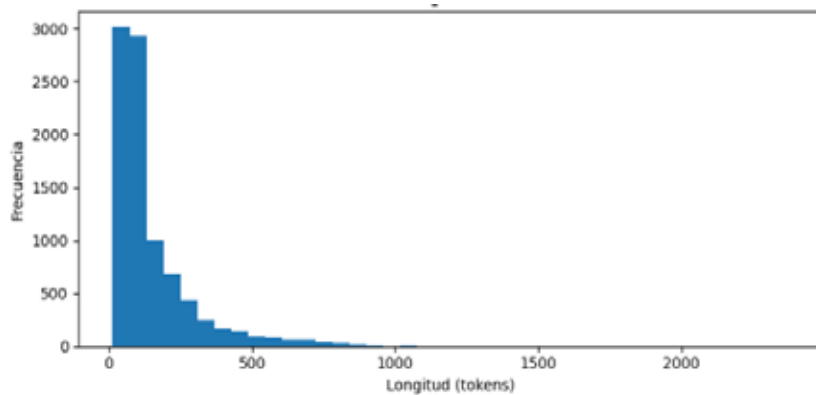
El análisis realizado sobre el corpus *Reuters Newswire Topics* permite aproximar un escenario que resulta pertinente para el estudio de la analítica textual en contextos empresariales. Al tratarse de noticias financieras y corporativas clasificadas en múltiples categorías, el conjunto de datos reproduce una situación cercana a la que enfrentan organizaciones que requieren procesar grandes volúmenes de información para identificar

tendencias, cambios sectoriales o eventos relevantes para la toma de decisiones. En este sentido, el problema no se limita a una tarea de clasificación automática, sino que se relaciona con la capacidad de transformar información dispersa en conocimiento útil para procesos de inteligencia competitiva y vigilancia del entorno.

Un primer hallazgo se relaciona con la estructura del conjunto de datos, cuya distribución de categorías presenta un comportamiento marcadamente heterogéneo: algunas clases concentran una alta proporción de documentos, mientras otras aparecen con una frecuencia reducida. Este patrón, común en problemas reales de clasificación textual, introduce una dificultad metodológica relevante, ya que los modelos tienden a aprender con mayor facilidad los patrones asociados a las clases dominantes y a mostrar menor sensibilidad frente a categorías minoritarias (Aggarwal, 2018). Desde una perspectiva empresarial, esta condición adquiere especial importancia, dado que los eventos menos frecuentes pueden ser estratégicamente más valiosos, por ejemplo, al anticipar cambios de mercado, riesgos reputacionales o transformaciones regulatorias. A esta situación se suma la variabilidad en la longitud de los documentos. La Figura 1 muestra una alta concentración de textos cortos y una disminución progresiva a medida que aumenta el número de tokens, lo que evidencia una distribución asimétrica característica de fuentes periodísticas económicas, donde coexisten noticias breves y documentos más extensos. Desde el punto de vista analítico, esta variabilidad requiere transformar los textos en representaciones de dimensión fija, dado que los modelos de aprendizaje automático no operan directamente sobre secuencias de longitud variable. En este contexto, la codificación multi-hot permite estandarizar la información de entrada, conservando la presencia léxica relevante de los documentos.

Figura 1.

Distribuciones de longitudes – Reuters Train

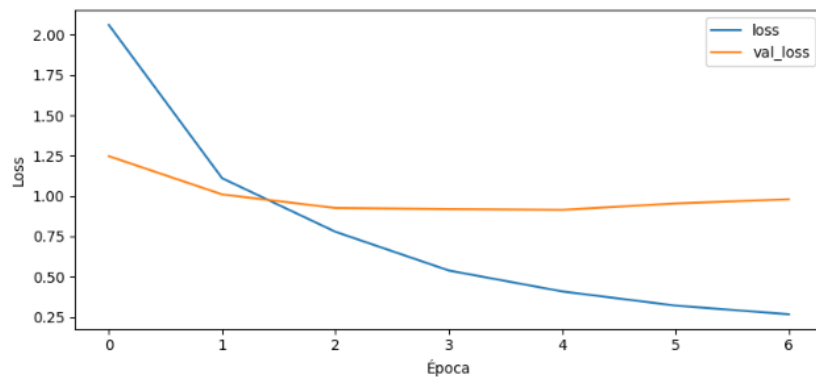


Nota: construcción propia

La Figura 2, muestra una disminución sostenida en el conjunto de entrenamiento, lo que indica que el modelo logra ajustarse progresivamente a los datos. Sin embargo, la pérdida de validación presenta una reducción inicial y posteriormente tiende a estabilizarse, evidenciando una desaceleración en la mejora fuera de muestra. Este comportamiento sugiere la presencia de un sobreajuste moderado, en el que el modelo comienza a especializarse en el conjunto de entrenamiento sin lograr una mejora proporcional en su capacidad de generalización.

Figura 2.

MLP baseline - Pérdida

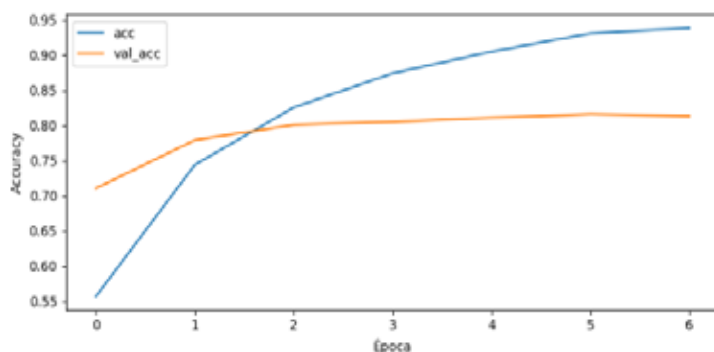


Nota: construcción propia

La Figura 3, centrada en la exactitud del modelo base, confirma el comportamiento observado en la función de pérdida. La exactitud en el conjunto de entrenamiento aumenta de forma sostenida, mientras que la exactitud de validación se estabiliza en etapas tempranas. Esta divergencia evidencia una brecha entre el ajuste interno del modelo y su desempeño fuera de muestra, lo que sugiere la presencia de un sobreajuste moderado. En consecuencia, aunque el modelo logra aprender patrones relevantes, no todos se traducen en una mejora equivalente en su capacidad de generalización, especialmente en contextos con alta dimensionalidad y desbalance entre clases (Chollet, 2021).

Figura 3.

MLP baseline - Exactitud



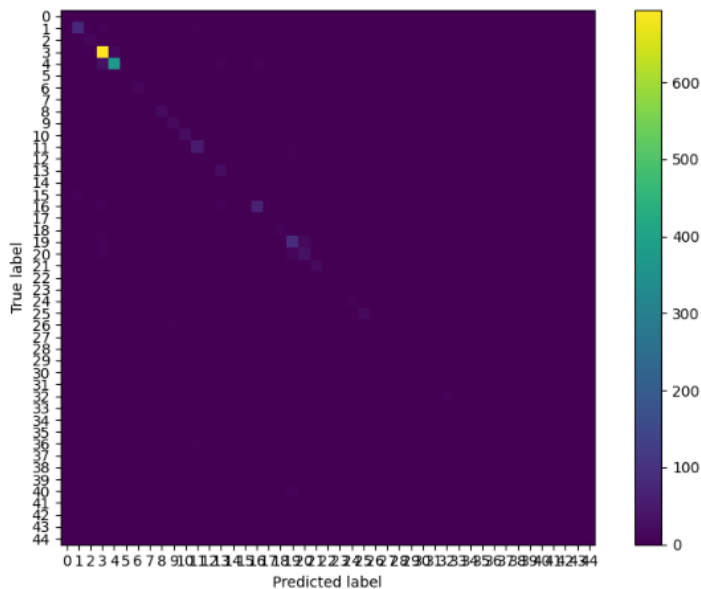
Nota: construcción propia

La Figura 4 presenta la matriz de confusión del modelo base, evidenciando una alta concentración de aciertos en la diagonal principal, junto con errores relevantes fuera de ella, especialmente en categorías minoritarias. Este comportamiento refleja un desempeño desigual entre clases, donde las más frecuentes son mejor clasificadas, mientras que las menos representadas presentan mayores dificultades, fenómeno típico en problemas con datos desbalanceados. En consecuencia, la exactitud global no logra capturar completamente estas limitaciones. Asimismo, se observan confusiones entre categorías con similitudes semánticas, lo que indica restricciones en la capacidad del modelo para discriminar ciertos patrones. Desde una perspectiva aplicada, estas limitaciones son críticas en contextos empresariales, donde la detección de eventos poco frecuentes resulta clave. Por ello, se

justifica el uso de métricas como la F1 macro para una evaluación más equilibrada del desempeño (Powers, 2020).

Figura 4.

Matriz de confusión – MLP Baseline



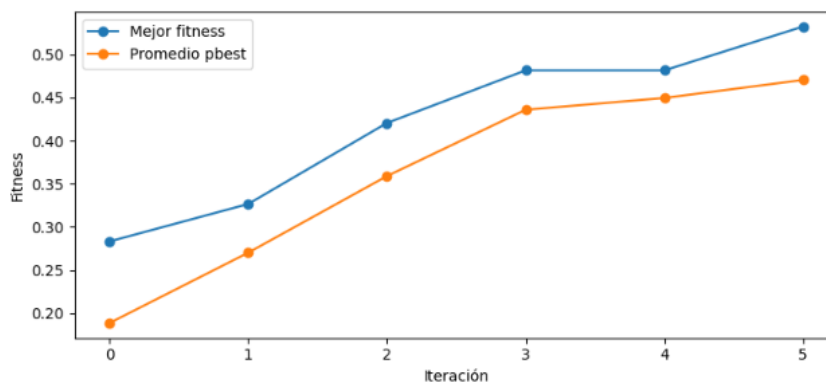
Nota: construcción propia

Con este punto de partida, se implementó la optimización de hiperparámetros mediante *Particle Swarm Optimization*. La Figura 5 permite observar la evolución del fitness a lo largo de las iteraciones del algoritmo. La tendencia creciente de esta curva muestra que el proceso de búsqueda no fue aleatorio ni errático, sino que logró orientar progresivamente el enjambre hacia configuraciones de mayor calidad. Este comportamiento es consistente con las propiedades reportadas en la literatura sobre PSO, donde la combinación entre exploración inicial y convergencia posterior favorece la identificación de regiones promisorias del espacio de búsqueda (Kennedy & Eberhart, 1995; Imran, et al. 2013). Lo interesante de este punto no es solo que el fitness aumente, sino lo que ese aumento representa. El algoritmo fue ajustando configuraciones asociadas a número de neuronas, *dropout*, tasa de aprendizaje, regularización y tamaño del lote, es decir, variables que inciden directamente en la complejidad del modelo y en la estabilidad de su entrenamiento. La lectura conjunta de estos

resultados sugiere que el PSO ayudó a encontrar un equilibrio más favorable entre capacidad de representación y control del sobreajuste. Desde una mirada práctica, esto significa que el modelo no mejoró simplemente por “entrenar más”, sino por entrenar bajo una configuración más apropiada para la naturaleza del problema.

Figura 5.

PSO – Fitness por iteración



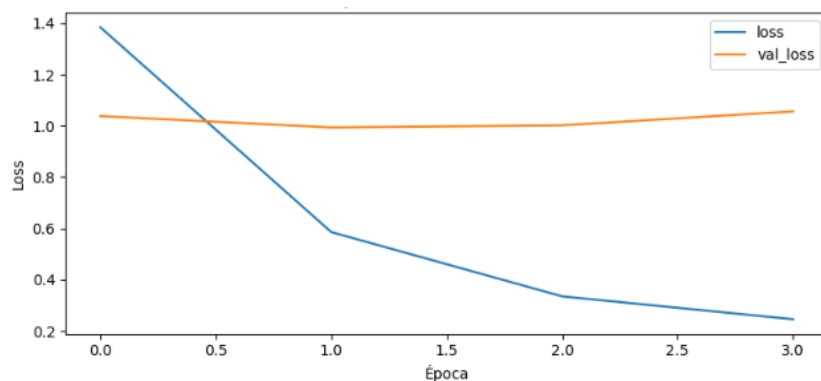
Nota: construcción propia

La Figura 6, correspondiente a la pérdida del modelo optimizado mediante PSO, permite observar una dinámica de aprendizaje más estable en comparación con el modelo base. La pérdida en el conjunto de entrenamiento disminuye de manera sostenida a lo largo de las épocas, lo que indica que el modelo continúa ajustándose progresivamente a los datos. Por su parte, la pérdida de validación muestra una leve reducción en las primeras iteraciones y posteriormente tiende a estabilizarse, sin presentar incrementos abruptos. Aunque persiste una diferencia moderada entre ambas curvas, esta resulta menos pronunciada, lo que sugiere una mejora en la capacidad de generalización del modelo. Este comportamiento contrasta con el observado en la fase inicial, donde la divergencia entre entrenamiento y validación era más evidente. En este caso, la trayectoria de la pérdida refleja un proceso de aprendizaje más equilibrado, en el que el modelo logra reducir el error sin comprometer significativamente su desempeño fuera de muestra. La estabilidad de la curva de validación, incluso en etapas posteriores del entrenamiento, indica que la configuración de hiperparámetros obtenida mediante PSO favorece un ajuste más controlado, evitando que el modelo se especialice en

exceso en los datos de entrenamiento. Desde una perspectiva analítica, estos resultados sugieren que la optimización no solo permitió mejorar el desempeño del modelo, sino también regular su comportamiento durante el proceso de aprendizaje. En lugar de forzar una reducción agresiva del error en entrenamiento, la configuración encontrada parece propiciar un equilibrio más adecuado entre capacidad de representación y control del sobreajuste.

Figura 6.

MLP optimizado con PSO - Pérdida

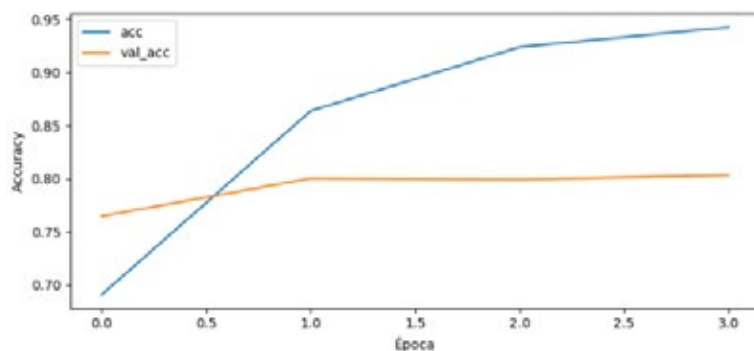


Nota: construcción propia

Por su parte, en la Figura 7 se observa que la exactitud en el conjunto de entrenamiento aumenta de forma sostenida a lo largo de las épocas, alcanzando valores cercanos a 0.95, lo que indica una adecuada capacidad del modelo para ajustarse a los datos. Por su parte, la exactitud en validación presenta una mejora inicial, pero tiende a estabilizarse alrededor de 0.80, evidenciando una brecha entre ambos comportamientos. Esta diferencia sugiere la presencia de un sobreajuste moderado, en el cual el modelo aprende patrones específicos del conjunto de entrenamiento que no se traducen completamente en una mejora equivalente en datos no vistos. No obstante, la estabilidad de la curva de validación indica que el modelo mantiene un desempeño consistente, lo que respalda la pertinencia de aplicar estrategias de optimización para mejorar su capacidad de generalización.

Figura 7.

MLP optimizado con PSO - Exactitud



Nota: construcción propia

La síntesis cuantitativa más clara de esta evolución se encuentra en la Tabla 1. Allí se observa que la exactitud global pasa de 0.7935 a 0.7990. A primera vista, el incremento puede parecer modesto. No obstante, la lectura más relevante no está en esa cifra aislada, sino en las métricas macro. La precisión macro aumenta de 0.5941 a 0.6516, el recall macro de 0.4108 a 0.4943 y la F1 macro de 0.4528 a 0.5338. Este último resultado es especialmente importante, porque refleja una mejora relativa cercana al 18 % en la capacidad del modelo para clasificar de manera más equilibrada el conjunto de categorías. En problemas desbalanceados, esta mejora tiene mayor valor analítico que una variación marginal en la exactitud global.

Tabla 1.

Comparación del modelo

Modelo	Accuracy	Precision_macro	Recall_macro	F1_macro
MLP optimizada con PSO	0.7990	0.6516	0.4943	0.5338
MLP baseline	0.7935	0.5941	0.4108	0.4528

Nota: construcción propia

Desde una perspectiva empresarial, estos hallazgos son significativos. Un sistema de clasificación textual orientado a apoyar procesos de análisis estratégico no puede juzgarse

únicamente por su capacidad para acertar en las categorías más frecuentes. Su valor también depende de qué tan bien logra reconocer temas menos dominantes, pero relevantes para la vigilancia del entorno. En este sentido, la mejora en F1 macro sugiere que la optimización de hiperparámetros fortalece la utilidad del modelo para contextos donde se requiere una lectura más balanceada de la información disponible. La comparación entre el modelo base y el modelo optimizado permite afirmar que el uso de PSO contribuyó a reducir parcialmente las limitaciones de la configuración inicial, mejorar la estabilidad del entrenamiento y elevar el desempeño en métricas que capturan con mayor precisión la complejidad del problema. Dicho de otra forma, el aporte principal del estudio no radica solo en demostrar que un modelo clasifica noticias financieras, sino en mostrar que la calidad de esa clasificación depende de manera importante de la estrategia con la que se aborda la optimización del modelo.

4. Conclusiones

Los resultados permiten concluir que el desempeño de los modelos de aprendizaje profundo aplicados a la clasificación de información textual depende no solo de la arquitectura utilizada, sino, en gran medida, de la adecuada configuración de sus hiperparámetros. Aunque el modelo base mostró capacidad para capturar patrones relevantes, evidenció limitaciones en términos de generalización y en la clasificación de categorías minoritarias, en línea con lo reportado en la literatura (Goodfellow et al. 2016; Chollet, 2021). La incorporación del algoritmo Particle Swarm Optimization permitió mejorar el equilibrio entre capacidad de representación y regularización, generando avances consistentes en el desempeño del modelo. Más allá de la exactitud global, la mejora en métricas como la F1 macro evidencia una mayor capacidad para clasificar de manera equilibrada las diferentes categorías, aspecto especialmente relevante en contextos empresariales donde la identificación de eventos poco frecuentes puede ser determinante para la toma de decisiones. Desde una perspectiva metodológica, los resultados destacan la complementariedad entre los métodos de optimización por gradiente y los algoritmos metaheurísticos, al combinar ajuste interno del modelo con exploración eficiente del espacio de soluciones. En conjunto, los hallazgos refuerzan la importancia de considerar la optimización de hiperparámetros como un componente central en el diseño de modelos de aprendizaje profundo, dado que su

adecuada implementación no solo mejora el desempeño técnico, sino también la utilidad práctica de los sistemas en el análisis automatizado de información empresarial.

Referencias

- Aggarwal, C. C. (2018). Machine learning for text. *Springer*. <https://doi.org/10.1007/978-3-319-94463-0>
- Alpaydin, E. (2021). *Machine learning* (4th ed.). MIT Press.
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798–1828. <https://doi.org/10.1109/TPAMI.2013.50>
- Bergstra, J. and Bengio, Y. (2012) Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13, 281-305
- Blum, C., & Roli, A. (2003). Metaheuristics. *ACM Computing Surveys*, 35(3), 268–308. <https://doi.org/10.1145/937503.937505>
- Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine*, 9(2), 48-57.
- Chen, H., Chiang, R., & Storey, V. (2012). Business intelligence. *MIS Quarterly*, 36(4), 1165–1188.
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21, 6. <https://doi.org/10.1186/s12864-019-6413-7>
- Chollet, F. (2021). *Deep learning with Python* (2nd ed.). Simon and Schuster.
- Clerc, M. (2006). *Particle swarm optimization*. John Wiley & Sons.
- Davenport, T. H., & Ronanki, R. (2018). Artificial Intelligence for the Real World. *Harvard Business Review*, 96(1), 108–116.
- Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. <https://doi.org/10.48550/arXiv.1810.04805>

- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78–87.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning: Adaptive Computation and Machine Learning series. MIT Press.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). Elements of statistical learning: Data Mining, Inference, and Prediction. Springer.
- Hinton, G., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*. <http://dx.doi.org/10.1162/neco.2006.18.7.1527>
- Hutter, F., Kotthoff, L., & Vanschoren, J. (2019). Automated Machine Learning: Methods, Systems, Challenges. Springer. https://doi.org/10.1007/978-3-030-05318-5_1
- Imran, M., Hashim, R., & Abd Khalid, N., (2013). An Overview of Particle Swarm Optimization Variants. 53 (491-496). *Research Management & Innovation Centre, Universiti Malaysia Perlis*. <https://doi.org/10.1016/j.proeng.2013.02.063>
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260. <https://doi.org/10.1126/science.aaa8415>
- Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization. IEEE. <https://doi.org/10.1109/ICNN.1995.488968>
- Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. <https://doi.org/10.48550/arXiv.1412.6980>
- Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60, 84 - 90.
- LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep Learning. *Nature*, 521, 436-444. <http://dx.doi.org/10.1038/nature14539>.
- Liu, B. (2020). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge University Press.
- McAfee, A.P., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, 90 10, 60-6, 68, 128.
- Murphy, K. (2012) Machine Learning. A Probabilistic Perspective. MIT Press, Cambridge, Massachusetts, London

- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011) Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- Poli, R., Kennedy, J. and Blackwell, T. (2007) Particle Swarm Optimization: An Overview. *Swarm Intelligence*, 1, 33- 57. <http://dx.doi.org/10.1007/s11721-007-0002-0>
- Powers, D.M. (2011) Evaluation: From Precision, Recall and f-Measure to Roc, Informedness, Markedness and Correlation. *Journal of Machine Learning Technologies*, 2, 37-63. <https://doi.org/10.48550/arXiv.2010.16061>.
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media.
- Raiaan, M. A. K., Sakib, S., Fahad, N. M., Mamun, A. A., Rahman, M. A., Shatabda, S., & Mukta, M. S. H. (2024). A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks. *Decision Analytics Journal*, 11, 1-32. Article 100470. <https://doi.org/10.1016/j.dajour.2024.100470>
- Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, 34(1), 1–47. <https://doi.org/10.1145/505282.505283>
- Snoek, J., Larochelle, H. and Adams, R.P. (2012) Practical Bayesian Optimization of Machine Learning Algorithms. *Advances in Neural Information Processing Systems*, 5, 2951-2959. <https://doi.org/10.48550/arXiv.1206.2944>
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Talbi, E. G. (2009). *Metaheuristics: from design to implementation*. John Wiley & Sons, Inc., Hoboken, New Jersey
- Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J., Dubey, R., & Childe, S. J. (2017). Big Data Analytics and Firm Performance: Effects of Dynamic Capabilities. *Journal of Business Research*, 70, 356-365. <https://doi.org/10.1016/j.jbusres.2016.08.009>.
- Yang, X. S. (2021). *Nature-Inspired Optimization Algorithms*. Elsevier. Second Edition.

- Young, T., Hazarika, D., Poria, S., & Cambria, E. (2017). Recent Trends in Deep Learning Based Natural Language Processing. *IEEE Comput. Intell. Magazine.*, 13, 55-75.
- Zhang, Y., Jin, R. and Zhou, Z.H., (2010). Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1-4), 43-52..
- Zhou, Z. H. (2012). *Ensemble Methods: Foundations and Algorithms*. CRC Press.
- Zeng B, Shang X, Lu R, Zhang Y (2025) Particle swarm optimization-based NLP methods for optimizing automatic document classification and retrieval. *PLoS One* 20(7): e0325851. <https://doi.org/10.1371/journal.pone.0325851>