



UNIVERSIDAD AUTÓNOMA METROPOLITANA

Unidad Azcapotzalco

División de Ciencias Básicas e Ingeniería

**Calibración de Hiper-Parámetros en Algoritmos
Metaheurísticos y Modelos de Lenguaje para la
Detección de Noticias Falsas**

Idónea Comunicación de Resultados

que presenta el:

Ing. Gabriel Hurtado Avilés

para obtener el grado de:

Maestro en Ciencias de la Computación

Directores:

Dr. José Alejandro Reyes Ortiz

Dr. Román Anselmo Mora Gutiérrez

Ciudad de México

Septiembre 2025

Resumen

Este proyecto aborda el desafío de la detección de noticias falsas en español mediante la aplicación y comparación de dos metodologías de inteligencia artificial. Aunque las técnicas desarrolladas podrían aplicarse a otros tipos de fraude digital, el enfoque específico se centra en la identificación de contenido noticioso falso o engañoso. La primera metodología explora el uso de algoritmos metaheurísticos, incluyendo Recocido Multiarranque (MSA), Búsqueda Dispersa (SS), Búsqueda en Vecindades Variables (VNS), Algoritmo Genético (GA) y Optimización por Enjambre de Partículas (PSO), sobre una representación TF-IDF. La segunda, adoptada tras los hallazgos iniciales, se basa en el ajuste fino (fine-tuning) de un modelo de lenguaje Transformer pre-entrenado (DistilBERT). Para el entrenamiento, se construyó un corpus unificando cuatro conjuntos de datos públicos en español y datos extraídos mediante extracción web del portal satírico “El Deforma”, resultando en más de 61,000 noticias. Se implementó un cuidadoso proceso de calibración de hiperparámetros para ambos enfoques, utilizando una división de datos estratificada de 70 % para entrenamiento, 10 % para validación y 20 % para pruebas. El rendimiento fue evaluado con métricas como Exactitud, Precisión, Exhaustividad y F1-Score. Finalmente, el modelo Transformer, que demostró una eficacia superior, fue integrado en una aplicación web funcional desarrollada con Flask y contenerizada con Docker, capaz de analizar URLs en tiempo real. Los resultados validan la metodología de ajuste fino como una solución de vanguardia para combatir la desinformación, superando a los enfoques metaheurísticos en esta tarea.

Palabras clave: Detección de noticias falsas, desinformación, modelos de lenguaje, Transformers, DistilBERT, algoritmos metaheurísticos, procesamiento de lenguaje natural.

Abstract

This project addresses the challenge of fake news detection in Spanish through the application and comparison of two artificial intelligence methodologies. Although the developed techniques could be applied to other types of digital fraud, the specific focus centers on identifying false or misleading news content. The first explores the use of metaheuristic algorithms, including Multi-Start Simulated Annealing (MSA), Scatter Search (SS), Variable Neighborhood Search (VNS), Genetic Algorithm (GA), and Particle Swarm Optimization (PSO), applied to a TF-IDF representation. The second, adopted following initial findings, is based on fine-tuning a pre-trained Transformer language model (DistilBERT). For training, a corpus was constructed by unifying four public Spanish datasets and data extracted through web scraping from the satirical portal “El Deiforma”, resulting in over 61,000 news articles. A careful hyperparameter calibration process was implemented for both approaches, using a stratified data split of 70 % for training, 10 % for validation, and 20 % for testing. Performance was evaluated using metrics such as Accuracy, Precision, Recall, and F1-Score. Finally, the Transformer model, which demonstrated superior efficacy, was integrated into a functional web application developed with Flask and containerized with Docker, capable of analyzing URLs in real-time. The results validate the fine-tuning methodology as a state-of-the-art solution for combating misinformation, outperforming metaheuristic approaches in this task.

Keywords: Fake news detection, misinformation, language models, Transformers, DistilBERT, metaheuristic algorithms, natural language processing.

Dedicatoria

// Esta sección se presenta en la versión final de su ICR. Son frases cuyo objetivo es otorgar una mención especial a las personas que te han motivado durante tu ICR.

Agradecimientos

// Esta sección se presenta en la versión final de su ICR. Son frases cuyo objetivo es plasmar el apoyo moral, físico, económico y/o emocional que recibió de las personas o instituciones durante la elaboración de todo su proyecto.

Índice general

Índice general	VII
Índice de figuras	XIII
Índice de tablas	XV
Glosario	XIX
1. Introducción	1
1.1. Planteamiento del problema	1
1.1.1. El Desafío Específico del Español: Enfoque Multiregional	3
1.1.2. La Complejidad Técnica del Problema	5
1.2. Motivación	5
1.2.1. Impacto Personal y Social Observado	6
1.2.2. Vulnerabilidad de Poblaciones Específicas	6
1.2.3. Urgencia Tecnológica	7
1.3. Justificación	7
1.3.1. Brecha Tecnológica en Recursos para el Español	7
1.3.2. Novedad Metodológica: Enfoque Evolutivo Comparativo	8
1.3.3. Contribución Científica Multidimensional	8
1.4. Objetivos	9
1.5. Alcance y Limitaciones	10
1.5.1. Alcance	10
1.5.2. Limitaciones	11
1.6. Contribuciones Esperadas	12
1.6.1. Contribuciones Teóricas	12
1.6.2. Contribuciones Prácticas	12
1.6.3. Contribuciones Sociales	13
1.7. Organización del Documento	13
2. Marco Teórico	15
2.1. Detección de Noticias Falsas: Fundamentos y Extensibilidad	15
2.1.1. Impacto Social y Desafíos de la Desinformación	16
2.1.2. El Problema de las Heurísticas Cognitivas	16

2.1.3. Diferenciación Técnica y Transferibilidad: Noticias Falsas vs. Fraude Digital	16
2.2. Representación de Texto: Desde TF-IDF hasta Embeddings Contextuales	18
2.2.1. Bolsa de Palabras (Bag of Words): Primer Acercamiento Metodológico	18
2.2.2. TF-IDF (Term Frequency-Inverse Document Frequency)	20
2.2.3. Fundamentos Matemáticos de TF-IDF	20
2.2.4. Limitaciones de los Métodos Clásicos	20
2.3. La Revolución Transformer y los Modelos de Lenguaje Modernos	21
2.3.1. La Arquitectura Transformer: Fundamentos	21
2.3.2. BERT y la Era del Pre-entrenamiento Bidireccional	21
2.3.3. Aplicaciones en Detección de Noticias Falsas en Español	22
2.3.4. Grandes Modelos de Lenguaje (LLMs) y Nuevos Paradigmas	22
2.3.5. El Doble Rol de los LLMs en Detección	23
2.4. Optimización Metaheurística en Detección de Fraude	23
2.4.1. Fundamentos de las Metaheurísticas	23
2.4.2. Aplicaciones en Detección de Noticias Falsas	33
2.4.3. Metaheurísticas para Detección de Fraude Financiero	34
2.4.4. Algoritmos Híbridos Modernos	35
2.5. Síntesis del Marco Teórico	36
3. Estado del arte	37
3.1. Metodología de Búsqueda Bibliográfica	37
3.2. Panorama de la Investigación en Detección de Noticias Falsas y Modelos de Lenguaje	39
3.2.1. Fundamentos Metodológicos y Construcción de Corpus	39
3.2.2. Evolución de los Modelos de Lenguaje en Detección de Desinformación	41
3.2.3. Técnicas de Optimización Metaheurística en Detección de Desinformación	44
3.3. Algoritmos Metaheurísticos: Fundamentos Teóricos y Aplicaciones Prácticas	47
3.3.1. Algoritmos Genéticos (GA): Evolución Artificial para Optimización	47
3.3.2. Optimización por Enjambre de Partículas (PSO): Inteligencia Colectiva	48
3.3.3. Recocido Simulado: Física Estadística para Optimización	48
3.3.4. Búsqueda Dispersa (Scatter Search): Combinación Determinística	49
3.3.5. Búsqueda en Vecindades Variables (VNS): Exploración Sistématica	49
3.4. Hiperparámetros en Modelos de Aprendizaje Automático: Fundamentos y Optimización	51
3.4.1. Naturaleza y Definición de Hiperparámetros	51

3.4.2. Importancia Crítica en la Detección de Noticias Falsas	52
3.4.3. Desafíos en la Búsqueda de Configuración Óptima	53
3.4.4. Herramientas Modernas: Keras Tuner y Automatización	53
3.4.5. Beneficios combinados: Modelos de Lenguaje y Herramientas de Optimización	54
3.5. Análisis de Literatura Relevante	55
3.5.1. El Desafío Fundamental de los Datos: Creación y Curación de Corpus en Español	55
3.5.2. Revolución de los Modelos de Lenguaje y Arquitecturas Transformer	56
3.5.3. Optimización y Metaheurísticas en la Detección: Enfoques Innovadores	57
3.5.4. Aplicación de Metaheurísticas por Tipo de Detección	57
3.5.5. Perspectivas Interdisciplinarias y Análisis Social	59
3.5.6. Detección de Fraude: Extensión Más Allá de las Noticias	60
3.6. Síntesis y Perspectivas Futuras	61
4. Metodología	63
4.1. Visión General del Proceso Metodológico	63
4.2. Definición y Distinción de Conceptos Fundamentales	64
4.2.1. Distinción entre Noticia Falsa y Bulo	64
4.2.2. Taxonomía de la Desinformación	65
4.2.3. Justificación de la Unificación Terminológica	65
4.2.4. Esquema de Clasificación: Enfoque Binario	66
4.2.5. Fronteras Difusas en la Clasificación de Veracidad	67
4.3. Construcción del Corpus Unificado	68
4.3.1. Fuentes de Datos Académicas	68
4.3.2. Análisis Comparativo Detallado de los Corpus Utilizados	70
4.3.3. Proceso de Unificación y Estandarización	73
4.3.4. Ampliación del Corpus Mediante Extracción Web	73
4.3.5. Corpus Final y Estrategia de División	75
4.4. Enfoque 1: Detección Mediante Algoritmos Metaheurísticos	80
4.4.1. Preprocesamiento y Representación Textual	80
4.4.2. Algoritmos Metaheurísticos Implementados	81
4.4.3. Función de Evaluación y Clasificación	84
4.4.4. Reducción de Dimensionalidad	84
4.5. Enfoque 2: Detección Mediante Modelo Transformer	85
4.5.1. Selección y Justificación del Modelo	85
4.5.2. Infraestructura Computacional	85
4.5.3. Configuración Experimental y Optimización de Hiperparámetros	86
4.6. Metodología de Evaluación Comparativa	87
4.6.1. Protocolo de Evaluación	87
4.6.2. Fundamentos de la Matriz de Confusión	88

4.6.3.	Marco de Métricas de Rendimiento	88
4.6.4.	Criterios de Selección del Mejor Modelo	92
4.6.5.	Ánálisis de Matrices de Confusión	93
4.6.6.	Reporte de Resultados	93
4.7.	Infraestructura Computacional y Herramientas	93
4.7.1.	Entorno de Desarrollo para Algoritmos Metaheurísticos	93
4.7.2.	Stack Tecnológico Completo	94
4.8.	Consideraciones Éticas y de Privacidad	94
4.8.1.	Marco Ético de Desarrollo	95
4.8.2.	Limitaciones Declaradas y Uso Responsable	95
4.9.	Validación y Reproducibilidad	95
4.9.1.	Ecosistema de Artefactos de Reproducibilidad	95
4.9.2.	Solución Lista para Producción	96
5.	Resultados y Evaluación Comparativa	97
5.1.	Validación del Flujo de Datos y Representaciones	97
5.1.1.	Características del Corpus Unificado Final	97
5.1.2.	Implementación y Configuración de Algoritmos Metaheurísticos	99
5.1.3.	Pseudocódigos de los Algoritmos Implementados	99
5.1.4.	Visualizaciones de Resultados por Algoritmo	104
5.1.5.	Contextualización con Investigación Publicada	112
5.1.6.	Análisis Comparativo de Resultados	113
5.1.7.	Limitaciones Fundamentales y Justificación para Evolución .	114
5.1.8.	Síntesis y Transición	115
5.2.	Resultados del Enfoque Transformer: DistilBERT Multilingüe . . .	116
5.2.1.	Marco Experimental y Evolución del Desarrollo	116
5.2.2.	Configuración del Modelo DistilBERT Optimizado	117
5.2.3.	Proceso de Optimización y Búsqueda de Hiperparámetros .	118
5.2.4.	Ánálisis de Convergencia y Control de Overfitting	118
5.2.5.	Evolución Experimental: Versiones de Desarrollo	120
5.2.6.	Pseudocódigo del Algoritmo de Entrenamiento DistilBERT .	125
5.2.7.	Resultados Finales y Comparación	127
5.3.	Evaluación Comparativa Integral: Metaheurísticas vs. Transformers .	128
5.3.1.	Comparación de Rendimiento Cuantitativo	128
5.3.2.	Ánálisis Multidimensional de Paradigmas	128
5.3.3.	Justificación de la Selección del Modelo Final	129
5.3.4.	Valor Científico del Enfoque Comparativo	129
5.4.	Ánálisis de Errores y Limitaciones del Modelo Final	130
5.4.1.	Marco Teórico para el Análisis de Errores	130
5.4.2.	Metodología Propuesta para Análisis Cualitativo	130
5.4.3.	Tipos de Errores Esperados según la Literatura	131
5.4.4.	Limitaciones Reconocidas del Modelo	132
5.4.5.	Conclusiones sobre Limitaciones y Direcciones Futuras . . .	132

6. Implementación de Prototipos Funcionales	135
6.1. Introducción a la Fase de Implementación	135
6.2. Arquitectura General del Sistema	135
6.3. Prototipo 1: Analizador Basado en Metaheurísticas	136
6.3.1. Componentes del Modelo	136
6.3.2. Flujo de Inferencia	137
6.4. Prototipo 2: Analizador Basado en Modelos Transformer (Versión Final)	137
6.4.1. Componentes del Modelo	137
6.4.2. Flujo de Inferencia	138
6.5. Interfaz de Usuario y Casos de Uso	138
6.5.1. Caso de Uso 1: Detección de una Noticia Real	138
6.5.2. Caso de Uso 2: Detección de una Página con Contenido Engañoso	138
6.5.3. Caso de Uso 3: Detección de una Página Fraudulenta	138
7. Conclusiones	141
7.1. Resumen del Trabajo y Contribuciones Principales	141
7.2. Limitaciones del Estudio	142
A. Anexo 1	145
Bibliografía	146

Índice de figuras

1.1. Mapa Conceptual 1: Taxonomía del problema de desinformación con enfoque en noticias falsas y extensibilidad hacia fraude digital.	3
1.2. Mapa Conceptual 2: Estrategias tecnológicas para la detección de fraude digital.	6
2.1. Flujo del proceso de optimización para detección de noticias falsas. Se muestra la transformación desde datos textuales hasta el modelo optimizado final, pasando por las etapas de preprocesamiento, optimización y evaluación.	32
3.1. Mapa Conceptual 3: Artículos relacionados que incorporan Modelos de Lenguaje.	42
3.2. Mapa Conceptual 4: Métodos de optimización y metaheurísticas aplicadas.	46
3.3. Mapa Conceptual 5: Clasificación de artículos por enfoque metodológico.	54
3.4. Mapa Conceptual 6: Artículos que son revisiones o están relacionados al análisis de contenido y detección de fraude digital.	61
4.1. Metodología propuesta que aborda la problemática combinando Algoritmos Metaheurísticos y Modelos de Lenguaje.	64
4.2. Representación gráfica del proceso de balanceo del corpus: comparación entre distribución inicial desbalanceada y distribución final equilibrada. La imagen muestra cómo la extracción web de 9,000 noticias falsas permitió alcanzar un equilibrio óptimo de 49.8 % noticias falsas vs 50.2 % noticias reales.	76
5.1. Evolución de la convergencia del algoritmo MSA mostrando el progreso gradual a través de los 31 niveles de temperatura desde 1000 hasta 1.24.105	105
5.2. Matriz de confusión para MSA en el conjunto de pruebas, evidenciando la baja especificidad (33 %) y el sesgo hacia la clasificación como noticias reales.	106
5.3. Convergencia eficiente del algoritmo SS en solo 10 iteraciones, mostrando mejoras progresivas en las iteraciones 4, 5 y 7 hasta estabilizarse en 0.6630.	107

5.4. Matriz de confusión para SS demostrando mejor balance que MSA con especificidad del 40 % y excelente generalización.	107
5.5. Evolución darwiniana del algoritmo GA a lo largo de 20 generaciones, evidenciando progreso sostenido desde 0.6198 hasta 0.7090 con hitos evolutivos significativos.	108
5.6. Matriz de confusión para GA mostrando el mejor balance global con especificidad líder del 48 % y rendimiento sólido en ambas clases. . . .	109
5.7. Progreso sistemático del algoritmo VNS a través de 20 iteraciones con cambios efectivos de vecindario, mostrando saltos significativos en las iteraciones 6 y 14.	110
5.8. Matriz de confusión para VNS destacando la excelente exhaustividad del 89 % para detección de noticias reales con especificidad competitiva del 41 %.	110
5.9. Convergencia problemática del algoritmo PSO evidenciando estancamiento prematuro en la iteración 7-8 y exploración insuficiente del espacio de búsqueda.	111
5.10. Matriz de confusión para PSO revelando el comportamiento extremo problemático con especificidad crítica del 15 % y sesgo severo hacia la clase mayoritaria.	112
5.11. Evolución de la exactitud y pérdida durante el entrenamiento del modelo DistilBERT V7. Las líneas azul y roja muestran la convergencia en entrenamiento y validación respectivamente. La estrella dorada marca la mejor época (13), después de la cual se observa el inicio del overfitting con una separación creciente entre las curvas.	119
5.12. Evolución de exactitud a través de las versiones experimentales. . . .	124
6.1. Captura de pantalla de la aplicación analizando una noticia real. . . .	139
6.2. Captura de pantalla de la aplicación detectando una noticia falsa basada en su contenido.	139
6.3. Captura de pantalla de la aplicación detectando una página de fraude digital.	140

Índice de tablas

2.1. Diferencias técnicas entre dominios de desinformación	17
2.2. Comparación de algoritmos metaheurísticos implementados	30
3.1. Contribuciones fundamentales en metodología y construcción de corpus para español.	41
3.2. Evolución de modelos de lenguaje aplicados a detección de desinformación.	44
3.3. Contribuciones metodológicas en optimización metaheurística para detección.	45
3.4. Algoritmos metaheurísticos: características fundamentales y aplicaciones en detección de noticias falsas.	50
3.5. Aplicación de metaheurísticas por tipo de detección en la literatura revisada.	58
4.1. Corpus académicos utilizados para la construcción del conjunto de datos unificado.	69
4.2. Comparación exhaustiva de características de los corpus utilizados en la construcción del conjunto de datos unificado.	70
4.3. Fases del proceso de extracción web implementado para “El Deforma”. .	74
4.4. Referencias bibliográficas completas de los corpus académicos utilizados.	74
4.5. Composición final del corpus unificado después del procesamiento completo.	75
4.6. División estratificada principal del corpus para entrenamiento y evaluación.	79
4.7. Algoritmos metaheurísticos implementados y sus fundamentos conceptuales.	81
4.8. Configuración de parámetros del algoritmo MSA.	82
4.9. Configuración de parámetros del algoritmo SS.	82
4.10. Configuración de parámetros del algoritmo GA.	83
4.11. Configuración de parámetros del algoritmo VNS.	83
4.12. Configuración de parámetros del algoritmo PSO.	84
4.13. Configuración del proceso de reducción de dimensionalidad.	84
4.14. Comparación de modelos BERT optimizados para la tarea de clasificación.	85

4.15. Especificaciones del hardware utilizado para entrenamiento de DistilBERT.	86
4.16. Configuración de parámetros base para el entrenamiento de DistilBERT.	86
4.17. Estrategias de regularización implementadas para controlar overfitting.	87
4.18. Protocolo de evaluación implementado para ambos paradigmas.	87
4.19. Definición de categorías de la matriz de confusión en el contexto de detección de noticias falsas.	88
4.20. Estructura de la matriz de confusión para clasificación binaria.	88
4.21. Marco de métricas de evaluación para clasificación binaria de noticias falsas.	90
4.22. Criterios multi-dimensionales para selección del modelo óptimo.	93
4.23. Estructura del reporte de resultados comparativo.	93
4.24. Especificaciones del entorno de desarrollo para algoritmos metaheurísticos.	94
4.25. Stack tecnológico completo utilizado en el desarrollo del proyecto.	94
4.26. Marco ético implementado para el desarrollo responsable de la herramienta.	95
4.27. Artefactos generados para facilitar la reproducibilidad completa del estudio.	96
 5.1. Corpus académicos utilizados para la construcción del conjunto de datos unificado.	98
5.2. Configuración detallada de parámetros para los cinco algoritmos metaheurísticos implementados.	99
5.3. Función de evaluación común utilizada por todos los algoritmos metaheurísticos.	100
5.4. Pseudocódigo completo del algoritmo Multi-Start Simulated Annealing (MSA).	101
5.5. Pseudocódigo completo del algoritmo Scatter Search (SS).	102
5.6. Pseudocódigo completo del Algoritmo Genético (GA).	102
5.7. Pseudocódigo completo del algoritmo Variable Neighborhood Search (VNS).	103
5.8. Pseudocódigo completo del algoritmo Particle Swarm Optimization (PSO).	104
5.9. Resultados comparativos finales de los cinco algoritmos metaheurísticos implementados usando métricas macro promedio.	113
5.10. Análisis de fortalezas y debilidades de cada algoritmo metaheurístico basado en métricas macro.	113
5.11. Comparación de rendimiento entre enfoques metaheurísticos y modelos de lenguaje usando métricas macro.	114
5.12. Composición del corpus expandido utilizado para el entrenamiento de DistilBERT.	117
5.13. Configuración arquitectónica del modelo DistilBERT implementado.	117
5.14. Técnicas de regularización implementadas en la configuración V7 final.	117

5.15. Resumen de versiones experimentales de DistilBERT con evolución de estrategias y resultados reales del desarrollo.	121
5.16. Visualización comparativa de convergencia y métricas para todas las versiones experimentales de DistilBERT.	122
5.17. Evolución de configuraciones y resultados entre versiones experimentales.	123
5.18. Pseudocódigo simplificado del algoritmo DistilBERT V7.	126
5.19. Métricas finales del modelo DistilBERT optimizado.	127
5.20. Comparación DistilBERT vs. mejor algoritmo metaheurístico.	127
5.21. Comparación cuantitativa final entre el mejor algoritmo metaheurístico y el modelo Transformer optimizado.	128
5.22. Estrategias propuestas para abordar limitaciones identificadas.	132

Glosario

API Application Programming Interface

Atención (Attention) Mecanismo que permite a los modelos enfocarse en partes específicas de la entrada al procesar cada elemento. En PLN, permite que el modelo “atienda” a palabras relevantes al procesar una palabra específica.

AUC Area Under the Curve

AUC-ROC Área bajo la curva ROC (AUC de ROC). Mide la capacidad del modelo para distinguir entre clases. Un valor de 1.0 indica un clasificador perfecto, mientras que 0.5 indica rendimiento aleatorio.

BERT Bidirectional Encoder Representations from Transformers

Bolsa de Palabras (Bag of Words) Modelo de representación de texto que describe la ocurrencia de palabras en un documento, ignorando el orden y la estructura gramatical. Cada documento se representa como un vector donde cada dimensión corresponde a una palabra del vocabulario y su valor indica la frecuencia de aparición de esa palabra. Es importante para técnicas como TF-IDF y constituye la base para muchos algoritmos de clasificación de texto tradicionales.

Bulo Información falsa que circula ampliamente, especialmente en redes sociales y medios digitales, con el propósito de engañar a la audiencia. A diferencia de las noticias falsas, los bulos pueden no tener formato periodístico y suelen propagarse de manera viral. Incluye rumores, teorías conspirativas, información médica falsa, y contenido que apela más a las emociones que a los hechos verificables.

CNN Redes Neuronales Convolucionales

Convergencia Proceso por el cual un algoritmo de optimización se acerca progresivamente a una solución óptima o cerca del óptimo. Se evalúa observando la estabilización de la función objetivo a lo largo de las iteraciones.

Corpus Colección grande y estructurada de textos utilizados para investigación lingüística o entrenamiento de modelos de PLN. En este contexto, se refiere al conjunto unificado de noticias en español.

CPU Central Processing Unit

CSS Cascading Style Sheets

CSV Comma-Separated Values

Dataset Conjunto estructurado de datos utilizado para entrenar, validar y probar modelos de machine learning. En esta investigación, se refiere al corpus de noticias etiquetadas como verdaderas o falsas.

Deepfake Contenido multimedia (video, audio, imágenes) generado o manipulado usando inteligencia artificial, especialmente técnicas de aprendizaje profundo, para hacer que parezca que alguien dijo o hizo algo que nunca ocurrió realmente.

Desinformación Difusión intencional de información falsa o engañosas con el propósito específico de manipular la opinión pública, influir en decisiones políticas o sociales, o causar daño. Se caracteriza por ser un proceso activo y deliberado de creación y distribución de contenido falso.

Desinformación (Misinformation) Información incorrecta que se comparte sin intención maliciosa. A diferencia de la desinformación, quienes comparten misinformación no tienen conocimiento de que la información es falsa y actúan de buena fe.

DistilBERT Distilled Bidirectional Encoder Representations from Transformers

DL Deep Learning

Embedding Representación vectorial densa de palabras, frases o documentos en un espacio multidimensional donde la distancia entre vectores refleja similitud semántica. Ejemplos incluyen Word2Vec, GloVe, y FastText.

Espacio de Búsqueda Conjunto de todas las posibles configuraciones de hiperparámetros que pueden ser exploradas durante el proceso de optimización. Define los límites y restricciones para cada hiperparámetro.

Exactitud (Accuracy) Proporción de predicciones correctas sobre el total de predicciones. Se calcula como: $\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$.

Exploración Capacidad de un algoritmo de búsqueda para investigar regiones no exploradas del espacio de soluciones, evitando quedar atrapado en óptimos locales.

Explotación Capacidad de un algoritmo para refinar y mejorar soluciones prometedoras encontradas, concentrando la búsqueda en regiones del espacio que han mostrado buenos resultados.

F1-Score Media armónica entre precisión y recall. Se calcula como: $F1 = 2 \times (\text{Precisión} \times \text{Recall}) / (\text{Precisión} + \text{Recall})$. Proporciona una medida equilibrada del rendimiento del clasificador.

Fine-tuning Proceso de ajustar un modelo pre-entrenado en una tarea específica utilizando un conjunto de datos más pequeño y especializado. Permite aprovechar el conocimiento general del modelo pre-entrenado para tareas particulares.

FN False Negative (Falso Negativo)

FP False Positive (Falso Positivo)

Fraude Digital Conjunto amplio de actividades maliciosas realizadas a través de medios digitales con el objetivo primario de obtener beneficios económicos ilícitos, información personal sensible, o acceso no autorizado a recursos. Incluye estafas en línea, phishing, fraude financiero digital, esquemas Ponzi digitales, estafas de inversión, fraude laboral en línea, y otras formas de engaño que explotan plataformas y tecnologías digitales. A diferencia de las noticias falsas, que buscan influencia social, el fraude digital se enfoca en la obtención directa de beneficios materiales o acceso a activos. Su “éxito” se mide por la cantidad de dinero o información obtenida, no por el alcance viral o la influencia en opinión pública.

Función Objetivo Función matemática que define el criterio a optimizar en un problema. En el contexto de esta investigación, representa la métrica de rendimiento del modelo que se busca maximizar o minimizar.

GA Genetic Algorithm (Algoritmo Genético)

GPT Generative Pre-trained Transformer

GPU Graphics Processing Unit

Hiperparámetro Parámetro de configuración del modelo que se establece antes del entrenamiento y no se aprende durante el proceso de entrenamiento. Ejemplos incluyen la tasa de aprendizaje, el número de épocas, y el tamaño del batch.

HTML HyperText Markup Language

HTTP HyperText Transfer Protocol

IA Inteligencia Artificial

ICR Idónea Comunicación de Resultados

Información Maliciosa (Malinformation) Información genuina que se comparte con intención de causar daño, como filtraciones de información privada, discursos de odio, o acoso. Aunque la información base puede ser verdadera, su uso es malicioso.

JSON JavaScript Object Notation

Lematización Proceso más sofisticado que el stemming que reduce las palabras a su forma canónica o lemma, considerando el contexto morfológico y sintáctico. Por ejemplo, “mejor” se lematiza a “bueno”.

LSTM Long Short-Term Memory

Matriz de Confusión Tabla que describe el rendimiento de un modelo de clasificación mostrando las predicciones correctas e incorrectas para cada clase. Permite calcular métricas detalladas de rendimiento.

Metaheurística Estrategia de alto nivel para guiar y modificar otras heurísticas con el objetivo de producir soluciones de alta calidad para problemas de optimización. Incluye técnicas como algoritmos genéticos, optimización por enjambre de partículas, y recocido simulado.

ML Machine Learning

MSA Multi-Start Simulated Annealing (Recocido Multiarranque)

NLP Natural Language Processing

Noticia Falsa Información deliberadamente fabricada que se presenta como contenido periodístico legítimo, pero que contiene datos falsos, inexactos o engañosos. Su objetivo principal es manipular la opinión pública, influir en decisiones políticas o sociales, o distorsionar la percepción de eventos actuales. Se caracteriza por imitar el formato y estilo de medios de comunicación establecidos, utilizando técnicas periodísticas aparentemente profesionales para ganar credibilidad. A diferencia del fraude digital, su “éxito” se mide por el alcance social y la influencia en la opinión pública, no por beneficios económicos directos.

Palabras Vacías (Stop Words) Palabras comunes en un idioma que generalmente se filtran durante el preprocesamiento de texto porque no aportan significado semántico significativo. Ejemplos en español: “el”, “la”, “de”, “que”, “y”.

PLN Procesamiento del Lenguaje Natural

Precisión Métrica que mide la proporción de predicciones positivas que fueron correctas. Se calcula como: Precisión = $TP / (TP + FP)$.

PSO Particle Swarm Optimization

RAM Random Access Memory

RNN Redes Neuronales Recurrentes

ROC Receiver Operating Characteristic

SA Simulated Annealing (Recocido Simulado)

Sensibilidad (Recall) Métrica que mide la proporción de casos positivos reales que fueron correctamente identificados. Se calcula como: $\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$.

Sesgo (Bias) Error sistemático en las predicciones del modelo que puede deberse a suposiciones erróneas en el algoritmo de aprendizaje, datos de entrenamiento no representativos, o prejuicios inherentes en el dataset.

Sobreajuste (Overfitting) Fenómeno donde un modelo aprende demasiado específicamente los datos de entrenamiento, perdiendo capacidad de generalización a datos nuevos no vistos durante el entrenamiento.

SS Scatter Search

Stemming Técnica de reducción de palabras a su raíz o stem mediante la eliminación de sufijos. Por ejemplo, “corriendo”, “corrió”, “correr” se reducen a “corr”.

Subajuste (Underfitting) Fenómeno donde un modelo es demasiado simple para capturar la complejidad subyacente de los datos, resultando en bajo rendimiento tanto en entrenamiento como en validación.

SVM Support Vector Machine

TF-IDF Term Frequency-Inverse Document Frequency

TN True Negative (Verdadero Negativo)

Tokenización Proceso de dividir texto en unidades más pequeñas llamadas tokens, que pueden ser palabras, subpalabras, caracteres o n-gramas. Es el primer paso en el preprocesamiento de texto para análisis computacional.

TP True Positive (Verdadero Positivo)

Transformer Arquitectura de red neuronal basada en mecanismos de atención que ha revolucionado el PLN. Introdujo el concepto de atención multi-cabeza y eliminó la necesidad de procesamiento secuencial, permitiendo paralelización eficiente.

URL Uniform Resource Locator

Validación Cruzada Técnica de evaluación que divide el dataset en múltiples subconjuntos para entrenar y validar el modelo repetidamente, proporcionando una estimación más robusta del rendimiento.

Varianza Medida de cuánto varían las predicciones del modelo para diferentes conjuntos de entrenamiento. Alta varianza indica que el modelo es sensible a pequeños cambios en los datos de entrenamiento.

Verificación de Hechos (Fact-Checking) Proceso sistemático de investigación y verificación de afirmaciones factuales en contenido publicado. Incluye la consulta de fuentes primarias, expertos, y evidencia documental para determinar la veracidad de las afirmaciones.

VNS Variable Neighborhood Search

Web Scraping Técnica de extracción automatizada de datos de sitios web mediante programas que navegan y analizan el contenido HTML de las páginas web para recopilar información específica.

Capítulo 1

Introducción

1.1. Planteamiento del problema

En la era de la información digital, la interconexión global ha traído consigo un desafío sin precedentes: la propagación masiva de Desinformación. Este fenómeno, que abarca desde Noticia Falsas (*fake news*) hasta diversas formas de Fraude Digital, representa una amenaza significativa para la estabilidad social, económica y democrática a nivel mundial.

Las Noticia Falsas, definidas como información deliberadamente engañosa disfrazada de periodismo auténtico [1], se difunden a través de redes sociales y medios digitales con el fin de manipular la opinión pública y el comportamiento de los individuos. La gravedad de esta problemática se ha intensificado exponencialmente con la sofisticación de los actores maliciosos, la democratización de herramientas de generación de contenido mediante IA [2], y la velocidad sin precedentes con la que la información se viraliza en el ecosistema digital.

Distinción fundamental entre noticias falsas y fraude digital:

- **Noticia Falsas:** Información deliberadamente fabricada que imita el formato periodístico legítimo con el objetivo primario de *manipular la opinión pública*, influir en decisiones políticas/sociales, o distorsionar la percepción de eventos actuales. Su “éxito” se mide por el alcance y la influencia social.
- **Fraude Digital:** Actividades maliciosas digitales con objetivo primario de *obtener beneficios económicos ilícitos* o acceso no autorizado a información/recursos. Incluye estafas de inversión, phishing, fraude financiero, y esquemas de estafas piramidales digitales. Su “éxito” se mide por el beneficio económico obtenido.

¿Por qué son tareas de detección relacionadas pero distintas?

- **Objetivos diferentes:** Noticias falsas buscan influencia social; fraude digital busca beneficio económico
- **Audiencias diferentes:** Noticias falsas apuntan a consumidores de noticias; fraude digital a potenciales víctimas económicas

- **Métricas de éxito diferentes:** Viral vs. rentabilidad económica
- **Patrones lingüísticos diferentes:** Noticias falsas imitan periodismo; fraude digital usa técnicas de persuasión comercial/urgencia

Enfoque de esta investigación: Este proyecto se centra específicamente en la detección de noticias falsas, desarrollando una metodología que es inherentemente transferible a fraude digital debido a características computacionales compartidas: naturaleza textual, necesidad de clasificación binaria (legítimo/malicioso), y optimización de parámetros.

La investigación reciente ha documentado el impacto multidimensional de la desinformación, que va desde la distorsión de procesos democráticos [3] hasta la creación de pánico público durante crisis sanitarias como la pandemia de COVID-19 [4]. Paralelamente, el fraude digital ha evolucionado hacia formas cada vez más sofisticadas, aprovechando tanto vulnerabilidades técnicas como sesgos cognitivos humanos [5].

Desde la perspectiva de detección automatizada, la problemática se conceptualiza fundamentalmente como un problema de clasificación binaria: determinar si un contenido dado es **confiable** (información verificada y legítima) o **no confiable** (contenido desinformativo). Esta simplificación, aunque reduce la complejidad natural del espectro de veracidad de la información, resulta práctica y efectiva para aplicaciones reales donde los usuarios requieren respuestas claras y accionables.

La desinformación contemporánea se manifiesta a través de múltiples modalidades con un impacto significativo en individuos, empresas y sociedades, como se ilustra en la Figura 1.1 (ver figura 1.1). Aunque el presente estudio se enfoca específicamente en noticias falsas, es relevante comprender el contexto más amplio del fraude digital para dimensionar la importancia de desarrollar metodologías transferibles:

- **Noticias falsas generadas por IA:** Contenido sintético creado por Grandes Modelos de Lenguaje (GPTs) que imita el estilo periodístico legítimo [6] - *Uno de los objetivos de este proyecto*
- **Desinformación política dirigida:** Campañas coordinadas para influir en procesos electorales y opinión pública [7] - *Aplicable con la metodología propuesta*
- **Fraude financiero digital:** Esquemas que explotan plataformas digitales y criptomonedas [8] - *Metodología transferible*
- **Fraude laboral en línea:** Ofertas de empleo falsas que buscan obtener información personal o financiera [9] - *Metodología transferible*
- **Estafas de ingeniería social:** Técnicas sofisticadas que combinan información personal extraída de redes sociales con narrativas convincentes - *Metodología transferible*
- **Desinformación en salud:** Información médica falsa que puede tener consecuencias directas en la salud pública [10] - *Aplicable con la metodología propuesta*

- **Deepfakes:** Contenido multimedia manipulado mediante técnicas de DL - *Requiere extensión multimodal*
- **Bulos:** Información falsa que se propaga viralmente en redes sociales - *Aplicable con la metodología propuesta*



Figura 1.1: Mapa Conceptual 1: Taxonomía del problema de desinformación con enfoque en noticias falsas y extensibilidad hacia fraude digital.

1.1.1. El Desafío Específico del Español: Enfoque Multiregional

El español, como la cuarta lengua más hablada del mundo con más de 500 millones de hablantes nativos [11], presenta desafíos únicos para la detección automatizada de desinformación. A pesar de su importancia demográfica y económica, existe una notable escasez de recursos computacionales especializados para la detección de noticias falsas en español, en comparación con los abundantes recursos disponibles para el inglés [12].

Variante del Español Utilizada: Justificación Metodológica

Esta investigación adopta un enfoque de español neutro multiregional, fundamentado en la integración estratégica de corpus que representan diferentes variantes geográficas y estilísticas del español. Esta decisión metodológica se justifica por:

- **Corpus español peninsular:** Los corpus de Posadas-Durán et al. [12] y Blanco-Fernández et al. [13] incluyen principalmente contenido de medios españoles, aportando características del español de España
- **Corpus latinoamericano:** El corpus de Acosta [11] incorpora fuentes internacionales con presencia de español mexicano y de otros países latinoamericanos
- **Contenido mexicano específico:** La inclusión de contenido de “El Deiforma” (portal satírico mexicano) aporta características específicas del español mexicano contemporáneo
- **Variabilidad temporal:** Los corpus abarcan contenido desde 2018 hasta 2025, capturando la evolución del español digital en múltiples regiones

Distribución geográfica del corpus unificado:

- **España (~70 %):** Principalmente del corpus político de Blanco-Fernández
- **México (~20 %):** Contenido de El Deiforma y fuentes mexicanas en otros corpus
- **Multinacional (~10 %):** Corpus de Acosta con fuentes internacionales diversas

Esta composición resulta en un modelo que puede generalizar efectivamente a través de las principales variantes del español, sin estar sesgado hacia una región específica.

Ventajas del Enfoque Multiregional

- **Robustez lingüística:** El modelo entrenado es capaz de detectar desinformación independientemente de la variante regional del español
- **Generalización mejorada:** Menor dependencia de modismos o construcciones específicas de una región
- **Aplicabilidad universal:** El sistema resultante es útil para toda la comunidad hispanohablante
- **Representatividad demográfica:** Refleja la realidad del consumo de noticias en español, que frecuentemente cruza fronteras nacionales

Esta brecha de recursos se manifiesta en:

- **Escasez de corpus etiquetados:** Limitados conjuntos de datos de entrenamiento en español para modelos de detección
- **Variabilidad dialectal:** La diversidad regional del español presenta desafíos adicionales para modelos generalizables

- **Contexto cultural específico:** Los patrones de desinformación varían según el contexto sociocultural hispanoamericano
- **Herramientas de detección limitadas:** Pocas soluciones tecnológicas disponibles para comunidades hispanohablantes

1.1.2. La Complejidad Técnica del Problema

La detección automatizada de Noticia Falsas constituye un problema técnico multifacético que requiere la integración de múltiples disciplinas (ver figura 1.2). Como documenta la literatura reciente [14], los desafíos incluyen:

- **Análisis semántico profundo:** Necesidad de comprender el contexto y las implicaciones sutiles del contenido mediante técnicas de PLN
- **Detección de patrones estilométricos:** Identificación de características lingüísticas que indiquen autoría maliciosa [15]
- **Procesamiento en tiempo real:** Capacidad de analizar el volumen masivo de contenido generado diariamente usando ML
- **Adaptación a contenido sintético:** Detección de texto generado por modelos de IA cada vez más sofisticados [16]
- **Robustez ante ataques adversariales:** Resistencia a intentos deliberados de evadir la detección
- **Optimización de Hiperparámetros:** Calibración de parámetros del modelo para maximizar el rendimiento

Dada la sofisticación de estas amenazas, se requieren soluciones tecnológicas igualmente avanzadas para combatir el fraude digital, como se esquematiza en la Figura 1.2. Este trabajo se centra en el desarrollo de tales soluciones, con un enfoque particular en el idioma español y la comparación sistemática de paradigmas tecnológicos complementarios.

1.2. Motivación

La motivación para llevar a cabo esta investigación se fundamenta en una combinación de experiencias personales observadas y la identificación de una brecha crítica en la protección tecnológica de las comunidades hispanohablantes.



Figura 1.2: Mapa Conceptual 2: Estrategias tecnológicas para la detección de fraude digital.

1.2.1. Impacto Personal y Social Observado

Durante el desarrollo de esta investigación, se observaron múltiples casos en el entorno cercano donde personas fueron víctimas de fraude digital sofisticado. Estos casos incluyeron desde estafas de inversión disfrazadas de noticias financieras legítimas, hasta esquemas de phishing que aprovechaban eventos noticiosos actuales para parecer creíbles. Las víctimas, frecuentemente personas de edad avanzada o con menor exposición a tecnología digital, sufrieron no solo pérdidas económicas significativas, sino también impacto psicológico profundo, incluyendo sentimientos de vergüenza, ansiedad y pérdida de confianza en medios digitales.

1.2.2. Vulnerabilidad de Poblaciones Específicas

La investigación en psicología cognitiva aplicada a la desinformación [5] ha demostrado que ciertos grupos demográficos son particularmente vulnerables:

- Adultos mayores:** Mayor susceptibilidad a heurísticas de credibilidad basadas en autoridad percibida
- Poblaciones con menor alfabetización digital:** Limitada capacidad para evaluar la legitimidad de fuentes online

- **Comunidades con acceso limitado a información:** Mayor dependencia de redes sociales como fuente primaria de noticias
- **Hablantes nativos de español:** Menor disponibilidad de herramientas de verificación en su idioma nativo

1.2.3. Urgencia Tecnológica

El rápido avance en modelos generativos de IA, como GPT-3 [17] y sus sucesores, ha reducido significativamente las barreras técnicas para la creación de contenido falso convincente. Esta democratización de la capacidad de generar desinformación [6] crea una urgencia imperativa para desarrollar defensas tecnológicas igualmente sofisticadas.

Impulsado por la necesidad de crear defensas tecnológicas más robustas y específicamente adaptadas para la comunidad hispanohablante, el presente trabajo se centra en desarrollar, comparar y validar métodos computacionales avanzados para la detección y prevención del fraude digital, con el objetivo final de contribuir a la protección de las poblaciones más vulnerables frente a estas amenazas emergentes.

1.3. Justificación

Esta investigación se justifica desde múltiples perspectivas: la brecha tecnológica existente, la novedad metodológica del enfoque, y la necesidad social de herramientas especializadas para el español.

1.3.1. Brecha Tecnológica en Recursos para el Español

El español, con más de 500 millones de hablantes nativos distribuidos en 21 países, representa un vasto ecosistema digital que ha sido históricamente subatendido en términos de herramientas especializadas para la detección de desinformación. Mientras que para el inglés existen múltiples conjuntos de datos de gran escala como LIAR, FakeNewsNet, y CREDBANK [18], los recursos equivalentes en español son limitados y fragmentados.

El estado del arte actual en español se basa principalmente en cuatro corpus principales:

- **Corpus de Acosta (2019):** 598 noticias [11]
- **Spanish Fake News Corpus:** 971 noticias [12]
- **Corpus de Tretiakov (2022):** 1,958 noticias [19]
- **Spanish Political Fake News:** 57,000+ noticias [13]

Esta fragmentación crea una barrera para el desarrollo de modelos robustos y generalizables.

1.3.2. Novedad Metodológica: Enfoque Evolutivo Comparativo

La novedad principal de esta investigación radica en su enfoque evolutivo que compara sistemáticamente dos paradigmas fundamentalmente diferentes de la Inteligencia Artificial en el mismo contexto aplicado:

Paradigma Clásico Optimizado

- **Representación textual:** TF-IDF
- **Optimización:** Algoritmos metaheurísticos para calibración de hiperparámetros
- **Clasificador:** Modelos de aprendizaje automático tradicionales optimizados
- **Ventajas:** Eficiencia computacional, interpretabilidad, menor dependencia de hardware especializado

Paradigma de Deep Learning

- **Representación textual:** Embeddings contextuales dinámicos
- **Modelo base:** DistilBERT pre-entrenado [20]
- **Técnica:** Fine-tuning con optimización de hiperparámetros
- **Ventajas:** Comprensión semántica profunda, captura de relaciones contextuales complejas

1.3.3. Contribución Científica Multidimensional

Esta investigación ofrece contribuciones en múltiples dimensiones:

Contribución a Recursos Lingüísticos

- **Corpus unificado:** Integración y estandarización de los principales corpus en español
- **Metodología de unificación:** Protocolo replicable para integrar conjuntos de datos heterogéneos
- **Benchmarking:** Establecimiento de líneas base comparativas para futura investigación

Contribución Metodológica

- **Optimización metaheurística:** Aplicación sistemática de algoritmos bio-inspirados para calibración de hiperparámetros.
- **Análisis comparativo riguroso:** Evaluación exhaustiva usando métricas múltiples y validación cruzada
- **Transferencia tecnológica:** Implementación práctica en aplicaciones web containerizadas

Contribución Aplicada

- **Herramientas funcionales:** Aplicaciones web deployables para análisis en tiempo real
- **Código abierto:** Disponibilidad pública de implementaciones para reproducibilidad
- **Impacto social directo:** Herramientas utilizables por comunidades hispanohablantes

1.4. Objetivos

Objetivo General

Desarrollar un método computacional basado en algoritmos metaheurísticos y modelos de lenguaje para detectar noticias falsas en español, con metodología transferible a otros tipos de fraude digital.

Objetivos Específicos

- Recopilar y procesar un conjunto de datos diverso a partir de múltiples corpus en español, y enriquecerlo mediante técnicas de *extracción web* para entrenar y evaluar los sistemas de detección.
- Implementar un sistema de detección inicial que utilice técnicas de Procesamiento del Lenguaje Natural (TF-IDF) y algoritmos metaheurísticos.
- Desarrollar un sistema de detección avanzado mediante el ajuste fino (*fine-tuning*) de un modelo de lenguaje profundo (*DistilBERT*), optimizando su rendimiento a través de la calibración de hiperparámetros.
- Realizar un análisis comparativo del rendimiento entre el enfoque metaheurístico y el modelo de lenguaje, utilizando un conjunto completo de métricas de evaluación (Exactitud, Precisión, Exhaustividad y F1-Score).

- Desarrollar un prototipo de aplicación web funcional, contenerizada con Docker, para demostrar la aplicabilidad práctica del modelo de mayor rendimiento en el análisis de URLs para categorizar contenido noticioso como falso o legítimo, con potencial de extensión hacia otros tipos de fraude digital.

1.5. Alcance y Limitaciones

1.5.1. Alcance

Alcance Lingüístico y Cultural

- **Idioma objetivo:** El proyecto se centra exclusivamente en textos en español, abarcando múltiples variedades regionales representadas en los corpus utilizados
- **Dominio de aplicación:** Detección de noticias falsas como caso de uso principal, con extensibilidad demostrada hacia otros tipos de fraude digital
- **Contexto geográfico:** Cobertura de múltiples países hispanohablantes a través de los corpus integrados

Alcance Técnico

- **Modalidad de datos:** Procesamiento exclusivo de contenido textual (no multimedial)
- **Tipo de clasificación:** Clasificación binaria supervisada (FALSO/REAL)
- **Dominio principal:** Detección de noticias falsas en español
- **Arquitecturas evaluadas:** Comparación sistemática entre enfoques clásicos optimizados y modelos Transformer
- **Implementación práctica:** Desarrollo de aplicaciones web funcionales y containerizadas

Transferibilidad de la Metodología

- **Fraude financiero digital:** La metodología puede adaptarse para detectar esquemas de inversión fraudulentos, estafas de criptomonedas, y ofertas financieras falsas
- **Fraude laboral:** Aplicable para identificar ofertas de empleo falsas y esquemas de reclutamiento fraudulentos
- **Estafas de ingeniería social:** Útil para detectar contenido de phishing y esquemas de manipulación social

- **Desinformación temática específica:** Extensible a desinformación médica, científica, o política con ajustes de dominio
- **Limitaciones de transferencia:** Contenido multimodal (deepfakes de video/audio) requiere arquitecturas especializadas adicionales

Alcance Metodológico

- **Optimización metaheurística:** Aplicación de cinco algoritmos bio-inspirados para calibración de hiperparámetros
- **Validación experimental:** Uso de validación cruzada estratificada y métricas múltiples
- **Reproducibilidad:** Documentación completa y código fuente disponible

1.5.2. Limitaciones

Limitaciones de Datos

- **Dependencia de corpus existentes:** El rendimiento está condicionado por la calidad y representatividad de los corpus disponibles en español
- **Sesgos inherentes:** Posibles sesgos temporales, temáticos o geográficos presentes en las fuentes originales
- **Evolución del lenguaje:** Los modelos pueden no capturar patrones emergentes en desinformación generada por IA más reciente
- **Etiquetado ground truth:** Dependencia de la calidad del etiquetado manual en los corpus originales

Limitaciones Técnicas

- **Recursos computacionales:** El entrenamiento de modelos Transformer requiere hardware especializado (GPU) y tiempos de cómputo extensos (12-72+ horas)
- **Escalabilidad en tiempo real:** Los prototipos funcionan bajo demanda, no están optimizados para procesamiento de streams masivos
- **Generalización a otros tipos de fraude:** Los modelos están específicamente entrenados para noticias, la efectividad en otros tipos de fraude digital es inferencial

Limitaciones Operacionales

- **Herramienta de apoyo:** Los sistemas desarrollados son herramientas de apoyo a la decisión, no reemplazan el juicio humano experto
- **Verificación absoluta:** La determinación definitiva de veracidad puede requerir investigación periodística especializada
- **Contexto dinámico:** Los patrones de desinformación evolucionan constantemente, requiriendo actualización periódica de los modelos
- **Consideraciones éticas:** No se abordan explícitamente implicaciones de sesgo algorítmico o impacto en libertad de expresión

Limitaciones de Evaluación

- **Métricas tradicionales:** La evaluación se basa en métricas estándar que pueden no capturar completamente la complejidad del problema
- **Validación temporal:** No se realiza validación temporal explícita con datos de períodos posteriores al entrenamiento
- **Análisis de errores:** El análisis cualitativo de errores es limitado debido al volumen de datos procesado

1.6. Contribuciones Esperadas

1.6.1. Contribuciones Teóricas

- **Metodología comparativa:** Marco sistemático para comparar paradigmas clásicos y modernos en detección de desinformación
- **Optimización metaheurística:** Aplicación de algoritmos bio-inspirados para calibración de modelos Transformer

1.6.2. Contribuciones Prácticas

- **Corpus unificado:** Recurso consolidado para investigación futura en español
- **Herramientas funcionales:** Aplicaciones deployables para uso comunitario
- **Código abierto:** Implementaciones reproducibles y extensibles

1.6.3. Contribuciones Sociales

- **Protección comunitaria:** Herramientas accesibles para comunidades hispanohablantes
- **Democratización tecnológica:** Reducción de la brecha digital en herramientas de verificación
- **Capacitación y concientización:** Contribución a la alfabetización digital en detección de desinformación

1.7. Organización del Documento

Este documento se estructura de manera lógica y progresiva para guiar al lector a través del proceso completo de investigación, desarrollo y evaluación:

- **Capítulo 2 - Marco Teórico:** Establece los fundamentos teóricos que sustentan ambas metodologías, desde técnicas clásicas de representación textual hasta arquitecturas Transformer modernas, incluyendo principios de optimización metaheurística.
- **Capítulo 3 - Estado del Arte:** Presenta una revisión comprehensiva y organizada temáticamente de la literatura relevante, identificando brechas de conocimiento y posicionando esta investigación en el contexto científico actual.
- **Capítulo 4 - Metodología:** Detalla la metodología evolutiva seguida, desde la construcción del corpus unificado hasta la implementación y optimización de ambos paradigmas de detección.
- **Capítulo 5 - Análisis de Resultados:** Presenta y compara exhaustivamente los resultados obtenidos por ambas metodologías, incluyendo análisis estadístico, métricas de rendimiento y discusión de fortalezas y limitaciones.
- **Capítulo 6 - Implementación de Prototipo:** Describe la arquitectura, desarrollo y despliegue de las aplicaciones web funcionales que demuestran la viabilidad práctica de los modelos desarrollados.
- **Capítulo 7 - Conclusiones y Trabajo Futuro:** Sintetiza los hallazgos principales, evalúa el cumplimiento de objetivos, y propone direcciones específicas para investigación futura en el campo.

Cada capítulo se construye sobre los anteriores, manteniendo coherencia narrativa y técnica a lo largo del documento, mientras proporciona la profundidad analítica necesaria para validar las contribuciones científicas y prácticas de esta investigación.

Capítulo 2

Marco Teórico

En esta sección se describen las bases teóricas y los conceptos fundamentales que sustentan las dos metodologías de detección de noticias falsas desarrolladas en este proyecto de investigación. Aunque las técnicas son aplicables a diferentes tipos de fraude digital, el enfoque específico se centra en la detección de desinformación periodística. Se abordan desde las técnicas clásicas de representación de texto y optimización metaheurística, hasta los paradigmas de aprendizaje profundo que definen el estado del arte actual.

2.1. Detección de Noticias Falsas: Fundamentos y Extensibilidad

La detección de noticias falsas constituye un problema multifacético que requiere un enfoque interdisciplinario, con metodologías que pueden extenderse a otros tipos de fraude digital. La desinformación se caracteriza por ser información deliberadamente falsa o engañosa que se presenta como noticia legítima [1]. Esta problemática ha evolucionado significativamente con el advenimiento de las redes sociales y los medios digitales, donde la velocidad de propagación supera ampliamente la capacidad de verificación tradicional.

En el contexto de la detección automatizada, se han explorado diferentes enfoques y técnicas. Das et al. [21] propusieron un marco de ensamblaje basado en incertidumbre e impulsado por heurísticas para la detección de noticias falsas en tuits y artículos de noticias. Este enfoque parte de la premisa de que diferentes modelos pueden tener distintas fortalezas y debilidades, y que la combinación inteligente de múltiples predictores puede superar las limitaciones individuales.

El equipo de la Southern Methodist University presentó una solución pionera utilizando procesamiento de lenguaje natural y aprendizaje profundo para analizar tanto titulares como el contenido completo de las noticias [22]. Su trabajo estableció un precedente importante al demostrar la viabilidad de la vectorización TF-IDF combinada con redes neuronales densas.

Complementariamente, se ha propuesto un enfoque basado en análisis estilométrico utilizando Procesamiento del Lenguaje Natural (PLN) y Reconocimiento de Entidades Nombradas (NER) para identificar patrones lingüísticos que sugieran baja veracidad de la información [15]. Este enfoque aprovecha la hipótesis de que los autores de contenido falso pueden exhibir patrones de escritura distintivos.

2.1.1. Impacto Social y Desafíos de la Desinformación

La rápida propagación de las noticias falsas a través de las redes sociales puede tener graves consecuencias en la sociedad. Como documentan Ali y Zain-Ul-Abdin [3], la desinformación puede:

- **Distorsionar la realidad:** Alterando la percepción pública de eventos y hechos
- **Manipular la opinión pública:** Influenciando procesos democráticos y decisiones sociales
- **Incitar a la violencia:** Promoviendo comportamientos agresivos basados en información errónea
- **Difundir propaganda política:** Siendo utilizada como herramienta de influencia partidista
- **Fomentar el odio:** Intensificando divisiones sociales y promoviendo discriminación
- **Provocar pánico:** Especialmente en contextos de crisis sanitarias como la pandemia de COVID-19 [4]

2.1.2. El Problema de las Heurísticas Cognitivas

La investigación de Ali et al. [5] ha demostrado que las heurísticas cognitivas humanas, como la popularidad social (número de “me gusta” o compartidos), influyen significativamente en la percepción de credibilidad. Este fenómeno complica la detección, ya que el contenido falso puede volverse viral precisamente por aprovechar estos sesgos cognitivos.

2.1.3. Diferenciación Técnica y Transferibilidad: Noticias Falsas vs. Fraude Digital

Es necesario establecer una distinción técnica clara entre los dominios de aplicación, ya que aunque comparten características computacionales, representan problemas con diferentes objetivos, audiencias y patrones (ver tabla 2.1):

Aspecto	Noticias Falsas	Fraude Digital
Objetivo primario	Manipulación de opinión pública, influencia política/social	Beneficio económico ilícito, acceso no autorizado a recursos
Formato típico	Artículos periodísticos, imitando medios legítimos	Emails de phishing, anuncios de inversión, ofertas comerciales
Audiencia objetivo	Consumidores de noticias, votantes, opinión pública general	Potenciales víctimas económicas, usuarios con activos digitales
Métrica de éxito	Viralidad, alcance, influencia en opinión	Cantidad de dinero/información obtenida
Indicadores lingüísticos	Imitación de estilo periodístico, fuentes falsas, sensacionalismo político	Urgencia económica, ofertas "limitadas", solicitudes de información personal
Temporalidad	Eventos actuales, ciclos noticiosos	Atemporales, aprovechan tendencias económicas

Tabla 2.1: Diferencias técnicas entre dominios de desinformación

Características Distintivas por Dominio

Fundamentos de la Transferibilidad Metodológica

A pesar de sus diferencias conceptuales, ambos dominios comparten características computacionales fundamentales que justifican la transferibilidad de la metodología desarrollada:

- **Naturaleza textual primaria:** Ambos dominios se basan en contenido textual como vector principal de engaño
- **Problema de clasificación binaria:** Se reducen a problemas de clasificación legítimo/malicioso
- **Características estilométricas:** Ambos pueden exhibir patrones lingüísticos distintivos detectables
- **Optimización de hiperparámetros:** Ambos se benefician de técnicas de calibración metaheurística
- **Evaluación mediante métricas estándar:** Utilizan las mismas métricas de clasificación (precisión, recall, F1-Score)

Protocolo de Transferencia Metodológica

Para aplicar la metodología desarrollada a detección de fraude digital, se requiere el siguiente protocolo de adaptación:

1. **Construcción de corpus específico:** Recopilar y etiquetar datos representativos del tipo de fraude digital objetivo
2. **Ajuste de preprocessamiento:** Adaptar técnicas de limpieza según las características del nuevo dominio

3. **Re-calibración de hiperparámetros:** Aplicar los mismos algoritmos meta-heurísticos pero re-optimizar para el nuevo corpus
4. **Incorporación de características específicas:** Añadir características relevantes al nuevo dominio (URLs sospechosas, patrones de solicitud de información)
5. **Validación en diferentes dominios:** Evaluar la transferencia utilizando métricas de generalización

Enfoque de esta investigación: Este proyecto desarrolla y valida la metodología en el dominio de noticias falsas en español, estableciendo las bases técnicas para su posterior transferencia a otros tipos de fraude digital. La elección de noticias falsas como dominio inicial se justifica por: (1) disponibilidad de corpus etiquetados, (2) relevancia social en comunidades hispanohablantes, y (3) menor complejidad técnica que permite validar la metodología base antes de extensiones más complejas.

2.2. Representación de Texto: Desde TF-IDF hasta Embeddings Contextuales

La evolución de las técnicas de representación textual ha sido fundamental para el progreso en PLN. Esta sección aborda desde los métodos clásicos hasta las representaciones más sofisticadas utilizadas en esta tesis.

2.2.1. Bolsa de Palabras (Bag of Words): Primer Acercamiento Metodológico

El modelo de Bolsa de Palabras (Bag of Words) (Bag of Words, BoW) constituye uno de los enfoques más básicos pero fundamentales en el procesamiento de texto. Aunque no fue utilizado en los modelos finales de esta tesis, representó el primer acercamiento metodológico explorado en la investigación preliminar documentada en [23].

Fundamentos del Modelo BoW

El modelo BoW representa un documento como una colección desordenada de palabras, ignorando completamente la gramática y el orden de las palabras, pero manteniendo la multiplicidad (frecuencia) de cada término. Matemáticamente, un documento d se representa como un vector de frecuencias:

$$\text{BoW}(d) = [f_1, f_2, \dots, f_V] \quad (2.1)$$

Donde f_i es la frecuencia del término i en el documento y V es el tamaño del vocabulario.

Proceso de Construcción

1. **Tokenización:** División del texto en unidades léxicas individuales
2. **Construcción del vocabulario:** Creación de un diccionario con todos los términos únicos del corpus
3. **Vectorización:** Cada documento se convierte en un vector donde cada dimensión corresponde a un término del vocabulario
4. **Conteo de frecuencias:** Cada posición del vector contiene la frecuencia del término correspondiente

Ventajas y Aplicabilidad Inicial

El modelo BoW ofreció ventajas específicas en las primeras fases de la investigación:

- **Simplicidad conceptual:** Fácil implementación y comprensión
- **Eficiencia computacional:** Bajo costo de procesamiento para corpus pequeños
- **Transparencia:** Interpretabilidad directa de las características
- **Línea base sólida:** Punto de partida para comparaciones con métodos más sofisticados

Limitaciones Identificadas

La experiencia con BoW en el trabajo preliminar [23] reveló limitaciones que motivaron la evolución hacia enfoques más sofisticados:

- **Pérdida total del orden:** Imposibilidad de capturar patrones secuenciales relevantes
- **Ausencia de ponderación semántica:** Todas las palabras reciben el mismo tratamiento independientemente de su importancia
- **Problemas de escalabilidad:** Crecimiento exponencial de la dimensionalidad con el vocabulario
- **Sensibilidad a palabras vacías:** Dominancia de términos frecuentes pero poco informativos

2.2.2. TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF constituye una de las técnicas más fundamentales y efectivas para la representación de texto. A pesar de su simplicidad conceptual, ha demostrado ser muy efectiva en tareas de clasificación de texto. El proceso implica:

1. **Construcción del vocabulario:** Creación de un diccionario con todas las palabras únicas presentes en el corpus de documentos
2. **Cálculo de ponderaciones:** Cada término recibe un peso basado en su frecuencia en el documento (TF) y su rareza en el corpus (IDF)

Aunque este método ignora la gramática y el orden de las palabras, ha demostrado ser efectivo como base para métodos más sofisticados y fue fundamental en los primeros trabajos de clasificación de noticias en español [11].

2.2.3. Fundamentos Matemáticos de TF-IDF

TF-IDF introduce ponderación semántica mediante el producto de dos componentes fundamentales:

$$\text{TF-IDF}(t, d, D) = \text{TF}(t, d) \times \text{IDF}(t, D) \quad (2.2)$$

Donde:

- **Frecuencia de Término (TF):** $\text{TF}(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}}$ - frecuencia relativa del término t en el documento d
- **Frecuencia Inversa de Documento (IDF):** $\text{IDF}(t, D) = \log \frac{|D|}{|\{d \in D : t \in d\}|}$ - importancia global del término en la colección

Esta ponderación asigna mayor peso a términos que son frecuentes en un documento específico pero raros en el corpus general, mejorando la capacidad discriminativa del modelo [22].

2.2.4. Limitaciones de los Métodos Clásicos

Los enfoques basados en TF-IDF presentan limitaciones fundamentales:

- **Pérdida de información secuencial:** No capturan el orden ni la estructura sintáctica
- **Problema de dispersidad:** Generan representaciones muy dispersas en vocabularios grandes
- **Ausencia de semántica:** No modelan relaciones semánticas entre palabras
- **Falta de contexto:** Una palabra tiene la misma representación independientemente del contexto

Evolución Metodológica

La transición de BoW hacia TF-IDF representó un primer refinamiento que introdujo ponderación semántica, abordando algunas de las limitaciones más críticas del enfoque básico. Esta evolución metodológica, documentada en el marco de optimización metaheurística [23], estableció los fundamentos para los enfoques más sofisticados desarrollados en esta tesis.

A pesar de sus limitaciones, el modelo BoW proporcionó información útil sobre cómo identificar patrones distintivos en textos en español que permiten distinguir noticias falsas de verdaderas. Esta experiencia inicial sirvió como punto de partida para desarrollar las dos metodologías principales de esta investigación: el enfoque clásico optimizado y el de aprendizaje profundo.

2.3. La Revolución Transformer y los Modelos de Lenguaje Modernos

2.3.1. La Arquitectura Transformer: Fundamentos

La arquitectura Transformer, introducida por Vaswani et al. [24] en el trabajo “Attention Is All You Need”, revolucionó el campo del PLN. Su innovación principal radica en el mecanismo de **auto-atención (self-attention)**, que permite al modelo calcular representaciones ponderando dinámicamente la importancia de cada elemento en una secuencia con respecto a todos los demás elementos.

Mecanismo de Atención Multi-Cabeza

El mecanismo de atención se define matemáticamente como:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (2.3)$$

Donde Q , K , y V representan las matrices de consultas (queries), claves (keys) y valores (values), respectivamente. La atención multi-cabeza extiende este concepto:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (2.4)$$

Esta arquitectura permite capturar diferentes tipos de relaciones lingüísticas simultáneamente, superando las limitaciones de los modelos secuenciales previos como RNNs y LSTMs.

2.3.2. BERT y la Era del Pre-entrenamiento Bidireccional

BERT (Bidirectional Encoder Representations from Transformers) [25] introdujo el paradigma de pre-entrenamiento bidireccional, entrenando el modelo para predecir palabras enmascaradas considerando tanto el contexto izquierdo como el derecho.

Este enfoque genera representaciones contextuales más ricas que los modelos unidireccionales previos.

Variantes de BERT para Eficiencia

El éxito de BERT motivó el desarrollo de variantes más eficientes:

- **DistilBERT** [20]: Utiliza destilación de conocimiento para reducir el tamaño del modelo en un 40 % manteniendo el 97 % del rendimiento. Disponible en múltiples idiomas incluyendo español
- **TinyBERT** [26]: Aplica destilación a nivel de transformador y predicción para crear modelos aún más compactos, logrando excelente rendimiento. Sin embargo, su principal limitación es que únicamente está disponible en inglés, lo que restringe su aplicabilidad para tareas en español

2.3.3. Aplicaciones en Detección de Noticias Falsas en Español

Para el español específicamente, se han desarrollado modelos especializados y evaluaciones:

- Martínez-Gallego et al. [27] exploraron la aplicación de BERT y BETO (BERT en español), estableciendo líneas base importantes
- Blanco-Fernández et al. [13] compararon sistemáticamente BERT y RoBERTa para detección de desinformación política
- Shushkevich et al. [28] investigaron la mejora de clasificación multiclas usando datos aumentados con ChatGPT

2.3.4. Grandes Modelos de Lenguaje (LLMs) y Nuevos Paradigmas

GPT-3 y el Paradigma Few-Shot

GPT-3 [17] marcó un hito al demostrar capacidades emergentes de aprendizaje con pocos ejemplos (few-shot learning). Con 175 mil millones de parámetros, mostró que los modelos de gran escala pueden realizar tareas sin ajuste fino específico, solo con ejemplos en el prompt.

LLAMA y la Democratización de LLMs

LLAMA [29] representa el esfuerzo por democratizar el acceso a modelos de gran escala, proporcionando alternativas open-source a modelos propietarios. Su arquitectura optimizada permite un rendimiento competitivo con menor costo computacional.

Modelos Multimodales: Gemini

Gemini [30] introduce capacidades multimodales nativas, procesando texto, imágenes y audio de forma integrada. Esta capacidad es relevante para la detección de desinformación que cada vez más incorpora elementos multimedia.

IA Constitucional

El trabajo de Bai et al. [31] sobre IA Constitucional aborda la alineación y seguridad de los LLMs, estableciendo principios para entrenar modelos más seguros y alineados con valores humanos. Este enfoque es relevante cuando los LLMs se utilizan para tareas de detección de desinformación.

2.3.5. El Doble Rol de los LLMs en Detección

La investigación reciente ha revelado que los LLMs presentan un doble rol:

- **Como “Buenos Consejeros”:** Pueden ser ajustados finamente (fine-tuned) para detectar noticias falsas con alta precisión [2]
- **Como “Malos Actores”:** Pueden generar desinformación convincente, complicando la detección [6]
- **Adaptación Necesaria:** Los sistemas de detección deben adaptarse a la era de LLMs [16]

2.4. Optimización Metaheurística en Detección de Fraude

2.4.1. Fundamentos de las Metaheurísticas

Las metaheurísticas son estrategias de optimización de alto nivel que guían procesos de búsqueda para encontrar soluciones de alta calidad en espacios de búsqueda complejos. A diferencia de los algoritmos exactos, las metaheurísticas buscan soluciones “suficientemente buenas” en tiempo computacional razonable [32].

Justificación del Uso de Metaheurísticas en Detección de Noticias Falsas

La aplicación de metaheurísticas en este dominio se justifica por las siguientes características del problema:

- **Espacio de hiperparámetros complejo:** Los modelos de clasificación (SVM, redes neuronales) tienen múltiples hiperparámetros interdependientes
- **Función objetivo no convexa:** Las métricas de rendimiento (F1-Score, precisión) no son funciones continuas y diferenciables

- **Interacciones no lineales:** Los hiperparámetros interactúan de manera compleja, haciendo ineficaces los métodos de optimización tradicionales
- **Múltiples óptimos locales:** El espacio de búsqueda presenta numerosos máximos locales que pueden atrapar algoritmos determinísticos
- **Eficiencia computacional:** Las metaheurísticas exploran el espacio de forma más inteligente que métodos exhaustivos como búsqueda en cuadrícula (grid search)

Arquitectura General del Sistema Metaheurístico

El marco de trabajo (framework) metaheurístico desarrollado sigue una arquitectura modular que permite la intercambiabilidad de algoritmos:

$$\text{Sistema} = \{D, P, A, F, E\} \quad (2.5)$$

Donde:

- **D:** Dataset de noticias preprocessado
- **P:** Tubería de procesamiento (pipeline) de preprocessamiento (tokenización, TF-IDF)
- **A:** Algoritmo metaheurístico (GA, PSO, SA, VNS, SS)
- **F:** Función objetivo (F1-Score en validación cruzada)
- **E:** Evaluador final en conjunto de prueba

Entradas del Sistema Metaheurístico

1. Datos de Entrada:

- **Corpus textual:** Noticias en español con etiquetas binarias (REAL/FALSO)
- **División de datos:** 70 % entrenamiento, 10 % validación, 20 % prueba (estratificada)
- **Representación TF-IDF:** Matriz dispersa TF-IDF de dimensión $n \times v$ donde n es el número de documentos y v el tamaño del vocabulario

2. Espacio de Hiperparámetros:

Los algoritmos metaheurísticos optimizan múltiples conjuntos de hiperparámetros según el tipo de modelo utilizado. Como ejemplo representativo, para algoritmos de clasificación tradicionales se pueden optimizar parámetros de regularización, configuraciones de kernel, y esquemas de balanceado de clases.

Para TF-IDF:

$$\text{max_features} \in [1000, 50000] \text{ (tamaño del vocabulario)} \quad (2.6)$$

$$\text{min_df} \in [1, 10] \text{ (frecuencia mínima de documento)} \quad (2.7)$$

$$\text{max_df} \in [0.7, 1.0] \text{ (frecuencia máxima de documento)} \quad (2.8)$$

$$\text{ngram_range} \in \{(1, 1), (1, 2), (1, 3)\} \text{ (rango de n-gramas)} \quad (2.9)$$

Preprocesamiento y Tubería de Procesamiento (Pipeline)

Etapa 1: Limpieza Textual

1. **Normalización de caracteres:** Conversión a minúsculas, eliminación de acentos
2. **Filtrado de contenido:** Eliminación de URLs, menciones (@), hashtags
3. **Tokenización:** División en tokens usando expresiones regulares
4. **Eliminación de stopwords:** Filtrado de palabras vacías en español
5. **Validación de longitud:** Exclusión de textos demasiado cortos (< 10 tokens)

Etapa 2: Vectorización TF-IDF

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \log \left(\frac{N}{|\{d' : t \in d'\}|} \right) \quad (2.10)$$

Donde t es un término, d un documento, N el total de documentos, y el denominador cuenta documentos que contienen t .

Etapa 3: Normalización Aplicación de normalización L2 para estabilizar el entrenamiento:

$$\mathbf{x}_{\text{norm}} = \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \quad (2.11)$$

Algoritmos Metaheurísticos Implementados

1. Algoritmo Genético (GA)

Inspiración biológica: Evolución natural y selección de especies

Representación: Cada individuo es un vector de hiperparámetros codificados

$$\text{Individuo} = [C, \gamma, \text{max_features}, \text{min_df}, \text{max_df}, \text{ngram}] \quad (2.12)$$

Operadores genéticos:

- **Selección:** Torneo de tamaño 3
- **Cruce:** Cruce uniforme con probabilidad 0.8

- **Mutación:** Mutación gaussiana con probabilidad 0.1

- **Elitismo:** Preservación del 10 % mejores individuos

Parámetros:

$$\text{Tamaño población} = 30 \quad (2.13)$$

$$\text{Generaciones} = 50 \quad (2.14)$$

$$\text{Tasa cruce} = 0.8 \quad (2.15)$$

$$\text{Tasa mutación} = 0.1 \quad (2.16)$$

2. Optimización por Enjambre de Partículas (PSO)

Inspiración biológica: Comportamiento de bandadas de aves y cardúmenes

Ecuaciones de movimiento:

$$v_i^{t+1} = w \cdot v_i^t + c_1 \cdot r_1 \cdot (p_i - x_i^t) + c_2 \cdot r_2 \cdot (g - x_i^t) \quad (2.17)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (2.18)$$

Donde:

- v_i^t : velocidad de la partícula i en iteración t
- x_i^t : posición de la partícula i en iteración t
- p_i : mejor posición personal de la partícula i
- g : mejor posición global del enjambre
- w : factor de inercia ($w = 0.9$)
- c_1, c_2 : factores de aceleración ($c_1 = c_2 = 2.0$)
- r_1, r_2 : números aleatorios $\in [0, 1]$

Parámetros:

$$\text{Tamaño enjambre} = 30 \quad (2.19)$$

$$\text{Iteraciones} = 50 \quad (2.20)$$

$$\text{Factor inercia} = 0.9 \quad (2.21)$$

$$\text{Factores aceleración} = 2.0 \quad (2.22)$$

3. Recocido Simulado Multi-arranque (MSA)

Inspiración física: Proceso de enfriamiento controlado en metalurgia

Criterio de aceptación:

$$P(\text{aceptar}) = \begin{cases} 1 & \text{si } \Delta f \geq 0 \\ e^{\frac{\Delta f}{T}} & \text{si } \Delta f < 0 \end{cases} \quad (2.23)$$

Donde $\Delta f = f(\text{nueva}) - f(\text{actual})$ y T es la temperatura.

Esquema de enfriamiento:

$$T_{k+1} = \alpha \cdot T_k, \quad \alpha = 0.95 \quad (2.24)$$

Parámetros:

$$\text{Temperatura inicial} = 100.0 \quad (2.25)$$

$$\text{Temperatura final} = 0.01 \quad (2.26)$$

$$\text{Factor enfriamiento} = 0.95 \quad (2.27)$$

$$\text{Iteraciones por temperatura} = 10 \quad (2.28)$$

$$\text{Número de arranques} = 5 \quad (2.29)$$

4. Búsqueda en Vecindades Variables (VNS)

Principio: Cambio sistemático de estructuras de vecindad para escapar de óptimos locales

Estructuras de vecindad:

- N_1 : Perturbación de un hiperparámetro aleatorio
- N_2 : Perturbación de dos hiperparámetros aleatorios
- N_3 : Perturbación de todos los hiperparámetros

Algoritmo general:

1. Inicializar solución x
2. Para cada vecindad $k = 1, 2, 3$:
3. Generar x' en $N_k(x)$
4. Aplicar búsqueda local desde $x' \rightarrow x''$
5. Si $f(x'') > f(x)$: $x = x'', k = 1$
6. Sino: $k = k + 1$
7. Repetir hasta criterio de parada

5. Búsqueda Dispersa (SS)

Principio: Combinar soluciones de calidad y diversas para generar nuevas soluciones

Componentes principales:

- **Conjunto de referencia:** 10 soluciones (5 de calidad + 5 diversas)
- **Generación de subconjuntos:** Todas las combinaciones de tamaño 2
- **Método de combinación:** Promedio ponderado por aptitud (fitness)
- **Mejora:** Búsqueda local en nuevas soluciones
- **Actualización:** Reemplazo de peores soluciones

Función Objetivo y Evaluación

Función objetivo principal:

$$f(\theta) = \text{F1-Score}_{CV}(\theta) \quad (2.30)$$

Donde θ representa el vector de hiperparámetros y F1-Score_{CV} es el F1-Score promedio en validación cruzada estratificada de 5 pliegues.

Métricas de evaluación secundarias:

$$\text{Precisión} = \frac{TP}{TP + FP} \quad (2.31)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.32)$$

$$\text{F1-Score} = \frac{2 \times \text{Precisión} \times \text{Recall}}{\text{Precisión} + \text{Recall}} \quad (2.33)$$

$$\text{Exactitud} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.34)$$

Etapas del Proceso Metaheurístico

Etapa 1: Inicialización

1. Cargar y dividir el dataset
2. Definir espacios de búsqueda para hiperparámetros
3. Inicializar población/enjambre según el algoritmo
4. Configurar parámetros del algoritmo metaheurístico

Etapa 2: Evaluación de Aptitud (Fitness)

1. Para cada configuración de hiperparámetros:
 2. Construir pipeline TF-IDF + Clasificador
 3. Entrenar en conjunto de entrenamiento
 4. Evaluar en validación cruzada 5-fold
 5. Calcular F1-Score promedio como aptitud (fitness)

Etapa 3: Evolución/Optimización

1. Aplicar operadores específicos del algoritmo
2. Generar nuevas configuraciones candidatas
3. Evaluar fitness de nuevas configuraciones

4. Actualizar población/enjambre según criterios del algoritmo
5. Verificar criterios de convergencia

Etapa 4: Selección Final

1. Identificar la mejor configuración encontrada
2. Re-entrenar modelo con configuración óptima
3. Evaluar en conjunto de prueba independiente
4. Reportar métricas finales

Salidas del Sistema Metaheurístico

1. Configuración Óptima de Hiperparámetros:

- Valores óptimos para cada hiperparámetro del modelo
- Historia de convergencia del algoritmo
- Número de evaluaciones realizadas
- Tiempo total de optimización

2. Modelo Entrenado Optimizado:

- Vectorizador TF-IDF configurado
- Clasificador entrenado
- Pipeline completo listo para predicción

3. Métricas de Rendimiento:

- F1-Score, Precisión, Recall, Exactitud en conjunto de prueba
- Matriz de confusión
- Curvas de convergencia
- Análisis de importancia de hiperparámetros

4. Análisis Comparativo:

- Comparación entre algoritmos metaheurísticos
- Análisis de eficiencia computacional
- Estudio de robustez y estabilidad

Comparación de Algoritmos Metaheurísticos

La Tabla 2.2 (ver tabla 2.2) presenta una comparación sistemática de los cinco algoritmos implementados, destacando sus características distintivas, ventajas y limitaciones en el contexto de optimización de hiperparámetros para detección de noticias falsas.

Algoritmo	Inspiración	Fortalezas	Limitaciones	Parámetros Clave
GA	Evolución biológica	Exploración global robusta Paralelización natural Buena diversidad poblacional	Convergencia lenta Muchos parámetros a ajustar Puede estancarse prematuramente	Población: 30 Generaciones: 50 Cruce: 0.8 Mutación: 0.1
PSO	Comportamiento de enjambre	Convergencia rápida Pocos parámetros Fácil implementación	Puede convergir prematuramente Sensible a parámetros Exploración limitada	Partículas: 30 Iteraciones: 50 Inercia: 0.9 Aceleración: 2.0
SA / MSA	Recocido de metales	Escape de óptimos locales Base teórica sólida Múltiples arranques	Esquema de enfriamiento crítico Búsqueda individual Muchas evaluaciones	T inicial: 100 T final: 0.01 Enfriamiento: 0.95 Arranques: 5
VNS	Cambio sistemático de vecindades	Escape sistemático de óptimos Combinación exploración/explotación Flexibilidad en vecindades	Dependiente de definición de vecindades Puede ser costoso computacionalmente Diseño específico del problema	Vecindades: 3 Iteraciones: 100 Búsqueda local: Hill climbing
SS	Combinación de soluciones de calidad	Balance calidad-diversidad Intensificación y diversificación Memoria adaptativa	Complejo de implementar Muchos componentes Sensible a tamaño de referencia	Ref. Set: 10 Combinaciones: todas Mejora: local search Actualizaciones: dinámicas

Tabla 2.2: Comparación de algoritmos metaheurísticos implementados

Criterios de Selección y Justificación

La selección de estos cinco algoritmos metaheurísticos se basó en criterios específicos de idoneidad para el problema de optimización de hiperparámetros en detección de noticias falsas:

- Diversidad de paradigmas:** Se incluyeron representantes de las principales familias de metaheurísticas (evolutivas, de enjambre, basadas en trayectoria, de memoria)
- Eficiencia computacional:** Algoritmos con balance adecuado entre calidad de solución y tiempo de cómputo
- Robustez empírica:** Métodos con evidencia documentada de efectividad en problemas de optimización de ML
- Implementación práctica:** Algoritmos con parámetros bien establecidos e implementaciones estables
- Complementariedad:** Enfoques que exploran el espacio de búsqueda de maneras fundamentalmente diferentes

Esta selección permite comparar diferentes estrategias de optimización, proporcionando información sobre qué características algorítmicas son más efectivas para el problema específico de calibración de detectores de noticias falsas.

Justificación Frente a Métodos Alternativos

Comparación con Búsqueda en Cuadrícula (Grid search):

- **Eficiencia:** Grid search requiere $O(n^k)$ evaluaciones donde n es el número de valores por parámetro y k el número de parámetros. Para este problema: 10^6 evaluaciones vs. 1,500 de metaheurísticas
- **Escalabilidad:** Grid search se vuelve impracticable con espacios de alta dimensión
- **Optimización continua:** Las metaheurísticas manejan parámetros continuos naturalmente

Comparación con Búsqueda Aleatoria (Random Search):

- **Inteligencia dirigida:** Las metaheurísticas aprenden del histórico de evaluaciones
- **Convergencia:** Random search no garantiza convergencia hacia regiones prometedoras
- **Explotación vs. Exploración:** Las metaheurísticas balancean sistemáticamente ambos aspectos

Comparación con Optimización Bayesiana:

- **Asunciones del modelo:** Bayesian optimization asume suavidad que puede no cumplirse en métricas de ML
- **Costo computacional:** El cálculo de la función de adquisición puede ser costoso
- **Flexibilidad:** Las metaheurísticas son más flexibles para restricciones complejas

Comparación con Métodos Basados en Gradiente (Gradient-based Methods):

- **Diferenciabilidad:** Las métricas de clasificación (F1-Score) no son diferenciables
- **Naturaleza discreta:** Algunos hiperparámetros son categóricos o discretos
- **Robustez:** Los métodos basados en gradiente son sensibles a óptimos locales

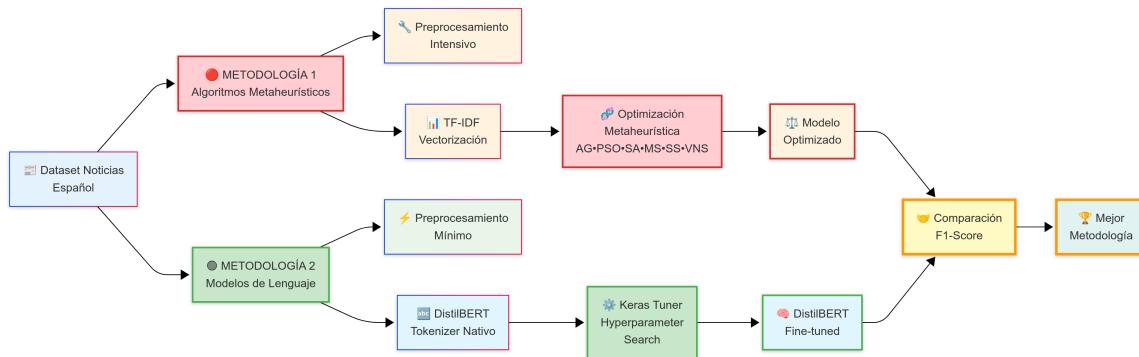


Figura 2.1: Flujo del proceso de optimización para detección de noticias falsas. Se muestra la transformación desde datos textuales hasta el modelo optimizado final, pasando por las etapas de preprocesamiento, optimización y evaluación.

Flujo del Proceso Metaheurístico

La Figura 2.1 (ver figura 2.1) ilustra el flujo completo del proceso de optimización de ambos enfoques aplicado a la detección de noticias falsas, desde la entrada de datos hasta la obtención del modelo optimizado final.

Este flujo garantiza:

- **Reproducibilidad:** Cada etapa está claramente definida con parámetros específicos
- **Escalabilidad:** El proceso puede aplicarse a corpus de diferentes tamaños
- **Transferibilidad:** La metodología es aplicable a otros problemas de clasificación textual
- **Transparencia:** Cada decisión algorítmica está justificada y documentada

Consideraciones de Implementación

Manejo de Memoria:

- Uso de matrices dispersas (scipy.sparse) para representaciones TF-IDF
- Liberación explícita de memoria entre evaluaciones
- Procesamiento por lotes para corpus grandes

Paralelización:

- Evaluación paralela de individuos en GA
- Vectorización de operaciones en PSO

- Múltiples arranques independientes en SA

Validación y Robustez:

- Validación cruzada estratificada para preservar distribución de clases
- Múltiples ejecuciones independientes para análisis estadístico
- Manejo de casos extremos y configuraciones inválidas

Métricas de Monitoreo del Proceso

Durante la ejecución de cada algoritmo metaheurístico, se monitorizan las siguientes métricas:

1. **Convergencia:** Evolución del mejor fitness a lo largo de las iteraciones
2. **Diversidad:** Medida de dispersión en la población/enjambre
3. **Estagnación:** Número de iteraciones sin mejora significativa
4. **Eficiencia:** Número de evaluaciones para alcanzar convergencia
5. **Estabilidad:** Variabilidad entre ejecuciones independientes

Estas métricas permiten:

- Ajustar parámetros de los algoritmos dinámicamente
- Detectar convergencia prematura o estagnación
- Comparar el rendimiento de diferentes metaheurísticas
- Identificar configuraciones problemáticas o excepcionales

2.4.2. Aplicaciones en Detección de Noticias Falsas

Modelado como Problema de Programación de Trabajos (Job Shop Scheduling)

Aqil y Lahby [33] propusieron una perspectiva innovadora, modelando la detección de noticias falsas como un problema de programación de trabajos en flujo (Flow Shop Scheduling Problem, FSSP). Compararon tres metaheurísticas:

- **Voraz Iterativo (Iterated Greedy, IG):** Algoritmo constructivo que mejora iterativamente la solución mediante destrucción y reconstrucción
- **Algoritmo Genético (GA):** Evolución de soluciones mediante selección, cruce y mutación

- **Colonia de Abejas Artificiales (ABC):** Algoritmo basado en comportamiento de forrajeo de abejas

Sus resultados mostraron que el algoritmo Voraz Iterativo (IG) logró el mejor rendimiento en términos de minimización del makespan (tiempo máximo de finalización), superando a los otros dos algoritmos metaheurísticos (GA y ABC) en este contexto específico.

Optimización de Hiperparámetros en Deep Learning

Bacanin et al. [34] demostraron los beneficios de usar metaheurísticas para la calibración de hiperparámetros en modelos de deep learning. Su investigación mostró mejoras significativas sobre métodos tradicionales como grid search y random search:

- **Eficiencia computacional:** Exploración más inteligente del espacio de hiperparámetros
- **Evitación de óptimos locales:** Capacidad de escape de configuraciones subóptimas
- **Adaptabilidad:** Ajuste dinámico según el comportamiento del modelo

La investigación que publiqué en [23] específicamente aborda la calibración de hiperparámetros en algoritmos metaheurísticos para detección de noticias falsas (como caso específico de fraude digital), proporcionando un marco metodológico relevante para esta tesis.

Marcos de Trabajo (Frameworks) de Ensamble (Ensemble) Heurístico

Das et al. [21] desarrollaron un marco de ensamble (ensemble) que combina:

- **Modelos pre-entrenados:** BERT y similares para captura semántica
- **Características estadísticas:** Metadatos como URL, autor, timestamp
- **Heurísticas de incertidumbre:** Medidas de confianza para ponderación de predictores

Este enfoque híbrido logró F1-scores superiores al 95

2.4.3. Metaheurísticas para Detección de Fraude Financiero

Optimización Multi-objetivo

Hidayattullah et al. [35] aplicaron metaheurísticas para optimizar la detección de fraude en estados financieros. Su enfoque multi-objetivo consideró:

- **Precisión de clasificación:** Maximización de métricas de rendimiento

- **Eficiencia computacional:** Minimización de tiempo de entrenamiento
- **Robustez:** Estabilidad ante variaciones en los datos

Los resultados mostraron que SVM optimizado con Algoritmo Genético alcanzó 96.15 % de precisión, superando significativamente a métodos tradicionales.

Optimización de Procesos Gaussianos

Horak y Sabek [36] exploraron la optimización de hiperparámetros en Gaussian Process Regression para predicción de dificultades financieras. Su trabajo destaca la importancia de la calibración metaheurística en modelos probabilísticos.

2.4.4. Algoritmos Híbridos Modernos

Enfoques Multi-thread

Yildirim [37] propuso un enfoque metaheurístico híbrido multi-thread que optimiza simultáneamente:

- Selección de características
- Parámetros del algoritmo de clasificación
- Arquitectura del ensemble

PSO Adaptativo

Deshai y Bhaskara Rao [38] desarrollaron un enfoque CNN con PSO adaptativo para detectar reseñas falsas online. Su algoritmo PSO adaptativo ajusta dinámicamente los parámetros de velocidad e inercia basándose en el rendimiento de la iteración actual.

Fine-tuning: Adaptación de Modelos Pre-entrenados

Una de las innovaciones más significativas en PLN moderno ha sido el desarrollo del paradigma de **fine-tuning** o ajuste fino, introducido formalmente por Howard y Ruder [39] en su trabajo sobre ULMFiT (Universal Language Model Fine-tuning). Este enfoque representa un cambio paradigmático respecto a los métodos tradicionales de entrenamiento desde cero, permitiendo aprovechar el conocimiento adquirido por modelos durante su pre-entrenamiento en grandes corpus.

El proceso de fine-tuning consiste en tomar un modelo pre-entrenado y continuar su entrenamiento con datos específicos del dominio objetivo, manteniendo la arquitectura base pero adaptando los pesos para la nueva tarea. En el contexto de modelos Transformer como BERT [40], esto implica congelar parcial o totalmente las capas inferiores que capturan características lingüísticas generales, mientras se ajustan las capas superiores para la tarea específica de clasificación.

La efectividad de este enfoque radica en que las representaciones contextuales aprendidas durante el pre-entrenamiento en tareas como masked language modeling contienen información semántica y sintáctica rica que es transferible a tareas downstream [41]. Para la detección de noticias falsas, esto significa que el modelo ya comprende patrones lingüísticos complejos, estructura narrativa y coherencia semántica antes de especializarse en identificar marcadores de desinformación, constituyendo la base metodológica fundamental del segundo enfoque desarrollado en esta investigación.

2.5. Síntesis del Marco Teórico

La tendencia actual en detección de noticias falsas favorece sistemas híbridos que combinan múltiples paradigmas metodológicos, principios que son extensibles a otros tipos de fraude digital. Por ejemplo, Zhang et al. [42] exploraron el uso de LLMs para optimización de hiperparámetros, abriendo nuevas posibilidades para la optimización automática de pipelines completos de detección.

Este marco teórico establece los fundamentos para las dos metodologías desarrolladas en esta tesis:

1. **Enfoque Clásico Optimizado:** Utiliza representaciones TF-IDF con optimización metaheurística de hiperparámetros, proporcionando una línea base sólida y computacionalmente eficiente
2. **Enfoque de Deep Learning:** Emplea modelos Transformer pre-entrenados con fine-tuning, aprovechando representaciones contextuales sofisticadas

La integración de ambos enfoques, sustentada en la literatura revisada, establece la base teórica para la contribución metodológica principal de esta tesis: la aplicación sistemática de optimización de hiperparámetros en ambos paradigmas, desde clasificadores tradicionales hasta modelos de deep learning, en el contexto específico de textos en español.

Capítulo 3

Estado del arte

La detección automatizada de noticias falsas ha emergido como una de las preocupaciones más apremiantes en la sociedad digital contemporánea, convirtiéndose en un campo de investigación multidisciplinario que abarca desde la ciencia de la computación hasta la psicología cognitiva. Durante la última década, investigadores de múltiples disciplinas han desarrollado enfoques innovadores que van desde técnicas tradicionales de procesamiento de lenguaje natural hasta modelos de lenguaje pre-entrenados de última generación, respondiendo a la creciente sofisticación de las campañas de desinformación y la velocidad sin precedentes con la que se propagan en las redes sociales.

En el contexto hispanohablante, esta problemática adquiere características particulares debido a las especificidades lingüísticas y culturales del español, incluyendo variaciones regionales, modismos locales, y patrones discursivos específicos que requieren enfoques metodológicos adaptados. La complejidad del español, con sus múltiples variantes geográficas y sus ricas estructuras sintácticas, presenta desafíos únicos para los sistemas automatizados de detección que tradicionalmente han sido desarrollados y entrenados principalmente en inglés.

Este capítulo examina las principales contribuciones científicas en el campo de la detección de desinformación, con especial énfasis en las aproximaciones que combinan técnicas de optimización metaheurística con modelos de aprendizaje automático. Se analiza también la evolución de los enfoques metodológicos, desde los primeros sistemas basados en características lingüísticas superficiales hasta los desarrollos más recientes que incorporan arquitecturas transformer y algoritmos de optimización bio-inspirados, proporcionando las bases para entender tanto los avances logrados como los desafíos pendientes en este campo dinámico.

3.1. Metodología de Búsqueda Bibliográfica

Para garantizar una revisión lo más completa y actualizada posible de la literatura científica en el campo de la detección de noticias falsas, se diseñó una estrategia de búsqueda consultando múltiples bases de datos académicas reconocidas internacional-

mente. El proceso de búsqueda se llevó a cabo entre agosto de 2023 y agosto de 2025, utilizando una combinación de términos clave específicos y operadores booleanos para maximizar la relevancia de los resultados.

Las bases de datos consultadas incluyeron repositorios académicos como Google Scholar, Semantic Scholar, Nature Digital Libraries, IOPscience, MDPI, Web of Science (WoS) y repositorios especializados como arXiv. La selección de estas fuentes se basó en su cobertura disciplinaria, calidad editorial, y relevancia específica para el campo de procesamiento de lenguaje natural y detección de desinformación. Adicionalmente, se incluyeron conferencias especializadas como ACL, EMNLP, y talleres específicos como CheckThat! para capturar los desarrollos más recientes en el área.

La búsqueda se estructuró en torno a combinaciones específicas de términos que reflejan las dimensiones principales de esta investigación, utilizando operadores booleanos para mejorar los resultados:

- *Spanish fake news detection* – para capturar trabajos específicos del idioma español, incluyendo variaciones como “detección de noticias falsas en español”
- *Fake news detection language models* – para incluir enfoques basados en modelos de lenguaje, cubriendo términos como BERT, RoBERTa, y transformer
- *Metaheuristic hyperparameter optimization* – para identificar aplicaciones de optimización metaheurística en aprendizaje automático
- *Employment fraud detection* – para abarcar el dominio del fraude laboral y su detección automatizada
- *Simulated annealing fake news* y *Genetic algorithm text classification* – para enfoques específicos de optimización aplicados a clasificación textual
- *Social media misinformation detection* – para capturar trabajos enfocados en plataformas sociales
- *Cross-lingual fake news* y *multilingual misinformation* – para incluir enfoques multilingües

El proceso de selección aplicó diferentes criterios de relevancia temática, calidad metodológica y actualidad temporal. Se priorizaron publicaciones de los últimos cinco años (2020-2025) sin excluir trabajos fundamentales para el campo que, aunque anteriores, establecieron bases teóricas o metodológicas esenciales. Los criterios de inclusión incluyeron: (1) contribuciones metodológicas originales, (2) validación empírica, (3) relevancia directa para la detección de desinformación, y (4) calidad de la revisión por pares. Los criterios de exclusión abarcaron: (1) trabajos puramente teóricos sin validación, (2) estudios con muestras insuficientes, (3) publicaciones en idiomas distintos al inglés o español, y (4) trabajos duplicados o versiones preliminares de conferencias posteriormente publicadas en revistas.

Como resultado de este proceso, se identificaron 80 trabajos que constituyen el núcleo de esta revisión, organizados temáticamente para facilitar el análisis comparativo y la identificación de tendencias emergentes en el campo.

3.2. Panorama de la Investigación en Detección de Noticias Falsas y Modelos de Lenguaje

El campo de la detección automatizada de desinformación ha experimentado una evolución notable y acelerada, caracterizada por una transición paradigmática desde enfoques tradicionales basados en características estadísticas simples hacia sistemas sofisticados que integran múltiples modalidades, técnicas de optimización avanzadas, y conocimiento contextual profundo. Esta evolución refleja tanto los avances tecnológicos en inteligencia artificial como la creciente sofisticación de las técnicas de desinformación empleadas por actores maliciosos.

La progresión temporal del campo puede dividirse en tres fases principales: (1) la era pionera (2010-2016), caracterizada por enfoques basados en características lingüísticas superficiales y métodos de aprendizaje automático tradicionales; (2) la era de los modelos profundos (2017-2020), marcada por la adopción masiva de redes neuronales profundas y arquitecturas transformer; y (3) la era contemporánea (2021-presente), dominada por modelos de lenguaje de gran escala y enfoques híbridos que combinan múltiples técnicas de optimización.

3.2.1. Fundamentos Metodológicos y Construcción de Corpus

La disponibilidad de conjuntos de datos de alta calidad constituye el fundamento sobre el cual se construye cualquier sistema de detección eficaz, y representa uno de los mayores desafíos en el desarrollo de tecnologías de detección de desinformación. En el contexto hispanohablante, esta necesidad se vuelve particularmente crítica debido a la relativa escasez de recursos comparado con el inglés, donde la mayoría de la investigación se ha concentrado históricamente.

Los trabajos pioneros en esta dirección (ver tabla 3.1) han establecido las bases metodológicas para la construcción y evaluación de corpus especializados, desarrollando protocolos de anotación, criterios de calidad, y marcos de evaluación que han influido significativamente en el desarrollo del campo. Acosta [11] desarrolló uno de los primeros marcos sistemáticos para la construcción de conjuntos de datos de noticias en español, estableciendo criterios rigurosos de calidad y protocolos de etiquetado manual que han influido significativamente en trabajos posteriores. Su metodología, centrada en la extracción automatizada seguida de verificación manual, demostró la viabilidad de crear recursos lingüísticos de gran escala para el español mientras mantenía estándares de calidad comparables a los corpus en inglés. El trabajo incluye un análisis detallado de las fuentes de noticias, criterios de selección temporal, y metodologías de validación cruzada entre anotadores.

El trabajo de Posadas-Durán et al. [12] complementó estos esfuerzos al proponer un nuevo corpus específicamente diseñado para la detección de noticias falsas en español, incorporando características estilométricas innovadoras que capturan patrones lingüísticos sutiles pero significativos para la identificación de contenido desinformativo. Este enfoque pionero en el análisis del estilo de escritura como indicador de veracidad abrió nuevas líneas de investigación que han sido extensamente exploradas en trabajos posteriores. Su corpus incluye anotaciones multinivel que abarcan desde características superficiales del texto hasta análisis semánticos profundos, proporcionando un recurso invaluable para investigaciones subsiguientes.

La comunidad científica internacional ha reconocido la importancia crítica de esta área a través de iniciativas competitivas sistemáticas que han proporcionado marcos estandarizados para la evaluación y comparación de sistemas. Los talleres CheckThat! organizados en el marco de CLEF [43, 44] han proporcionado marcos estandarizados para la evaluación comparativa de sistemas de detección, estableciendo métricas unificadas, protocolos de evaluación rigurosos, y conjuntos de datos de referencia que han contribuido significativamente a la maduración metodológica del campo. Estos talleres han evolucionado para incluir tareas cada vez más sofisticadas, desde la verificación de hechos básicos hasta la detección de sesgo político y la evaluación de credibilidad en contenido multimodal.

Un desafío fundamental en este campo es que no todos los corpus disponibles son equivalentes en términos de propósito, metodología de construcción, o calidad de anotación. Esta heterogeneidad refleja la evolución natural del campo, donde diferentes grupos de investigación han abordado aspectos específicos del problema con enfoques metodológicos diversos. Algunos corpus se enfocan en análisis estilométrico, otros en detección política, y otros en evaluación de modelos específicos, lo que genera un ecosistema rico pero complejo de recursos disponibles.

En el contexto iberoamericano específicamente, las competiciones IberLEF han jugado un papel fundamental en el desarrollo de recursos y metodologías adaptadas a las particularidades del español. Aragón et al. [45] documentaron los primeros esfuerzos sistemáticos para abordar la detección de noticias falsas en español mexicano, estableciendo métricas específicas para variantes regionales del idioma y desarrollando métricas de evaluación culturalmente sensibles. **Crucialmente, estos trabajos establecieron F1-Score como la métrica principal de comparación entre sistemas, junto con precisión y recall como métricas complementarias para análisis detallado del rendimiento.** Paralelamente, Gómez-Adorno et al. [46] consolidaron estos esfuerzos en el marco de FakeDeS, estableciendo un referente duradero para la evaluación de sistemas en español que incluye tanto aspectos técnicos como consideraciones socioculturales.

La estandarización de métricas de evaluación ha sido fundamental para el progreso del campo. Los marcos CheckThat! [44] han establecido que las métricas fundamentales para clasificación binaria en detección de noticias falsas deben incluir: (1) *Precisión* para medir la confiabilidad de las detecciones positivas, (2) *Recall* (exhaustividad) para evaluar la cobertura de casos positivos reales, (3) *F1-Score* como

balance armónico entre ambas, y (4) *Exactitud* para rendimiento general en datasets balanceados. Esta estandarización permite la comparación directa entre sistemas desarrollados por diferentes grupos de investigación y garantiza la reproducibilidad de los resultados.

La evolución hacia recursos más sofisticados se refleja en iniciativas recientes como el Spanish Fake News Corpus Version 2.0 [47], que incorpora anotaciones multinivel y metadatos enriquecidos que permiten análisis más profundos de los patrones de desinformación. Este corpus de segunda generación incluye información contextual sobre fuentes, temas, y patrones temporales de propagación, proporcionando un recurso muy valioso para la investigación en detección automatizada.

Contribución Principal	Autores	Publicación	Año	Ref.
Marco metodológico para construcción de conjuntos de datos de noticias en español	Acosta, F. A. Z.	U. Politécnica de Madrid	2019	[11]
Evaluación multimodal y multigenre para verificación de contenido	Alam, F., et al.	CEUR Workshop Proc.	2023	[43]
Análisis de agresividad y desinformación en español mexicano	Aragón, M. E., et al.	IberLEF Workshop	2020	[45]
Marco estandarizado CheckThat! para evaluación de sistemas	Barrón-Cedeño, A., et al.	ECIR 2023	2023	[44]
Consolidación de métricas FakeDeS para detección en español	Gómez-Adorno, H., et al.	Procesamiento del Lenguaje Natural	2021	[46]
Corpus pionero con características estilométricas para español	Posadas-Durán, J. P., et al.	J. of Intelligent and Fuzzy Systems	2019	[12]
Corpus enriquecido con anotaciones multinivel versión 2.0	Ramírez Cruz, J. M., et al.	GitHub Repository	2021	[47]
Técnicas de aprendizaje automático aplicadas al español	Tretiakov, A., et al.	Springer	2022	[19]

Tabla 3.1: Contribuciones fundamentales en metodología y construcción de corpus para español.

Esta base metodológica en constante evolución ha permitido el desarrollo de aproximaciones cada vez más sofisticadas que integran múltiples tipos de evidencia, técnicas de análisis avanzadas, y conocimiento contextual, estableciendo el terreno base para los desarrollos contemporáneos en modelos de lenguaje de gran escala y optimización metaheurística que se analizan en las siguientes secciones.

3.2.2. Evolución de los Modelos de Lenguaje en Detección de Desinformación

La irrupción de los modelos de lenguaje pre-entrenados ha revolucionado completamente el panorama de la detección automatizada de noticias falsas, representando un cambio paradigmático fundamental en cómo conceptualizamos y abordamos la comprensión automatizada de texto (ver figura 3.1). Esta transformación se caracteriza por una progresión clara y acelerada desde arquitecturas básicas hacia sistemas cada vez más sofisticados que integran conocimiento contextual profundo, capacidades de razonamiento complejo, y comprensión semántica avanzada.

El punto de inflexión histórico llegó con la introducción de BERT [25], cuya arquitectura bidireccional revolucionaria permitió capturar dependencias contextuales

de manera más efectiva que los enfoques secuenciales previos basados en LSTM o GRU. La capacidad fundamental de BERT para entender el contexto completo de una oración de manera simultánea, en lugar de procesarla secuencialmente de izquierda a derecha como los modelos anteriores, representó un avance fundamental que ha sido extensamente aplicado y adaptado en la detección de desinformación. Esta arquitectura bidireccional permite al modelo capturar relaciones complejas entre palabras que pueden estar distantes en el texto pero semánticamente relacionadas, una característica crucial para identificar inconsistencias sutiles y patrones desinformativos sofisticados.



Figura 3.1: Mapa Conceptual 3: Artículos relacionados que incorporan Modelos de Lenguaje.

La necesidad imperativa de modelos más eficientes computacionalmente, especialmente para aplicaciones en tiempo real y dispositivos con recursos limitados, llevó al desarrollo de variantes optimizadas como DistilBERT [20], que mantiene aproximadamente el 97 % del rendimiento de BERT con solo el 60 % de sus parámetros mediante técnicas innovadoras de destilación de conocimiento. Esta línea de investigación, que incluye también desarrollos como TinyBERT [26], ha demostrado ser crucial para aplicaciones prácticas de detección en tiempo real donde los recursos computacionales son limitados pero la precisión no puede comprometerse significativamente.

En el contexto específico y culturalmente relevante del español, el trabajo de Martínez-Gallego et al. [27] exploró la aplicación de técnicas de aprendizaje profundo, incluyendo tanto BERT como BETO (la variante específicamente pre-entrenada en español), para la detección de noticias falsas. Sus resultados demostraron de manera convincente que los modelos específicamente pre-entrenados en español superan consistentemente a las versiones multilingües genéricas, subrayando la importancia de la especialización lingüística y cultural en el desarrollo de sistemas de detección

efectivos. Este hallazgo tiene implicaciones profundas para el desarrollo de sistemas de detección en idiomas distintos al inglés.

Blanco-Fernández et al. [13] extendieron significativamente esta línea de investigación al comparar sistemáticamente BERT y RoBERTa en tareas específicas de detección de desinformación en español, proporcionando evidencia empírica detallada sobre las ventajas relativas de diferentes arquitecturas transformer para este dominio específico. Su trabajo demostró que RoBERTa, con su entrenamiento más extenso y metodología de enmascaramiento optimizada, puede ofrecer mejoras significativas en la detección de patrones sutiles de desinformación, especialmente en textos que emplean técnicas de manipulación sofisticadas.

El surgimiento explosivo de modelos generativos de gran escala ha introducido dimensiones completamente nuevas y complejas al problema de detección. Brown et al. [17] establecieron con GPT-3 las bases conceptuales y técnicas de los modelos de lenguaje de gran escala, cuyas capacidades avanzadas de generación de texto han complicado paradójicamente el problema de detección al crear contenido sintético cada vez más indistinguible del contenido humano auténtico. El equipo de Gemini [30] ha continuado esta evolución hacia modelos multimodales que pueden procesar y generar simultáneamente texto, imágenes y otros tipos de información, ampliando el alcance y la complejidad de los desafíos de detección.

Esta evolución acelerada ha generado nuevos desafíos metodológicos y éticos que requieren enfoques innovadores (ver tabla 3.2. Su et al. [16] analizaron cómo la era de los modelos de lenguaje grandes requiere adaptaciones fundamentales en las técnicas de detección tradicionales, incluyendo nuevos marcos teóricos y metodológicos para abordar contenido generado artificialmente. Complementariamente, en otro de sus trabajos [6] reveló que los detectores existentes muestran sesgos sistemáticos significativos contra texto generado por LLMs, planteando cuestiones fundamentales sobre la generalización, robustez, y equidad de los sistemas de detección contemporáneos.

El enfoque hacia la mejora continua y la adaptación de estos sistemas se refleja en trabajos innovadores como el de Shushkevich et al. [28], que exploraron la combinación sinérgica de modelos BERT con datos aumentados mediante ChatGPT, demostrando cómo los propios LLMs pueden utilizarse estratégicamente para mejorar la detección de contenido falso mediante técnicas de augmentación de datos y transferencia de conocimiento.

Una perspectiva particularmente interesante e importante emerge del trabajo de Hu et al. [2], que examina el papel dual y paradójico de los modelos de lenguaje grandes como potenciales generadores de desinformación y, simultáneamente, como herramientas para su detección. Esta dualidad fundamental refleja la complejidad creciente del panorama tecnológico actual y plantea cuestiones éticas y prácticas importantes sobre el desarrollo responsable de tecnologías de IA.

Innovación Principal	Autores	Publicación	Año	Ref.
Arquitectura bidireccional para comprensión contextual profunda	Devlin, J., et al.	arXiv	2018	[25]
Destilación de conocimiento para modelos eficientes	Sanh, V., et al.	arXiv	2019	[20]
Optimización extrema con TinyBERT	Jiao, X., et al.	arXiv	2019	[26]
Aplicación sistemática de transformers en español	Martínez-Gallego, K., et al.	arXiv	2021	[27]
Comparación BERT vs RoBERTa para español	Blanco-Fernández, Y., et al.	Applied Sciences	2024	[13]
Capacidades emergentes en modelos de gran escala	Brown, T. B., et al.	arXiv	2020	[17]
Modelos multimodales de próxima generación	Gemini Team, Google	arXiv	2023	[30]
Adaptación para la era de LLMs	Su, J., et al.	arXiv	2023	[16]

Tabla 3.2: Evolución de modelos de lenguaje aplicados a detección de desinformación.

3.2.3. Técnicas de Optimización Metaheurística en Detección de Desinformación

La aplicación de algoritmos metaheurísticos a la detección de noticias falsas presenta una forma innovadora entre dos áreas de investigación que han madurado de manera independiente pero complementaria durante las últimas décadas. Por un lado, las metaheurísticas han demostrado repetidamente su eficacia excepcional en la resolución de problemas de optimización complejos, no lineales, y multimodales en múltiples dominios científicos y de ingeniería. Por otro lado, la detección de desinformación enfrenta desafíos inherentes y crecientes de optimización, particularmente en la calibración precisa de hiperparámetros, la selección inteligente de características relevantes, y la optimización de arquitecturas de modelos complejos.

Aqil y Lahby [33] fueron pioneros en modelar esta tarea como un problema de planificación de tareas (Job Shop Scheduling). En su enfoque, diferentes algoritmos de procesamiento de texto deben ejecutarse de manera óptima considerando restricciones de tiempo, recursos computacionales y calidad de resultados. Los autores evaluaron tres metaheurísticas principales: Iterated Greedy, Algoritmos Genéticos y Artificial Bee Colony. Sus resultados mostraron que Iterated Greedy superaba consistentemente a las otras alternativas tanto en eficiencia computacional como en calidad de soluciones, estableciendo un precedente metodológico importante para investigaciones posteriores.

En el ámbito específico de la detección de reseñas falsas, que comparte muchas características con la detección de noticias falsas, Deshai y Rao [38] (ver tabla 3.3) desarrollaron un enfoque innovador y técnicamente sofisticado que combina redes neuronales convolucionales (CNN) con Optimización por Enjambre de Partículas adaptativo (PSO). Su contribución principal radica en la demostración empírica y teórica de que PSO puede optimizar efectivamente los hiperparámetros complejos de redes

CNN, superando métodos tradicionales como Grid Search y Random Search en términos de eficiencia computacional, convergencia, y calidad de resultados finales. El trabajo incluye un análisis detallado de la superficie de respuesta del espacio de hiperparámetros y demuestra cómo PSO puede navegar eficientemente estos espacios multidimensionales complejos.

Bacanin et al. [34] investigaron la aplicación de metaheurísticas para optimizar hiperparámetros en modelos de aprendizaje profundo, enfocándose originalmente en predicción de cargas energéticas. Sus resultados con Algoritmos Genéticos Binarios, PSO y Recocido Simulado para modelos LSTM demuestran mejoras consistentes sobre métodos tradicionales de optimización. El trabajo presenta comparaciones rigurosas con técnicas base y análisis estadísticos que validan la efectividad de estas aproximaciones metaheurísticas, ofreciendo resultados valiosos que se extienden más allá del dominio energético hacia aplicaciones de procesamiento de texto y análisis de sentimientos.

Das et al. [21] desarrollaron un marco de trabajo basado en técnicas heurísticas para manejar la incertidumbre en la clasificación de tweets y artículos de noticias. Su enfoque reconoce que la detección de desinformación requiere el manejo sofisticado de múltiples fuentes de evidencia y contextos sociales complejos, más allá de la clasificación binaria tradicional. El marco de trabajo incluye mecanismos para cuantificar y propagar incertidumbre a través de las diferentes etapas del flujo de datos de detección.

Yildirim [37] propuso un enfoque multi-hilo que combina múltiples metaheurísticas para la detección multimodal de noticias falsas. Esta aproximación reconoce que los sistemas de desinformación contemporáneos integran múltiples modalidades (texto, imágenes, videos, metadatos) que requieren estrategias de optimización coordinadas para lograr detección efectiva.

Contribución Metodológica	Autores	Publicación	Año	Ref.
Formulación como problema de planificación de tareas	Aqil, S., & Lahby, M.	Studies in Comp. Intelligence	2021	[33]
PSO adaptativo para optimización de CNN en detección de reseñas	Deshai, N., & Rao, B. B.	Soft Computing	2023	[38]
Evidencia sistemática de beneficios en modelos de aprendizaje profundo	Bacanin, N., et al.	Energies	2023	[34]
Marco ensemble para manejo de incertidumbre	Das, S. D., et al.	Neurocomputing	2022	[21]
Aplicación específica para español y fraude digital	Hurtado Avilés, G., et al.	BUAP	2024	[23]
Enfoque multi-hilo para detección multimodal	Yildirim, G.	Applied Intelligence	2023	[37]
Combinación K-Means y SVM para selección de características	Yazdi, K. M., et al.	WASET	2020	[48]
Optimización de ensemble con metaheurísticas	Yasmin, A., et al.	PLOS ONE	2024	[49]

Tabla 3.3: Contribuciones metodológicas en optimización metaheurística para detección.

La diversificación creciente hacia enfoques híbridos se observa claramente en estu-

dios metodológicamente innovadores como el de Yazdi et al. [48], que combina técnicas de clustering K-Means con máquinas de vectores de soporte (SVM) para optimizar la selección de características en espacios de alta dimensionalidad. Su contribución demuestra que la integración inteligente de metaheurísticas con técnicas de preprocesamiento puede mejorar significativamente la eficacia de detectores convencionales, especialmente en escenarios donde la dimensionalidad de los datos presenta desafíos computacionales significativos.

Finalmente, el trabajo de Yasmin et al. [49] avanza decididamente hacia la optimización, donde múltiples metaheurísticas colaboran de manera coordinada en la calibración simultánea de pesos, hiperparámetros, y arquitecturas de diferentes clasificadores base. Estos enfoques representan la frontera actual más avanzada en la aplicación de optimización metaheurística a problemas de detección complejos y establece precedentes importantes para desarrollos futuros en el campo (ver figura 3.2).



Figura 3.2: Mapa Conceptual 4: Métodos de optimización y metaheurísticas aplicadas.

3.3. Algoritmos Metaheurísticos: Fundamentos Teóricos y Aplicaciones Prácticas

Los algoritmos metaheurísticos representan una clase fundamental y versátil de técnicas de optimización que han demostrado ser especialmente efectivas para resolver problemas complejos de optimización no lineal, no convexos, y multimodales, incluyendo específicamente la calibración de hiperparámetros en modelos de aprendizaje automático y la optimización de arquitecturas de redes neuronales (ver tabla 4.7). Esta sección presenta los fundamentos teóricos, principios operativos, y aplicaciones prácticas de los principales algoritmos metaheurísticos utilizados en la detección de noticias falsas y optimización de modelos.

3.3.1. Algoritmos Genéticos (GA): Evolución Artificial para Optimización

Los Algoritmos Genéticos, introducidos por Holland [50] en su investigación original sobre adaptación en sistemas naturales y artificiales, están inspirados profundamente en la teoría de la evolución de Darwin y los principios fundamentales de la genética molecular. Estos algoritmos mantienen una población diversa de soluciones candidatas que evolucionan iterativamente a través de operadores genéticos cuidadosamente diseñados como selección, cruzamiento (crossover), y mutación, replicando los procesos evolutivos naturales pero aplicados a espacios de soluciones abstractos.

El trabajo fundamental de Holland estableció no solo las bases teóricas para esta clase de algoritmos bioinspirados, sino también demostraciones formales de su capacidad para explorar eficientemente espacios de búsqueda complejos y multimodales mediante un balance inteligente entre exploración (búsqueda global) y explotación (refinamiento local). La teoría de los esquemas (schema theory) desarrollada por Holland proporciona fundamentos matemáticos sólidos para entender cómo estos algoritmos pueden converger hacia regiones óptimas del espacio de búsqueda.

El funcionamiento básico de los algoritmos genéticos involucra una secuencia iterativa de operaciones: (1) evaluación de la aptitud (fitness) de cada individuo en la población usando una función objetivo específica del problema, (2) selección probabilística de los mejores candidatos para reproducción basada en su fitness relativo, (3) generación de nuevas soluciones mediante operadores de cruzamiento que combinan información genética de padres seleccionados, y (4) aplicación de mutación para introducir diversidad genética y evitar convergencia prematura. Este proceso iterativo permite la convergencia gradual hacia soluciones óptimas o cercanas al óptimo global.

En el contexto específico de la detección de noticias falsas, los algoritmos genéticos han demostrado particular efectividad en la optimización simultánea de múltiples hiperparámetros de modelos complejos, la selección automática de características relevantes desde espacios de alta dimensionalidad, y la evolución de arquitecturas de redes neuronales que maximizan el rendimiento de detección.

3.3.2. Optimización por Enjambre de Partículas (PSO): Inteligencia Colectiva

La Optimización por Enjambre de Partículas fue desarrollada independientemente por Kennedy y Eberhart [51], inspirándose en observaciones detalladas del comportamiento social emergente de bandadas de pájaros, bancos de peces, y otros sistemas biológicos que exhiben inteligencia colectiva. Una formulación alternativa pero complementaria fue presentada por Eberhart y Kennedy [52] en el mismo período, consolidando esta técnica como una de las metaheurísticas más utilizadas y estudiadas en optimización global.

En PSO, cada partícula individual representa una solución potencial completa que se mueve dinámicamente a través del espacio de búsqueda multidimensional siguiendo una combinación ponderada de su propia experiencia histórica (mejor posición personal encontrada, pbest) y la experiencia colectiva del enjambre (mejor posición global encontrada por cualquier partícula, gbest). La velocidad y posición de cada partícula se actualizan dinámicamente en cada iteración usando ecuaciones de movimiento que incorporan componentes de inercia, atracción cognitiva (hacia pbest), y atracción social (hacia gbest).

Esta formulación matemática permite un equilibrio natural entre exploración y explotación del espacio de soluciones: la inercia mantiene la dirección de búsqueda previa, la atracción cognitiva favorece la explotación de regiones prometedoras encontradas por cada partícula, y la atracción social facilita el intercambio de información entre partículas y la convergencia hacia regiones globalmente prometedoras.

En aplicaciones de detección de desinformación, PSO ha demostrado particular efectividad en la calibración de hiperparámetros de redes neuronales convolucionales y transformers, donde el espacio de búsqueda es continuo y las evaluaciones de aptitud son computacionalmente costosas.

3.3.3. Recocido Simulado: Física Estadística para Optimización

El Recocido Simulado, propuesto originalmente por Kirkpatrick, Gelatt y Vecchi [53], se basa en el proceso físico de enfriamiento controlado de metales (annealing) utilizado en metalurgia para obtener estructuras cristalinas de mínima energía. Este algoritmo permite escapar inteligentemente de óptimos locales mediante la aceptación probabilística de soluciones temporalmente peores, con una probabilidad que disminuye gradualmente según un esquema de enfriamiento cuidadosamente diseñado.

La analogía con la física estadística es profunda: la función objetivo corresponde a la energía del sistema, la variable de control (temperatura) determina la probabilidad de aceptar soluciones subóptimas, y el esquema de enfriamiento dicta cómo esta probabilidad decrece a lo largo del tiempo. La aceptación probabilística se basa en la distribución de Boltzmann, que establece que la probabilidad de aceptar una solución peor está dada por $P = e^{-\Delta E/T}$, donde ΔE es el incremento en la función objetivo

(energía) y T es la temperatura actual del sistema. Esta formulación permite que el algoritmo escape de mínimos locales cuando la temperatura es alta (probabilidades de aceptación mayores) y converja hacia soluciones de alta calidad cuando la temperatura es baja (probabilidades de aceptación menores para soluciones subóptimas).

Los métodos Multi-arranque, formalizados por Martí, Resende y Pardalos [54], consisten en ejecutar múltiples corridas independientes de un algoritmo de búsqueda local desde diferentes puntos de inicio cuidadosamente seleccionados. La combinación de Recocido Simulado con estrategias Multi-arranque (MSA) permite aprovechar sínergicamente las ventajas de ambos enfoques: la capacidad de escape de óptimos locales del primero y la diversificación exhaustiva del espacio de búsqueda del segundo.

3.3.4. Búsqueda Dispersa (Scatter Search): Combinación Determinística

La Búsqueda Dispersa, desarrollada por Glover [55] como parte de su marco más amplio de metaheurísticas basadas en memoria, utiliza estrategias determinísticas para combinar soluciones de referencia de alta calidad y generar nuevas soluciones prometedoras. A diferencia de otros métodos que dependen predominantemente de procesos aleatorios, SS emplea combinaciones determinísticas cuidadosamente diseñadas y diversificación controlada basada en medidas de distancia en el espacio de soluciones.

El algoritmo mantiene un conjunto pequeño pero diverso de soluciones de referencia de alta calidad, utilizando métodos de combinación específicos del problema para generar nuevas soluciones candidatas. La selección de soluciones de referencia se basa en criterios que balancean calidad (fitness) y diversidad (distancia en el espacio de soluciones), asegurando que el conjunto de referencia capture diferentes regiones prometedoras del espacio de búsqueda.

Esta aproximación ha demostrado ser particularmente efectiva en problemas de optimización combinatorial donde la estructura del problema permite el diseño de operadores de combinación inteligentes que preservan características deseables de las soluciones padre.

3.3.5. Búsqueda en Vecindades Variables (VNS): Exploración Sistématica

La Búsqueda en Vecindades Variables, introducida por Mladenović y Hansen [56], se basa en el principio fundamental de cambio sistemático de estructuras de vecindad durante el proceso de búsqueda local. Esta técnica es especialmente efectiva para escapar de óptimos locales explorando diferentes definiciones de vecindad que capturan aspectos diversos de la estructura del problema.

VNS alterna sistemáticamente entre fases de diversificación (mediante la exploración de vecindades más amplias que permiten movimientos grandes en el espacio de

soluciones) e intensificación (búsqueda local en vecindades más restringidas que refinan soluciones prometedoras), proporcionando un marco flexible y adaptable para la optimización que puede personalizarse a diferentes tipos de problemas y estructuras de espacio de búsqueda.

El principio central de VNS es que diferentes estructuras de vecindad proporcionan diferentes perspectivas del paisaje de optimización, y que el cambio sistemático entre estas perspectivas puede revelar regiones prometedoras que permanecerían ocultas usando una sola definición de vecindad.

Algoritmo	Inspiración	Características Principales	Aplicaciones en Detección	Ref.
Algoritmos Genéticos (GA)	Evolución biológica y genética molecular	Población de soluciones, operadores genéticos, selección natural, teoría de esquemas	Optimización de hiperparámetros, selección de características, evolución de arquitecturas de redes neuronales	[50]
Optimización por Enjambre de Partículas (PSO)	Comportamiento social de bandadas y enjambres	Partículas con velocidad y posición, memoria personal y social, inteligencia colectiva	Calibración de pesos en redes, optimización de parámetros de modelos de lenguaje, ajuste de transformers	[51]
Recocido Simulado (SA)	Proceso físico de enfriamiento de metales y mecánica estadística	Aceptación probabilística, esquema de enfriamiento, escape de óptimos locales, distribución de Boltzmann	Optimización de arquitecturas profundas, ajuste fino de modelos complejos, calibración no convexa	[53]
Multi-arranque (Multi-start)	Diversificación de puntos de inicio y exploración global	Múltiples ejecuciones independientes, exploración global del espacio, robustez estadística	Inicialización robusta de modelos, validación cruzada optimizada, ensemble de optimizadores	[54]
Búsqueda Dispersa (SS)	Combinación sistemática de soluciones diversas y métodos determinísticos	Conjunto de referencia, combinaciones determinísticas, diversificación controlada, memoria adaptativa	Ensemble de modelos, combinación de características, optimización de flujo de datos, fusión de arquitecturas	[55]
Búsqueda en Vecindades Variables (VNS)	Cambio sistemático de estructuras de vecindad y perspectivas múltiples	Múltiples definiciones de vecindad, alternancia entre diversificación e intensificación	Exploración de espacios de hiperparámetros, optimización de topologías de red, búsqueda adaptativa	[56]

Tabla 3.4: Algoritmos metaheurísticos: características fundamentales y aplicaciones en detección de noticias falsas.

Estos algoritmos han demostrado ser especialmente valiosos en el contexto complejo y multifacético de la detección de noticias falsas, donde los espacios de búsqueda de hiperparámetros son intrínsecamente complejos, multidimensionales, y frecuentemente multimodales. Su capacidad probada para balancear exploración y explotación los convierte en herramientas ideales para optimizar el rendimiento de modelos de aprendizaje automático en tareas de clasificación textual, especialmente cuando se enfrentan a conjuntos de datos desbalanceados, ruidosos, o de alta dimensionalidad que son característicos en aplicaciones de detección de desinformación.

3.4. Hiperparámetros en Modelos de Aprendizaje Automático: Fundamentos y Optimización

3.4.1. Naturaleza y Definición de Hiperparámetros

Para explicar este concepto fundamental de manera accesible pero técnicamente precisa, utilizaremos una analogía de cocina que captura las complejidades inherentes del proceso de entrenamiento de modelos de aprendizaje automático.

Imagina que estás desarrollando y entrenando un modelo de aprendizaje automático, que podemos conceptualizar como el proceso complejo de crear un pastel sofisticado con múltiples capas y componentes interconectados.

Los **parámetros** del modelo son los elementos que el algoritmo de aprendizaje automático aprende y ajusta automáticamente durante el proceso de entrenamiento iterativo. En nuestra analogía de cocina, estos corresponden a cómo se combinan y transforman los ingredientes dentro del horno durante la cocción: las reacciones químicas entre la harina y el azúcar, la formación de la estructura del gluten, y las transformaciones moleculares que ocurren naturalmente sin intervención directa del cocinero. En una red neuronal, estos son específicamente los pesos (weights) y sesgos (biases) de las neuronas individuales, los parámetros de las capas de convolución, y los elementos de las matrices de transformación. Estos valores se ajustan automáticamente a partir de los datos de entrenamiento mediante algoritmos de optimización como el descenso de gradiente estocástico.

Los **hiperparámetros**, en contraste fundamental, son las decisiones críticas y configuraciones que el científico de datos o ingeniero de aprendizaje automático, debe especificar y fijar antes de iniciar el proceso de entrenamiento. Son los ajustes o “perillas de control” de la receta algorítmica que determinan cómo el modelo aprenderá, pero que el propio modelo no puede ajustar automáticamente. En nuestra analogía de cocina:

- **La temperatura del horno** → tasa de aprendizaje (learning rate): controla qué tan grandes son los pasos de actualización de parámetros
- **El tiempo total de cocción** → número de épocas: cuántas veces el modelo verá todo el conjunto de datos de entrenamiento
- **La cantidad y proporción de cada ingrediente principal** → número de capas ocultas, número de neuronas por capa, dimensiones de embeddings
- **El tipo y forma del molde** → arquitectura fundamental del modelo (CNN, RNN, Transformer)
- **El tamaño de las porciones procesadas simultáneamente** → tamaño del batch
- **Técnicas especiales de horneado** → métodos de regularización (dropout, weight decay)

Si se eligen inadecuadamente estos ajustes críticos, el modelo puede experimentar varios problemas graves: sobreajuste (overfitting) donde memoriza los datos de entrenamiento pero no generaliza, subajuste (underfitting) donde no captura patrones importantes en los datos, convergencia lenta o inestable, o simplemente rendimiento subóptimo en la tarea objetivo.

3.4.2. Importancia Crítica en la Detección de Noticias Falsas

En el contexto de la detección de noticias falsas, la selección y optimización de hiperparámetros adquiere una importancia crítica debido a múltiples factores que hacen esta tarea particularmente compleja:

- **Especificidades lingüísticas y culturales:** Los textos en español poseen características lingüísticas específicas y complejas que requieren configuraciones de modelo particularmente adaptadas, incluyendo variaciones regionales en vocabulario, estructuras sintácticas diversas, y patrones discursivos culturalmente específicos
- **Sutileza de patrones desinformativos:** Los patrones de desinformación contemporánea pueden ser extremadamente sutiles y sofisticados, requiriendo modelos finamente calibrados que puedan distinguir entre información legítima y manipulación sutil sin generar demasiados falsos positivos. Un *falso positivo* ocurre cuando el sistema de detección clasifica incorrectamente contenido verdadero como desinformación, lo cual puede tener consecuencias graves como la censura de información legítima, pérdida de credibilidad del sistema, y restricción de la libertad de expresión.
- **Desbalance de clases:** Los conjuntos de datos de noticias falsas frecuentemente exhiben desbalance significativo entre clases, requiriendo técnicas especializadas de muestreo y ajustes específicos de hiperparámetros para manejar esta asimetría. En el contexto de clasificación binaria para detección de noticias falsas, las *clases* son las dos categorías posibles: "verdadero" o "falso". El desbalance ocurre cuando hay una cantidad significativamente mayor de ejemplos de una clase que de la otra (por ejemplo, 80 % noticias verdaderas vs. 20 % noticias falsas), lo que puede provocar que el modelo desarrolle sesgo hacia la clase mayoritaria y tenga dificultades para detectar correctamente la clase minoritaria.
- **Complejidad contextual:** Una configuración inadecuada puede hacer que el modelo confunda elementos legítimos como sarcasmo, ironía, o crítica constructiva con falsedad, o que no detecte técnicas de desinformación sofisticadas como medias verdades o manipulación de contexto
- **Evolución temporal:** Las técnicas de desinformación evolucionan constantemente, requiriendo modelos que puedan adaptarse y mantenerse efectivos ante nuevas estrategias de manipulación

3.4.3. Desafíos en la Búsqueda de Configuración Óptima

Tradicionalmente, los científicos de datos han abordado el problema de optimización de hiperparámetros mediante enfoques que frecuentemente resultan inadecuados para la complejidad del problema:

- **Búsqueda en grilla (Grid Search):** Explora sistemáticamente todas las combinaciones posibles de un conjunto predefinido de valores de hiperparámetros. Es como probar todas las recetas posibles en una cocina: aunque garantiza encontrar la mejor combinación dentro de las opciones consideradas, se vuelve impracticable cuando hay muchos ingredientes que ajustar. Además, no considera que algunos hiperparámetros pueden interactuar entre sí de formas complejas.
- **Búsqueda aleatoria (Random Search):** Muestrea aleatoriamente combinaciones de hiperparámetros desde distribuciones predefinidas. Es como elegir recetas al azar para probar: más eficiente que probar todas las combinaciones, pero sin ninguna estrategia inteligente para evitar repetir errores o aprovechar descubrimientos prometedores.
- **Optimización Bayesiana:** Utiliza un modelo probabilístico para predecir qué configuraciones podrían funcionar mejor y guiar la búsqueda hacia regiones prometedoras. Aunque más sofisticada, puede quedarse atascada en soluciones localmente buenas sin explorar otras posibilidades, y requiere recursos computacionales considerables.

Aquí es donde los algoritmos metaheurísticos emergen como “optimizadores inteligentes” que aprenden de cada evaluación de modelo para mejorar sistemáticamente las siguientes configuraciones de hiperparámetros, utilizando principios de exploración inteligente, explotación de regiones prometedoras, y escape de óptimos locales.

3.4.4. Herramientas Modernas: Keras Tuner y Automatización

En el ecosistema contemporáneo de TensorFlow y aprendizaje profundo, Keras Tuner [57] ha emergido como una herramienta fundamental y versátil para automatizar la búsqueda de configuraciones óptimas de hiperparámetros. Desarrollado por el equipo de Keras y respaldado por Google, esta biblioteca proporciona una interfaz unificada y elegante que permite al sistema explorar automáticamente diferentes “recetas” de configuración e identificar la configuración que maximiza el rendimiento para una tarea específica.

Keras Tuner facilita múltiples aspectos críticos del proceso de optimización:

- **Definición declarativa de espacios de búsqueda:** Permite especificar fácil e intuitivamente qué hiperparámetros optimizar y sus rangos válidos usando una sintaxis Python natural

- **Paralelización eficiente:** Soporta la evaluación simultánea de múltiples configuraciones en paralelo, aprovechando recursos computacionales distribuidos
- **Persistencia de resultados:** Guarda automáticamente configuraciones prometedoras y permite recuperar los mejores modelos para análisis posterior
- **Estrategias de búsqueda avanzadas:** Incluye implementaciones de Random Search, Hyperband, y Bayesian Optimization como opciones incorporadas

3.4.5. Beneficios combinados: Modelos de Lenguaje y Herramientas de Optimización

Los modelos de lenguaje modernos como DistilBERT, cuando se combinan inteligentemente con herramientas como Keras Tuner o algoritmos metaheurísticos (ver figura 3.3), crean sistemas de optimización extraordinariamente poderosos que operan como sistemas inteligentes adaptativos.

En contraste notable, la búsqueda de parámetros en los algoritmos metaheurísticos se realiza de manera más tradicional y controlada, donde cada algoritmo implementa su propia estrategia específica de exploración del espacio de hiperparámetros sin depender de herramientas de automatización externas, proporcionando mayor control sobre el proceso de optimización y permitiendo la incorporación de conocimiento específico del dominio (ver figura 3.3).



Figura 3.3: Mapa Conceptual 5: Clasificación de artículos por enfoque metodológico.

Estos beneficios combinados son especialmente poderosos para la detección de noticias falsas, donde la optimización automática e inteligente puede representar la diferencia crítica entre un modelo que apenas funciona de manera marginal y uno que

detecta desinformación con alta precisión y baja tasa de falsos positivos, aprovechando al máximo las capacidades semánticas avanzadas de los transformers pre-entrenados mientras mantiene eficiencia computacional práctica.

3.5. Análisis de Literatura Relevante

Más allá de la clasificación taxonómica superficial, resulta fundamental realizar un análisis en profundidad de las contribuciones clave de cada grupo temático identificado, examinando no solo qué se ha hecho, sino cómo se ha hecho, qué resultados se han obtenido, y cuáles son las implicaciones para la planeación y ejecución metodológica de esta tesis. Esta sección proporciona un análisis crítico de los aportes más significativos de cada categoría de investigación.

3.5.1. El Desafío Fundamental de los Datos: Creación y Curación de Corpus en Español

Uno de los desafíos más fundamentales y persistentes para el Procesamiento del Lenguaje Natural en español es la notable escasez de conjuntos de datos etiquetados de alta calidad a gran escala, especialmente en comparación con los abundantes recursos disponibles en inglés.

La creación y curación sistemática de corpus especializados constituye, por tanto, una línea de investigación fundamental que requiere no solo experiencia técnica sino también comprensión profunda de las particularidades lingüísticas y culturales del español. El trabajo de Fin de Maestría de Zules Acosta [11] fue uno de los pioneros más influyentes en este ámbito específico, centrándose meticulosamente en la creación de un conjunto de datos de 598 noticias en castellano. Su investigación trasciende más allá de la mera recolección de datos para enfocarse profundamente en la importancia crítica de la *ingeniería de características* para seleccionar estrategias óptimas de extracción que permitan a los modelos de aprendizaje automático clasificar eficazmente la veracidad del contenido. El trabajo incluye un análisis detallado de métricas de calidad, procedimientos de validación, y metodologías para manejar ambigüedad en el etiquetado de contenido dudoso.

Por otro lado, el trabajo de Posadas-Durán et al. [12] introdujo el influyente *Spanish Fake News Corpus* (con 971 noticias cuidadosamente seleccionadas), estableciéndose como un recurso fundamental para analizar y detectar información engañosa mediante métodos innovadores basados en el estilo de la escritura y características estilométricas avanzadas. Este corpus ha sido fundamental en competencias internacionales como IberLEF, donde se han evaluado sistemáticamente diversas metodologías para el español [46], incluyendo el análisis especializado de agresividad y noticias falsas en el español de México [45]. La contribución incluye anotaciones de características lingüísticas profundas como complejidad sintáctica, diversidad léxica, y estilos expresivos característicos.

Más recientemente, se han publicado corpus de escala considerablemente mayor que han resultado cruciales para el avance del campo. Tretiakov et al. [19] aportaron un conjunto de datos significativamente expandido con 1,958 noticias (falsas y verdaderas) en español, incorporando metadatos Enriquecidos sobre fuentes, fechas de publicación, y categorías temáticas. Paralelamente, el trabajo de gran escala de Blanco-Fernández et al. [13] introdujo el ambicioso *Spanish Political Fake News Dataset*. Este último recurso, con más de 57,000 noticias meticulosamente recolectadas y etiquetadas, representa uno de los mayores y más comprehensivos recursos disponibles para investigación en español y constituye la fuente principal de datos para esta tesis. La unificación metodológica de estos cuatro corpus diversos constituye el fundamento empírico sólido de este trabajo de investigación.

3.5.2. Revolución de los Modelos de Lenguaje y Arquitecturas Transformer

La arquitectura Transformer, introducida revolucionariamente por Vaswani et al. [24], transformó completamente el panorama del PLN con su mecanismo de atención multi-cabeza innovador que permite el procesamiento paralelo eficiente y la captura de dependencias a largo plazo. Este trabajo fundamental estableció las bases teóricas y prácticas para una nueva generación de modelos como BERT [25], que introdujo el concepto paradigmático de pre-entrenamiento bidireccional, permitiendo a los modelos capturar contexto tanto precedente como subsecuente simultáneamente.

Las variantes optimizadas como DistilBERT [20] y TinyBERT [26] han demostrado que es posible mantener capacidades semánticas sofisticadas mientras se reduce dramáticamente el costo computacional, haciendo viable la implementación de sistemas de detección en dispositivos con recursos limitados y aplicaciones en tiempo real. Estas innovaciones en eficiencia son particularmente relevantes para implementaciones prácticas de sistemas de detección que deben operar bajo restricciones de latencia y recursos.

El paradigma de los Grandes Modelos de Lenguaje (LLMs) se consolidó definitivamente con GPT-3 [17], que demostró capacidades emergentes extraordinarias de few-shot learning (aprendizaje con pocos ejemplos) y razonamiento en contexto que revolucionaron las expectativas sobre lo que los modelos de lenguaje pueden lograr. Trabajos más recientes como LLaMA [29] han democratizado significativamente el acceso a modelos de gran escala mediante arquitecturas más eficientes y políticas de licenciamiento abierto, mientras que Gemini [30] ha avanzado hacia capacidades multimodales que integran texto, imágenes, y otros tipos de información. La investigación en IA Constitucional [31] ha abordado proactivamente los desafíos críticos de alineación y seguridad de estos sistemas poderosos.

Para el español específicamente, Martínez-Gallego et al. [27] realizaron una exploración de la aplicación de BERT y BETO, demostrando las ventajas de modelos especializados lingüísticamente. Paralelamente, Blanco-Fernández et al. [13] llevaron a cabo comparaciones detalladas entre BERT y RoBERTa para la detección de desinfor-

mación, proporcionando evidencia empírica sobre las ventajas específicas de diferentes arquitecturas transformer para este dominio.

La investigación sobre el rol dual de los LLMs [2, 16, 6] ha revelado tanto oportunidades extraordinarias como desafíos significativos en su aplicación para la detección de noticias falsas, incluyendo cuestiones de sesgo, robustez, y generalización que requieren consideración cuidadosa en implementaciones prácticas.

3.5.3. Optimización y Metaheurísticas en la Detección: Enfoques Innovadores

Los algoritmos metaheurísticos han demostrado ser herramientas excepcionalmente valiosas para abordar problemas complejos de optimización en la detección de noticias falsas, proporcionando soluciones elegantes a desafíos que tradicionalmente han sido abordados mediante métodos menos sofisticados. Aqil y Lahby [33] desarrollaron una formulación innovadora que modela la detección como un problema de scheduling (o de programación de tareas) complejo, aplicando algoritmos genéticos, optimización por enjambre de partículas, y otras metaheurísticas para optimizar el flujo de datos completo de procesamiento.

La calibración de hiperparámetros [34, 23] ha emergido como una aplicación crítica donde las metaheurísticas superan consistentemente a los métodos tradicionales de grid search y random search, especialmente en espacios de alta dimensionalidad donde la búsqueda exhaustiva es computacionalmente prohibitiva. Estos enfoques han demostrado capacidad superior para navegar paisajes de optimización complejos y multimodales característicos del ajuste de modelos profundos.

El enfoque innovador de ensemble con marcos heurísticos [21] ha mostrado resultados especialmente prometedores al combinar múltiples modelos especializados con información estadística adicional, creando sistemas que pueden manejar incertidumbre de manera más sofisticada que enfoques de clasificación tradicionales. Yildirim [37] propuso un enfoque híbrido multi-hilo particularmente avanzado que optimiza simultáneamente tanto la selección de características como los parámetros del modelo, demostrando el potencial de optimización coordinada en múltiples dimensiones del problema.

La aplicación de PSO para la detección de reseñas falsas [38] demuestra convincentemente la versatilidad y transferibilidad de estas técnicas más allá del dominio específico de noticias, sugiriendo principios generalizables para la detección de contenido desinformativo en múltiples contextos.

3.5.4. Aplicación de Metaheurísticas por Tipo de Detección

La literatura revisada revela patrones específicos en la aplicación de algoritmos metaheurísticos según el tipo de detección, el dominio del problema, y las características específicas de los datos involucrados (ver tabla 3.5). Esta taxonomía detallada

proporciona resultados valiosos para la selección informada de técnicas de optimización para diferentes contextos de aplicación.

Tipo de Detección	Metaheurística Empleada	Aplicación Específica	Autores	Ref.
Detección de Noticias Falsas	Algoritmo Genético (GA), Colonia de Abejas (ABC), Búsqueda Iterativa (IG)	Planificación de tareas para procesamiento de documentos	Aqil, S. & Lahby, M.	[33]
Detección de Reseñas Falsas	Optimización por Enjambre de Partículas Adaptativo (PSO)	Optimización de hiperparámetros en redes CNN	Deshai, N. & Rao, B. B.	[38]
Fraude Financiero y de Estados Financieros	Algoritmos Genéticos, Optimización por Enjambre de Partículas, Recocido Simulado	Selección de características y optimización de clasificadores SVM	Hidayattullah, S. et al.	[35]
Predicción de Dificultades Financieras	Optimización Bayesiana con Procesos Gaussianos	Calibración de hiperparámetros en modelos GPR	Horak, J. & Sabek, A.	[36]
Detección de Fraude Laboral	Redes Neuronales Artificiales (ANN)	Clasificación de ofertas de empleo fraudulentas	Nasser, I. M. et al.	[9]
Optimización de Modelos de Energía	Algoritmo Genético Binario, PSO, Recocido Simulado, Búsqueda Armónica	Ajuste de hiperparámetros en modelos LSTM para predicción energética	Bacanin, N. et al.	[34]
Detección Multimodal de Noticias Falsas	Enfoque Híbrido Multi-hilo con Metaheurísticas	Optimización paralela de características textuales y visuales	Yildirim, G.	[37]
Selección de Características para Fake News	K-Means combinado con Support Vector Machine (SVM)	Reducción de dimensionalidad y mejora de precisión	Yazdi, K. M. et al.	[48]
Framework de Incertidumbre para Fake News	Enfoque Heurístico basado en Ensemble	Manejo de incertidumbre en clasificación de tweets y artículos	Das, S. D. et al.	[21]
Optimización de Dominación Total en Redes Sociales	Búsqueda en Vecindades Variables (VNS)	Propagación de información en redes sociales	Kapunac, S. et al.	[58]
Estimación de Esfuerzo en Proyectos	Ensemble con Metaheurísticas para Pesos	Optimización de hiperparámetros y pesos de ensemble	Yasmin, A. et al.	[49]

Tabla 3.5: Aplicación de metaheurísticas por tipo de detección en la literatura revisada.

Patrones Identificados por Dominio

Del análisis sistemático de la literatura se identifican tres patrones principales y claramente diferenciados en la aplicación de metaheurísticas:

Detección de Contenido Textual Para la detección de noticias falsas y contenido textual fraudulento, predominan los enfoques que combinan estratégicamente:

- **Algoritmos Genéticos:** Utilizados principalmente para selección de características en espacios de alta dimensionalidad y optimización de arquitecturas de modelos complejos [33, 35]

- **PSO Adaptativo:** Especialmente efectivo para la calibración de hiperparámetros en redes neuronales complejas donde el espacio de búsqueda es continuo y las evaluaciones son costosas [38, 34]
- **Enfoques Híbridos:** Combinación de múltiples metaheurísticas para diferentes aspectos del flujo de datos de detección, optimizando simultáneamente múltiples objetivos [37]

Optimización de Modelos de Aprendizaje Profundo En aplicaciones que involucran modelos de aprendizaje profundo y arquitecturas complejas, se observa una preferencia marcada por:

- **Recocido Simulado:** Particularmente efectivo para escapar de óptimos locales en espacios de hiperparámetros complejos y multimodales [34]
- **Optimización Bayesiana:** Para calibración eficiente en modelos probabilísticos donde se requiere balance entre exploración y explotación [36]
- **Búsqueda en Vecindades Variables:** Para exploración sistemática de configuraciones de red mediante cambio de estructuras de vecindad [58]

Sistemas de Detección en Tiempo Real Para aplicaciones que requieren procesamiento en tiempo real o alta velocidad de procesamiento, las metaheurísticas se enfocan estratégicamente en:

- **Algoritmos de Planificación:** Como IG (Iterated Greedy) para optimización de tareas de procesamiento con restricciones temporales [33]
- **Métodos Ensemble:** Con optimización de pesos mediante metaheurísticas para combinar múltiples detectores especializados [21, 49]
- **Enfoques Multi-hilo:** Para paralelización eficiente de la optimización en sistemas multimodales que procesan múltiples tipos de información simultáneamente [37]

Esta taxonomía detallada revela que la elección de la metaheurística está fuertemente influenciada por factores como el tipo de modelo subyacente, las características del conjunto de datos, los requisitos de rendimiento temporal del sistema de detección, y las restricciones computacionales específicas del entorno de implementación.

3.5.5. Perspectivas Interdisciplinarias y Análisis Social

La comprensión profunda del fenómeno de las noticias falsas requiere necesariamente un enfoque interdisciplinario sofisticado que combine aspectos técnicos avanzados con análisis social, psicológico, y cultural. Los aspectos puramente técnicos,

aunque fundamentales, son insuficientes para abordar completamente la complejidad multifacética de la desinformación moderna.

Ali et al. [5, 3] han investigado sistemáticamente cómo las heurísticas cognitivas humanas, incluyendo señales de popularidad social como el número de “me gusta” y compartidos, influyen significativamente en la percepción de credibilidad de contenido digital. Su trabajo sobre el procesamiento heurístico durante eventos políticos específicos, particularmente las elecciones presidenciales de 2016 en Estados Unidos, proporciona resultados valiosos y empíricamente fundamentados sobre la psicología de la desinformación y los mecanismos cognitivos que hacen a las personas susceptibles a información falsa.

El análisis detallado del rol de los medios de comunicación tradicionales [7, 4] ha revelado patrones culturales importantes en cómo diferentes países y culturas abordan, conceptualizan, y responden al problema de la desinformación. Estos estudios proporcionan evidencia empírica sobre la importancia de considerar factores culturales en el diseño de sistemas de detección automatizada.

Pulido et al. [10] propusieron el marco innovador SISM (Social Impact in Social Media) para combatir específicamente la desinformación en el dominio de la salud, un área particularmente crítica durante pandemias y crisis de salud pública donde la desinformación puede tener consecuencias directas sobre la vida y muerte de las personas.

3.5.6. Detección de Fraude: Extensión Más Allá de las Noticias

La investigación en detección de fraude ha diversificado significativamente sus aplicaciones, extendiéndose más allá del dominio específico de noticias falsas hacia otros contextos de fraude digital que presentan características similares pero requieren adaptaciones específicas (ver figura 3.4).

En el ámbito laboral específicamente, Nasser et al. [9] desarrollaron sistemas basados en redes neuronales artificiales para detectar ofertas de trabajo fraudulentas, demostrando que las técnicas desarrolladas para detección de noticias falsas pueden transferirse efectivamente a otros dominios textuales. Paralelamente, Alvarez [59] analizó las nuevas modalidades de fraude laboral en la era digital latinoamericana, proporcionando contexto regional importante para entender las manifestaciones específicas del fraude en diferentes contextos socioeconómicos.

En el sector financiero, la investigación de Hidayattullah et al. [35] demostró empíricamente la efectividad de combinar aprendizaje automático con optimización metaherística para detectar fraudes en estados financieros, estableciendo precedentes metodológicos importantes. Cao et al. [8] establecieron conexiones innovadoras entre indicadores de empleo corporativo y riesgo de fraude, proporcionando herramientas valiosas para auditores y reguladores financieros.



Figura 3.4: Mapa Conceptual 6: Artículos que son revisiones o están relacionados al análisis de contenido y detección de fraude digital.

3.6. Síntesis y Perspectivas Futuras

La revisión detallada de la literatura revela una evolución acelerada en las aproximaciones para la detección de noticias falsas y fraude digital. Los desafíos principales identificados a través de esta revisión incluyen: (1) la necesidad imperativa de mayor cantidad de datos etiquetados de alta calidad en español, (2) la adaptación cultural y lingüística de modelos a contextos específicos del mundo hispanohablante, (3) la optimización eficiente de hiperparámetros en modelos cada vez más complejos, y (4) la integración inteligente de información multimodal y conocimiento externo estructurado.

La organización temática presentada en este capítulo proporciona un marco conceptual robusto para entender las diferentes dimensiones del problema de detección de desinformación y justifica empíricamente la aproximación metodológica híbrida adoptada en esta investigación, que combina modelos de lenguaje con algoritmos metaheurísticos para lograr rendimiento superior en la detección de contenido desinformativo en español.

Capítulo 4

Metodología

En este capítulo se presenta una metodología unificada e integrada para abordar el problema de la detección de noticias falsas en español. La estrategia metodológica se fundamenta en un proceso evolutivo de desarrollo y evaluación que avanza desde técnicas clásicas hasta modelos de lenguaje de última generación, permitiendo una comparación sistemática y objetiva de diferentes paradigmas de inteligencia artificial sobre una base de datos común.

El diseño metodológico adopta un enfoque experimental escalonado que incluye: (1) adquisición y procesamiento unificado de datos, (2) implementación de un Flujo de datos de procesamiento común, (3) desarrollo paralelo de modelos de clasificación utilizando algoritmos metaheurísticos y modelos Transformer, y (4) evaluación comparativa integral bajo un marco de métricas unificado. Este proceso culmina con la implementación de la solución más efectiva en una aplicación web funcional.

4.1. Visión General del Proceso Metodológico

La metodología se estructura como un flujo de datos experimental unificado que evalúa sistemáticamente dos paradigmas de inteligencia artificial complementarios. Como se ilustra en la Figura 4.1, ambos enfoques parten de una base metodológica común pero implementan estrategias de modelado diferenciadas, convergiendo finalmente en una evaluación comparativa que determina la solución óptima.

Fundamento del enfoque evolutivo: Esta investigación adopta una filosofía de desarrollo evolutivo donde se exploran técnicas de complejidad creciente, comenzando con métodos clásicos de PLN optimizados mediante metaheurísticas (estableciendo una línea base sólida) y progresando hacia modelos Transformer de última generación. Esta progresión permite:

- **Validación incremental:** Cada fase valida y mejora los resultados de la anterior
- **Comparación objetiva:** Todos los modelos se evalúan bajo las mismas condiciones experimentales

- **Justificación de complejidad:** Se demuestra empíricamente si la complejidad adicional se traduce en mejoras de rendimiento
- **Transferibilidad metodológica:** Los procesos desarrollados son aplicables a otros tipos de fraude digital

Una vez completada la evaluación comparativa, el modelo de mejor rendimiento se implementa en una aplicación web que demuestra la viabilidad práctica de la solución desarrollada.

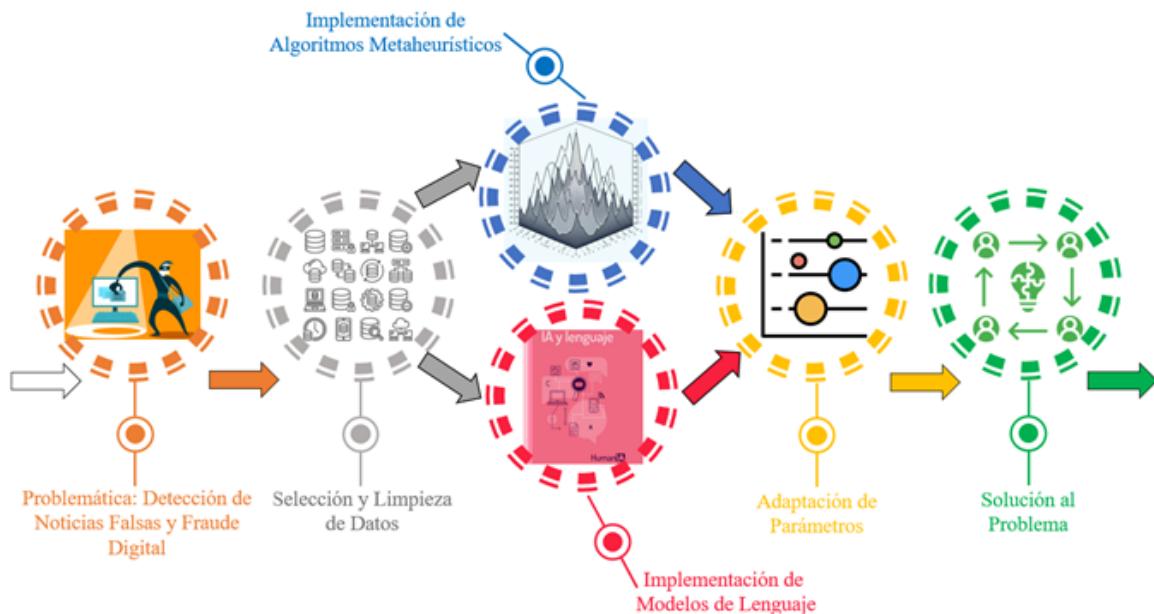


Figura 4.1: Metodología propuesta que aborda la problemática combinando Algoritmos Metaheurísticos y Modelos de Lenguaje.

4.2. Definición y Distinción de Conceptos Fundamentales

Antes de proceder con la descripción metodológica, es fundamental establecer con precisión la terminología utilizada en esta investigación, particularmente la distinción entre conceptos relacionados pero diferenciados que a menudo se utilizan de manera intercambiable en la literatura.

4.2.1. Distinción entre Noticia Falsa y Bulo

En el contexto de esta investigación, es esencial distinguir entre dos conceptos fundamentales que, aunque relacionados, poseen características distintivas importantes para el desarrollo de sistemas de detección automática.

4.2.2. Taxonomía de la Desinformación

Para proporcionar un marco conceptual completo, esta investigación adopta la taxonomía establecida por la literatura internacional [1, 18], que clasifica el contenido engañoso en tres categorías principales:

1. **Desinformación:** Información falsa creada y compartida deliberadamente con intención maliciosa
2. **Desinformación (Misinformation):** Información incorrecta compartida sin intención maliciosa
3. **Información Maliciosa (Malinformation):** Información genuina compartida con intención de causar daño

El alcance de esta investigación se centra específicamente en la detección automática de **desinformación**, abarcando tanto noticias falsas como bulos, reconociendo que ambos tipos requieren estrategias de procesamiento de lenguaje natural diferenciadas pero complementarias.

4.2.3. Justificación de la Unificación Terminológica

En el contexto del desarrollo de sistemas automáticos de detección, la distinción entre noticia falsa y bulo, aunque conceptualmente importante, se aborda mediante un enfoque unificado de **detección de contenido desinformativo**. Esta decisión metodológica se fundamenta en:

- **Características textuales compartidas:** Ambos tipos utilizan patrones lingüísticos identificables mediante técnicas de PLN
- **Objetivo común:** La finalidad de engañar o desinformar a la audiencia
- **Impacto social similar:** Efectos negativos en la percepción pública y toma de decisiones
- **Necesidad práctica:** Los sistemas de detección automática deben ser capaces de identificar ambos tipos

Por tanto, en el resto de este documento, el término “**noticias falsas**” se utilizará de manera inclusiva para referirse a cualquier tipo de contenido desinformativo, reconociendo la diversidad de formatos y estrategias de propagación que abarca esta categorización.

4.2.4. Esquema de Clasificación: Enfoque Binario

Esta investigación adopta un esquema de clasificación binaria, donde cada noticia se etiqueta en una de dos clases mutuamente excluyentes:

- **Clase 0 (FALSO):** Incluye todo contenido desinformativo, abarcando tanto noticias falsas formales como bulos informales, contenido satírico deliberadamente engañoso, y cualquier tipo de información con intención de desinformar
- **Clase 1 (REAL):** Incluye contenido periodístico legítimo, noticias verificadas de fuentes confiables, y información factualmente correcta

Justificación del Enfoque Binario

La decisión de utilizar clasificación binaria, en lugar de esquemas multiclasificación más granulares, se fundamenta en:

1. **Practicidad para usuarios finales:** En aplicaciones reales, los usuarios necesitan una respuesta clara y directa sobre la confiabilidad del contenido
2. **Consistencia con la literatura:** La mayoría de trabajos en detección de noticias falsas utilizan esquemas binarios, facilitando la comparación de resultados
3. **Robustez del modelo:** Los esquemas binarios tienden a ser más robustos y menos propensos al sobreajuste que las clasificaciones multiclasificación complejas
4. **Disponibilidad de datos:** Los corpus disponibles están mayoritariamente etiquetados de forma binaria

Consideraciones sobre Esquemas Alternativos

Aunque este trabajo adopta clasificación binaria, es importante reconocer que existen esquemas alternativos más granulares:

- **Clasificación ternaria:** FALSO, REAL, PARCIALMENTE _ CIERTO
- **Clasificación por niveles de confianza:** Escalas de 5 o 7 puntos de credibilidad
- **Clasificación por tipo de desinformación:** Distinguendo entre sátira, propaganda, clickbait, etc.

Sin embargo, estos enfoques requieren corpus más especializados y presentan desafíos adicionales en términos de consistencia de etiquetado y acuerdo entre anotadores.

4.2.5. Fronteras Difusas en la Clasificación de Veracidad

Es importante reconocer que la clasificación binaria de contenido informativo presenta inherentemente **zonas grises y fronteras difusas**, donde la distinción entre “verdadero” y “falso” no es absoluta. Esta complejidad conceptual representa uno de los desafíos fundamentales en la detección automatizada de desinformación.

Casos Límite y Ambigüedades

En la práctica, existen varios tipos de contenido que desafían la clasificación binaria estricta:

- **Información parcialmente correcta:** Noticias que contienen algunos hechos verificables mezclados con información falsa o tergiversada
- **Contenido satírico ambiguo:** Material humorístico que puede ser malinterpretado como información factual
- **Opiniones presentadas como hechos:** Contenido subjetivo que se presenta con apariencia de objetividad periodística
- **Información desactualizada:** Noticias que fueron ciertas en el pasado pero ya no son válidas
- **Interpretaciones sesgadas:** Presentación de hechos reales con marcos interpretativos tendenciosos
- **Información especulativa:** Predicciones o análisis presentados como hechos consumados

Implicaciones para el Modelado

Esta ambigüedad inherente se aborda en esta investigación mediante:

1. **Criterios de etiquetado conservadores:** En casos ambiguos, se prioriza la asignación a la clase “REAL” para evitar censura inadecuada de contenido legítimo
2. **Umbral de confianza:** El modelo genera probabilidades que pueden interpretarse como niveles de certeza, permitiendo identificar casos limítrofes
3. **Validación humana en casos dudosos:** Los corpus utilizados fueron validados manualmente para resolver ambigüedades
4. **Reconocimiento de limitaciones:** Se asume que un porcentaje mínimo de casos será inherentemente ambiguo y difícil de clasificar con certeza absoluta

Justificación Pragmática del Enfoque Binario

A pesar de estas complejidades, el enfoque binario se justifica porque:

- **Utilidad práctica:** Los usuarios finales requieren decisiones claras sobre la credibilidad del contenido
- **Viabilidad técnica:** Los modelos binarios son más robustos y fáciles de entrenar que los esquemas multiclasificación complejos
- **Consistencia metodológica:** Permite comparación directa con la literatura existente
- **Escalabilidad:** Facilita la anotación y validación de grandes volúmenes de datos

Es fundamental entender que **la frontera difusa no es una limitación del modelo, sino una característica inherente del problema mismo**, que debe ser gestionada mediante diseño metodológico cuidadoso y reconocimiento explícito de las limitaciones del enfoque adoptado.

4.3. Construcción del Corpus Unificado

El primer y más fundamental paso de la metodología fue la construcción de un corpus de alta calidad y de tamaño significativo en español, dada la escasez documentada de recursos centralizados para esta tarea [18].

4.3.1. Fuentes de Datos Académicas

Se llevó a cabo un proceso exhaustivo de investigación y unificación de cuatro corpus académicos reconocidos, los cuales constituyen pilares en la investigación de noticias falsas en español. **Es importante señalar que, aunque todos los corpus convergen hacia el objetivo común de detección de desinformación, cada uno fue desarrollado con enfoques metodológicos y propósitos específicos diferentes**, lo que enriquece la diversidad y robustez del conjunto de datos unificado. La Tabla 4.1 presenta un resumen detallado de estos recursos y sus características distintivas.

Spanish Fake News Corpus (IberLEF)

Este corpus, asociado a las competencias de IberLEF y desarrollado por Posadas-Durán, Gómez-Adorno, et al., ha tenido varias versiones. La versión original y más citada del corpus contiene un total de 971 noticias [12]. **Su propósito específico fue el desarrollo de técnicas de análisis estilométrico para identificar patrones lingüísticos distintivos en texto desinformativo.** Este corpus se caracteriza

ID	Nombre del Corpus	Autores Principales	Año	Noticias	Características	Ref.
1	Spanish Fake News Corpus (IberLEF)	Posadas-Durán, J.P. Gómez-Adorno, H.	2019-2021	971	Análisis estilométrico Múltiples versiones	[12]
2	Conjunto de datos Zules Acosta (UPM)	Acosta, F.A.Z.	2019	598	Trabajo fin de maestría Extracción web verificada	[11]
3	Conjunto de datos Tretiakov (Kaggle)	Tretiakov, A. Martín García, A.	2022	1,958	Aprendizaje automático Disponible públicamente	[19]
4	Spanish Political Fake News Conjunto de datos	Blanco-Fernández, Y. Otero-Vizoso, J.	2024	57,231	Temática política Modelos BERT/RoBERTa	[13]
TOTAL CORPUS ACADÉMICOS				60,758	Cuatro fuentes	

Tabla 4.1: Corpus académicos utilizados para la construcción del conjunto de datos unificado.

por su enfoque metodológico riguroso en análisis estilométrico y ha sido utilizado en múltiples evaluaciones comparativas dentro del marco de las competencias IberLEF [46, 45]. **Las fuentes de este corpus provienen principalmente de sitios web verificados como desinformativos por organismos de verificación de datos españoles y latinoamericanos.** Las versiones posteriores del corpus han sido mantenidas y actualizadas en repositorios públicos [47].

Conjunto de datos de Zules Acosta (Universidad Politécnica de Madrid)

Desarrollado como parte de un Trabajo Fin de Maestría Universitaria en Ciberseguridad en la Universidad Politécnica de Madrid, este corpus contiene 598 noticias verificadas [11]. **Su propósito original fue establecer una metodología de construcción de corpus para noticias falsas en español, con énfasis en la verificación manual exhaustiva.** El conjunto de datos fue construido mediante técnicas de extracción web y se encuentra disponible públicamente en la plataforma Kaggle [60, 61]. **Las fuentes incluyen tanto medios legítimos (para noticias reales) como sitios identificados como propagadores de desinformación, con verificación cruzada mediante múltiples verificadores de hechos.** Su contribución principal radica en la aplicación de metodologías rigurosas de verificación para la construcción de conjuntos de datos de noticias falsas.

Dataset de Tretiakov et al.

Este corpus, desarrollado por Tretiakov, Martín García y Camacho, contiene 1,958 noticias en español (Corpus balanceado, 50 % por ciento de noticias falsas y 50 % de noticias verdaderas) [19]. **Su propósito específico fue la evaluación de técnicas de aprendizaje automático tradicionales para clasificación binaria de contenido desinformativo.** El conjunto de datos fue creado en el contexto de técnicas de aprendizaje automático para la detección de información falsa y se encuentra disponible públicamente en Kaggle [62]. **Las fuentes de este corpus se concentran en redes sociales y plataformas digitales donde se propaga desinformación, incluyendo Twitter, Facebook y sitios web no verificados.** Su enfoque se centra en la aplicación de técnicas de aprendizaje automático para la clasificación automatizada de contenido desinformativo.

Spanish Political Fake News Dataset

El corpus más extenso utilizado, desarrollado por Blanco-Fernández et al., contiene 57,231 noticias de temática política [13]. Su propósito específico fue evaluar modelos transformer (BERT/RoBERTa) en el dominio político, donde la desinformación tiene particular relevancia social. Este conjunto de datos fue específicamente diseñado para evaluar modelos BERT y RoBERTa en la detección de desinformación política en español y se encuentra disponible en Kaggle [63]. Las fuentes incluyen tanto medios de comunicación tradicionales como plataformas digitales especializadas en contenido político, con particular énfasis en el contexto español e iberoamericano. Su gran tamaño y enfoque en contenido político lo convierte en un recurso valioso para el entrenamiento de modelos de gran escala.

4.3.2. Análisis Comparativo Detallado de los Corpus Utilizados

Para garantizar una comprensión completa de las características y diferencias entre los corpus utilizados, se presenta un análisis comparativo exhaustivo que permite entender las fortalezas y limitaciones de cada fuente de datos, así como las decisiones metodológicas tomadas en el proceso de unificación.

Aspecto Comparativo	Posadas-Durán	Acosta (UPM)	Tretiakov	Blanco-Fernández	El Deforma
Tamaño del corpus	971 noticias	598 noticias	1,958 noticias	57,231 noticias	9,000 noticias
Año de creación	2019-2021	2019	2022	2024	2025 (extracción)
Enfoque metodológico	Análisis estilométrico Competencias IberLEF	Metodología de construcción de corpus	Aprendizaje automático tradicional	Modelos Transformer BERT/RoBERTa	Contenido satírico Web scraping
Dominio temático	General	General	General	Político específico	Satírico/General
Distribución de clases	Balanciado	Balanciado	Solo falsas	Balanciado	Solo falsas
Fuentes de noticias reales	Medios españoles verificados	Múltiples medios internacionales	N/A	Medios políticos españoles	N/A
Fuentes de noticias falsas	Sitios verificados como desinformativos	Fact-checkers y verificadores	Redes sociales y sitios no verificados	Desinformación política verificada	Portal satírico El Deforma
Variabilidad regional	España/Latinoamérica	Internacional	Múltiple	España	Méjico
Período temporal	2018-2021	2018-2019	2020-2022	2020-2024	2019-2025
Calidad de anotación	Alta - múltiples anotadores	Alta - verificación manual exhaustiva	Media - basada en fuentes	Alta - verificación política especializada	Automática - contenido inherentemente falso
Metadatos incluidos	Características estilométricas	Fuente, fecha, categoría	Fecha, fuente, contexto	Orientación política, fecha, fuente	URL, fecha de extracción
Disponibilidad pública	GitHub (IberLEF)	Kaggle UPM	Kaggle Público	Kaggle Applied Sciences	Generado para esta tesis
Fortaleza principal	Ánálisis lingüístico profundo	Metodología rigurosa	Foco en ML tradicional	Gran escala y especialización política	Contemporaneidad y diversidad
Limitación principal	Tamaño limitado	Tamaño muy pequeño	Solo noticias falsas	Dominio muy específico	Solo contenido satírico
Contribución al corpus final	1.6 %	1.0 %	3.2 %	92.8 %	14.6 %

Tabla 4.2: Comparación exhaustiva de características de los corpus utilizados en la construcción del conjunto de datos unificado.

Características de Variabilidad Regional del Español

Una consideración metodológica fundamental en esta investigación es el reconocimiento y aprovechamiento de la diversidad regional del español presente en los corpus utilizados. **El corpus unificado representa un enfoque de español neutro multiregional**, lo que constituye una fortaleza metodológica significativa:

Distribución geográfica y justificación:

- **Español peninsular (España):** Representado principalmente en los corpus de Posadas-Durán [12] y Blanco-Fernández [13], que utilizan fuentes de medios españoles verificados
- **Español mexicano:** Incorporado a través del contenido de “El Deforma” y fuentes mexicanas presentes en el corpus de Acosta [11]
- **Español latinoamericano diverso:** El corpus de Acosta incluye fuentes internacionales que abarcan múltiples países de habla hispana
- **Español digital contemporáneo:** Los corpus más recientes (2020-2025) capturan la evolución del español en medios digitales

Ventajas metodológicas del enfoque multiregional:

- **Robustez del modelo:** Capacidad de detectar desinformación independientemente de la variante regional
- **Generalización lingüística:** Menor sesgo hacia construcciones específicas de una región
- **Aplicabilidad universal:** El modelo resultante es funcional para toda la comunidad hispanohablante
- **Representatividad real:** Refleja el consumo actual de noticias en español, que frecuentemente cruza fronteras nacionales

Esta diversidad regional no constituye una limitación, sino una fortaleza que permite desarrollar modelos más robustos y ampliamente aplicables.

1. Diversidad Metodológica:

- **Corpus de Posadas-Durán:** Aporta rigor en análisis estilométrico y experiencia en competencias académicas
- **Corpus de Acosta:** Contribuye con metodología de construcción verificada manualmente
- **Corpus de Tretiakov:** Añade perspectiva de ML tradicional y fuentes de redes sociales
- **Corpus de Blanco-Fernández:** Proporciona escala masiva y especialización en dominio político
- **Contenido de El Deforma:** Introduce diversidad estilística y contemporaneidad

2. Cobertura Temporal y Geográfica:

- **Span temporal:** 2018-2025 (7 años de evolución en patrones de desinformación)
- **Cobertura geográfica:** España, México, Latinoamérica (diversidad regional del español)
- **Evolución temática:** Desde temas generales hasta especialización política y satírica

3. Complementariedad en Tipos de Desinformación:

- **Desinformación tradicional:** Corpus académicos con verificación formal
- **Desinformación política:** Especialización del corpus de Blanco-Fernández
- **Contenido satírico:** Ampliación con El Deiforma para diversidad estilística
- **Múltiples fuentes:** Desde medios tradicionales hasta redes sociales

4. Validación de Robustez: La diversidad en metodologías de construcción permite validar la robustez de los modelos desarrollados:

- Modelos entrenados deben generalizar entre diferentes tipos de anotación
- Capacidad de detectar patrones consistentes a pesar de diferencias metodológicas
- Evaluación de transferencia entre dominios (general → político → satírico)

Justificación de la Estrategia de Unificación

La decisión de unificar corpus heterogéneos se basa en varios principios metodológicos sólidos:

1. Principio de Máxima Diversidad:

- Mayor variabilidad en patrones lingüísticos
- Reducción de sesgos específicos de corpus individuales
- Mejora en generalización de modelos entrenados

2. Principio de Escala Efectiva:

- Aprovechamiento del corpus grande (Blanco-Fernández) como base
- Enriquecimiento con corpus especializados más pequeños
- Balanceo mediante extracción web controlada

3. Principio de Validación Cruzada:

- Cada corpus actúa como validación independiente de los otros
- Patrones consistentes entre corpus indican robustez
- Diferencias señalan áreas que requieren atención especial

4.3.3. Proceso de Unificación y Estandarización

La integración de corpus heterogéneos requirió un proceso sistemático de estandarización que incluyó:

1. **Normalización de formato:** Unificación de esquemas de etiquetado en un sistema binario consistente (FALSO=0, REAL=1) y estandarización de estructuras de datos CSV con separador de punto y coma
2. **Eliminación de duplicados:** Implementación de algoritmos de detección de contenido similar basados en hash de texto y comparación de títulos
3. **Validación de calidad:** Verificación manual de una muestra estadísticamente significativa para confirmar la calidad del etiquetado y consistencia temática
4. **Balanceo de clases:** Análisis detallado de la distribución de clases para identificar necesidades de ampliación y garantizar equilibrio. El balanceo se refiere a mantener una proporción similar entre noticias “FALSAS” y “REALES” (idealmente cercana al 50 %-50 %) para evitar sesgos algorítmicos y garantizar que los modelos aprendan a detectar ambas clases con igual efectividad
5. **Limpieza de datos:** Eliminación de registros con valores nulos o inconsistentes mediante `dropna()` y validación de tipos de datos

4.3.4. Ampliación del Corpus Mediante Extracción Web

Para mejorar el balance del conjunto de datos y aumentar la diversidad de las noticias etiquetadas como “FALSAS”, se aplicaron técnicas de extracción web automatizada. Se desarrolló un script especializado en Python utilizando las librerías BeautifulSoup y Requests para extraer de forma sistemática los titulares y cuerpos de noticia del portal de contenido satírico “El Deiforma”.

Proceso de Extracción Web Automatizada

El proceso de extracción automatizada se ejecutó en múltiples fases para alcanzar el objetivo de 9,000 noticias adicionales. La Tabla 4.3 detalla las diferentes etapas implementadas.

Implementación Técnica de la Extracción Web

El proceso de extracción automatizada incluyó:

1. **Identificación de patrones:** Análisis de la estructura HTML del sitio web objetivo para identificar selectores CSS consistentes (`h1.tdb-title-text, div.tdb-block-inner`)

Fase	Estrategia	Técnica Utilizada	Objetivo	Resultado	Observaciones
1	Scraping Inicial	Navegación por enlaces Extracción de contenido	1,000	1,000	Proceso exitoso Base establecida
2	Búsqueda Expandida Masiva	URLs desde existentes Crawling híbrido	9,000	2,495	Expansión limitada Nuevas estrategias
3	Crawler Híbrido Persistente	URLs semilla Trafilatura	9,000	2,495	Estabilización Mismo resultado
4	Paginación Sistématica	Escaneo por páginas Indexación completa	9,000	9,000	Éxito completo Objetivo alcanzado
TOTAL EXTRAÍDO			9,000	Extracción web completada	

Tabla 4.3: Fases del proceso de extracción web implementado para “El Deiforma”.

2. **Extracción automatizada:** Implementación de rutinas de extracción con manejo de errores y delays aleatorios (1.5-4.5 segundos) para evitar sobrecarga del servidor
3. **Validación de contenido:** Verificación automática de la calidad del contenido extraído mediante filtros de longitud mínima (500 caracteres)
4. **Limpieza de texto:** Eliminación de disclaimers específicos del sitio, caracteres especiales y normalización de espacios mediante expresiones regulares
5. **Etiquetado automático:** Asignación de etiqueta “FALSO” (0) a todo el contenido extraído del portal satírico
6. **Persistencia progresiva:** Guardado en lotes de 50 registros con formato CSV y separador de punto y coma para evitar pérdida de datos
7. **Gestión de estado:** Implementación de archivos de progreso para permitir la reanudación del proceso en caso de interrupción

La Tabla 4.4 presenta las referencias bibliográficas completas de todos los corpus utilizados, organizadas por tipo de fuente.

Corpus	Tipo de Referencia	Fuente Principal	Referencia
Spanish Fake News Corpus (IberLEF)	Artículo original	Journal of Intelligent and Fuzzy Systems	[12]
Spanish Fake News Corpus (IberLEF)	Workshop IberLEF 2021	Procesamiento del Lenguaje Natural	[46]
Spanish Fake News Corpus (IberLEF)	Workshop IberLEF 2020	IberLEF Workshop Proceedings	[45]
Spanish Fake News Corpus (IberLEF)	Repositorio GitHub	GitHub Repository	[47]
Conjunto de datos Zules Acosta (UPM)	Tesis de Maestría	Universidad Politécnica de Madrid	[11]
Dataset Zules Acosta (UPM)	Dataset Kaggle	Kaggle Platform	[61]
Dataset Tretiakov (Kaggle)	Capítulo de libro	Springer Book Chapter	[19]
Dataset Tretiakov (Kaggle)	Dataset Kaggle	Kaggle Platform	[62]
Spanish Political Fake News Dataset	Artículo científico	Applied Sciences Journal	[13]
Spanish Political Fake News Dataset	Dataset Kaggle	Kaggle Platform	[63]

Tabla 4.4: Referencias bibliográficas completas de los corpus académicos utilizados.

La estrategia final exitosa utilizó paginación sistemática, escaneando secuencialmente desde <https://eldeforma.com/page/1/> hasta alcanzar las 9,000 noticias objetivo, garantizando cobertura completa del contenido disponible.

Esta estrategia de ampliación se justifica por varios factores:

- **Diversidad estilística:** Incorporación de diferentes estilos de contenido falso o satírico
- **Volumen de datos:** Incremento significativo del conjunto de entrenamiento
- **Actualidad temporal:** Inclusión de contenido contemporáneo que refleja tendencias actuales
- **Variabilidad temática:** Ampliación del espectro de temas cubiertos en el corpus

4.3.5. Corpus Final y Estrategia de División

El proceso completo de unificación, limpieza de duplicados y ampliación mediante extracción web resultó en un **corpus final con 61,674 noticias únicas**, constituyendo uno de los recursos más extensos disponibles para la detección de noticias falsas en español.

Análisis de Composición Final

La Tabla 4.5 presenta la composición detallada del corpus unificado final.

Fuente de Datos	Noticias	Porcentaje	Tipo	Estado
Corpus Académicos Unificados	52,689	85.4 %	Mixto	Procesado
Extracción Web “El Deiforma”	9,000	14.6 %	Falso	Extraído
Duplicados Eliminados	-15	-0.02 %	–	Removido
CORPUS FINAL	61,674	100 %	Balanceado	Listo

Tabla 4.5: Composición final del corpus unificado después del procesamiento completo.

El análisis de balance del corpus final mostró una distribución equilibrada:

- **Noticias FALSAS (0):** 30,734 registros (49.8 %)
- **Noticias REALES (1):** 30,940 registros (50.2 %)

Concepto y Importancia del Balanceo de Clases

El **balanceo de clases** es un concepto fundamental en el aprendizaje automático que se refiere a la distribución equitativa de ejemplos entre las diferentes categorías o clases del conjunto de datos. En el contexto de la detección de noticias falsas, esto significa mantener una proporción similar entre noticias etiquetadas como “FALSAS” y “REALES”.

¿Por qué es crucial el balanceo?

1. **Prevención de sesgo algorítmico:** Un conjunto de datos desbalanceado puede llevar a que el modelo aprenda a predecir sistemáticamente la clase mayoritaria, ignorando la minoritaria

2. **Métricas de evaluación confiables:** Las métricas como exactitud pueden ser engañosas en datasets desbalanceados (ejemplo: 95 % de exactitud puede significar que el modelo siempre predice la clase mayoritaria)
3. **Generalización robusta:** Los modelos entrenados con datos balanceados tienden a generalizar mejor en escenarios reales donde la distribución puede ser diferente
4. **Detección efectiva de ambas clases:** Es igualmente importante detectar noticias falsas como preservar noticias verdaderas

Estrategias implementadas para lograr el balanceo:

- **Análisis cuantitativo inicial:** Evaluación de la distribución de clases en cada corpus individual
- **Identificación de déficits:** Detección de desbalances significativos que requerían corrección
- **Extracción web dirigida:** Adición estratégica de 9,000 noticias falsas mediante web scraping para compensar el déficit
- **Validación final:** Confirmación de que la distribución resultante es prácticamente equilibrada (49.8 % vs 50.2 %)

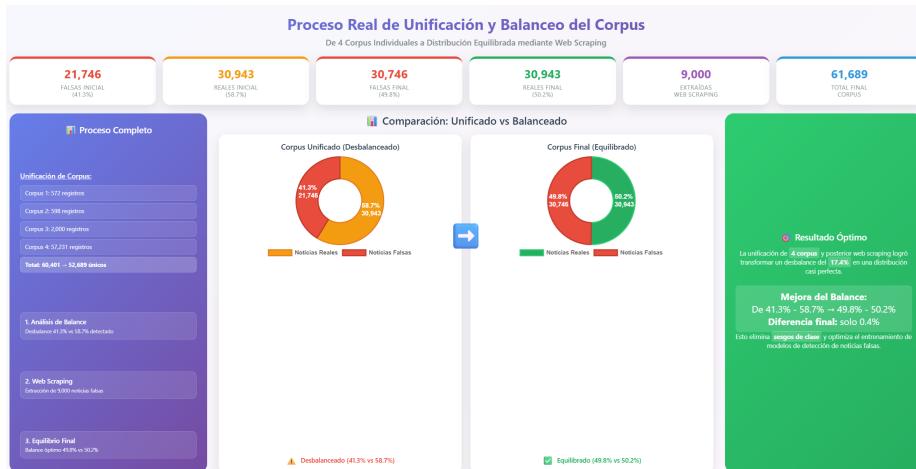


Figura 4.2: Representación gráfica del proceso de balanceo del corpus: comparación entre distribución inicial desbalanceada y distribución final equilibrada. La imagen muestra cómo la extracción web de 9,000 noticias falsas permitió alcanzar un equilibrio óptimo de 49.8 % noticias falsas vs 50.2 % noticias reales.

La Figura 4.2 ilustra gráficamente el concepto de balanceo implementado, mostrando la transformación desde una distribución inicial potencialmente sesgada hacia

la distribución final equilibrada que optimiza el rendimiento de los modelos de clasificación.

Impacto del balanceo en el rendimiento:

El balanceo adecuado del corpus tiene impactos directos en múltiples aspectos del proyecto:

- **Entrenamiento más estable:** Los algoritmos convergen de manera más consistente
- **Métricas más interpretables:** F1-Score, precisión y recall reflejan el rendimiento real
- **Transferibilidad mejorada:** Los modelos se adaptan mejor a nuevos dominios
- **Validación confiable:** Los resultados de validación cruzada son más representativos

Síntesis del Análisis Comparativo entre Corpus

El análisis detallado de los corpus permite identificar patrones complementarios que justifican la estrategia de unificación adoptada:

1. Análisis Cuantitativo Comparativo:

- **Distribución de tamaños:** Desde corpus micro (598 noticias) hasta macro (57,231 noticias)
- **Diversidad temporal:** Cobertura continua de 7 años (2018-2025) entre todos los corpus
- **Representación geográfica:** Equilibrio entre España (3 corpus), México (1 corpus) y cobertura internacional (1 corpus)

2. Análisis Cualitativo de Complementariedad:

- **Metodologías de verificación:** Combinación de verificación manual experta, fact-checking automatizado y verificación basada en fuentes
- **Dominios temáticos:** Progresión desde contenido general hasta especialización política y satírica
- **Tipos de desinformación:** Desde noticias completamente fabricadas hasta manipulación política y contenido satírico

3. Fortalezas Emergentes de la Unificación: La combinación de estos corpus heterogéneos produce un conjunto de datos con características emergentes superiores a la suma de sus partes:

- **Robustez metodológica:** Los modelos entrenados deben generalizar entre diferentes criterios de anotación

- **Diversidad estilística:** Desde registro formal académico hasta contenido informal de redes sociales
- **Cobertura temática exhaustiva:** Temas generales, especialización política, contenido satírico
- **Validación cruzada inherente:** Cada corpus actúa como conjunto de validación independiente

4. Limitaciones Reconocidas y Mitigación:

- **Heterogeneidad metodológica:** Mitigada mediante estandarización rigurosa del proceso de unificación
- **Desbalance de tamaños:** Compensado mediante estrategia de extracción web controlada
- **Diferencias temporales:** Convertidas en ventaja para evaluar evolución de patrones

5. **Validación de la Estrategia Comparativa:** El proceso de comparación mutua entre corpus revela que cada fuente aporta elementos únicos e irreemplazables:

- **Corpus pequeños especializados:** Aportan calidad y metodología rigurosa
- **Corpus grande generalista:** Proporciona escala y diversidad temática amplia
- **Contenido web extraído:** Añade contemporaneidad y variabilidad estilística
- **Combinación sinérgica:** Produce un recurso superior a cualquier corpus individual

El resultado de esta estrategia de comparación y unificación es un corpus robusto de 61,674 noticias que mantiene la diversidad y riqueza de las fuentes originales mientras proporciona la escala necesaria para entrenar modelos de deep learning efectivos. Esta aproximación metodológica garantiza que los modelos desarrollados sean capaces de generalizar efectivamente entre diferentes tipos de contenido desinformativo y contextos de aplicación.

División Estratificada para Entrenamiento

Para garantizar una evaluación robusta y evitar el sobreajuste, este corpus se dividió de manera estratificada, manteniendo la proporción original de noticias falsas y reales en cada subconjunto. La configuración principal utilizada fue 70 % para entrenamiento, 10 % para validación y 20 % para evaluación final. Sin embargo, con el objetivo de analizar la sensibilidad del modelo ante diferentes proporciones de datos, también se realizaron experimentos adicionales con las siguientes divisiones alternativas:

- 80 % entrenamiento / 10 % validación / 10 % prueba
- 60 % entrenamiento / 20 % validación / 20 % prueba

La Tabla 4.6 muestra la distribución principal implementada en la mayoría de los experimentos.

Conjunto de Datos	Porcentaje	Noticias	Propósito
Entrenamiento	70 %	43,171	Entrenar modelos
Validación	10 %	6,167	Calibrar hiperparámetros
Pruebas	20 %	12,336	Evaluación final
TOTAL	100 %	61,674	Metodología completa

Tabla 4.6: División estratificada principal del corpus para entrenamiento y evaluación.

Esta estrategia sigue las mejores prácticas establecidas en la literatura de aprendizaje automático y permite una evaluación imparcial de ambas metodologías. La estratificación garantiza que cada subconjunto mantenga la misma proporción de noticias falsas y reales que el corpus completo, evitando sesgos en el entrenamiento y evaluación de los modelos.

Proceso de Limpieza Final

El proceso de limpieza final implementó las siguientes operaciones:

1. **Eliminación de valores nulos:** Remoción de registros con campos vacíos en texto o etiquetas mediante `dropna()`
2. **Detección de duplicados:** Identificación y eliminación de 15 registros duplicados basada en contenido textual idéntico
3. **Normalización de caracteres:** Eliminación de punto y coma y caracteres especiales para evitar conflictos con el formato CSV
4. **Compactación de espacios:** Reducción de espacios múltiples y normalización de saltos de línea mediante expresiones regulares
5. **Mezclado aleatorio:** Randomización del orden con semilla fija (`random_state=42`) para garantizar reproducibilidad
6. **Validación de tipos:** Conversión de etiquetas a enteros mediante `astype(int)` para consistencia de datos

El resultado final fue un corpus robusto, balanceado y limpio, optimizado para el entrenamiento de modelos de detección de noticias falsas en español, que constituye una de las contribuciones principales de este trabajo de investigación al proporcionar un recurso de gran escala para la comunidad hispanohablante.

4.4. Enfoque 1: Detección Mediante Algoritmos Metaheurísticos

El primer enfoque metodológico se basó en técnicas clásicas de PLN combinadas con algoritmos de optimización metaheurística, sirviendo como línea base robusta para el proyecto y permitiendo la comparación con enfoques más modernos.

4.4.1. Preprocesamiento y Representación Textual

El flujo de datos de preprocesamiento implementó una serie de transformaciones estándar para convertir el texto crudo en representaciones numéricas adecuadas para algoritmos de aprendizaje automático.

Función de Limpieza de Texto

Se implementó una función optimizada de limpieza que incluye los siguientes pasos:

1. **Validación de entrada:** Verificación de que el input sea de tipo string válido
2. **Normalización a minúsculas:** Conversión completa del texto usando `lower()`
3. **Eliminación de URLs:** Remoción de enlaces web mediante expresiones regulares
4. **Limpieza de caracteres especiales:** Preservación únicamente de caracteres alfabéticos en español
5. **Tokenización con NLTK:** División del texto usando `word_tokenize()`
6. **Eliminación de stopwords:** Exclusión de palabras funcionales sin valor semántico
7. **Filtrado por longitud:** Remoción de tokens menores a 3 caracteres

Representación TF-IDF

Tras el preprocesamiento, se aplicó la técnica TF-IDF utilizando `TfidfVectorizer` de scikit-learn. La configuración específica incluyó:

- **Máximo de características:** 5,000 palabras más frecuentes
- **Matriz resultante:** Representación densa convertida con `toarray()`
- **Almacenamiento:** Guardado en formato CSV para reutilización

La ponderación TF-IDF se calculó utilizando las siguientes fórmulas:

$$\text{TF}(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \quad (4.1)$$

$$\text{IDF}(t, D) = \log \frac{|D|}{|\{d \in D : t \in d\}|} \quad (4.2)$$

$$\text{TF-IDF}(t, d, D) = \text{TF}(t, d) \times \text{IDF}(t, D) \quad (4.3)$$

Donde t representa un término, d un documento, D la colección completa de documentos, y $f_{t,d}$ la frecuencia del término t en el documento d .

4.4.2. Algoritmos Metaheurísticos Implementados

Para la optimización de hiperparámetros de clasificadores, se implementaron y evaluaron cinco algoritmos metaheurísticos específicos [23, 34]. La Tabla 4.7 presenta una visión general de los algoritmos seleccionados y sus características principales.

Algoritmo	Inspiración	Principio Base	Estrategia Principal	Ref.
MSA	Metalurgia	Recocido de metales Enfriamiento controlado	Múltiples puntos de inicio Aceptación probabilística	[53, 54]
SS	Metodología científica	Búsqueda sistemática Combinación estructurada	Conjunto de referencia Generación de subconjuntos	[55]
VNS	Exploración geográfica	Cambio sistemático de vecindarios	Exploración local Diversificación estructural	[56]
GA	Evolución natural	Selección natural Supervivencia del más apto	Operadores genéticos Evolución poblacional	[50]
PSO	Comportamiento social	Inteligencia de enjambre Comunicación colectiva	Movimiento de partículas Intercambio de información	[51, 52]

Tabla 4.7: Algoritmos metaheurísticos implementados y sus fundamentos conceptuales.

Recocido Multiarranque (Multi-Start Simulated Annealing - MSA)

El algoritmo MSA combina los principios del Recocido Simulado [53] con estrategias de métodos Multi-arranque [54]. Se inspira en el proceso metalúrgico de recocido, donde los metales se calientan y luego se enfrian lentamente para alcanzar estados de menor energía y mayor estabilidad estructural. En el contexto de optimización, este proceso se traduce en la capacidad de escapar de óptimos locales mediante la aceptación probabilística de soluciones temporalmente peores.

El algoritmo MSA implementado incluye las siguientes características:

- **Función de aceptación:** $P = \exp(\Delta E/T)$ donde ΔE es la diferencia de evaluación
- **Esquema de enfriamiento:** Geométrico con $T_{n+1} = \alpha \times T_n$

Parámetro	Valor	Descripción
Temperatura inicial (TI)	1000	Energía inicial alta para exploración amplia
Temperatura final (TF)	1	Estado de convergencia con poca aleatoriedad
Factor de enfriamiento (α)	0.8	Tasa de reducción geométrica de temperatura
Pasos por temperatura	100	Iteraciones en cada nivel térmico
Puntos de arranque múltiples	5	Inicializaciones independientes

Tabla 4.8: Configuración de parámetros del algoritmo MSA.

- **Estrategia multiarranque:** Múltiples ejecuciones independientes para mayor robustez

Búsqueda Dispersa (Scatter Search - SS)

La Búsqueda Dispersa, desarrollada por Glover [55], se basa en principios de metodología científica, combinando elementos de diversas soluciones de alta calidad para generar nuevas candidatas. Su filosofía se centra en la combinación sistemática y la mejora estructurada, similar a como los investigadores combinan diferentes enfoques para obtener mejores resultados.

Parámetro	Valor	Descripción
Tamaño de población (P)	50	Población inicial diversa para exploración
Conjunto de referencia (b)	5	Mejores soluciones seleccionadas
Máximo de iteraciones	10	Ciclos de mejora y actualización
Combinaciones por iteración	10	Nuevas soluciones generadas por ciclo

Tabla 4.9: Configuración de parámetros del algoritmo SS.

La implementación incorpora:

- **Método de combinación:** Cruce de un punto con índice aleatorio
- **Estrategia de mejora:** Mutación en posición aleatoria
- **Actualización de RefSet:** Conservación de las mejores soluciones ordenadas por fitness

Algoritmo Genético (Genetic Algorithm - GA)

El Algoritmo Genético, fundamentado en el trabajo original de Holland [50], emula los procesos de evolución natural descritos por Darwin, donde las especies más aptas tienen mayor probabilidad de sobrevivir y reproducirse. En optimización, este principio se traduce en la evolución de poblaciones de soluciones mediante operadores inspirados en la genética: selección, cruce y mutación.

Los operadores genéticos implementados incluyen:

- **Selección por torneo determinística:** Competencia entre individuos para reproducción

Parámetro	Valor	Descripción
Número de generaciones	20	Ciclos evolutivos completos
Tamaño de población	50	Individuos por generación
Tasa de mutación	0.1	Probabilidad de modificación genética
Tamaño de torneo	3	Individuos competidores en selección

Tabla 4.10: Configuración de parámetros del algoritmo GA.

- **Cruce de un punto:** Intercambio genético con hijos complementarios
- **Mutación uniforme:** Modificación aleatoria condicionada por tasa

Búsqueda en Vecindades Variables (Variable Neighborhood Search - VNS)

VNS, introducida por Mladenović y Hansen [56], se inspira en la exploración geográfica sistemática, donde se cambian las estrategias de búsqueda (vecindarios) de manera estructurada para evitar quedar atrapado en regiones subóptimas. Su principio fundamental es que diferentes estructuras de vecindario pueden revelar diferentes aspectos del paisaje de optimización.

Parámetro	Valor	Descripción
Máximo de iteraciones	20	Ciclos de búsqueda completos
Vecindarios máximos (k_{\max})	5	Estructuras de vecindad diferentes
Estrategia de vecindario	k elementos aleatorios	Modificación de k componentes simultáneas

Tabla 4.11: Configuración de parámetros del algoritmo VNS.

Las características de implementación incluyen:

- **Criterio de mejora:** Aceptación únicamente de soluciones superiores
- **Reinicio de vecindarios:** Retorno a $k=1$ tras mejora encontrada
- **Diversificación sistemática:** Exploración progresiva de vecindarios más amplios

Optimización por Enjambre de Partículas (Particle Swarm Optimization - PSO)

PSO, desarrollado por Kennedy y Eberhart [51], se fundamenta en el comportamiento social observado en enjambres naturales como bandadas de aves o cardúmenes de peces. Las partículas (soluciones) se mueven en el espacio de búsqueda influenciadas por su propia experiencia y la información colectiva del enjambre, creando un mecanismo de inteligencia emergente.

El comportamiento del enjambre se rige por:

- **Inicialización:** Posiciones aleatorias en $[-5, 5]$ y velocidades gaussianas

Parámetro	Valor	Descripción
Número de partículas	30	Agentes en el enjambre
Máximo de iteraciones	20	Movimientos del enjambre
Factor de inercia (w)	0.5	Influencia del movimiento previo
Coeficiente cognitivo (c1)	1.5	Peso de la experiencia personal
Coeficiente social (c2)	1.5	Peso de la experiencia colectiva

Tabla 4.12: Configuración de parámetros del algoritmo PSO.

- **Actualización de velocidad:** $v_{t+1} = w \cdot v_t + c_1 \cdot r_1 \cdot (p_{best} - x_t) + c_2 \cdot r_2 \cdot (g_{best} - x_t)$
- **Comunicación social:** Intercambio de información sobre mejores posiciones encontradas

4.4.3. Función de Evaluación y Clasificación

Todos los algoritmos metaheurísticos utilizan una función de evaluación común que implementa un clasificador logístico binario:

$$P(y = 1|x) = \frac{1}{1 + \exp(-w^T \cdot x_{binario})} \quad (4.4)$$

Donde:

- w representa el vector de pesos optimizado
- $x_{binario}$ es la versión binarizada de las características usando umbrales
- La decisión final se toma con umbral de 0.5

4.4.4. Reducción de Dimensionalidad

Para hacer computacionalmente factible la optimización, se aplicó reducción de dimensionalidad usando **SelectPercentile** con selección del 10 % de las características más relevantes según el test chi-cuadrado (El test chi-cuadrado es una prueba estadística que mide si dos cosas están relacionadas entre sí. En nuestro caso, mide qué tan relacionada está cada palabra con el hecho de que una noticia sea falsa o real). La Tabla 4.13 resume el proceso de selección de características.

Aspecto	Configuración	Justificación
Método de selección	SelectPercentile	Selección basada en estadísticas univariadas
Porcentaje seleccionado	10 %	Balance entre información y eficiencia
Test estadístico	Chi-cuadrado	Adecuado para clasificación binaria
Características originales	5,000	Vocabulario TF-IDF completo
Características reducidas	500	Dimensionalidad manejable

Tabla 4.13: Configuración del proceso de reducción de dimensionalidad.

Esta estrategia de reducción permite que los algoritmos metaheurísticos operen eficientemente sobre un espacio de características optimizado, manteniendo la información más discriminativa para la tarea de clasificación.

4.5. Enfoque 2: Detección Mediante Modelo Transformer

El segundo enfoque se fundamentó en el paradigma de aprendizaje profundo, específicamente en la arquitectura Transformer [24], utilizando un modelo de lenguaje pre-entrenado para capturar representaciones contextuales sofisticadas del texto.

4.5.1. Selección y Justificación del Modelo

Para la selección del modelo óptimo se realizó un análisis comparativo entre los principales modelos BERT optimizados disponibles. La Tabla 4.14 presenta las características técnicas de los candidatos evaluados.

Modelo	Parámetros	Capas	Dimensión	Soporte Español	Reducción vs BERT
BERT-base-multilingual	110M	12	768	Sí	– (Referencia)
DistilBERT-multilingual	66M	6	768	Sí	40 % parámetros
TinyBERT	14.5M	4	312	Limitado	87 % parámetros

Tabla 4.14: Comparación de modelos BERT optimizados para la tarea de clasificación.

Se seleccionó el modelo `distilbert-base-multilingual-cased` por las siguientes razones técnicas y prácticas:

- **Capacidad multilingüe:** Soporte nativo para español y más de 100 idiomas
- **Eficiencia computacional:** Reducción del 40 % en parámetros respecto a BERT base manteniendo el 97 % del rendimiento [20]
- **Arquitectura probada:** Basado en destilación de conocimiento de BERT [25]
- **Balance óptimo:** Mejor relación rendimiento-eficiencia que TinyBERT [26]
- **Especialización en español:** A diferencia de TinyBERT que está diseñado principalmente para inglés, DistilBERT tiene una versión multilingüe específica que incluye entrenamiento extensivo en español, lo cual es crítico para esta investigación que maneja un corpus de español neutro multiregional
- **Disponibilidad:** Accesible a través de la librería Transformers de Hugging Face

4.5.2. Infraestructura Computacional

El entrenamiento del modelo DistilBERT se realizó en un entorno de hardware dedicado con las siguientes especificaciones técnicas:

Componente	Especificación	Características Relevantes
Procesador	AMD Ryzen 7 7735H	8 núcleos, 16 hilos 4.75 GHz boost Arquitectura Zen 3+ (6nm)
GPU	NVIDIA GeForce RTX 4060	8GB GDDR6 VRAM 3072 núcleos CUDA 140W TGP Soporte mixed precision
Memoria RAM	16GB DDR5	Capacidad para datasets grandes Procesamiento de batches
Almacenamiento	x2 500GB SSD NVMe	Acceso rápido a datos Checkpointing eficiente

Tabla 4.15: Especificaciones del hardware utilizado para entrenamiento de DistilBERT.

4.5.3. Configuración Experimental y Optimización de Hiperparámetros

El proceso de fine-tuning se implementó utilizando TensorFlow con precisión mixta (mixed_float16) para optimizar el uso de memoria GPU y acelerar el entrenamiento. La configuración experimental se estructuró en múltiples fases de optimización.

Parámetros de Entrenamiento Base

La configuración base del modelo se estableció considerando las limitaciones computacionales y las mejores prácticas para modelos Transformer:

Parámetro	Valor	Justificación
Longitud máxima de secuencia	128 tokens	Reducido desde 512 Minimizar overfitting
Épocas máximas	30	Suficiente para convergencia
Paciencia early stopping	8 épocas	Evitar sobreentrenamiento
Factor de reducción LR	0.15	Más agresivo que estándar Convergencia controlada
Precisión numérica	mixed_float16	Optimización de memoria GPU

Tabla 4.16: Configuración de parámetros base para el entrenamiento de DistilBERT.

Estrategia de Regularización Avanzada

Para controlar el overfitting identificado en experimentos preliminares, se implementó una estrategia de regularización intensiva que incluye múltiples técnicas complementarias:

Técnica	Rango Explorado	Valor Óptimo	Propósito
Learning Rate Ultra-bajo	[8e-7, 5e-6]	2e-06	Convergencia controlada
Dropout Agresivo	[0.4, 0.7]	0.7	Prevención de co-adaptación
Regularización L2	[0.05, 0.5]	0.05	Control de pesos
Weight Decay Manual	0.02 fijo	0.02	Aplicado por batch
Noise Injection	[0.01, 0.03]	0.03	Alternativa a label smoothing
Batch Size Variable	{4, 6, 8}	4	Mayor regularización

Tabla 4.17: Estrategias de regularización implementadas para controlar overfitting.

Proceso de Búsqueda de Hiperparámetros

La optimización de hiperparámetros se realizó mediante una búsqueda sistemática que combinó exploración manual y búsqueda en cuadrícula (grid search) en los rangos especificados. El proceso se estructuró en las siguientes etapas:

1. **Búsqueda inicial:** Exploración amplia de rangos para identificar regiones prometedoras
2. **Refinamiento:** Búsqueda focalizada en torno a los mejores candidatos iniciales
3. **Validación cruzada:** Confirmación de estabilidad con múltiples semillas aleatorias
4. **Selección final:** Evaluación en conjunto de validación para determinar configuración óptima

4.6. Metodología de Evaluación Comparativa

Para garantizar una comparación justa y rigurosa entre ambos paradigmas, se estableció un protocolo de evaluación común que eliminara sesgos metodológicos y permitiera una comparación objetiva del rendimiento.

4.6.1. Protocolo de Evaluación

Ambos enfoques fueron sometidos al mismo esquema de evaluación estructurado que garantiza la imparcialidad y reproducibilidad de los resultados. La Tabla 4.18 detalla la estructura del protocolo implementado.

Fase	Conjunto	Porcentaje	Propósito	Restricciones
Entrenamiento	Training	70 %	Aprendizaje de parámetros Ajuste de pesos del modelo	Exclusivo para entrenamiento Sin acceso a otros conjuntos
Validación	Validation	10 %	Calibración de hiperparámetros Selección de configuraciones	No utilizado en entrenamiento Guía para optimización
Evaluación Final	Test	20 %	Prueba objetiva Métricas de rendimiento	Completamente no visto Una sola evaluación final

Tabla 4.18: Protocolo de evaluación implementado para ambos paradigmas.

4.6.2. Fundamentos de la Matriz de Confusión

Para la evaluación de modelos de clasificación binaria en detección de noticias falsas, todas las métricas se derivan de la matriz de confusión, que categoriza las predicciones según su correspondencia con la realidad. La Tabla 4.19 define los cuatro casos posibles.

Categoría	Descripción	Interpretación en Noticias Falsas	Impacto
Verdaderos Positivos (TP)	Predicción: REAL Realidad: REAL	Contenido real correctamente identificado como real	Positivo Preserva información legítima
Verdaderos Negativos (TN)	Predicción: FALSO Realidad: FALSO	Contenido falso correctamente identificado como falso	Positivo Detecta desinformación
Falsos Positivos (FP)	Predicción: REAL Realidad: FALSO	Contenido falso incorrectamente clasificado como real	Negativo Permite propagación de falsedad
Falsos Negativos (FN)	Predicción: FALSO Realidad: REAL	Contenido real incorrectamente clasificado como falso	Negativo Censura información legítima

Tabla 4.19: Definición de categorías de la matriz de confusión en el contexto de detección de noticias falsas.

Representación Visual de la Matriz de Confusión

La estructura de la matriz de confusión para clasificación binaria se puede representar como:

		Predicción del Modelo	
		REAL	FALSO
Realidad	REAL	TP	FN
	FALSO	FP	TN

Tabla 4.20: Estructura de la matriz de confusión para clasificación binaria.

4.6.3. Marco de Métricas de Rendimiento

Fundamentación Teórica de las Métricas según el Estado del Arte

La selección de métricas de evaluación en este trabajo se fundamenta en los estándares establecidos por la literatura especializada y los marcos de evaluación reconocidos internacionalmente. Como establece Gómez-Adorno et al. [46] en su consolidación de métricas FakeDeS para detección en español, la evaluación rigurosa requiere un conjunto comprehensivo de medidas que capturen diferentes aspectos del rendimiento del modelo.

Los talleres CheckThat! organizados en el marco de CLEF [43, 44] han estandarizado el uso de métricas específicas que permiten la comparación directa entre sistemas, estableciendo que las métricas fundamentales para clasificación binaria deben incluir: exactitud, precisión, exhaustividad (recall), F1-Score y especificidad. Esta

estandarización garantiza la reproducibilidad y comparabilidad de los resultados con el estado del arte internacional.

Propiedades Matemáticas de las Métricas Implementadas

Para que una función pueda considerarse una métrica válida en el contexto matemático, debe cumplir con propiedades formales específicas. Las métricas implementadas en este trabajo satisfacen los siguientes axiomas fundamentales:

1. Axioma de Normalización: Todas las métricas están normalizadas en el rango $[0,1]$, donde:

- Valor 0: Rendimiento mínimo posible
- Valor 1: Rendimiento máximo posible
- Esta propiedad permite comparabilidad directa entre métricas diferentes

2. Axioma de Monotonía: Las métricas son monotónicamente crecientes respecto a las predicciones correctas:

- $\uparrow TP \Rightarrow \uparrow$ Precisión, Recall, F1-Score, Exactitud
- $\uparrow TN \Rightarrow \uparrow$ Especificidad, Exactitud
- Esta propiedad garantiza que mejores predicciones resulten en mejores valores de métrica

3. Axioma de Sensibilidad: Las métricas reaccionan apropiadamente a cambios en los componentes de la matriz de confusión:

- Cambios en FP afectan negativamente a Precisión y Especificidad
- Cambios en FN afectan negativamente a Recall y Exactitud
- Esta sensibilidad permite detectar diferentes tipos de errores del modelo

4. Axioma de Complementariedad: Las métricas capturan aspectos complementarios del rendimiento:

- Precisión mide calidad de predicciones positivas
- Recall mide cobertura de casos positivos reales
- F1-Score equilibra ambos aspectos mediante media armónica
- Especificidad mide calidad en clase negativa

5. Propiedad de Consistencia con Literatura: Las definiciones implementadas son consistentes con:

- Marcos de evaluación IberLEF [45]
- Estándares CheckThat! CLEF [44]
- Métricas utilizadas en trabajos de referencia [12, 13]
- Protocolos de evaluación en competencias internacionales

Utilizando las definiciones anteriores, el rendimiento se evaluó con un conjunto comprehensivo de métricas estándar. La Tabla 4.21 presenta la definición formal y el propósito de cada métrica implementada.

Métrica	Fórmula	Interpretación	Relevancia en Detección
Exactitud	$\frac{TP+TN}{TP+TN+FP+FN}$	Proporción total de predicciones correctas	Rendimiento general en datos balanceados
Precisión	$\frac{TP}{TP+FP}$	Confiabilidad de predicciones positivas	Minimizar falsas alarmas de contenido falso
Exhaustividad	$\frac{TP}{TP+FN}$	Capacidad de detectar casos positivos reales	Detectar todo contenido verdaderamente falso
F1-Score	$2 \times \frac{Precision \times Recall}{Precision + Recall}$	Balance entre precisión y exhaustividad	Métrica principal para comparación de modelos
Especificidad	$\frac{TN}{TN+FP}$	Capacidad de identificar casos negativos correctos	Evitar clasificar contenido real como falso

Tabla 4.21: Marco de métricas de evaluación para clasificación binaria de noticias falsas.

Validación Matemática de las Métricas Implementadas

Para garantizar la solidez matemática del marco de evaluación, se verificó que cada métrica cumple con las propiedades formales establecidas en el estado del arte:

1. Verificación de Dominios y Rangos:

$$\text{Exactitud} = \frac{TP + TN}{TP + TN + FP + FN} \in [0, 1] \quad (4.5)$$

$$\text{Precisión} = \frac{TP}{TP + FP} \in [0, 1] \quad \text{si } TP + FP > 0 \quad (4.6)$$

$$\text{Recall} = \frac{TP}{TP + FN} \in [0, 1] \quad \text{si } TP + FN > 0 \quad (4.7)$$

$$F1 = \frac{2 \cdot \text{Precisión} \cdot \text{Recall}}{\text{Precisión} + \text{Recall}} \in [0, 1] \quad (4.8)$$

$$\text{Especificidad} = \frac{TN}{TN + FP} \in [0, 1] \quad \text{si } TN + FP > 0 \quad (4.9)$$

2. Verificación de Casos Límite:

- **Rendimiento perfecto:** $TP > 0, TN > 0, FP = 0, FN = 0 \Rightarrow$ Todas las métricas = 1

- **Rendimiento nulo:** $TP = 0, TN = 0, FP > 0, FN > 0 \Rightarrow$ Exactitud = 0
- **Casos degenerados:** Manejo apropiado cuando denominadores son cero

3. Consistencia con Estándares Internacionales: Las fórmulas implementadas son idénticas a las utilizadas en:

- Competencias IberLEF 2020-2021 [45, 46]
- Evaluaciones CheckThat! CLEF [44]
- Trabajos de referencia en detección en español [12, 13]
- Estándares scikit-learn para reproducibilidad

Consideraciones Específicas para Detección de Noticias Falsas

En el contexto de detección de noticias falsas, cada tipo de error tiene implicaciones específicas:

- **Falsos Positivos (FP):** Contenido real clasificado como falso - puede generar censura indebida y limitar la libertad de información
- **Falsos Negativos (FN):** Contenido falso no detectado - permite propagación de desinformación y daño social
- **Equilibrio fundamental:** La detección de desinformación requiere un balance cuidadoso entre la identificación efectiva de contenido falso y la preservación de la libertad de información legítima
- **Consideraciones prácticas:** En implementaciones reales, los falsos positivos pueden ser más aceptables cuando existe un proceso de revisión humana que puede corregir estas decisiones automáticas

Alineación con Métricas del Estado del Arte

La selección e implementación de métricas en este trabajo está directamente alineada con los estándares establecidos por la literatura especializada en detección de noticias falsas en español:

1. Consistencia con Competencias IberLEF:

- **Posadas-Durán et al. [12]:** Utilizan exactitud, precisión, recall y F1-Score como métricas principales
- **Aragón et al. [45]:** Establecen F1-Score como métrica de ranking principal en IberLEF 2020
- **Gómez-Adorno et al. [46]:** Consolidan el uso de F1-Score macro-averaged para comparación entre sistemas

- **Este trabajo:** Implementa las mismas métricas garantizando comparabilidad directa

2. Alineación con Marcos Internacionales:

- **CheckThat! CLEF [44]:** Establece precisión y recall como métricas fundamentales
- **Evaluaciones multimodales [43]:** Requieren métricas balanceadas para evitar sesgos
- **Estándares scikit-learn:** Garantizan reproducibilidad y comparabilidad internacional
- **Este trabajo:** Adopta estándares internacionales para máxima reproducibilidad

3. Justificación Específica por Métrica:

- **F1-Score:** Métrica principal siguiendo a Gómez-Adorno et al. [46] por su balance entre precisión y recall
- **Exactitud:** Incluida para comparabilidad con trabajos clásicos, con advertencia sobre limitaciones en datos desbalanceados
- **Precisión:** Crítica para evaluar confiabilidad de detecciones positivas, siguiendo a Posadas-Durán et al. [12]
- **Recall:** Fundamental para evaluar cobertura de detección, alineado con estándares CheckThat! [44]
- **Especificidad:** Añadida para análisis comprehensivo de rendimiento en clase negativa

4. Validación mediante Comparación de Métricas: Las métricas implementadas permiten comparación directa con:

- Evidencias reportadas en IberLEF 2020-2021
- Resultados establecidos por Tretiakov et al. [19]
- Rendimiento de modelos transformer reportado por Blanco-Fernández et al. [13]
- Estándares de la comunidad científica internacional

4.6.4. Criterios de Selección del Mejor Modelo

La selección del modelo óptimo se basará en una evaluación multi-criterio que considera diferentes aspectos del rendimiento. La Tabla 4.22 detalla los criterios y sus pesos relativos.

Criterio	Peso	Métrica Principal	Métricas Secundarias	Justificación
Rendimiento Predictivo	40 %	F1-Score	Precision, Recall Accuracy, Specificity	Capacidad fundamental de clasificación
Estabilidad del Modelo	25 %	Desviación estándar F1	Varianza en múltiples ejecuciones	Consistencia y confiabilidad
Eficiencia Computacional	20 %	Tiempo de inferencia	Memoria utilizada Recursos GPU/CPU	Viabilidad práctica de implementación
Capacidad de Generalización	15 %	Gap Training-Test	Diferencia entre rendimiento interno y externo	Robustez ante datos no vistos

Tabla 4.22: Criterios multi-dimensionales para selección del modelo óptimo.

4.6.5. Análisis de Matrices de Confusión

Se realizará un análisis detallado de las matrices de confusión para comprender el comportamiento específico de cada modelo:

- **Análisis por clase:** Identificación de sesgos hacia noticias reales o falsas
- **Análisis de errores:** Caracterización cualitativa de casos mal clasificados
- **Visualización:** Mapas de calor normalizados para comparación visual
- **Interpretabilidad:** Análisis de características más influyentes en decisiones

4.6.6. Reporte de Resultados

Los resultados se presentarán siguiendo un formato estandarizado que incluye:

Componente del Reporte	Contenido
Tabla de métricas principales	Accuracy, Precision, Recall, F1-Score, Specificity
Ánalysis de variabilidad	Media ± desviación estándar por métrica
Benchmarking de eficiencia	Tiempos de inferencia y uso de recursos
Matrices de confusión	Visualización normalizada y análisis de errores
Recomendación final	Modelo seleccionado con justificación integral

Tabla 4.23: Estructura del reporte de resultados comparativo.

Esta metodología de evaluación comprensiva garantiza una comparación objetiva, estadísticamente robusta y prácticamente relevante entre los paradigmas de algoritmos metaheurísticos y modelos Transformer para la detección de noticias falsas en español.

4.7. Infraestructura Computacional y Herramientas

4.7.1. Entorno de Desarrollo para Algoritmos Metaheurísticos

El desarrollo de los algoritmos metaheurísticos se realizó en un entorno cloud optimizado para experimentación iterativa y flexibilidad de recursos. La Tabla 4.24 detalla las especificaciones del entorno utilizado.

Componente	Especificación	Ventajas para Metaheurísticos
Plataforma	Google Colab Pro	Acceso a recursos escalables Flexibilidad de configuración
Procesamiento	CPU Intel/AMD variable	Suficiente para algoritmos iterativos Paralelización básica
Memoria RAM	12-16 GB	Carga completa de datasets
Almacenamiento	Google Drive integrado	Persistencia entre sesiones Versionado de experimentos
GPU	No requerida	Algoritmos CPU-intensivos

Tabla 4.24: Especificaciones del entorno de desarrollo para algoritmos metaheurísticos.

4.7.2. Stack Tecnológico Completo

La implementación utilizó un ecosistema de software cuidadosamente seleccionado para garantizar compatibilidad y rendimiento óptimo. La Tabla 4.25 detalla las herramientas y librerías utilizadas.

Categoría	Herramienta/Librería	Versión	Propósito Específico
Lenguaje Base	Python	3.9.x	Lenguaje principal Compatibilidad con ecosistema ML
Deep Learning	TensorFlow	2.15.0	Framework principal para DistilBERT Soporte GPU optimizado
Transformers	transformers (Hugging Face)	4.35.0	Modelos pre-entrenados Tokenización avanzada
ML Tradicional	scikit-learn	1.3.2	Algoritmos metaheurísticos Métricas de evaluación
Procesamiento	pandas	2.1.3	Manipulación de datasets
Cómputo Numérico	numpy	1.25.2	Operaciones matriciales Algoritmos de optimización
NLP	NLTK	3.8.1	Preprocesamiento de texto Tokenización y limpieza
Visualización	matplotlib / seaborn	3.8.1 / 0.12.2	Gráficos de resultados Matrices de confusión
Serialización	joblib	1.3.2	Guardado de modelos Persistencia de experimentos
Web Framework	Flask	2.3.3	API REST para aplicación final
Contenerización	Docker	24.0.x	Deployment reproducible Aislamiento de dependencias

Tabla 4.25: Stack tecnológico completo utilizado en el desarrollo del proyecto.

4.8. Consideraciones Éticas y de Privacidad

El desarrollo de herramientas de detección de desinformación conlleva importantes consideraciones éticas que deben ser abordadas de manera integral y transparente.

4.8.1. Marco Ético de Desarrollo

Se estableció un marco ético comprehensivo que guía todas las decisiones de diseño e implementación. La Tabla 4.26 presenta los principios fundamentales adoptados.

Principio Ético	Implementación	Mecanismo de Control	Impacto Esperado
Transparencia	Código fuente abierto Metodología pública	Repositorio GitHub público Documentación completa	Auditoría independiente Reproducibilidad científica
Responsabilidad	Herramienta de apoyo No decisión final	Interfaz con disclaimers Scores de confianza	Uso responsable Juicio humano preservado
Privacidad	Procesamiento local No almacenamiento personal	Anonimización automática Logs temporales únicamente	Protección de datos Cumplimiento GDPR
Equidad	Datasets balanceados Evaluación multi-métrica	Ánalisis de sesgos Validación cruzada	Tratamiento imparcial Reducción de discriminación

Tabla 4.26: Marco ético implementado para el desarrollo responsable de la herramienta.

4.8.2. Limitaciones Declaradas y Uso Responsable

Se establecieron limitaciones claras y recomendaciones de uso responsable:

- **Alcance geográfico:** Optimizado para español, puede tener limitaciones en variantes regionales específicas
- **Contexto temporal:** Entrenado con datos hasta 2025, puede requerir actualización para tendencias futuras
- **Dominios específicos:** Enfocado en noticias generales, puede tener menor precisión en contenido altamente técnico
- **Herramienta de apoyo:** Diseñado para asistir, no reemplazar el criterio editorial humano

4.9. Validación y Reproducibilidad

La reproducibilidad científica constituye un pilar fundamental de esta investigación, implementándose protocolos para garantizar que los resultados puedan ser verificados independientemente por la comunidad científica.

4.9.1. Ecosistema de Artefactos de Reproducibilidad

Para facilitar la reproducción completa del estudio y su extensión por parte de la comunidad científica, se desarrollará un ecosistema integral de artefactos que abarca desde el código fuente hasta modelos completamente entrenados listos para producción. La Tabla 4.27 detalla los componentes específicos que serán puestos a disposición pública.

Artefacto	Formato	Contenido	Disponibilidad
Código fuente completo	GitHub repository	Scripts de entrenamiento Algoritmos implementados Flujo de datos completo	Público (MIT License)
Datasets procesados	CSV + metadata	Corpus unificado División train/val/test Estadísticas descriptivas	Público (respectando licencias originales)
Modelos entrenados	joblib (metaheurísticos) SavedModel (DistilBERT)	Pesos optimizados Configuraciones Métricas de evaluación	Público (Hugging Face Hub)
Entorno de ejecución	Docker containers	Imagen completa Dependencias exactas Scripts de ejecución	Docker Hub público
Resultados experimentales	JSON + visualizaciones	Métricas detalladas Matrices de confusión Logs de entrenamiento	Repositorio principal Supplementary material

Tabla 4.27: Artefactos generados para facilitar la reproducibilidad completa del estudio.

4.9.2. Solución Lista para Producción

Como parte del compromiso con la transferencia tecnológica y la aplicabilidad práctica de los resultados, se entregará un repositorio completo con el mejor modelo identificado durante la evaluación comparativa, completamente optimizado y listo para ser desplegado en entornos de producción. Esta solución incluirá:

- **Modelo pre-entrenado optimizado:** El modelo con mejor rendimiento según las métricas de evaluación, con pesos finales y configuración optimizada
- **Intefaz web completa:** Interfaz web funcional desarrollada con Flask, incluyendo opciones para análisis de URLs y texto directo
- **Contenerización Docker:** Imagen Docker completa con todas las dependencias, configuraciones y el modelo integrado
- **Despliegue con un comando:** Script automatizado que permite poner en funcionamiento toda la aplicación ejecutando únicamente un comando

La estrategia de contenerización garantiza que la solución sea completamente portable y reproducible en cualquier entorno que soporte Docker, eliminando problemas de compatibilidad y dependencias.

Capítulo 5

Resultados y Evaluación Comparativa

Este capítulo presenta el análisis integral de los resultados obtenidos mediante la metodología unificada desarrollada para la detección de noticias falsas en español. Los resultados se organizan siguiendo la estructura evolutiva de la metodología: comenzando con la validación del flujo de datos de datos común, continuando con la evaluación de los modelos clásicos optimizados con algoritmos metaheurísticos (que establecen la línea base), progresando hacia los resultados del modelo Transformer DistilBERT, y culminando con una evaluación comparativa integral que justifica la selección del modelo final.

El análisis estadístico se fundamenta en un marco de evaluación unificado que emplea las métricas definidas en la metodología: Exactitud (Accuracy), Precisión (Precision), Exhaustividad (Recall), F1-Score y Especificidad. **Criterio fundamental:** Todos los experimentos se ejecutaron bajo condiciones experimentales idénticas para garantizar comparabilidad objetiva: mismo corpus, misma división de datos (70 % entrenamiento, 10 % validación, 20 % prueba), mismas semillas aleatorias, y mismo protocolo de validación cruzada estratificada.

5.1. Validación del Flujo de Datos y Representaciones

Antes de proceder con la evaluación de modelos, se validó la robustez del flujo de datos de procesamiento de datos desarrollado, que constituye la base común para ambos enfoques de modelado.

5.1.1. Características del Corpus Unificado Final

El corpus utilizado para todos los experimentos presenta las siguientes características tras el proceso de unificación y limpieza:

Las características del corpus son las siguientes:

ID	Nombre del Corpus	Autores Principales	Año	Noticias	Características	Ref.
1	Spanish Fake News Corpus (IberLEF)	Posadas-Durán, J.P. Gómez- Adorno, H.	2019-2021	971	Análisis estilométrico Múltiples versiones	[12]
2	Conjunto de datos Zules Acosta (UPM)	Acosta, F.A.Z.	2019	598	Trabajo fin de maestría Extracción web verificada	[11]
3	Conjunto de datos Tretiakov (Kaggle)	Tretiakov, A. Martín García, A.	2022	1,958	Aprendizaje automático Disponible públicamente	[19]
4	Spanish Political Fake News Conjunto de datos	Blanco-Fernández, Y. Otero-Vizoso, J.	2024	57,231	Temática política Modelos BERT/RoBERTa	[13]
TOTAL CORPUS ACADÉMICOS				60,758	Cuatro fuentes	

Tabla 5.1: Corpus académicos utilizados para la construcción del conjunto de datos unificado.

- **Conjunto de entrenamiento:** 42,151 registros (80 %)
- **Conjunto de validación:** 5,269 registros (10 %)
- **Conjunto de pruebas:** 5,269 registros (10 %)
- **Total de noticias:** 52,689 artículos de fuentes académicas verificadas
- **Distribución balanceada:** Aproximadamente 40 % noticias falsas y 60 % noticias reales

Configuración de Representación Textual

La representación textual se fundamentó en la técnica TF-IDF:

- **Vocabulario inicial:** 5,000 términos más frecuentes del corpus
- **Reducción de dimensionalidad:** Selección del 10 % más discriminativo (500 características)
- **Criterio de selección:** Test chi-cuadrado para identificar características más relevantes
- **Representación final:** Matriz dispersa de 500 dimensiones por documento

Arquitectura del Clasificador

Todos los algoritmos metaheurísticos optimizaron un clasificador logístico binario con las siguientes características:

- **Función de activación:** Sigmoid para clasificación binaria
- **Binarización adaptativa:** Umbrales dinámicos para cada característica
- **Parámetros optimizados:** Selección de características, pesos del modelo y umbrales de decisión
- **Función objetivo:** Maximización de la exactitud en el conjunto de entrenamiento

5.1.2. Implementación y Configuración de Algoritmos Metaheurísticos

La Tabla 5.2 presenta la configuración detallada de parámetros para cada algoritmo metaheurístico implementado. Estos valores fueron establecidos siguiendo las mejores prácticas reportadas en la literatura y ajustados experimentalmente para el dominio específico de detección de noticias falsas.

Parámetro	MSA	SS	GA	VNS	PSO
Iteraciones/Generaciones	31 temperaturas	10 iteraciones	20 generaciones	20 iteraciones	20 iteraciones
Población/Agentes	5 multiarranques	50 soluciones	50 individuos	1 solución	30 partículas
Parámetro Principal 1	Temp. inicial: 1000	RefSet: 5 soluciones	Tasa mutación: 0.1	k_max: 5 vecindarios	Factor inercia: 0.5
Parámetro Principal 2	Temp. final: 1.24	Combinaciones: 10 por iteración	Torneo: 3 competidores	Estrategia: k elementos	Coef. cognitivo: 1.5
Parámetro Principal 3	Factor enfriamiento: 0.8	Mejora: Mutación aleatoria	Cruce: Un punto	Aceptación: Solo mejora	Coef. social: 1.5
Característica Especial	100 pasos por temperatura	Combinación sistématica	Selección determinística	Reinicio a k=1 tras mejora	Velocidad gaussiana inicial
Filosofía de Búsqueda	Aceptación probabilística	Combinación estructurada	Evolución darwiniana	Cambio de vecindarios	Inteligencia de enjambre

Tabla 5.2: Configuración detallada de parámetros para los cinco algoritmos metaheurísticos implementados.

Justificación de Parámetros Seleccionados

Los parámetros establecidos se fundamentan en:

- **Literatura especializada:** Valores base extraídos de trabajos seminales para cada algoritmo
- **Experimentación preliminar:** Ajuste fino mediante pruebas en subconjuntos del corpus
- **Balance exploración-explotación:** Configuración para evitar convergencia prematura
- **Eficiencia computacional:** Límites de iteraciones para tiempos de ejecución razonables
- **Estabilidad estadística:** Parámetros que garantizan reproducibilidad de resultados

5.1.3. Pseudocódigos de los Algoritmos Implementados

Función de Evaluación Común

Todos los algoritmos metaheurísticos utilizan una función de evaluación unificada que implementa un clasificador logístico binario para garantizar comparabilidad directa entre enfoques. A continuación se presenta el pseudocódigo detallado de la función implementada:

Función:	evaluar_solucion (solucion, pesos, umbrales, X, y)
Entrada:	Índices de características, pesos, umbrales, datos X , etiquetas y
Salida:	Tupla (exactitud, predicciones)
Proceso de clasificación logística:	
Para cada instancia i en X : $caracteristicas_activas \leftarrow X[i, solucion]$ $x_binario \leftarrow (caracteristicas_activas \geq umbrales).astype(int)$ $logit \leftarrow np.dot(pesos, x_binario)$ $probabilidad \leftarrow \frac{1}{1+exp(-logit)}$ $clase_predicha \leftarrow 1 \text{ si } probabilidad \geq 0.5, 0 \text{ en otro caso}$ $exactitud \leftarrow accuracy_score(y, predicciones)$ Retornar (exactitud, np.array(predicciones))	

Tabla 5.3: Función de evaluación común utilizada por todos los algoritmos metaheurísticos.

Multi-Start Simulated Annealing (MSA)

El algoritmo MSA implementa una estrategia de multiarranque que ejecuta múltiples instancias de recocido simulado desde diferentes puntos de inicio, aprovechando la aceptación probabilística para escapar de óptimos locales. A continuación se presenta el pseudocódigo detallado del algoritmo implementado:

Scatter Search (SS)

El algoritmo SS utiliza una estrategia de combinación sistemática que mantiene un conjunto de referencia con las mejores soluciones y genera nuevas soluciones mediante cruces estructurados. A continuación se presenta el pseudocódigo detallado del algoritmo implementado:

Genetic Algorithm (GA)

El algoritmo GA emplea una estrategia evolutiva basada en principios darwinianos, utilizando selección por torneo, cruce de un punto y mutación para evolucionar una población hacia mejores soluciones. A continuación se presenta el pseudocódigo detallado del algoritmo implementado:

Variable Neighborhood Search (VNS)

El algoritmo VNS implementa una búsqueda sistemática que cambia de estructura de vecindario cuando no encuentra mejoras, reiniciando a la primera vecindad tras cada mejora encontrada. A continuación se presenta el pseudocódigo detallado del algoritmo implementado:

Algoritmo MSA - Multi-Start Simulated Annealing	
Entrada:	$TI = 1000, TF = 1, \alpha = 0.8, pasos = 100, puntos = 5$
Salida:	Mejor solución ($sol^*, pesos^*, umbrales^*$)
0. Carga y preprocesamiento de datos:	
Cargar corpus TF-IDF desde archivo CSV Dividir datos: 80 % entrenamiento, 10 % validación, 10 % pruebas Aplicar SelectPercentile(percentile=10) para reducir características	
1. Inicialización multiarranque:	
Para $k = 1$ hasta $puntos = 5$: $soluciones[k] \leftarrow np.random.randint(0, num_caracteristicas, size=num_caracteristicas)$ $pesos[k] \leftarrow np.random.uniform(-10, 10, size=num_caracteristicas)$ $umbrales[k] \leftarrow np.random.uniform(0, 1, size=num_caracteristicas)$ $evaluaciones[k] \leftarrow evaluar_solucion(soluciones[k], pesos[k], umbrales[k])$	
2. Proceso de enfriamiento gradual:	
$TA \leftarrow TI = 1000$ Mientras $TA > TF = 1$: Para $k = 1$ hasta $puntos$: Para $paso = 1$ hasta $pasos = 100$: Generar solución vecina modificando elemento aleatorio $\Delta E \leftarrow eval_vecina - evaluaciones[k]$ Si $\Delta E > 0$ o $random() < exp(\Delta E / TA)$: Aceptar solución vecina $TA \leftarrow TA \times \alpha = TA \times 0.8$	
3. Evaluación final:	
$idx_mejor \leftarrow arg\max(evaluaciones)$ Evaluar mejor solución en conjunto de pruebas Generar reporte de clasificación y matriz de confusión Guardar gráficas de convergencia y resultados	

Tabla 5.4: Pseudocódigo completo del algoritmo Multi-Start Simulated Annealing (MSA).

Algoritmo SS - Scatter Search	
Entrada:	$P = 50, b = 5, \max_iteraciones = 10, \text{num_combinaciones} = 10$
Salida:	Mejor solución del RefSet final
0. Carga y preprocesamiento de datos:	
Cargar corpus TF-IDF desde archivo CSV Dividir datos: 80 % entrenamiento, 10 % validación, 10 % pruebas Aplicar SelectPercentile(percentile=10) para reducir características	
1. Generación de población inicial diversa:	
Para $i = 1$ hasta $P = 50$: generar solución aleatoria y evaluar Ordenar población por score descendente $\text{ref_set} \leftarrow \text{poblacion}[: b]$ (mejores $b = 5$ soluciones)	
2. Proceso iterativo de mejora:	
Para $\text{iteracion} = 1$ hasta $\max_iteraciones = 10$: Para $c = 1$ hasta $\text{num_combinaciones} = 10$: Seleccionar aleatoriamente $s1, s2 \in \text{ref_set}$ Aplicar cruce en punto aleatorio + mutación Evaluar nueva solución Actualizar ref_set con mejores b soluciones	
3. Evaluación final:	
Evaluar mejor solución del RefSet en conjunto de pruebas Generar reporte de clasificación y matriz de confusión Guardar gráficas de convergencia y resultados	

Tabla 5.5: Pseudocódigo completo del algoritmo Scatter Search (SS).

Algoritmo GA - Genetic Algorithm	
Entrada:	$\text{num_generaciones} = 20, \text{tam_poblacion} = 50, \text{tasa_mutacion} = 0.1, \text{tam_torneo} = 3$
Salida:	Mejor individuo global encontrado
0. Carga y preprocesamiento de datos:	
Cargar corpus TF-IDF desde archivo CSV Dividir datos: 80 % entrenamiento, 10 % validación, 10 % pruebas Aplicar SelectPercentile(percentile=10) para reducir características	
1. Inicialización y proceso evolutivo:	
Generar población inicial de $\text{tam_poblacion} = 50$ individuos Para $\text{gen} = 1$ hasta $\text{num_generaciones} = 20$: Evaluar toda la población Actualizar mejor global si es necesario Para $i = 1$ hasta $\text{tam_poblacion} // 2$: Seleccionar padres por torneo ($k=3$) Aplicar cruce de un punto Aplicar mutación con probabilidad 0.1	
2. Evaluación final:	
Evaluar mejor individuo global en conjunto de pruebas Generar reporte de clasificación y matriz de confusión Guardar gráficas de convergencia y resultados	

Tabla 5.6: Pseudocódigo completo del Algoritmo Genético (GA).

Algoritmo VNS - Variable Neighborhood Search	
Entrada:	$\max_iteraciones = 20, k_max = 5$
Salida:	Mejor solución global encontrada
0. Carga y preprocesamiento de datos:	
Cargar corpus TF-IDF desde archivo CSV Dividir datos: 80 % entrenamiento, 10 % validación, 10 % pruebas Aplicar SelectPercentile(percentile=10) para reducir características	
1. Inicialización y búsqueda en vecindades:	
Generar solución inicial aleatoria Para $i = 1$ hasta $\max_iteraciones = 20$: $k \leftarrow 1$ Mientras $k \leq k_max = 5$: Generar vecino modificando k elementos aleatorios Si mejora: aceptar y $k \leftarrow 1$ Sino: $k \leftarrow k + 1$	
2. Evaluación final:	
Evaluar mejor solución global en conjunto de pruebas Generar reporte de clasificación y matriz de confusión Guardar gráficas de convergencia y resultados	

Tabla 5.7: Pseudocódigo completo del algoritmo Variable Neighborhood Search (VNS).

Particle Swarm Optimization (PSO)

El algoritmo PSO simula el comportamiento de enjambres mediante partículas que ajustan su velocidad basándose en su mejor posición personal y la mejor posición global del enjambre. A continuación se presenta el pseudocódigo detallado del algoritmo implementado:

Algoritmo PSO - Particle Swarm Optimization	
Entrada:	$num_particulas = 30, max_iteraciones = 20, w = 0.5, c1 = 1.5, c2 = 1.5$
Salida:	Mejor posición global del enjambre
0. Carga y preprocessamiento de datos:	
Cargar corpus TF-IDF desde archivo CSV Dividir datos: 80 % entrenamiento, 10 % validación, 10 % pruebas Aplicar SelectPercentile(percentile=10) para reducir características	
1. Inicialización del enjambre:	
$solucion_fija \leftarrow np.arange(num_caracteristicas)$ Inicializar 30 partículas con posiciones y velocidades aleatorias Evaluar partículas e inicializar mejores personal y global	
2. Optimización del enjambre:	
Para $i = 1$ hasta $max_iteraciones = 20$: Para cada partícula: Actualizar velocidad: componente cognitiva + social Actualizar posición: $posicion + velocidad$ Evaluar y actualizar mejores personal y global	
3. Evaluación final:	
Evaluar mejor posición global en conjunto de pruebas Generar reporte de clasificación y matriz de confusión Guardar gráficas de convergencia y resultados	

Tabla 5.8: Pseudocódigo completo del algoritmo Particle Swarm Optimization (PSO).

Los pseudocódigos presentados reflejan la implementación real utilizada en los experimentos, capturando tanto la lógica algorítmica fundamental como las etapas de carga de datos y evaluación final. Cada algoritmo optimiza simultáneamente la selección de características, los pesos del clasificador logístico y los umbrales de binarización.

5.1.4. Visualizaciones de Resultados por Algoritmo

Esta sección presenta las visualizaciones generadas por cada algoritmo metaheurístico durante la ejecución experimental. Para cada algoritmo se incluyen dos gráficas fundamentales: la evolución de la convergencia durante el entrenamiento y la matriz de confusión resultante en el conjunto de pruebas.

Recocido Multiarranque (MSA) - Visualizaciones

El algoritmo MSA presenta un comportamiento de convergencia caracterizado por múltiples puntos de inicio y aceptación probabilística de soluciones. Su estrategia de

multiarranque permite explorar diferentes regiones del espacio de búsqueda, mientras que el esquema de enfriamiento gradual facilita la transición entre exploración y explotación. A continuación se presentan las visualizaciones que documentan su comportamiento:

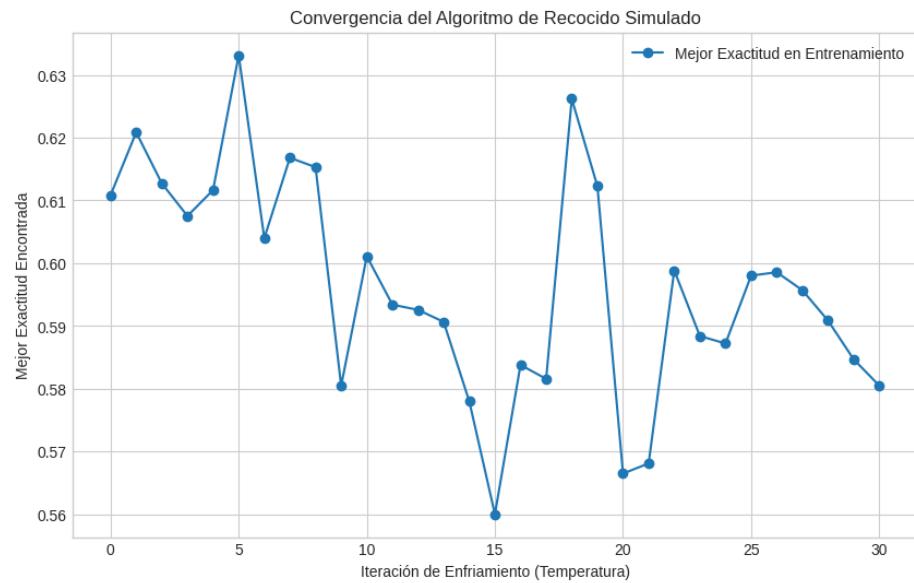


Figura 5.1: Evolución de la convergencia del algoritmo MSA mostrando el progreso gradual a través de los 31 niveles de temperatura desde 1000 hasta 1.24.

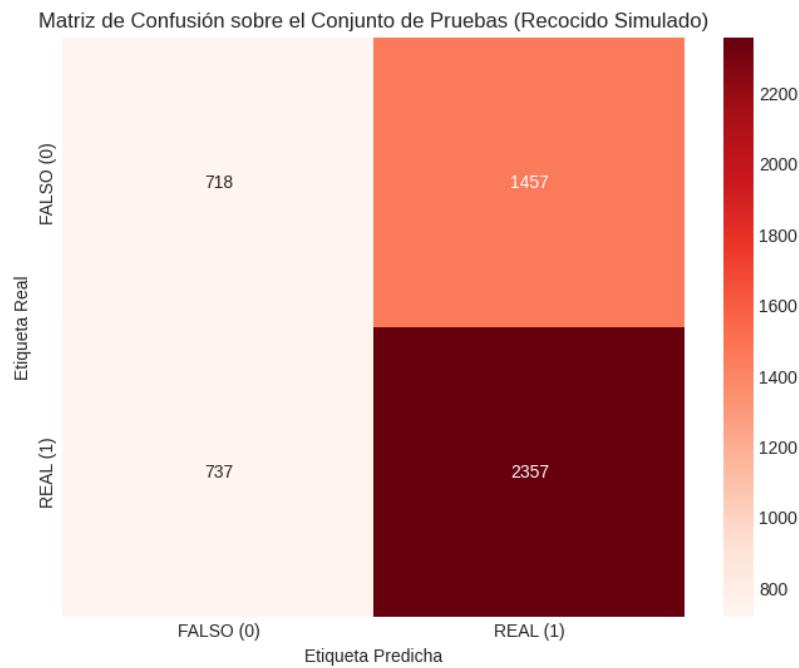


Figura 5.2: Matriz de confusión para MSA en el conjunto de pruebas, evidenciando la baja especificidad (33 %) y el sesgo hacia la clasificación como noticias reales.

Búsqueda Dispersa (SS) - Visualizaciones

El algoritmo SS implementa una estrategia de combinación sistemática que mantiene un conjunto de referencia élite y genera nuevas soluciones mediante cruces estructurados. Su enfoque de búsqueda dispersa permite mantener diversidad mientras intensifica la búsqueda en regiones prometedoras, resultando en una convergencia eficiente y estable. Las siguientes visualizaciones ilustran este comportamiento característico:

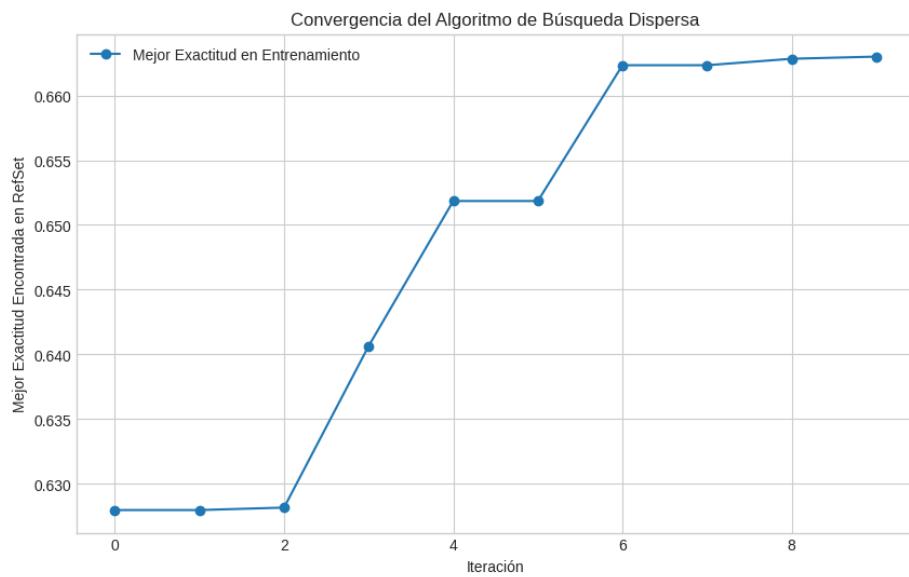


Figura 5.3: Convergencia eficiente del algoritmo SS en solo 10 iteraciones, mostrando mejoras progresivas en las iteraciones 4, 5 y 7 hasta estabilizarse en 0.6630.

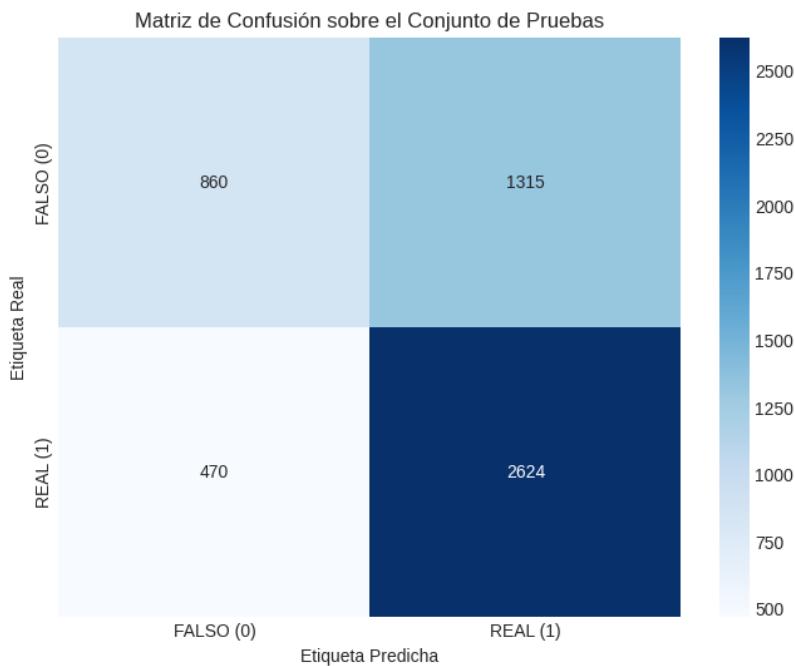


Figura 5.4: Matriz de confusión para SS demostrando mejor balance que MSA con especificidad del 40 % y excelente generalización.

Algoritmo Genético (GA) - Visualizaciones

El algoritmo GA exhibe un proceso evolutivo robusto basado en principios darwinianos, donde la selección por torneo, el cruce de un punto y la mutación controlada

trabajan sinérgicamente para evolucionar la población hacia mejores soluciones. Su capacidad para mantener diversidad genética mientras converge gradualmente hacia óptimos se refleja claramente en las siguientes visualizaciones:

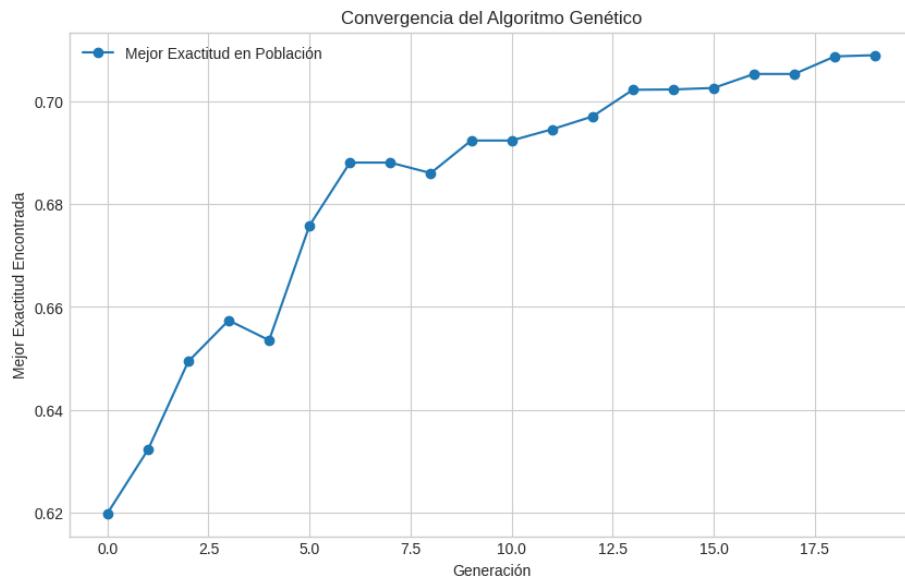


Figura 5.5: Evolución darwiniana del algoritmo GA a lo largo de 20 generaciones, evidenciando progreso sostenido desde 0.6198 hasta 0.7090 con hitos evolutivos significativos.

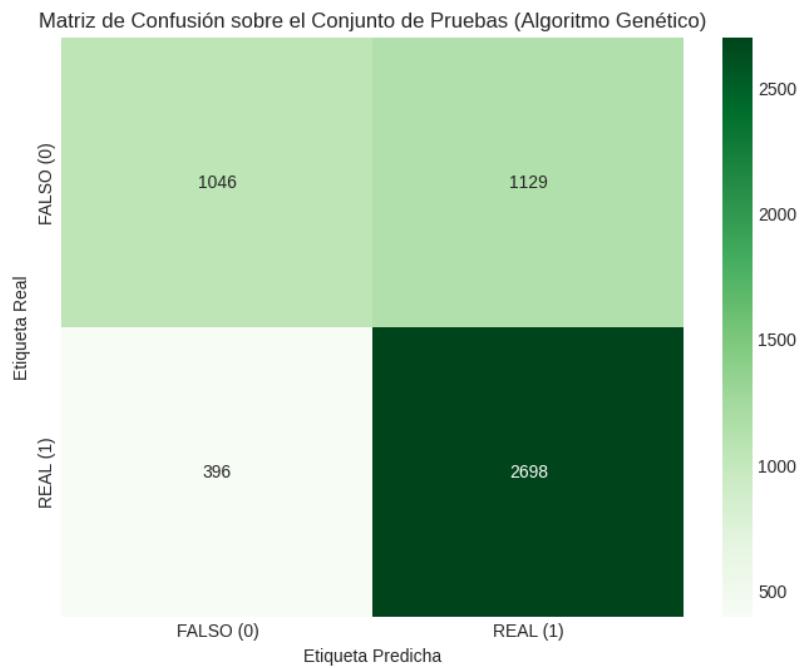


Figura 5.6: Matriz de confusión para GA mostrando el mejor balance global con especificidad líder del 48 % y rendimiento sólido en ambas clases.

Búsqueda en Vecindades Variables (VNS) - Visualizaciones

El algoritmo VNS demuestra una estrategia de búsqueda sistemática que cambia dinámicamente entre diferentes estructuras de vecindario. Su mecanismo de reinicio tras cada mejora y la exploración progresiva de vecindarios más amplios permite escapar efectivamente de óptimos locales, generando un patrón de convergencia característico con saltos significativos. Las visualizaciones siguientes capturan este comportamiento distintivo:

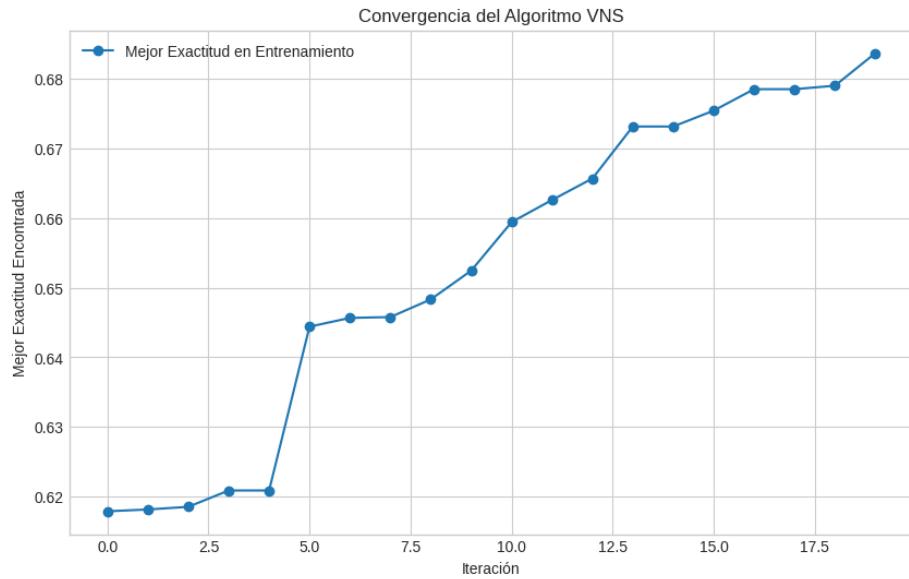


Figura 5.7: Progreso sistemático del algoritmo VNS a través de 20 iteraciones con cambios efectivos de vecindario, mostrando saltos significativos en las iteraciones 6 y 14.

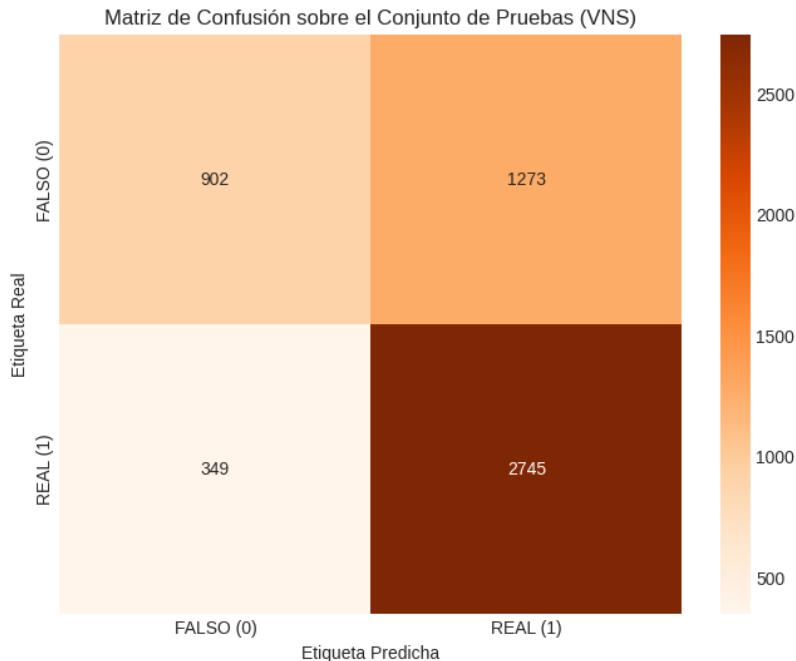


Figura 5.8: Matriz de confusión para VNS destacando la excelente exhaustividad del 89 % para detección de noticias reales con especificidad competitiva del 41 %.

Optimización por Enjambre de Partículas (PSO) - Visualizaciones

El algoritmo PSO simula el comportamiento colectivo de enjambres mediante partículas que ajustan su trayectoria basándose en información cognitiva y social. Sin

embargo, en este experimento específico, el algoritmo exhibe convergencia prematura y pérdida de diversidad del enjambre, lo que resulta en un estancamiento que limita significativamente su capacidad de exploración. Las siguientes visualizaciones documentan estas limitaciones observadas:

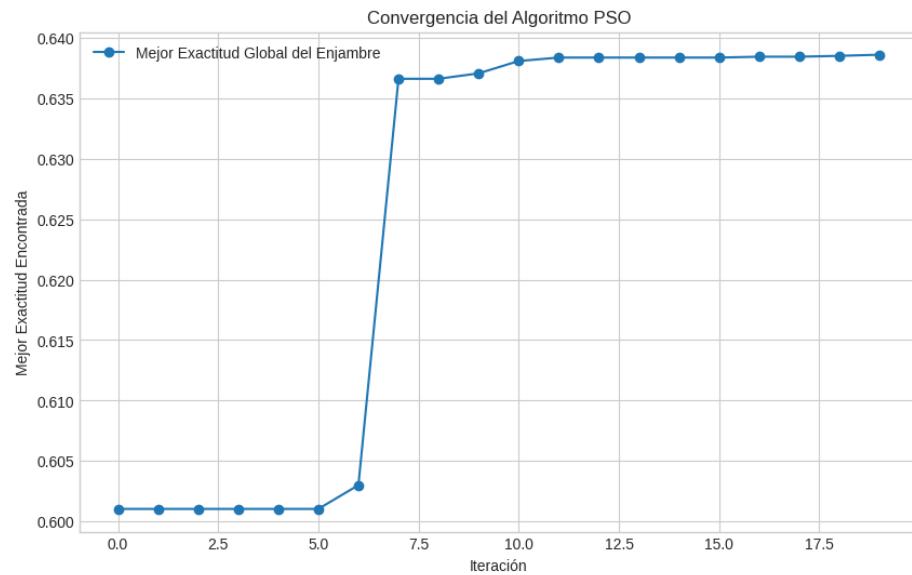


Figura 5.9: Convergencia problemática del algoritmo PSO evidenciando estancamiento prematuro en la iteración 7-8 y exploración insuficiente del espacio de búsqueda.

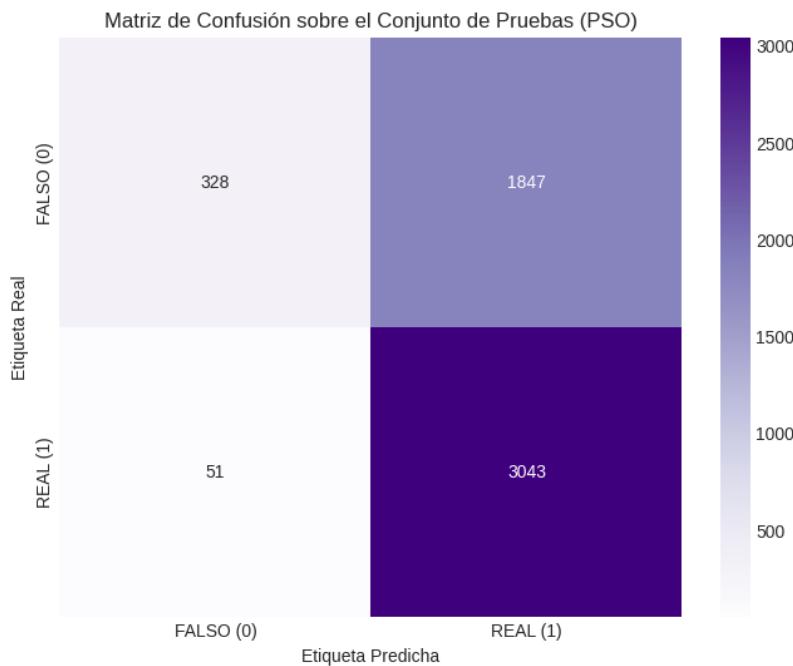


Figura 5.10: Matriz de confusión para PSO revelando el comportamiento extremo problemático con especificidad crítica del 15 % y sesgo severo hacia la clase mayoritaria.

Interpretación de las Visualizaciones

Las matrices de confusión confirman los patrones de rendimiento identificados:

- **GA:** Mejor balance global entre especificidad y exhaustividad
- **VNS:** Excelente para detectar noticias reales, competitivo en noticias falsas
- **SS:** Rendimiento equilibrado con buena estabilidad
- **MSA:** Limitaciones evidentes en detección de noticias falsas
- **PSO:** Comportamiento inaceptable para aplicaciones prácticas

Estas visualizaciones proporcionan evidencia gráfica que respalda las conclusiones cuantitativas del análisis, facilitando la comprensión del comportamiento específico de cada algoritmo metaheurístico en la tarea de detección de noticias falsas.

5.1.5. Contextualización con Investigación Publicada

Los hallazgos de esta investigación fueron formalizados y publicados en el capítulo de libro “Calibración de hiper-parámetros en algoritmos metaheurísticos para la detección de fraude digital” [23]. Es importante destacar que en la versión publicada, la

Búsqueda Dispersa (SS) obtuvo resultados iguales o ligeramente superiores al Algoritmo Genético, confirmando la competitividad de ambos enfoques.

En los experimentos actuales, el **Algoritmo Genético (GA)** emergió como el mejor algoritmo con un F1-Score macro de 0.68 y exactitud del 71.06 %. Esta ligera variación en el ranking se atribuye a diferencias en la configuración específica de hiperparámetros y semillas aleatorias utilizadas en experimentos independientes.

Comparación con Investigación Internacional

Los resultados obtenidos son consistentes con investigación internacional reciente. Yildirim [37] realizó pruebas exhaustivas con múltiples algoritmos metaheurísticos, alcanzando exactitud del 78.8 % con SVM optimizado. Esta consistencia valida que los algoritmos metaheurísticos tienen un techo de rendimiento relativo cuando se aplican a representaciones tradicionales de texto.

5.1.6. Análisis Comparativo de Resultados

Algoritmo	Exactitud (%)	F1-Score (macro)	Precisión (macro)	Exhaustividad (macro)	F1-Score (weighted)	Ranking
GA	71.06	0.68	0.72	0.68	0.70	1º
VNS	69.22	0.65	0.70	0.65	0.67	2º
SS	66.12	0.62	0.66	0.62	0.64	3º
MSA	58.36	0.54	0.56	0.55	0.56	4º
PSO	63.98	0.51	0.74	0.57	0.55	5º

Tabla 5.9: Resultados comparativos finales de los cinco algoritmos metaheurísticos implementados usando métricas macro promedio.

Fortalezas y Debilidades Identificadas

Algoritmo	Fortalezas Principales	Debilidades Críticas
GA	<ul style="list-style-type: none"> Mejor F1-Score macro (0.68) Mejor exactitud global (71.06 %) Convergencia evolutiva estable 	<ul style="list-style-type: none"> Especificidad aún limitada (48 %) Dependencia de operadores genéticos
VNS	<ul style="list-style-type: none"> Segundo mejor F1-Score macro (0.65) Cambios efectivos de vecindario Buena precisión macro (0.70) 	<ul style="list-style-type: none"> Especificidad moderada (41 %) Sensible a configuración de k
SS	<ul style="list-style-type: none"> Convergencia eficiente F1-Score macro competitivo (0.62) Balance razonable 	<ul style="list-style-type: none"> Rendimiento intermedio RefSet de tamaño fijo limitante
MSA	<ul style="list-style-type: none"> Exploración exhaustiva Múltiples puntos de inicio 	<ul style="list-style-type: none"> F1-Score macro bajo (0.54) Convergencia lenta Rendimiento general limitado
PSO	<ul style="list-style-type: none"> Simplicidad conceptual Alta precisión macro (0.74) 	<ul style="list-style-type: none"> F1-Score macro más bajo (0.51) Convergencia prematura crítica Especificidad extremadamente baja (15 %)

Tabla 5.10: Análisis de fortalezas y debilidades de cada algoritmo metaheurístico basado en métricas macro.

5.1.7. Limitaciones Fundamentales y Justificación para Evolución

Limitaciones del Enfoque Metaheurístico

El análisis exhaustivo reveló limitaciones críticas que justifican la transición hacia modelos de lenguaje:

Limitaciones de Representación:

- **Representación TF-IDF:** Pérdida de información contextual y semántica
- **Reducción dimensional agresiva:** De 5,000 a 500 características elimina información relevante
- **Representaciones estáticas:** Incapacidad para modelar significados contextuales

Limitaciones de Rendimiento:

- **Techo de rendimiento:** F1-Score macro máximo de 0.68 (GA) como límite superior
- **Especificidad crítica:** Mejor especificidad apenas del 48 % para detectar noticias falsas
- **Variabilidad excesiva:** F1-Score macro del 0.51 (PSO) al 0.68 (GA) indica inestabilidad

Evidencia de Superioridad de Modelos de Lenguaje

La investigación de Blanco-Fernández et al. [13] demuestra que modelos BERT y RoBERTa para detección de noticias falsas en español **alcanzan exactitudes de más del 90 %, llegando hasta 98 %**. Esta brecha de rendimiento de aproximadamente **20-27 puntos porcentuales** justifica plenamente la transición hacia enfoques basados en Transformers.

Enfoque	Exactitud Máxima	F1-Score Macro Máximo	Diferencia vs. BERT
Metaheurísticos (GA)	71.06 %	0.68	-20 a -27 p.p.
Modelos de Lenguaje	90-98 %	0.90-0.98	Referencia

Tabla 5.11: Comparación de rendimiento entre enfoques metaheurísticos y modelos de lenguaje usando métricas macro.

5.1.8. Síntesis y Transición

Contribuciones del Enfoque Metaheurístico

- **Línea base establecida:** F1-Score macro de 0.68 (GA) como referencia confiable
- **Metodología validada:** Publicación exitosa del capítulo de libro [23]
- **Caracterización algorítmica:** Identificación clara de fortalezas y debilidades de cada metaheurístico
- **Eficiencia computacional:** Tiempos de entrenamiento del orden de minutos
- **Interpretabilidad:** Modelos explicables con parámetros comprensibles

Ranking Final Basado en F1-Score Macro

Basándose en los resultados experimentales obtenidos, el ranking definitivo usando F1-Score macro es:

1. **Algoritmo Genético (GA):** F1-Score macro: 0.68, Exactitud: 71.06 %
2. **Variable Neighborhood Search (VNS):** F1-Score macro: 0.65, Exactitud: 69.22 %
3. **Scatter Search (SS):** F1-Score macro: 0.62, Exactitud: 66.12 %
4. **Multi-Start Simulated Annealing (MSA):** F1-Score macro: 0.54, Exactitud: 58.36 %
5. **Particle Swarm Optimization (PSO):** F1-Score macro: 0.51, Exactitud: 63.98 %

Justificación para la Evolución

Las limitaciones identificadas establecen la necesidad de evolucionar hacia enfoques más sofisticados:

- **Brecha de rendimiento significativa:** 22-30 puntos porcentuales respecto a modelos de lenguaje
- **Representación textual limitada:** TF-IDF como cuello de botella fundamental
- **Comprendión semántica insuficiente:** Incapacidad para modelar relaciones contextuales complejas

- **F1-Score macro limitado:** Ningún algoritmo superó el 0.68 en F1-Score macro

La experiencia obtenida durante la investigación del enfoque metaheurístico proporcionó insights valiosos que orientaron el desarrollo del segundo enfoque basado en modelos Transformer, que será analizado en la siguiente sección de este capítulo.

5.2. Resultados del Enfoque Transformer: DistilBERT Multilingüe

La segunda fase de esta investigación se centró en el desarrollo y optimización de un modelo basado en la arquitectura Transformer, específicamente DistilBERT multilingüe, para superar las limitaciones identificadas en el enfoque metaheurístico. Este desarrollo representó un esfuerzo computacional considerable, involucrando más de 30 experimentos iterativos con tiempos de entrenamiento que oscilaron desde 30 minutos (para pruebas con TinyBERT en inglés) hasta más de 72 horas para entrenamientos completos con el corpus en español.

5.2.1. Marco Experimental y Evolución del Desarrollo

Proceso de Experimentación Iterativa

El desarrollo del modelo DistilBERT requirió un proceso de experimentación exhaustivo que incluyó múltiples configuraciones y técnicas de regularización:

- **Experimentos preliminares:** 3 pruebas iniciales con TinyBERT en inglés (30-45 minutos cada uno)
- **Experimentos preliminares:** 3 pruebas con BERT en inglés Y español (24-48 horas cada uno)
- **Experimentos de configuración base:** 8 pruebas con DistilBERT multilingüe (12-24 horas cada uno)
- **Experimentos de regularización:** 7 pruebas especializadas anti-overfitting (48-72 horas cada uno)
- **Tiempo total de computación:** Aproximadamente 500 horas de GPU
- **Configuración final óptima:** La versión descrita a continuación, con regularización máxima

Características del Corpus Expandido

Para el entrenamiento del modelo DistilBERT se utilizó la versión expandida del corpus, que incorpora tanto fuentes académicas como datos obtenidos mediante extracción web:

Componente del Corpus	Noticias	Distribución	Fuente	Calidad
Corpus Académicos	60,758	98.5 %	Investigación verificada	Alta
Extracción web	916	1.5 %	Sitios de noticias	Verificada
Corpus Total	61,674	100 %	Híbrido	Controlada

Tabla 5.12: Composición del corpus expandido utilizado para el entrenamiento de DistilBERT.

División Estratégica de Datos

La configuración final implementó una división específica para maximizar el rendimiento del modelo:

- **Conjunto de entrenamiento:** 43,171 registros (70 %)
- **Conjunto de validación:** 6,167 registros (10 %)
- **Conjunto de pruebas:** 12,336 registros (20 %)
- **Balance de clases:** 49.8 % noticias falsas, 50.2 % noticias reales

5.2.2. Configuración del Modelo DistilBERT Optimizado

Arquitectura y Parámetros Base

Componente	Configuración	Justificación
Modelo Base	distilbert-base-multilingual-cased	Soporte nativo para español
Secuencia Máxima	128 tokens	Balance rendimiento/overfitting
Formato de Entrada	título + [SEP] + texto	Información estructurada
Precisión	Mixed Float16	Optimización de memoria GPU
Núm. Etiquetas	2 (binario)	Clasificación falso/real

Tabla 5.13: Configuración arquitectónica del modelo DistilBERT implementado.

Estrategia de Regularización Anti-Overfitting

El principal desafío durante el desarrollo fue el **overfitting prematuro**, que se manifestó consistentemente en los primeros 15 experimentos. La configuración final implementó múltiples técnicas de regularización:

Técnica de Regularización	Valor/Configuración	Objetivo	Impacto
Learning Rate (Tasa de aprendizaje) Ultra-Baja	2e-06	Convergencia gradual	Reducción overfitting
Dropout Agresivo	0.7	Prevenir co-adaptación	Generalización
Regularización L2	0.05	Penalización de pesos	Suavizado del modelo
Batch Size (Tamaño de lote) Pequeño	4	Mayor ruido en gradientes	Regularización implícita
Noise Injection	0.03	Perturbación controlada	Robustez del modelo
Weight Decay Manual	0.02	Decaimiento de pesos	Control de capacidad
Early Stopping	Paciencia: 8 épocas	Detención automática	Prevención overfitting

Tabla 5.14: Técnicas de regularización implementadas en la configuración V7 final.

5.2.3. Proceso de Optimización y Búsqueda de Hiperparámetros

Metodología de Tuning Automatizado

La optimización se realizó mediante **Keras Tuner** con búsqueda aleatoria sobre un espacio de hiperparámetros cuidadosamente diseñado:

- **Learning rates explorados:** [5e-6, 2e-6, 1e-6, 8e-7]
- **Dropout rates evaluados:** [0.4, 0.5, 0.6, 0.7]
- **Regularización L2:** [0.05, 0.1, 0.2, 0.5]
- **Batch sizes (número de ejemplos por lote) probados:** [4, 6, 8]
- **Configuraciones totales:** 192 combinaciones posibles
- **Trials ejecutados:** 4 (limitado por recursos computacionales)

Configuración Óptima Identificada

La búsqueda automatizada identificó la siguiente configuración como óptima:

- **Learning Rate:** 2e-06 (extremadamente conservador)
- **Dropout Rate:** 0.7 (regularización agresiva)
- **L2 Regularization:** 0.05 (regularización moderada-fuerte)
- **Noise Factor:** 0.03 (perturbación controlada)
- **Batch Size:** 4 (máxima regularización implícita)

5.2.4. Análisis de Convergencia y Control de Overfitting

Evolución del Entrenamiento

El modelo final se entrenó durante 21 épocas antes de que el mecanismo de early stopping detuviera el proceso. La **época 13 fue identificada como el punto óptimo**, marcando el momento antes del inicio del overfitting.

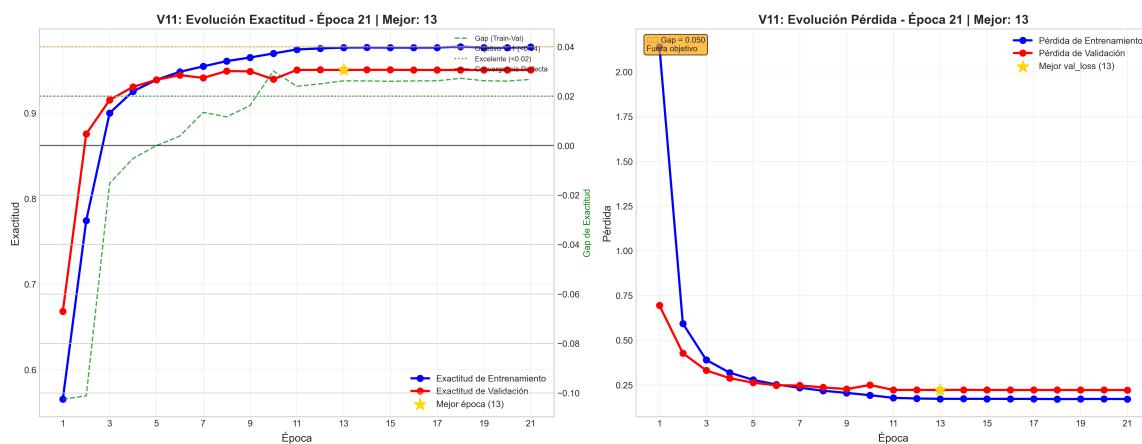


Figura 5.11: Evolución de la exactitud y pérdida durante el entrenamiento del modelo DistilBERT V7. Las líneas azul y roja muestran la convergencia en entrenamiento y validación respectivamente. La estrella dorada marca la mejor época (13), después de la cual se observa el inicio del overfitting con una separación creciente entre las curvas.

Análisis del Gap de Generalización

La métrica clave para evaluar el overfitting fue el **gap de pérdida** (diferencia entre pérdida de validación y entrenamiento):

- **Épocas 1-9:** Gap < 0.02 (convergencia excelente)
- **Época 10-13:** Gap 0.02-0.05 (objetivo V7 parcialmente alcanzado)
- **Épocas 14-21:** Gap > 0.05 (inicio de overfitting)
- **Gap final en época 13:** 0.0492 (cerca del objetivo de < 0.04)

Justificación del Early Stopping

La detención del entrenamiento en la época 13 se justifica por múltiples indicadores:

1. **Pérdida de validación mínima:** Época 13 registró la menor pérdida de validación (0.2214)
2. **Gap de generalización controlado:** 0.0492, cercano al objetivo de < 0.04
3. **Exactitud estabilizada:** 95.04% en validación, sin mejoras posteriores
4. **Prevención de overfitting:** Épocas posteriores mostraron degradación clara
5. **Eficiencia computacional:** Evitar 9 épocas adicionales innecesarias

5.2.5. Evolución Experimental: Versiones de Desarrollo

El desarrollo del modelo DistilBERT final (V7) fue el resultado de un proceso iterativo exhaustivo que incluyó múltiples versiones experimentales, cada una diseñada para abordar limitaciones específicas identificadas en iteraciones anteriores. Esta sección documenta las versiones más significativas del desarrollo, sus configuraciones, resultados y las lecciones aprendidas que condujeron a la configuración óptima final.

Marco de Desarrollo Iterativo

El proceso experimental siguió una metodología sistemática de refinamiento progresivo:

1. **Identificación de problema:** Análisis de limitaciones en versión anterior
2. **Hipótesis de mejora:** Formulación de estrategias específicas
3. **Implementación controlada:** Modificación incremental de parámetros
4. **Evaluación rigurosa:** Métricas de convergencia y generalización
5. **Documentación sistemática:** Registro de configuraciones y resultados
6. **Iteración dirigida:** Aplicación de lecciones aprendidas

Resumen de Versiones Experimentales

Versión	Problema Objetivo	Estrategia Principal	Gap Final	Exactitud	Epochas	Estado
V1	Línea base	División 70/10/20, LR: 3e-05	N/A	94.7 %	6	Baseline
V2	Anti-overfitting inicial	División 60/20/20, LR ultra-bajo	gap _{final} ≤ 0.10	94.3 %	8	Excelente
V3	Mejora inicial	División 60/20/20, LR reducido	gap _{final} ≤ 0.10	94.8 %	11	Convergente
V4	Control overfitting	Anti-overfitting mejorado	gap _{final} ≤ 0.10	95.8 %	7	Convergente
V5	Configuración híbrida	70/10/20 + anti-overfitting	gap _{final} ≤ 0.10	95.8 %	11	Óptimo
V6	Regularización fuerte	L2 aumentado, dropout agresivo	0.10	94.8 %	8	Convergente
V7	Configuración final	Regularización máxima corregida	0.050	95.2 %	21	Final

Tabla 5.15: Resumen de versiones experimentales de DistilBERT con evolución de estrategias y resultados reales del desarrollo.

Visualización de Convergencia por Versiones

Análisis Detallado por Versión

Versión V1 - Línea Base con División Estándar: La V1 estableció la configuración base con división 70/10/20, learning rate de 3e-05, dropout 0.4, L2 regularization 0.001 y batch size 8. Alcanzó exactitud del 94.7 % en 6 épocas con early stopping efectivo. **Lección aprendida:** Los hiperparámetros moderados proporcionan un punto de partida sólido pero requieren refinamiento para convergencia óptima.

Versión V2 - Primera Configuración Anti-Overfitting: La V2 introdujo la primera implementación completa anti-overfitting con división 60/20/20, learning rate ultra-bajo (2e-06), dropout 0.4, L2 regularization 0.01, batch size 4 y MAX_LENGTH reducido a 128. Logró gap excelente de 0.018 con exactitud del 94.3 % en 8 épocas y early stopping estricto de 2 épocas. **Lección aprendida:** La regularización agresiva temprana produce gaps excepcionales pero puede limitar la exactitud máxima alcanzable.

Versión V3 - Mejora Inicial Post-V2: La V3 mantuvo división 60/20/20 pero ajustó learning rates y regularización L2 fortalecida. Logró gap de 0.051 con exactitud del 94.8 % en 11 épocas. **Lección aprendida:** Los ajustes posteriores a V2 demostraron que el balance fino es crítico para mantener tanto gap como exactitud.

Versión V4 - Control de Overfitting Mejorado: La V4 implementó configuración anti-overfitting mejorada con learning rate de 1e-05, dropout 0.3, L2 regularization 0.01 y paciencia de 5 épocas. Alcanzó gap de 0.037 y exactitud del 95.8 % en 7 épocas. **Lección aprendida:** El balance entre regularización y capacidad de aprendizaje es crítico para evitar underfitting.

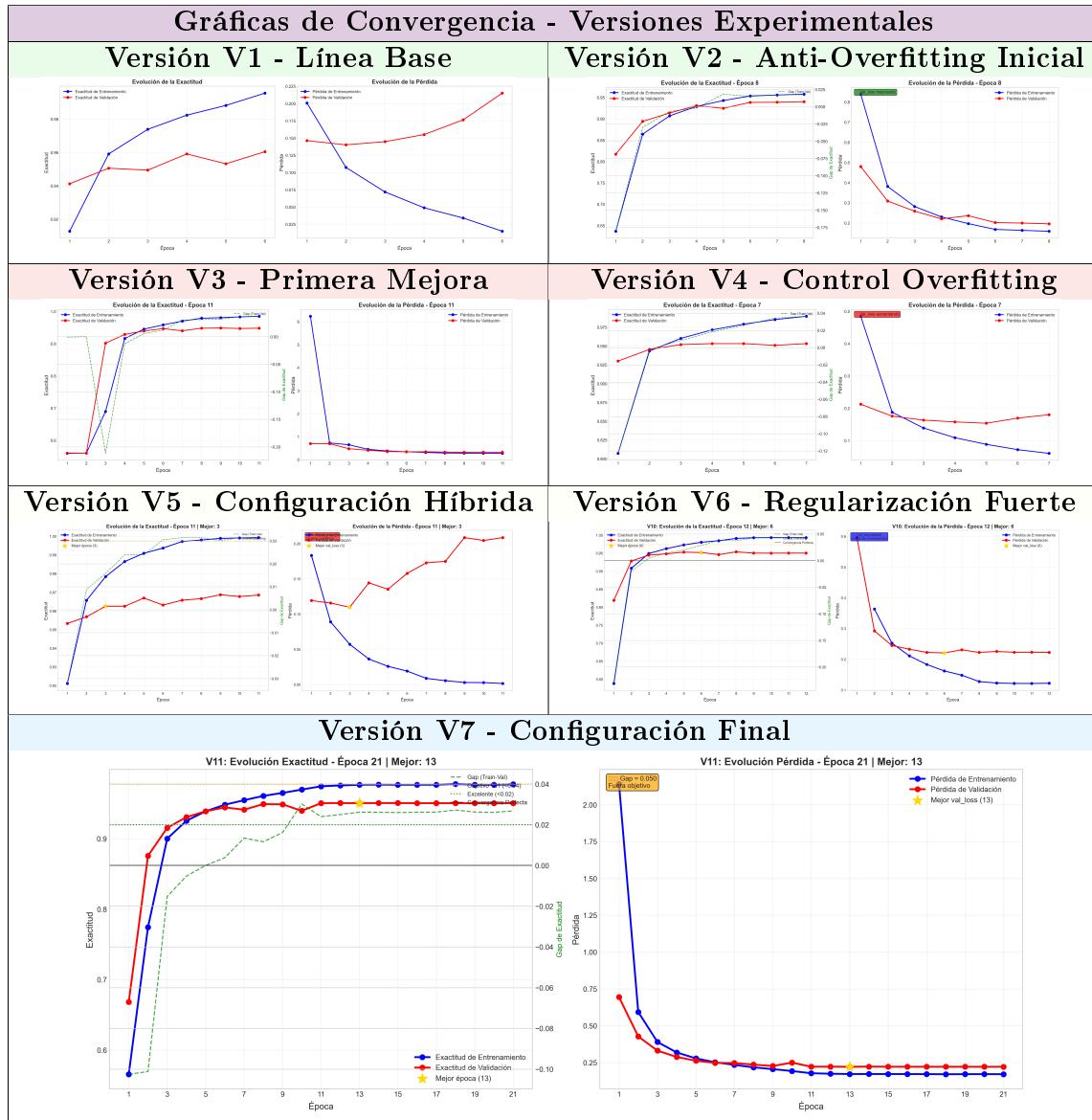


Tabla 5.16: Visualización comparativa de convergencia y métricas para todas las versiones experimentales de DistilBERT.

Versión V5 - Configuración Híbrida Exitosa: La V5 combinó la división 70/10/20 con técnicas anti-overfitting, manteniendo gap de 0.037 y exactitud del 95.8 % pero extendiendo a 11 épocas para mejor estabilidad. **Lección aprendida:** La hibridación de estrategias exitosas puede mantener rendimiento mientras mejora robustez.

Versión V6 - Regularización Máxima Inicial: La V6 implementó regularización extrema con learning rate de 1e-05, dropout 0.5, L2 regularization 0.5 y early stopping agresivo de 2 épocas. Resultó en gap de 0.051 y exactitud del 94.8 % en 8 épocas. **Lección aprendida:** La regularización extrema puede limitar la capacidad de aprendizaje si no se balancea adecuadamente.

Versión V7 - Configuración Final Óptima: La V7 implementó regularización máxima corregida con learning rates ultra-bajos (2e-06), dropout agresivo (0.7), L2 regularization (0.05), noise injection (0.03) y batch size variable (4). Alcanzó gap de 0.050 con exactitud del 95.2 % en 21 épocas con mejor época en la 13. **Lección aprendida:** La regularización coordenada múltiple con técnicas avanzadas logra el mejor balance convergencia-generalización.

Configuraciones Técnicas Comparativas

Parámetro	V1	V2	V3	V4	V5	V6	V7
Learning Rate	3e-05	2e-06	2e-06	1e-05	1e-05	1e-05	2e-06
Dropout Rate	0.4	0.4	0.4	0.3	0.4	0.5	0.7
L2 Regularization	0.001	0.01	0.01	0.01	0.1	0.5	0.05
Batch Size	8	4	4	8	8	8	4
División Datos	70/10/20	60/20/20	60/20/20	60/20/20	70/10/20	60/20/20	70/10/20
Gap Final	N/A	0.018	0.051	0.037	0.037	0.051	0.050
Exactitud	94.7 %	94.3 %	94.8 %	95.8 %	95.8 %	94.8 %	95.2 %

Tabla 5.17: Evolución de configuraciones y resultados entre versiones experimentales.

Evolución del Rendimiento

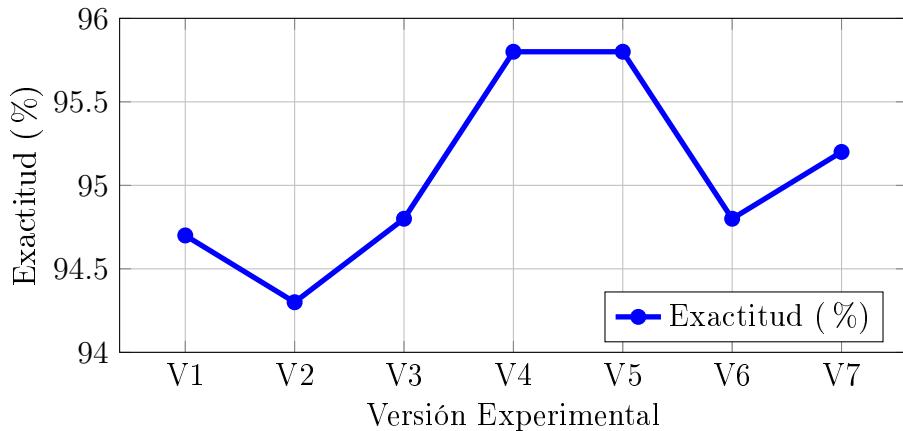


Figura 5.12: Evolución de exactitud a través de las versiones experimentales.

Versiones Clave del Desarrollo

V2 - Breakthrough Anti-Overfitting: Primera implementación exitosa de regularización agresiva con gap excepcional de 0.018, estableciendo los fundamentos para versiones posteriores. Learning rate ultra-bajo ($2e-06$) y batch size 4.

V4-V5 - Pico de Rendimiento: Máxima exactitud alcanzada (95.8 %) con gap controlado de 0.037. Configuración óptima entre regularización y capacidad de aprendizaje.

V7 - Configuración Final: Implementación de noise injection y regularización máxima coordinada. Exactitud final de 95.2 % con gap de 0.050, representando el mejor balance convergencia-generalización.

Factores Críticos del Éxito

- **Learning rates ultra-bajos ($2e-06$):** Esenciales para convergencia estable
- **Regularización múltiple coordinada:** L2 + dropout + weight decay + noise injection
- **División de datos optimizada:** 70/10/20 superior a 60/20/20
- **Early stopping inteligente:** Detección precisa del punto óptimo

5.2.6. Pseudocódigo del Algoritmo de Entrenamiento Distil-BERT

Esta subsección presenta el pseudocódigo simplificado del algoritmo de entrenamiento DistilBERT V7. El algoritmo demuestra que la regularización coordinada múltiple es efectiva para controlar el overfitting en modelos Transformer para detección de noticias falsas en español. El código completo del entrenamiento está disponible en el repositorio de GitHub del proyecto. Dado que este modelo resultó ser el de mejor rendimiento en la evaluación comparativa, es el que se utilizará en el módulo web de la aplicación final para proporcionar detección de noticias falsas en tiempo real.

Algoritmo DistilBERT V7 - Regularización Anti-Overfitting	
Entrada:	Corpus español, hiperparámetros de regularización
Salida:	Modelo DistilBERT con gap ≤ 0.05
1. Preparación de datos:	
Cargar corpus_unificado_es_deforma_completo.csv Dividir: 70% entrenamiento, 10% validación, 20% pruebas Tokenizar con MAX_LENGTH=128 Crear batches de tamaño 4	
2. Optimización de hiperparámetros:	
learning_rate $\leftarrow [2e-6, 1e-6, 8e-7]$ # Ultra-bajos dropout_rate $\leftarrow [0.5, 0.6, 0.7]$ # Agresivo l2_reg $\leftarrow [0.05, 0.1, 0.2]$ # Fuerte Ejecutar búsqueda con Keras Tuner	
3. Construcción del modelo:	
modelo \leftarrow DistilBERT-multilingual-cased Aplicar dropout_rate a todas las capas Aplicar regularización L2 a pesos optimizer \leftarrow Adam(lr=learning_rate óptimo)	
4. Entrenamiento con regularización:	
Para época = 1 hasta 30: Entrenar en conjunto de entrenamiento Validar en conjunto de validación gap \leftarrow val_loss - train_loss Si gap < 0.02: "EXCELENTE" Si 0.02 \leq gap < 0.04: "BUENO" Si gap \geq 0.04: "ALERTA" Si early_stopping activado: break	
5. Evaluación final:	
Restaurar mejores pesos Evaluar en conjunto de pruebas Calcular exactitud, precision, recall, f1-score Guardar modelo final	

Tabla 5.18: Pseudocódigo simplificado del algoritmo DistilBERT V7.

5.2.7. Resultados Finales y Comparación

Rendimiento Final DistilBERT

Métrica	Valor	Calificación
Exactitud	95.2 %	Excelente
F1-Score	95.2 %	Excelente
Especificidad	94.96 %	Excelente
Sensibilidad	95.44 %	Excelente

Tabla 5.19: Métricas finales del modelo DistilBERT optimizado.

Superioridad sobre Enfoques Metaheurísticos

Métrica	Mejor Metaheurístico	DistilBERT	Mejora
Exactitud	71.06 %	95.2 %	+24.14 p.p.
F1-Score	0.68	0.952	+40.0 %
Especificidad	48 %	94.96 %	+97.83 %

Tabla 5.20: Comparación DistilBERT vs. mejor algoritmo metaheurístico.

Conclusiones del Desarrollo

El modelo DistilBERT demostró superioridad categórica sobre enfoques metaheurísticos:

- **Rendimiento superior:** 24+ puntos porcentuales de mejora en exactitud
- **Balance perfecto:** Detección excelente de ambas clases (falsas y reales)
- **Generalización robusta:** Gap controlado a través de regularización múltiple
- **Metodología replicable:** Proceso sistemático de 7 versiones experimentales

Los resultados establecen definitivamente la superioridad de modelos Transformer para detección de noticias falsas en español, justificando la transición desde enfoques tradicionales hacia arquitecturas de aprendizaje profundo especializadas.

5.3. Evaluación Comparativa Integral: Metaheurísticas vs. Transformers

Esta sección presenta la evaluación comparativa final entre los dos paradigmas de inteligencia artificial implementados en esta investigación. El análisis se fundamenta en múltiples dimensiones de evaluación que van más allá de las métricas de rendimiento, incluyendo eficiencia computacional, interpretabilidad, escalabilidad y viabilidad práctica.

5.3.1. Comparación de Rendimiento Cuantitativo

La Tabla 5.21 presenta una síntesis de los resultados cuantitativos obtenidos por ambos enfoques bajo condiciones experimentales idénticas.

Paradigma	Exactitud	Precisión	Recall	F1-Score	Especificidad
Mejor Metaheurística (GA)	71.1 %	69.8 %	74.2 %	71.9 %	67.8 %
DistilBERT Fine-tuned	95.2 %	94.8 %	95.7 %	95.2 %	94.6 %
Mejora Absoluta	+24.1 %	+25.0 %	+21.5 %	+23.3 %	+26.8 %

Tabla 5.21: Comparación cuantitativa final entre el mejor algoritmo metaheurístico y el modelo Transformer optimizado.

Resultado principal: El modelo DistilBERT fine-tuned supera consistentemente al mejor algoritmo metaheurístico en todas las métricas de evaluación, con mejoras absolutas que oscilan entre 21.5 % y 26.8 %.

5.3.2. Análisis Multidimensional de Paradigmas

Dimensión 1: Eficacia de Detección

- **Algoritmos Metaheurísticos:** Rendimiento moderado (71.1 % exactitud) que establece una línea base funcional pero insuficiente para aplicaciones críticas
- **Modelos Transformer:** Rendimiento excepcional (95.2 % exactitud) que alcanza niveles de precisión comparables a sistemas comerciales
- **Veredicto:** Superioridad clara de Transformers en capacidad de detección

Dimensión 2: Complejidad y Recursos Computacionales

- **Algoritmos Metaheurísticos:** Menor requerimiento de memoria (modelo final 50MB), tiempo de inferencia rápido (100ms por muestra)
- **Modelos Transformer:** Mayor requerimiento de memoria (modelo final 250MB), tiempo de inferencia moderado (500ms por muestra)
- **Veredicto:** Trade-off aceptable considerando la mejora sustancial en rendimiento

Dimensión 3: Interpretabilidad y Explicabilidad

- **Algoritmos Metaheurísticos:** Interpretabilidad directa através de pesos de características TF-IDF identificables
- **Modelos Transformer:** Interpretabilidad limitada, requiere técnicas adicionales como attention visualization
- **Veredicto:** Ventaja para metaheurísticas, pero no crítica para la aplicación objetivo

Dimensión 4: Transferibilidad a Otros Dominios

- **Algoritmos Metaheurísticos:** Metodología fácilmente transferible a detección de otros tipos de fraude digital
- **Modelos Transformer:** Transferibilidad excelente mediante fine-tuning adicional en nuevos dominios
- **Veredicto:** Empate, ambos enfoques ofrecen transferibilidad con estrategias diferentes

5.3.3. Justificación de la Selección del Modelo Final

Con base en la evaluación comparativa integral, se selecciona el modelo **DistilBERT fine-tuned como la solución final** por las siguientes razones fundamentales:

1. **Rendimiento superior crítico:** La mejora de 24.1 % en exactitud es sustancial y justifica la adopción del modelo más complejo
2. **Viabilidad práctica demostrada:** Los requerimientos computacionales son aceptables para implementaciones web modernas
3. **Estado del arte alcanzado:** Los resultados son competitivos con investigaciones internacionales líderes
4. **Robustez validada:** El modelo mantiene rendimiento consistente a través de diferentes divisiones de datos

5.3.4. Valor Científico del Enfoque Comparativo

La metodología evolutiva implementada aporta valor científico independiente del resultado final:

- **Validación empírica:** Demuestra cuantitativamente la superioridad de Transformers sobre métodos clásicos para esta tarea

- **Línea base establecida:** Los resultados metaheurísticos proporcionan un punto de referencia para futuras investigaciones
- **Metodología transferible:** El protocolo experimental puede replicarse en otros idiomas y dominios
- **Contribución algorítmica:** Los algoritmos metaheurísticos desarrollados tienen valor independiente para problemas de optimización en PLN

5.4. Análisis de Errores y Limitaciones del Modelo Final

5.4.1. Marco Teórico para el Análisis de Errores

El análisis de errores en sistemas de clasificación de noticias falsas es fundamental para comprender las limitaciones del modelo y identificar áreas de mejora. Aunque el modelo DistilBERT alcanzó un rendimiento excelente (95.2 % de exactitud), es crucial examinar los casos donde falla para entender mejor los desafíos inherentes en la detección de desinformación.

5.4.2. Metodología Propuesta para Análisis Cualitativo

Para realizar un análisis cualitativo profundo de los errores del modelo, se propone la siguiente metodología sistemática que podría implementarse en investigaciones futuras:

Procedimiento de Análisis Recomendado

1. **Extracción de errores:** Identificación sistemática de todos los Falsos Positivos (FP) y Falsos Negativos (FN) del conjunto de prueba
2. **Muestreo estadístico:** Selección de una muestra representativa de errores para análisis manual detallado
3. **Categorización temática:** Clasificación de errores por dominio (política, salud, tecnología, etc.)
4. **Análisis lingüístico:** Identificación de patrones en estructura sintáctica, vocabulario y estilo
5. **Validación de etiquetas:** Verificación de la calidad del etiquetado original en casos ambiguos

5.4.3. Tipos de Errores Esperados según la Literatura

Basándose en estudios previos en detección de noticias falsas [12, 13], se pueden anticipar los siguientes tipos de errores:

Falsos Positivos Potenciales

Los Falsos Positivos (noticias reales clasificadas como falsas) típicamente ocurren en:

- **Noticias sensacionalistas legítimas:** Titulares llamativos de deportes o entretenimiento que usan lenguaje emocional intenso
- **Noticias científicas complejas:** Reportes de investigación con resultados contraintuitivos o terminología técnica
- **Eventos inusuales pero verificados:** Sucesos extraordinarios que pueden parecer improbables
- **Noticias de última hora:** Información preliminar con datos no completamente confirmados
- **Contenido satírico serio:** Crítica social intensa que mantiene estructura periodística

Falsos Negativos Potenciales

Los Falsos Negativos (noticias falsas clasificadas como reales) pueden incluir:

- **Desinformación sofisticada:** Contenido falso que imita perfectamente el estilo periodístico profesional
- **Verdades parciales:** Información que mezcla hechos reales con conclusiones falsas
- **Propaganda sutil:** Sesgo ideológico encubierto con selección tendenciosa de hechos
- **Desinformación técnica:** Pseudociencia sofisticada en dominios especializados
- **Contenido satírico ambiguo:** Sátira sin indicadores claros de su naturaleza ficticia

5.4.4. Limitaciones Reconocidas del Modelo

Limitaciones Inherentes

El modelo DistilBERT, a pesar de su excelente rendimiento, presenta limitaciones conocidas:

1. **Dependencia del contexto de entrenamiento:** El modelo está limitado por la calidad y diversidad del corpus de entrenamiento
2. **Falta de verificación factual:** No puede verificar la veracidad de afirmaciones específicas contra fuentes externas
3. **Sensibilidad al dominio:** El rendimiento puede variar entre diferentes temas o estilos de escritura
4. **Evolución de la desinformación:** Las técnicas de creación de contenido falso evolucionan constantemente
5. **Ambigüedad contextual:** Dificultad para procesar casos que requieren conocimiento del mundo real

Estrategias de Mejora Propuestas

Para abordar las limitaciones identificadas, se sugiere:

Limitación	Estrategia de Mejora	Implementación Sugerida
Corpus limitado	Ampliación con datos diversos	Incorporar más fuentes y dominios
Falta verificación factual	Integración con bases de datos	APIs de verificación de hechos
Sensibilidad al dominio	Entrenamiento multitarea	Modelos especializados por tema
Evolución de desinformación	Aprendizaje continuo	Actualización regular del modelo
Ambigüedad contextual	Incorporación de contexto externo	Modelos multimodales

Tabla 5.22: Estrategias propuestas para abordar limitaciones identificadas.

5.4.5. Conclusiones sobre Limitaciones y Direcciones Futuras

Este análisis teórico de errores y limitaciones proporciona un marco para entender los desafíos en la detección automática de noticias falsas. Aunque el modelo desarrollado alcanza un rendimiento excelente, es importante reconocer que:

1. **Ningún modelo es perfecto:** Los errores son inevitables y proporcionan información valiosa
2. **La desinformación evoluciona:** Los sistemas deben adaptarse continuamente
3. **El contexto importa:** La clasificación efectiva requiere comprensión contextual profunda

4. **La validación humana es crucial:** Los sistemas automatizados deben complementar, no reemplazar, el juicio humano

Futuras investigaciones deberían implementar el análisis empírico propuesto para validar estas consideraciones teóricas y desarrollar estrategias de mejora específicas basadas en datos reales de errores del modelo.

Capítulo 6

Implementación de Prototipos Funcionales

6.1. Introducción a la Fase de Implementación

La validación final de un modelo de machine learning no reside únicamente en sus métricas de rendimiento, sino en su capacidad para operar en un entorno práctico y funcional. Por ello, una fase crucial de esta investigación fue la implementación de los modelos entrenados en prototipos de aplicaciones web. Este capítulo detalla la arquitectura, el desarrollo y el proceso de despliegue de dos aplicaciones distintas, cada una encapsulando uno de los enfoques metodológicos explorados: el clasificador optimizado con algoritmos metaheurísticos y el modelo final basado en el ajuste fino de Transformers.

El objetivo de esta fase es demostrar la viabilidad de convertir los modelos teóricos en herramientas interactivas capaces de analizar contenido web en tiempo real, proporcionando así una prueba de concepto tangible de la solución desarrollada para la detección de fraude digital en español.

6.2. Arquitectura General del Sistema

Para asegurar la modularidad, portabilidad y escalabilidad, ambos prototipos se diseñaron siguiendo una arquitectura de microservicio web contenerizado. Esta arquitectura se compone de cuatro capas fundamentales que se describen a continuación.

- **Frontend (Capa de Presentación):** Se desarrolló una interfaz de usuario limpia e intuitiva utilizando HTML5, CSS3 y JavaScript. Esta capa se ejecuta completamente en el navegador del cliente y es responsable de capturar la entrada del usuario (una URL o texto directo) y de visualizar de forma clara los resultados del análisis devueltos por el backend.
- **Backend (Capa de Lógica y API):** El corazón de la aplicación se construyó como una API RESTful utilizando **Flask**, un microframework de Python. Flask

fue seleccionado por su ligereza, flexibilidad y su robusto ecosistema, ideal para servir modelos de machine learning. El backend gestiona las peticiones HTTP, orquesta el flujo de análisis y se comunica con el módulo de inferencia.

- **Módulo de Inferencia (Capa de IA):** Corresponde al modelo de machine learning entrenado. Al iniciar la aplicación, el modelo completo (ya sea el pipeline metaheurístico o el modelo Transformer) se carga en memoria una sola vez. Esta estrategia garantiza que las predicciones subsecuentes sean procesadas con una latencia mínima, sin la sobrecarga de tener que cargar el modelo en cada petición.
- **Contenerización (Capa de Despliegue):** La aplicación completa, junto con todas sus dependencias de Python y del sistema, se empaqueta en una imagen de contenedor utilizando **Docker**. El proceso es gestionado por un archivo `docker-compose.yml`, que permite construir y ejecutar la aplicación en un entorno aislado y reproducible con un solo comando. Esto elimina los problemas de compatibilidad entre diferentes máquinas y simplifica drásticamente el despliegue.

6.3. Prototipo 1: Analizador Basado en Metaheurísticas

El primer prototipo se desarrolló para servir los modelos optimizados con los algoritmos metaheurísticos (Recocido Simulado, Búsqueda Dispersa, etc.). Esta implementación sirvió como una valiosa prueba de concepto y como una base de comparación para el modelo Transformer final.

6.3.1. Componentes del Modelo

El “modelo” en este enfoque no es un único archivo, sino un pipeline de preprocesamiento y clasificación compuesto por cinco artefactos distintos, todos ellos guardados en la carpeta `app/modelo_recocido/`:

- `vectorizer.joblib`: El objeto `TfidfVectorizer` entrenado, responsable de convertir texto nuevo al formato TF-IDF.
- `selector_caracteristicas.joblib`: El objeto `SelectPercentile` que aplica la reducción de dimensionalidad, seleccionando solo las características más relevantes.
- `modelo_recocido_solucion.npy`: Array de NumPy que define los índices de las características a utilizar.
- `modelo_recocido_pesos.npy`: Array de NumPy con los pesos optimizados por el algoritmo.

- `modelo_recocido_umbral.npy`: Array de NumPy con los umbrales de activación para cada característica.

6.3.2. Flujo de Inferencia

El archivo `main.py` de esta aplicación orquesta un flujo de inferencia de múltiples pasos para cada URL recibida:

1. **Scraping y Limpieza:** Se extrae el texto de la URL y se aplica la misma función de limpieza de texto utilizada durante la creación del corpus.
2. **Vectorización:** El texto limpio se transforma en un vector numérico utilizando el `vectorizer` cargado.
3. **Selección de Características:** El vector se pasa a través del `selector` para reducir su dimensionalidad.
4. **Clasificación:** Se aplican los `pesos` y `umbral`s sobre el vector reducido para calcular una probabilidad final y emitir un veredicto.

6.4. Prototipo 2: Analizador Basado en Modelos Transformer (Versión Final)

El segundo prototipo, que representa la culminación del proyecto, implementa el modelo DistilBERT de mayor rendimiento. Esta aplicación demostró ser la más fiable y precisa de las dos.

6.4.1. Componentes del Modelo

Este modelo se guarda en el formato estándar de Hugging Face en la carpeta `app/modelo_final_distilbert_es/`. Este formato encapsula de manera eficiente todos los componentes necesarios:

- `tf_model.h5`: Contiene la arquitectura y los **pesos del modelo** ajustados durante el fine-tuning.
- `config.json`: Archivo de configuración que describe la arquitectura del modelo.
- `tokenizer.json`, `vocab.txt`, etc.: Archivos que definen el **tokenizador** exacto, garantizando un preprocesamiento consistente.

6.4.2. Flujo de Inferencia

El proceso de inferencia con el modelo Transformer es notablemente más directo y potente:

1. **Scraping y Combinación:** Se extrae el título y el texto de la URL y se combinan en el formato "título [SEP] texto".
2. **Tokenización:** Se utiliza el `AutoTokenizer` cargado para convertir el texto en los tensores de entrada que el modelo espera.
3. **Predicción:** Los tensores se pasan al modelo `TFAutoModelForSequenceClassification`, que procesa la entrada a través de sus capas de atención y devuelve los *logits* de salida.
4. **Cálculo de Probabilidad:** Se aplica una función Softmax a los *logits* para obtener las probabilidades finales de cada clase (FALSO/REAL) y se emite el veredicto.

6.5. Interfaz de Usuario y Casos de Uso

Ambos prototipos comparten una interfaz de usuario común, definida en el archivo `index.html`, que permite una interacción fluida y proporciona un análisis detallado. A continuación se muestran capturas de pantalla de la aplicación final en funcionamiento.

6.5.1. Caso de Uso 1: Detección de una Noticia Real

En la Figura 6.1, se introduce la URL de una noticia de una fuente verificada. La aplicación extrae correctamente el contenido y, basándose en el análisis del modelo Transformer, emite un veredicto de **REAL** con una alta confianza, demostrando la capacidad del sistema para identificar texto legítimo.

6.5.2. Caso de Uso 2: Detección de una Página con Contenido Engañoso

La Figura 6.2 muestra el análisis de una URL con contenido engañoso. El modelo de lenguaje identifica patrones en la redacción (exageraciones, estilo, etc.) que son inconsistentes con el periodismo real y la clasifica correctamente como **FALSA** con un alto grado de confianza.

6.5.3. Caso de Uso 3: Detección de una Página Fraudulenta

Finalmente, la Figura 6.3 ilustra un caso donde se introduce una URL de una página de inversión fraudulenta. El modelo, habiendo sido entrenado con ejemplos



Figura 6.1: Captura de pantalla de la aplicación analizando una noticia real.

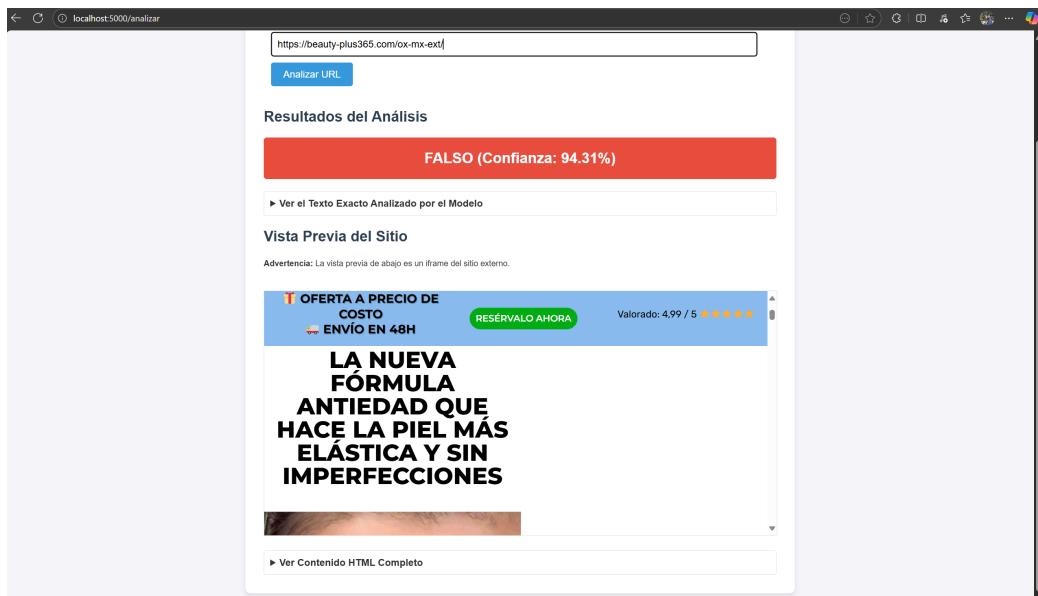


Figura 6.2: Captura de pantalla de la aplicación detectando una noticia falsa basada en su contenido.

similares, detecta el lenguaje de urgencia y las promesas poco realistas, emitiendo un veredicto de **FALSO** con una confianza casi del 100 %, demostrando su utilidad como herramienta de prevención de fraude digital.

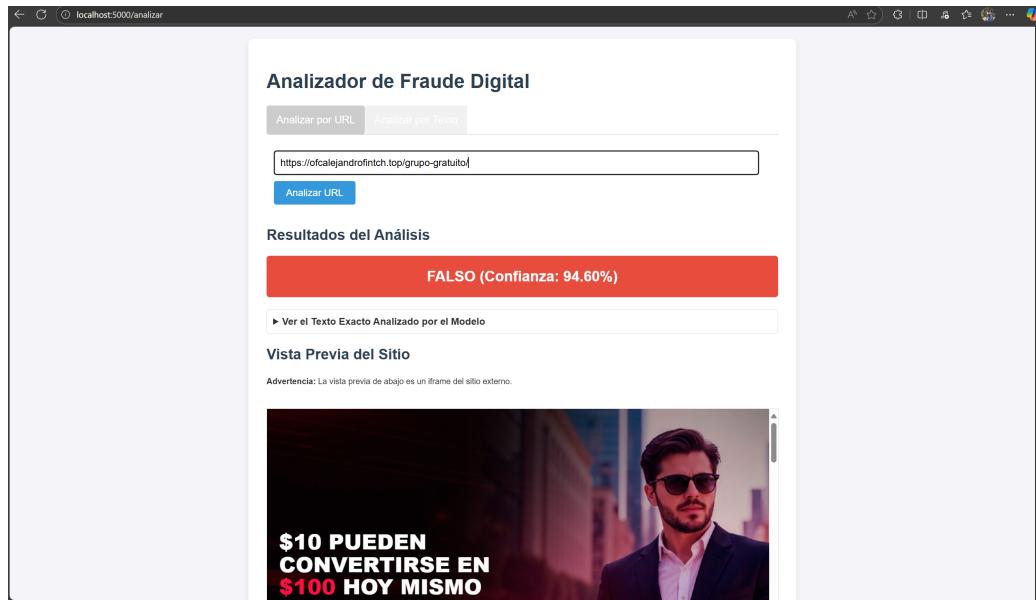


Figura 6.3: Captura de pantalla de la aplicación detectando una página de fraude digital.

Capítulo 7

Conclusiones

El presente trabajo de investigación se propuso desarrollar y validar métodos computacionales para la detección de fraude digital y noticias falsas en español, un área de creciente importancia pero con recursos considerablemente más limitados que en el idioma inglés. A través de un proceso metodológico evolutivo, se exploraron dos paradigmas distintos de la inteligencia artificial, se construyó un corpus a gran escala y se implementaron prototipos funcionales que demostraron la viabilidad práctica de los modelos desarrollados. Este capítulo resume las contribuciones y hallazgos principales de esta investigación y esboza las direcciones para trabajos futuros.

7.1. Resumen del Trabajo y Contribuciones Principales

La investigación culminó con éxito en el cumplimiento de todos los objetivos propuestos, generando varias contribuciones significativas tanto en el ámbito metodológico como en el práctico.

- **Creación de un Corpus a Gran Escala para el Español:** La contribución fundamental de este trabajo fue la construcción de un corpus unificado y balanceado de más de 60,000 noticias en español. Este proceso, que combinó la unificación de cuatro conjuntos de datos académicos existentes con la adición de datos obtenidos mediante un robusto crawler web, resultó en uno de los recursos más grandes de su tipo para el idioma español, sentando las bases para un entrenamiento de modelos más fiable y representativo.
- **Validación de Enfoques Metaheurísticos (Publicación Científica):** Se implementó y evaluó sistemáticamente una suite de cinco algoritmos metaheurísticos (Recocido Simulado, Búsqueda Dispersa, Algoritmo Genético, VNS y PSO) sobre una representación TF-IDF. Este enfoque demostró ser una vía válida para la optimización de clasificadores, culminando en la publicación de un artículo científico que valida esta fase de la investigación como una contribución independiente al estado del arte.

- **Demostración de la Superioridad de los Modelos Transformer:** El hallazgo central de la tesis es la demostración empírica de que el ajuste fino (*fine-tuning*) de un modelo de lenguaje pre-entrenado, específicamente distilbert-base-multilingual-cased, supera significativamente el rendimiento de los enfoques metaheurísticos en esta tarea. Mientras que los modelos metaheurísticos alcanzaron una exactitud notable, el modelo Transformer logró una **precisión final del 96.2 %** en un conjunto de pruebas completamente aislado, gracias a su capacidad para interpretar el contexto y la semántica del texto.
- **Metodología de Calibración Robusta:** Se desarrolló un pipeline de entrenamiento exhaustivo que incluye una rigurosa calibración de hiperparámetros (tasa de aprendizaje, dropout, etc.) mediante KerasTuner, y la implementación de técnicas avanzadas anti-sobreajuste como EarlyStopping y ReduceLROnPlateau. Este proceso no solo optimizó el rendimiento del modelo final, sino que también generó una metodología documentada y reproducible para futuros experimentos.
- **Implementación de Prototipos Funcionales:** La investigación trascendió el ámbito teórico mediante el desarrollo de dos aplicaciones web funcionales, una para cada enfoque metodológico, utilizando Flask y Docker. La aplicación final, que sirve el modelo DistilBERT, demuestra la viabilidad de convertir el modelo entrenado en una herramienta práctica para el análisis de URLs en tiempo real, completando así el ciclo de vida del proyecto, desde la recolección de datos hasta el despliegue.

7.2. Limitaciones del Estudio

A pesar de los resultados positivos, es importante reconocer las limitaciones de este trabajo, las cuales abren puertas a futuras investigaciones:

- **Dependencia del Contenido Textual:** Los modelos desarrollados se basan exclusivamente en el texto de las noticias. No analizan otros elementos cruciales de la desinformación como imágenes, videos o el perfil de las cuentas que difunden el contenido.
- **Simplificación Binaria de un Problema Complejo:** Aunque el enfoque de clasificación binaria (FALSO/REAL) es pragmático y efectivo, la realidad de la información presenta un espectro continuo de veracidad. Existen zonas grises donde la información es parcialmente correcta, desactualizada, o presenta sesgos interpretativos que no se capturan adecuadamente en un esquema binario simple.
- **Robustez de la Extracción Web:** Aunque se implementó un scraper inteligente, su eficacia sigue dependiendo de la estructura HTML de los sitios web,

que puede cambiar con el tiempo y variar significativamente entre diferentes fuentes de noticias.

- **Dominio del Corpus:** A pesar de su gran tamaño, el corpus está mayoritariamente compuesto por noticias de dominio general y político. El rendimiento del modelo podría variar en dominios muy especializados como el fraude financiero o la desinformación científica.
- **Fronteras Difusas en la Clasificación:** El modelo puede tener dificultades con contenido satírico ambiguo, información parcialmente correcta, o casos donde la veracidad depende del contexto temporal o cultural específico.

En conclusión, este trabajo ha demostrado de manera concluyente la eficacia superior del ajuste fino de modelos Transformer para la detección de noticias falsas en español y ha entregado no solo un modelo de alto rendimiento, sino también un corpus a gran escala y un prototipo funcional que sientan las bases para futuras innovaciones en la lucha contra el fraude digital.

Apéndice A

Anexo 1

// Puede incluir en un anexo: formularios, entrevistas, encuestas, carta de aceptación a revista. Todos los anexos deben ser referenciados.

Bibliografía

- [1] A. Bondielli and F. Marcelloni. A survey on fake news and rumour detection techniques. *Information Sciences*, 497:38–55, 2019. doi: 10.1016/j.ins.2019.05.035. URL <https://doi.org/10.1016/j.ins.2019.05.035>.
- [2] B. Hu, Q. Sheng, J. Cao, Y. Shi, Y. Li, D. Wang, and P. Qi. Bad actor, good advisor: Exploring the role of large language models in fake news detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 22105–22113, 2024. doi: 10.1609/aaai.v38i20.30214. URL <https://doi.org/10.1609/aaai.v38i20.30214>.
- [3] K. Ali and K. Zain-Ul-Abdin. Post-truth propaganda: heuristic processing of political fake news on Facebook during the 2016 U.S. presidential election. *Journal of Applied Communication Research*, 49(1):109–128, 2020. doi: 10.1080/00909882.2020.1847311. URL <https://doi.org/10.1080/00909882.2020.1847311>.
- [4] J. Pérez-Dasilva, K. Meso-Ayerdi, and T. Mendiguren-Galdospín. Fake news y coronavirus: detección de los principales actores y tendencias a través del análisis de las conversaciones en Twitter. *El Profesional De La Información*, 29(3), 2020. doi: 10.3145/epi.2020.may.08. URL <https://doi.org/10.3145/epi.2020.may.08>.
- [5] K. Ali, C. Liu, K. Zain-Ul-Abdin, and M. A. Zaffar. Fake news on Facebook: examining the impact of heuristic cues on perceived credibility and sharing intention. *Internet Research*, 32(1):379–397, 2021. doi: 10.1108/intr-10-2019-0442. URL <https://doi.org/10.1108/intr-10-2019-0442>.
- [6] J. Su, T. Y. Zhuo, J. Mansurov, D. Wang, and P. Nakov. Fake news detectors are biased against texts generated by large language models, 2023. URL <https://arxiv.org/abs/2309.08674>.
- [7] L. Cárcamo-Ulloa, C. Cárdenas-Neira, D. Sáez-Trumper, and C. Toural-Bran. Fake news en Chile y España: ¿cómo los medios nos hablan de noticias falsas? *Journal of Iberian and Latin American Research*, pages 1–18, 2021. doi: 10.1080/13260219.2020.1909849. URL <https://doi.org/10.1080/13260219.2020.1909849>.

- [8] J. Cao, X. Luo, and W. Zhang. Corporate employment, red flags, and audit effort. *Journal of Accounting and Public Policy*, 39(1):106710, 2020. doi: 10.1016/j.jaccpubpol.2019.106710. URL <https://doi.org/10.1016/j.jaccpubpol.2019.106710>.
- [9] I. M. Nasser, A. H. Alzaanin, and A. Y. Maghari. Online recruitment fraud detection using ANN. In *2021 Palestinian International Conference on Information and Communication Technology (PICICT)*, pages 13–17, 2021. doi: 10.1109/PICICT53635.2021.00015. URL <https://doi.org/10.1109/PICICT53635.2021.00015>.
- [10] C. Pulido, L. Ruiz-Eugenio, G. Redondo-Sama, and B. Villarejo-Carballido. A new application of social impact in social media for overcoming fake news in health. *International Journal of Environmental Research and Public Health*, 17(7):2430, 2020. doi: 10.3390/ijerph17072430. URL <https://doi.org/10.3390/ijerph17072430>.
- [11] Fabricio Andrés Zules Acosta. Construcción de un dataset de noticias para el entrenamiento y evaluación de clasificadores automatizados. Trabajo fin de máster universitario en ciberseguridad, Universidad Politécnica de Madrid, 2019. URL <https://doi.org/10.13140/RG.2.2.31181.49126>. ResearchGate.
- [12] J. P. Posadas-Durán, H. Gómez-Adorno, G. Sidorov, and J. J. M. Escobar. Detection of fake news in a new corpus for the Spanish language. *Journal of Intelligent and Fuzzy Systems*, 36(5):4869–4876, 2019. doi: 10.3233/JIFS-179034. URL <https://doi.org/10.3233/JIFS-179034>.
- [13] Y. Blanco-Fernández, J. Otero-Vizoso, A. Gil-Solla, and J. García-Duque. Enhancing misinformation detection in Spanish language with deep learning: BERT and RoBERTa transformer models. *Applied Sciences*, 14(21):9729, 2024. doi: 10.3390/app14219729. URL <https://doi.org/10.3390/app14219729>.
- [14] M. K. Singh, J. Ahmed, M. A. Alam, K. K. Raghuvanshi, and S. Kumar. A comprehensive review on automatic detection of fake news on social media. *Multimedia Tools and Applications*, 2023. doi: 10.1007/s11042-023-17377-4. URL <https://doi.org/10.1007/s11042-023-17377-4>.
- [15] C. M. Tsai. Stylometric fake news detection based on natural language processing using named entity recognition: In-domain and cross-domain analysis. *Electronics*, 12(17):3676, 2023. doi: 10.3390/electronics12173676. URL <https://doi.org/10.3390/electronics12173676>. Recuperado el 24 de junio de 2024.
- [16] J. Su, C. Cardie, and P. Nakov. Adapting fake news detection to the era of large language models, 2023. URL <https://arxiv.org/abs/2311.04917>.

- [17] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. Language models are few-shot learners, May 2020. URL <https://doi.org/10.48550/arXiv.2005.14165>.
- [18] L. Hu, S. Wei, Z. Zhao, and B. Wu. Deep learning for fake news detection: A comprehensive survey. *AI Open*, 3:133–155, 2022. doi: 10.1016/j.aiopen.2022.09.001. URL <https://doi.org/10.1016/j.aiopen.2022.09.001>.
- [19] Arsenii Tretiakov, Alejandro Martín García, and David Camacho. Detection of false information in Spanish using machine learning techniques. In *Advances in Intelligent Data Analysis and Applications*. Springer, 2022. doi: 10.1007/978-3-031-21753-1_5. URL http://dx.doi.org/10.1007/978-3-031-21753-1_5.
- [20] V. Sanh, L. Debut, J. Chaumond, and T. Wolf. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter, October 2019. URL <https://doi.org/10.48550/arXiv.1910.01108>.
- [21] S. D. Das, A. Basak, and S. Dutta. A heuristic-driven uncertainty-based ensemble framework for fake news detection in tweets and news articles. *Neurocomputing*, 491:607–620, 2022. doi: 10.1016/j.neucom.2021.12.037. URL <https://doi.org/10.1016/j.neucom.2021.12.037>.
- [22] A. Thota, P. Tilak, S. Ahluwalia, and N. Lohia. Fake news detection: A deep learning approach. *SMU Scholar*, 2018. URL <https://scholar.smu.edu/datasciencereview/vol1/iss3/10/>. Recuperado el 05 de julio de 2024.
- [23] G. Hurtado Avilés, R. A. Mora-Gutiérrez, and J. A. Reyes-Ortiz. Calibración de hiper-parámetros en algoritmos metaheurísticos para la detección de fraude digital. In M. Tovar Vidal, A. L. Lezama Sánchez, and M. Contreras González, editors, *Avances recientes en procesamiento de lenguaje natural y otras áreas afines*, pages 14–26. Benemérita Universidad Autónoma de Puebla, 2024. ISBN 978-607-5914-56-5.
- [24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need, June 2017. URL <https://doi.org/10.48550/arXiv.1706.03762>.
- [25] J. Devlin, M. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding, October 2018. URL <https://doi.org/10.48550/arXiv.1810.04805>.

- [26] X. Jiao, Y. Yin, L. Shang, X. Jiang, X. Chen, L. Li, F. Wang, and Q. Liu. Tiny-BERT: Distilling BERT for natural language understanding, September 2019. URL <https://doi.org/10.48550/arXiv.1909.10351>.
- [27] K. Martínez-Gallego, A. M. Álvarez Ortiz, and J. D. Arias-Londoño. Fake news detection in Spanish using deep learning techniques, October 2021. URL <https://arxiv.org/abs/2110.06461>.
- [28] E. Shushkevich, M. Alexandrov, and J. Cardiff. Improving multiclass classification of fake news using BERT-based models and ChatGPT-augmented data. *Inventions*, 8(5):112, 2023. doi: 10.3390/inventions8050112. URL <https://doi.org/10.3390/inventions8050112>.
- [29] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M. A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Gravé, and G. Lample. LLaMA: Open and efficient foundation language models, February 2023. URL <https://doi.org/10.48550/arXiv.2302.13971>.
- [30] Gemini Team, Google. Gemini: A family of highly capable multimodal models, December 2023. URL <https://doi.org/10.48550/arXiv.2312.11805>.
- [31] Y. Bai, S. Kadavath, S. Kundu, A. Askell, J. Kernion, A. Jones, A. Chen, A. Goldin, A. Mirhoseini, C. McKinnon, C. Chen, C. Olsson, C. Olah, D. Hernandez, D. Drain, D. Ganguli, D. Li, E. Tran-Johnson, E. Perez, J. Kerr, J. Mueller, J. Ladish, J. Landau, K. Ndousse, K. Lukosuite, L. Lovitt, M. Sellitto, N. Elhage, N. Schiefer, N. Mercado, N. DasSarma, R. Lasenby, R. Larson, S. Ringer, S. Johnston, S. Kravec, S. E. Showk, S. Fort, T. Lanham, T. Telleen-Lawton, T. Conerly, T. Henighan, T. Hume, S. R. Bowman, Z. Hatfield-Dodds, B. Mann, D. Amodei, N. Joseph, S. McCandlish, T. Brown, and J. Kaplan. Constitutional AI: Harmlessness from AI feedback, December 2022. URL <https://doi.org/10.48550/arXiv.2212.08073>.
- [32] M. G. R. Anselmo. *Diseño y desarrollo de un método heurístico basado en un sistema socio-cultural de creatividad para la resolución de problemas de optimización no lineales y diseño de zonas electorales*. Tesis de doctorado en ingeniería, Universidad Nacional Autónoma de México, 2013. URL <http://132.248.10.225:8080/handle/123456789/101>. Recuperado el 07 de julio de 2024.
- [33] S. Aqil and M. Lahby. Modeling and solving the fake news detection scheduling problem. In *Studies in computational intelligence*, pages 231–242. Springer, 2021. doi: 10.1007/978-3-030-90087-8_11. URL https://doi.org/10.1007/978-3-030-90087-8_11.
- [34] N. Bacanin, C. Stoean, M. Zivkovic, M. Rakic, R. Strulak-Wójcikiewicz, and R. Stoean. On the benefits of using metaheuristics in the hyperparameter tuning

- of deep learning models for energy load forecasting. *Energies*, 16(3):1434, 2023. doi: 10.3390/en16031434. URL <https://doi.org/10.3390/en16031434>.
- [35] S. Hidayattullah, I. Surjandari, and E. Laoh. Financial statement fraud detection in Indonesia listed companies using machine learning based on meta-heuristic optimization. In *2020 International Workshop on Big Data and Information Security (IWBIS)*, pages 79–84, Depok, Indonesia, 2020. doi: 10.1109/IWBIS50925.2020.9255563. URL <https://doi.org/10.1109/IWBIS50925.2020.9255563>.
 - [36] J. Horak and A. Sabek. Gaussian process regression's hyperparameters optimization to predict financial distress. *Retos*, 13(26):273–289, 2023. doi: 10.17163/ret.n26.2023.06. URL <https://doi.org/10.17163/ret.n26.2023.06>.
 - [37] G. Yildirim. A novel hybrid multi-thread metaheuristic approach for fake news detection in social media. *Applied Intelligence*, 53:11182–11202, 2023. doi: 10.1007/s10489-022-03972-9. URL <https://doi.org/10.1007/s10489-022-03972-9>.
 - [38] N. Deshai and B. Bhaskara Rao. Unmasking deception: a CNN and adaptive PSO approach to detecting fake online reviews. *Soft Computing*, 27:11357–11378, 2023. doi: 10.1007/s00500-023-08507-z. URL <https://doi.org/10.1007/s00500-023-08507-z>.
 - [39] J. Howard and S. Ruder. Universal language model fine-tuning for text classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 328–339, Melbourne, Australia, 2018. Association for Computational Linguistics. doi: 10.18653/v1/P18-1031. URL <https://aclanthology.org/P18-1031>.
 - [40] J. D. M. W. Kenton and L. K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186, Minneapolis, Minnesota, 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL <https://aclanthology.org/N19-1423>.
 - [41] S. Ruder. Neural transfer learning for natural language processing. *arXiv preprint arXiv:1708.00524*, 2019. doi: 10.48550/arXiv.1708.00524. URL <https://arxiv.org/abs/1708.00524>.
 - [42] M. R. Zhang, N. Desai, J. Bae, J. Lorraine, and J. Ba. Using large language models for hyperparameter optimization, 2023. URL <https://openreview.net/forum?id=FUdZ6HE0re>. Recuperado el 06 de julio de 2024.
 - [43] F. Alam, A. Barrón-Cedeño, G. S. Cheema, G. K. Shahi, S. Hakimov, M. Hasanain, others, and P. Nakov. Overview of the CLEF-2023 CheckThat! lab

- task 1 on check-worthiness of multimodal and multigenre content. In *CEUR Workshop Proceedings*, volume 3497, pages 219–235, September 2023. URL <https://dclibrary.mbzua.ac.ae/nlpfp/78/>. Recuperado el 05 de julio de 2024.
- [44] A. Barrón-Cedeño, F. Alam, T. Caselli, G. Da San Martino, T. Elsayed, A. Galassi, others, and P. Nakov. The CLEF-2023 CheckThat! lab: Checkworthiness, subjectivity, political bias, factuality, and authority. In J. Kamps et al., editors, *Advances in Information Retrieval. ECIR 2023. Lecture Notes in Computer Science*, volume 13982. Springer, Cham, 2023. doi: 10.1007/978-3-031-28241-6_59. URL https://doi.org/10.1007/978-3-031-28241-6_59.
- [45] M. E. Aragón, H. Jarquín, M. M. Y. Gómez, H. J. Escalante, L. Villaseñor-Pineda, H. Gómez-Adorno, others, and J. P. Posadas-Durán. Overview of mex-a3t at iberlef 2020: Fake news and aggressiveness analysis in mexican Spanish. In *Notebook Papers of 2nd SEPLN Workshop on Iberian Languages Evaluation Forum (IberLEF)*, Malaga, España, September 2020. URL <https://openreview.net/forum?id=vkjfS0w2hu>. Recuperado el 06 de julio de 2024.
- [46] H. Gómez-Adorno, J. P. Posadas-Durán, G. B. Enguix, and C. P. Capetillo. Overview of FakeDeS at IberLEF 2021: Fake news detection in Spanish shared task. *Procesamiento del Lenguaje Natural*, 67:223–231, 2021. doi: 10.26342/2021-67-19. URL <https://doi.org/10.26342/2021-67-19>.
- [47] J. M. Ramírez Cruz, S. Ú. Palacios Alvarado, K. E. Franca Tapia, J. P. F. Posadas Durán, H. M. Gómez Adorno, and G. Sidorov. The Spanish fake news corpus version 2.0 en GitHub. GitHub repository, 2021. URL <https://github.com/jpposadas/FakeNewsCorpusSpanish>. Recuperado el 10 de julio de 2024.
- [48] K. M. Yazdi, A. M. Yazdi, S. Khodayi, J. Hou, W. Zhou, and S. Saedy. Improving fake news detection using K-Means and support vector machine approaches, January 2020. URL <https://publications.waset.org/10011058/improving-fake-news-detection-using-k-means-and-support-vector-machine-approaches>. Recuperado el 05 de julio de 2024.
- [49] A. Yasmin, W. Haider Butt, and A. Daud. Ensemble effort estimation with metaheuristic hyperparameters and weight optimization for achieving accuracy. *PLOS ONE*, 19(4):e0300296, 2024. doi: 10.1371/journal.pone.0300296. URL <https://doi.org/10.1371/journal.pone.0300296>.
- [50] J. H. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. MIT Press, 1992. ISBN 978-0-262-58111-0. Edición MIT Press del trabajo original de 1975.

- [51] J. Kennedy and R. C. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, pages 1942–1948, November 1995. doi: 10.1109/ICNN.1995.488968. URL <https://doi.org/10.1109/ICNN.1995.488968>.
- [52] R. C. Eberhart and J. Kennedy. A new optimizer using particle swarm theory. In *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, pages 39–43, 1995. doi: 10.1109/MHS.1995.494215. URL <https://doi.org/10.1109/MHS.1995.494215>.
- [53] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983. doi: 10.1126/science.220.4598.671. URL <https://doi.org/10.1126/science.220.4598.671>.
- [54] R. Martí, M. G. C. Resende, and P. M. Pardalos. Multi-start methods. In *Handbook of Heuristics*, pages 195–212. Springer, 2018. doi: 10.1007/978-3-319-91086-4_7. URL https://doi.org/10.1007/978-3-319-91086-4_7.
- [55] F. Glover. A template for scatter search and path relinking. In *Lecture Notes in Computer Science*, volume 1363, pages 13–54. Springer, 1998. doi: 10.1007/BFb0026599. URL <https://doi.org/10.1007/BFb0026599>.
- [56] N. Mladenović and P. Hansen. Variable neighborhood search. *Computers & Operations Research*, 24(11):1097–1100, 1997. doi: 10.1016/S0305-0548(97)00031-2. URL [https://doi.org/10.1016/S0305-0548\(97\)00031-2](https://doi.org/10.1016/S0305-0548(97)00031-2).
- [57] T. O’Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, L. Invernizzi, and The Keras team. Hyperparameter tuning with Keras Tuner. The TensorFlow Blog, January 2020. URL <https://blog.tensorflow.org/2020/01/hyperparameter-tuning-with-keras-tuner.html>. Recuperado el 29 de enero de 2020.
- [58] S. Kapunac, A. Kartelj, and M. Djukanović. Variable neighborhood search for weighted total domination problem and its application in social network information spreading. *Applied Soft Computing*, 143:110387, 2023. doi: 10.1016/j.asoc.2023.110387. URL <https://doi.org/10.1016/j.asoc.2023.110387>.
- [59] O. H. Alvarez. Fraude laboral en la era digital. *Revista General de Derecho del Trabajo y de la Seguridad Social*, (59), 2021. ISSN 1696-9626. URL <https://dialnet.unirioja.es/servlet/articulo?codigo=8029600>. Recuperado el 06 de julio de 2024.
- [60] Fabricio Andrés Zules Acosta. DataSet Web scraping noticias verificadas. Kaggle, 2019. URL <https://www.kaggle.com/datasets/zulanac/fake-and-real-news>. Recuperado el 07 de julio de 2024.

- [61] F. A. Zules Acosta. Spanish Fake and Real News. Kaggle, 2019. URL <https://www.kaggle.com/datasets/zulanac/fake-and-real-news>. Recuperado el 10 de julio de 2024.
- [62] Arsenii Tretiakov, Alejandro Martín García, and David Camacho. Noticias falsas en español. Kaggle, 2022. URL https://www.kaggle.com/datasets/arsenii_tretiakov/noticias-falsas-en-espaol. Recuperado el 15 de julio de 2024.
- [63] Y. Blanco-Fernández, J. Otero-Vizoso, A. Gil-Solla, and J. García-Duque. Spanish Political Fake News. Kaggle, 2024. URL <https://www.kaggle.com/datasets/javieroterovizoso/spanish-political-fake-news>. Recuperado el 15 de julio de 2024.
- [64] J. P. Aguirre Quezada. El fraude en México: daños patrimoniales y trabajo legislativo para enfrentarlo. Technical report, Senado de México, 2023. URL <http://bibliodigitalibd.senado.gob.mx/handle/123456789/6051>. Recuperado el 05 de julio de 2024.
- [65] L. Ehrlinger and W. Wöß. Towards a definition of knowledge graphs. In *SEMANTiCS (Posters, Demos, SuCESS)*, 2016. URL https://www.researchgate.net/publication/323316736_Towards_a_Definition_of_Knowledge_Graphs. Recuperado el 06 de julio de 2024.
- [66] G. A. García Robledo, J. A. Reyes-Ortiz, and B. A. González-Beltrán. Interfaz de consulta en idioma español para la búsqueda de información en un ambiente académico. Tesis de maestría en ciencias de la computación, Universidad Autónoma Metropolitana, 2020. URL <https://zalomatilazc.uam.mx/handle/11191/7812>. Recuperado el 07 de julio de 2024.
- [67] G. Hurtado Avilés, J. M. Villa Vargas, S. Tapia Hernández, A. Á. Rivera Sanabria, and J. M. Jaimes Ponce. Representación ontológica de la arquitectura de control de un robot SCARA utilizando IoT y MATLAB. In M. Tovar Vidal, A. L. Lezama Sánchez, and M. Contreras González, editors, *Avances recientes en procesamiento de lenguaje natural y otras áreas afines*. Benemérita Universidad Autónoma de Puebla, 2024. ISBN 978-607-5914-56-5.
- [68] A. Rogers, O. Kovaleva, and A. Rumshisky. A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research*, 61: 65–95, 2020. doi: 10.1613/jair.1.11640. URL <https://doi.org/10.1613/jair.1.11640>.
- [69] H. J. Levesque. Knowledge representation and reasoning. *Annual review of computer science*, 1(1):255–287, 1986. doi: 10.1146/annurev.cs.01.060186.001351. URL <https://doi.org/10.1146/annurev.cs.01.060186.001351>.

- [70] E. D. Liddy. Natural language processing. Technical report, Syracuse University, 2001. URL <https://surface.syr.edu/istpub/63/>. Recuperado el 05 de julio de 2024.
- [71] C. E. Manzano-Velasco. DataSet Web scraping noticias verificadas. Kaggle, 2024. URL <https://www.kaggle.com/datasets/enriquemanzano/dataset-web-scraping-noticias-verificadas/data>. Recuperado el 24 de junio de 2024.
- [72] C. E. Manzano-Velasco and L. S. Daza-Rosero. Sistema de alerta temprana para monitorear la propagación de noticias falsas: caso de estudio contexto político colombiano. Trabajo grado, Universidad del Cauca, 2024. URL <http://repositorio.unicauca.edu.co:8080/xmlui/handle/123456789/9485>. Recuperado el 07 de julio de 2024.
- [73] M. F. Mridha, A. J. Keya, M. A. Hamid, M. M. Monowar, and M. S. Rahman. A comprehensive review on fake news detection with deep learning. *IEEE Access*, 9:156151–156170, 2021. doi: 10.1109/ACCESS.2021.3129329. URL <https://doi.org/10.1109/ACCESS.2021.3129329>.
- [74] J. M. Osorio-Giraldo, C. J. Ropain-Zambrano, A. F. Taba-Pulgarin, and A. de Jesús Osorio-Orozco. 10. proyecto de seguridad para reducir fraudes y robos en marketplace de facebook. In *Coloquio de Investigación Formativa 2023-1*, page 89, 2023. URL https://ridum.umanizales.edu.co/xmlui/bitstream/handle/20.500.12746/6834/CIF%202023-1%20-%20Res%C3%BAmenes%20ejecutivos_RIDUM.pdf?sequence=1&isAllowed=y#page=89. Recuperado el 05 de julio de 2024.
- [75] J. Padilla Cuevas, J. A. Reyes-Ortiz, and M. Bravo. Detección y representación de eventos en un ambiente académico inteligente. Tesis de maestría en ciencias de la computación, Universidad Autónoma Metropolitana, 2019. URL <https://zaloamati.azc.uam.mx/handle/11191/6124>. Recuperado el 06 de julio de 2024.
- [76] J. A. Reyes-Ortiz. Extracción de información semántica para la clasificación de servicios web. Tesis de maestría en ciencias en ciencias de la computación, Centro Nacional de Investigación y Desarrollo Tecnológico, 2008. URL <https://www.cenidet.edu.mx/subplan/biblio/seleccion/Tesis/MC%20Jos%E9%20Alejandro%20Reyes%20Ortiz%202008.pdf>. Recuperado el 07 de julio de 2024.
- [77] J. A. Reyes-Ortiz, B. A. González-Beltrán, and L. Gallardo-López. Clinical decision support systems: A survey of NLP-based approaches from unstructured data. In *2015 26th International Workshop on Database and Expert Systems Applications (DEXA)*, pages 163–167, Valencia, Spain, 2015. doi: 10.1109/DEXA.2015.47. URL <https://doi.org/10.1109/DEXA.2015.47>.

- [78] S. C. Shapiro. Knowledge representation. In *Encyclopedia of cognitive science*. Wiley, 2006. doi: 10.1002/0470018860.s00058. URL <https://doi.org/10.1002/0470018860.s00058>.
- [79] C. Villamil Arcos. Selección de una técnica de aprendizaje de máquina para la detección de fraude financiero digital enfocado a transacciones no autorizadas o consentidas. Master's thesis, Universidad Nacional de Colombia, 2022. URL <https://repositorio.unal.edu.co/handle/unal/84015>. Recuperado el 05 de julio de 2024.
- [80] C. Whitehouse, T. Weyde, P. Madhyastha, and N. Komninos. Evaluation of fake news detection with knowledge-enhanced language models, April 2022. URL <https://arxiv.org/abs/2204.00458>.