Tipo de Artículo: *Artículo de revisión*
Área: *Ingeniería*

Análisis De Sentimiento y Minería De Opiniones En Entornos Digitales Mediante Deep Learning: evolución desde modelos secuenciales hasta modelos de lenguaje de gran escala (LLMs)

Sentiment Analysis and Opinion Mining in Digital Environments Using Deep Learning: Evolution from Sequential Models to Large Language Models (LLMs)

Jaime David Camacho Castillo[1], Ciro Leonidas Flores Arcos[2], Miguel Ángel Tierra Moyon[3], Fanny Maricris Orozco Oña[4], Ángeles Karla Arteaga Pilataxi[5], Milton Josué Damian Aguilar[6]

Cita:
Jaime David Camacho Castillo et al. (2026). Análisis De Sentimiento y Minería De Opiniones En Entornos Digitales Mediante Deep Learning: evolución desde modelos secuenciales hasta modelos de lenguaje de gran escala (LLMs). *TESLA Revista Científica*, 6(2), e726. https://doi.org/10.55204/trc.v6i2.e726

Recibido: 23/06/2026
Revisado: 07/07/2026–28/07/2026
Corregido: 05/08/2026
Aceptado: 18/08/2026
Publicado: 08/09/2026

[1] Escuela Superior Politécnica de Chimborazo (ESPOCH), Riobamba, Ecuador.
[2–6] Investigador externo, Ecuador.

[1] jaimed.camacho@espoch.edu.ec, [2] cirox.flores@gmail.com, [3] tierram928@gmail.com, [4] mariorozco032@gmail.com, [5] angeles34arteaga@gmail.com, [6] josue.damian371@gmail.com

[1] 0000-0002-9110-6585, [2] 0009-0003-8647-2307, [3] 0009-0005-8752-8792, [4] 0009-0003-2824-5114, [5] 0009-0008-6212-3133, [6] 0009-0005-5741-6861

Resumen: El análisis de sentimiento en entornos digitales ha experimentado una rápida evolución metodológica, transitando desde modelos secuenciales clásicos hacia arquitecturas de aprendizaje profundo (Deep Learning) capaces de comprender el contexto lingüístico humano. Para examinar esta progresión, se realizó una revisión sistemática aplicando el protocolo PRISMA, evaluando estudios empíricos de alto impacto (2020-2025) extraídos de Scopus, IEEE Xplore, Web of Science, etc. Los hallazgos validan la eficacia de estas tecnologías en escenarios reales, destacando su uso en el monitoreo de satisfacción en plataformas de comercio electrónico y la evaluación de tendencias en salud pública. No obstante, la revisión revela un vacío de investigación prioritario: la urgencia de integrar la explicabilidad algorítmica (XAI) desde la concepción del sistema y democratizar estas herramientas hacia lenguas de bajos recursos. Se concluye que el estado del arte actual reside en la combinación de Transformers y Modelos de Lenguaje de Gran Escala (LLMs) con técnicas de adaptación eficiente de parámetros (como LoRA), lo cual ofrece a las organizaciones un equilibrio óptimo entre la máxima precisión predictiva y la viabilidad computacional.

Palabras clave: *Análisis de sentimiento, Deep Learning, Metodología PRISMA, Modelos de lenguaje de gran escala, explicabilidad algorítmica.*

Abstract: Sentiment analysis in digital environments has undergone a rapid methodological evolution, transitioning from classic sequential models to deep learning architectures capable of understanding profound human linguistic context. To examine this progression, a systematic review was conducted applying the PRISMA protocol, evaluating high-impact empirical studies (2020-2025) extracted from Scopus, IEEE Xplore, Web of Science, etc. The findings validate the efficacy of these technologies in real-world scenarios, highlighting their use in e-commerce satisfaction monitoring and public health trend assessment. However, the review reveals a priority research gap: the urgency of integrating algorithmic explainability (XAI) from the system's conception and democratizing these tools for low-resource languages. It is concluded that the current state of the art lies in combining Transformers and Large Language Models (LLMs) with parameter-efficient adaptation techniques (such as LoRA), offering organizations an optimal balance between maximum predictive accuracy and computational viability.

Keywords: *Sentiment analysis, Deep Learning, PRISMA methodology, Large language models, algorithmic explainability.*

INTRODUCCIÓN

El crecimiento exponencial de datos no estructurados en internet, generados diariamente en redes sociales, plataformas de comercio electrónico y foros, ha transformado radicalmente la manera en que las organizaciones comprenden el comportamiento y las preferencias humanas. Este flujo masivo de información textual requiere herramientas automatizadas para decodificar la opinión pública, convirtiendo al análisis de sentimiento en un pilar estratégico del procesamiento del lenguaje natural (PLN).

La verdadera dificultad de esta tarea radica en la complejidad intrínseca del lenguaje humano, donde la comunicación digital está saturada de ironía, sarcasmo, variaciones dialectales y señales contextuales implícitas que superan ampliamente las capacidades de los enfoques tradicionales. Los sistemas basados en diccionarios léxicos o en el aprendizaje automático clásico fallan al no comprender que una misma palabra puede invertir su polaridad dependiendo de la estructura de la oración o del dominio de aplicación.

Frente a esta rápida transición tecnológica, el objetivo central de esta revisión es sintetizar críticamente la evolución desde los modelos de aprendizaje profundo secuencial (como RNN y LSTM) hasta los actuales Modelos de Lenguaje de Gran Escala (LLMs).

La contribución principal de este trabajo reside en ofrecer una interpretación integrada que conecta las decisiones arquitectónicas de los modelos con los problemas lingüísticos específicos que intentan resolver. A partir de este análisis crítico, se identifican líneas de investigación futuras que responden de manera coherente a los desafíos aún no resueltos por la comunidad científica.

METODOLOGÍA

Para revisar la transparencia, exhaustividad y replicabilidad del estudio, esta investigación se adhiere a las directrices del protocolo PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), la cual se fundamenta en su capacidad para aportar rigor científico al proceso de revisión, permitiendo auditar cada decisión de selección, minimizar los sesgos cognitivos o de confirmación por parte de los investigadores, y sentar una base estructurada que facilite la reproducibilidad del estudio por parte de terceros.

El proceso de identificación de las fuentes de información se llevó a cabo mediante búsquedas en bases de datos académicas como: Scopus, IEEE Xplore, Web of Science, etc., para capturar la intersección del objeto de estudio, se diseñó la siguiente cadena de búsqueda utilizando operadores booleanos: ("sentiment analysis" OR "opinion mining") AND "Deep Learning" AND ("Transformer" OR "LLM"). Las búsquedas se delimitaron mediante filtros: documentos publicados entre 2020 y 2025, indexados en revistas de los cuartiles Q1 o Q2 (según Scimago Journal Rank o JCR), redactados en idioma inglés y que constituyeran estudios de naturaleza empírica.

La selección del corpus documental se rigió por parámetros que permitan detectar la evidencia técnica, actual y de alto rigor, estos criterios son:

Tabla 1

Criterios de inclusión y exclusión

Criterio	Inclusión	Exclusión
Tipo de estudio	Empírico, con métricas cuantitativas comprobables.	Conceptual, revisiones, o propuestas sin validación experimental.
Dominio	Análisis de sentimiento y minería de opiniones mediante *Deep Learning*.	Otras tareas de PLN (ej. traducción, resumen) sin foco en sentimiento.
Calidad	Artículos en revistas de alto impacto (Q1/Q2).	Conferencias de bajo impacto, blogs o artículos sin revisión por pares.
Idioma	Inglés (para asegurar un estándar de evaluación global).	Otros idiomas (debido a limitaciones de acceso, traducción y estandarización).

Nota: Elaboración propia con base en la información analizada.

El cribado de los artículos se ejecutó en etapas: en primer lugar, los resultados exportados de las bases de datos se sometieron a un gestor de fuentes (ej. Zotero y métodos computacionales), después, se llevó a cabo el cribado inicial leyendo los títulos y resúmenes de los registros únicos para descartar literatura temática o metodológicamente irrelevante.

Los artículos preseleccionados pasaron a la fase de revisión a texto completo, donde se registró meticulosamente el motivo de cada exclusión (ej. falta de métricas, ausencia de validación empírica, metodologías opacas), cualquier discrepancia fue resuelta mediante el debate y la búsqueda de consenso; en los casos necesarios.

Para la extracción de información, se construyó una matriz estandarizada en la que se recopilaron variables críticas de los estudios finales: arquitectura del modelo empleado (ej. LSTM, CNN, RoBERTa, LLaMA), conjunto de datos evaluado, métricas de rendimiento (Accuracy, Precision, Recall, F1-Score), dominio de aplicación (ej. e-commerce, redes sociales) y las limitaciones declaradas por los autores. El enfoque de síntesis aplicado fue de carácter temático-comparativo, diseñado para describir los resultados aislados y para identificar y argumentar los patrones evolutivos y los saltos arquitectónicos que definen la transición metodológica del área.

La ejecución del protocolo PRISMA ilustra el flujo de depuración, donde las bases de datos arrojaron un total de 35 registros. Tras eliminar 8 duplicados, se procedió al cribado de 27 artículos, para la evaluación a texto completo. La síntesis de este flujo se documenta en la Tabla 2.

Tabla 2

Síntesis del proceso de selección bibliográfica

Etapa del Proceso	Número de Registros	Razones de Exclusión
Identificación	35	N/A
Duplicados removidos	8	Coincidencia en DOI o título.
Evaluación a texto completo	27	N/A

Nota: Elaboración propia con base en la información analizada.

APLICACIONES PRÁCTICAS O FUTURAS LÍNEAS DE INVESTIGACIÓN

Los avances metodológicos revelados en esta revisión sistemática poseen implicaciones operativas directas para múltiples sectores que requieren tomar decisiones basadas en datos no estructurados:

• Comercio electrónico y analítica de productos:

La implementación de sistemas de Análisis de Sentimiento Orientado a Aspectos (ABSA) ha revolucionado la inteligencia de negocios, al extraer entidades específicas, las organizaciones ya no solo miden si un producto es "bueno" o "malo", sino que logran discriminar que, en un mismo smartphone, la "batería" genera rechazo mientras que la "pantalla" produce satisfacción (Daza et al., 2024). Como lo evidencian los hallazgos, la integración de modelos híbridos y arquitecturas basadas en Transformers (p. ej., BERT combinado con BiLSTM) ha reportado precisiones superiores al 97% en clasificación binaria, lo cual confirma su total viabilidad técnica para ser implementados en dashboards de monitoreo de reseñas en tiempo real (Lakshmi et al., 2025).

• Salud pública y epidemiología digital:

La revisión demuestra que el análisis temporal de sentimientos en redes sociales es una herramienta crítica para el monitoreo del bienestar emocional poblacional durante crisis, como se observó durante la pandemia de COVID-19 (Rodríguez-Ibáñez et al.,

2023). La aplicación trasciende el análisis descriptivo al integrar conceptos de causalidad algorítmica, como la Causalidad de Granger, esto permite a las instituciones sanitarias pasar de la simple correlación a la predicción de tendencias, anticipando picos de ansiedad social o rechazo a medidas de salud a partir de alteraciones tempranas en el discurso digital.

• Gobernanza digital y políticas públicas:

En el ámbito gubernamental, estas herramientas facilitan el rastreo continuo de la opinión pública, permitiendo la detección temprana de polarización social y la evaluación inmediata del impacto de nuevas políticas. Sin embargo, su despliegue operativo exige la implementación de estrictas salvaguardas éticas y auditorías de datos para evitar que los sesgos de interpretación del algoritmo marginen las voces de las minorías o malinterpreten dialectos locales (Venkit et al., 2023).

Futuras líneas de investigación a partir de las limitaciones y tensiones metodológicas detectadas:

• Recursos para lenguas de bajos recursos y brecha digital.

Los LLMs modernos están altamente sesgados hacia el idioma inglés, reduciendo su eficacia en dialectos o lenguas minoritarias, por ello, la investigación futura debe enfocarse en la democratización tecnológica mediante el preentrenamiento específico en lenguas de bajos recursos, combinado con técnicas de Ajuste Fino Eficiente de Parámetros (PEFT), como la Adaptación de Rango Bajo (LoRA). Esto permitirá a investigadores y empresas adaptar modelos masivos con una fracción mínima del costo computacional (Mabokela et al., 2024).

• Explicabilidad algorítmica y cajas negras.

La alta complejidad de los modelos actuales limita la transparencia, lo que frena su adopción en sectores críticos, es por ello que, los futuros estudios deben transitar hacia un Procesamiento de Lenguaje Natural Explicable (XNLP), donde la integración de Inteligencia Artificial Explicable (XAI) desde el diseño de la arquitectura (mediante atención supervisada o racionalización), pero no como un simple añadido posterior (post-hoc). Esto es vital para comprender por qué el modelo clasifica un texto de una manera u otra (Mohammadi et al., 2025; Adak et al., 2022).

• Evaluación multimodal y causalidad profunda.

La comunicación humana rara vez es exclusivamente textual, por ello, el campo debe avanzar hacia arquitecturas de Análisis de Sentimiento Multimodal (MSA) que integren de manera simultánea el texto, las inflexiones de voz y la gesticulación facial (Baberwal et al., 2025). Asimismo, se necesitan nuevos marcos de evaluación que superen la simple correlación estadística entre etiquetas, desarrollando métricas que certifiquen un razonamiento causal genuino por parte del modelo ante el sarcasmo o la ironía.

RESULTADOS Y DISCUSIÓN

La ejecución sistemática de la metodología PRISMA permitió una selección de literatura rigurosa, auditable y basada en evidencia. El flujo real del proceso partió con la identificación de 35 registros en Scopus, IEEE Xplore, Web of Science, aplicando las ecuaciones de búsqueda y los filtros temporales de 2020 a 2025. El cribado inicial por título y resumen abarcó 35 artículos, de los cuales 8 fueron descartados por irrelevancia temática o por carecer de un enfoque en aprendizaje profundo. Finalmente, el corpus se consolidó en 27 estudios primarios de alto impacto.

Tabla 3

Síntesis del proceso de selección PRISMA

Etapa del proceso	Número de registros	Razones de exclusión principal
Identificación	35	N/A
Duplicados eliminados	8	Coincidencias exactas en título/DOI
Inclusión	27	N/A

Nota: Elaboración propia con base en la información analizada.

Esta metodología mejoró la transparencia del estudio, mitigando el sesgo de confirmación ("cherry-picking") mediante el cual un investigador podría elegir solo los artículos que respaldan su hipótesis. Las decisiones más difíciles se presentaron en la fase de elegibilidad a texto completo, específicamente al evaluar artículos que proponían arquitecturas teóricas complejas pero carecían de benchmarks estandarizados para probar su rendimiento frente a modelos base; excluir estos trabajos requirió un fuerte consenso del equipo para priorizar el rigor empírico sobre la innovación puramente teórica.

Evolución arquitectónica: de secuenciales a LLMs:

• De RNN/LSTM al problema del vanishing gradient: La transición de métodos estadísticos a redes secuenciales respondió a la necesidad de recordar el orden de las palabras, las redes recurrentes (RNN) procesaban el texto de forma unidireccional, pero chocaron con el problema del desvanecimiento del gradiente (vanishing gradients): al leer reseñas extensas, el modelo "olvidaba" el contexto del inicio. Las capas LSTM mitigaron esto usando mecanismos de puertas para decidir qué recordar a largo plazo, resolviendo computacionalmente la dependencia temporal de las oraciones.

• Transformers y la atención global: Las LSTM fallaban en paralelización y en capturar dependencias a muy larga distancia, la arquitectura Transformer resuelve esto introduciendo el mecanismo de atención global (self-attention), permitiendo que el modelo analice todas las palabras de un documento simultáneamente y pondere cómo una palabra al inicio de la reseña modifica a un adjetivo al final. Lingüísticamente, esto permitió a las máquinas comprender por primera vez el contexto bidireccional y la polisemia.

• LLMs + PEFT y el equilibrio adaptativo: La creación de modelos gigantes (LLMs) puede traer capacidades generalistas y aprendizaje en contexto (zero-shot). Sin embargo, su tamaño hace inviable el entrenamiento para dominios específicos, por ello, la introducción de técnicas de Adaptación Eficiente de Parámetros (PEFT), como LoRA, puede resolver este cuello de botella: al congelar la red original y entrenar solo matrices de bajo rango, se logra un equilibrio perfecto entre la retención del conocimiento lingüístico mundial y la especialización económica en el dominio objetivo (ej. jerga local o dialectos).

Evaluación del desempeño: métricas, función de ajuste y grado de predicción.

Tabla 4
Comparación del desempeño arquitectónico

Estudio	Arquitectura	Dataset	Métrica Principal
Tan et al.	LSTM (Secuencial)	Amazon Reviews (Binario)	Accuracy
Ali et al.	BERT (Transformer)	Amazon Reviews (Binario)	Accuracy / F1
Shobayo et al.	GooglePaLM (LLM)	Amazon Fashion (Binario)	Accuracy Positiva
Bernal-Beltrán	MarIA-RoBERTa (Fine-tuned)	Rest-Mex (5 Clases)	F1-Score (Macro)

Nota: Elaboración propia con base en la información analizada.

En Deep Learning, el modelo "ajuste de función" se realiza mediante optimización iterativa de parámetros para minimizar una función de pérdida (loss function), usando retropropagación y descenso de gradiente.

En un caso concreto: en un LSTM para análisis de sentimiento, la función de ajuste aprende a ponderar palabras clave (como 'excelente' o 'terrible') en su contexto exacto, actualizando constantemente los pesos sinápticos durante el entrenamiento para reducir la diferencia (el error) entre su predicción y la etiqueta real otorgada por el usuario, dicho ajuste iterativo es el núcleo del éxito en esta disciplina, ya que permite capturar dependencias no lineales y contextuales (como una doble negación) que los modelos matemáticos lineales o de regresión tradicional simplemente no pueden representar.

Grado de predicción de nivel 'x': El grado de predicción se refiere a la capacidad del modelo para generalizar sus aciertos a datos no vistos previamente, medido típicamente mediante la precisión o el F1-Score en un conjunto de datos de prueba aislado. A futuro, la integración de técnicas como el Aprendizaje por Refuerzo a partir de Retroalimentación Humana (RLHF) y la incorporación de razonamiento causal podría elevar el grado de predicción hacia una verdadera comprensión pragmática del usuario, superando la mera correlación estadística entre palabras.

Argumentación de hallazgos

• Los modelos consistentemente exhiben mayor precisión al detectar sentimientos negativos frente a los positivos, esto ocurre porque el lenguaje humano de insatisfacción suele utilizar marcadores léxicos muy explícitos (adjetivos peyorativos, quejas fuertes), mientras que los sentimientos positivos o neutrales suelen ser expresados de forma más sutil o con ironía. En aplicaciones reales, este sesgo implica que un sistema de monitoreo corporativo podría generar falsas alarmas, sobredimensionando las quejas por encima de las valoraciones reales.

• Al adaptar el modelo al dominio pesa más que escalar sus billones de parámetros. Por ejemplo, modelos locales más pequeños preentrenados específicamente en un idioma o temática (como GreekBERT o MarIA para el español) superan con creces a modelos multilingües masivos como mBERT o GPT en configuraciones zero-shot al aplicarse en ese idioma. Por lo tanto, en la industria debe priorizarse la adaptación de modelos de tamaño medio (fine-tuning) por encima del despliegue ciego de LLMs gigantescos que elevan el costo sin garantizar el entendimiento de la jerga particular.

Vacíos de investigación identificados

• Existe una crítica falta de benchmarks estandarizados y marcos de evaluación para el Análisis de Sentimiento Multimodal (MSA), aunque los humanos se comunican fusionando texto, tono de voz y microexpresiones, los conjuntos de datos actuales no logran consolidar de manera fiable estas tres fuentes.

• Hay una falta de integración de herramientas de auditoría de sesgos y salvaguardas éticas en los pipelines de producción, en la que, modelos extremadamente precisos siguen replicando prejuicios históricos o discriminando lenguajes dialectales, operando como "cajas negras".

• Se evidencia una crisis de reproducibilidad ocasionada por la falta de publicación de código fuente abierto, métricas detalladas y datos de entrenamiento compartidos en la literatura, lo que dificulta la validación cruzada y el avance colaborativo.

CONCLUSIONES

En conclusión, la integración de arquitecturas basadas en transformers con técnicas de adaptación eficiente de parámetros (como LoRA y el preentrenamiento continuo) representa el actual estado del arte en el análisis de sentimiento. Esta convergencia tecnológica logra un equilibrio óptimo entre un alto rendimiento contextual y la viabilidad computacional necesaria para su implementación en entornos operativos reales, democratizando el acceso a la inteligencia artificial avanzada sin los costos prohibitivos de entrenar modelos masivos desde cero.

Los sesgos sistemáticos en la detección de polaridades (que suelen favorecer la identificación de sentimientos negativos), las brechas de precisión para lenguas minoritarias y la falta de reproducibilidad en los experimentos exigen ir más allá de los avances puramente técnicos, en la que, resulta imperativo establecer marcos de evaluación estandarizados y promover prácticas de investigación abierta que permitan validar y replicar los hallazgos.

La integración de la Inteligencia Artificial Explicable (XAI) desde la concepción misma de la arquitectura, y no como un simple añadido posterior o post-hoc, es un requisito fundamental para facilitar la auditoría algorítmica, permitir la mejora iterativa de los sistemas y garantizar la confianza del usuario, aspectos innegociables al desplegar estas herramientas en aplicaciones de alto impacto. Para enriquecer este panorama, futuros estudios podrían y deberían ampliar las estrategias de búsqueda hacia repositorios regionales, conferencias locales o preprints, con el objetivo de capturar una diversidad metodológica y lingüística que a menudo escapa a los circuitos académicos tradicionales.

FINANCIACIÓN

La presente investigación no recibió financiamiento específico de agencias públicas, instituciones académicas o entidades privadas. El trabajo se desarrolló con recursos de los autores en el marco de actividades académicas de formación investigativa.

CONFLICTO DE INTERESES

Se declara que no existe conflicto de intereses, financiero, personal o profesional, que pudiera haber influido en la interpretación o presentación de los resultados de esta revisión sistemática.

CONTRIBUCIÓN DE AUTORÍA

Tabla de contribución

Contribución de autoría según la taxonomía CRediT

Participar activamente en:	Camacho	Flores	Tierra	Orozco	Arteaga	Damian
Conceptualización	X	X	X	–	–	–
Análisis formal	X	–	X	X	X	–
Adquisición de fondos	–	–	–	–	–	–
Investigación	–	X	X	X	X	X
Metodología	X	X	–	–	X	X
Administración del proyecto	X	X	–	–	–	X
Recursos	–	X	–	X	X	X
Redacción – borrador original	X	X	X	X	X	X
Redacción – revisión y edición	X	X	X	X	X	X
Discusión de los resultados	X	X	X	X	X	X
Revisión y aprobación de la versión final	X	X	X	X	X	X

Nota: Distribución declarada conforme a la taxonomía CRediT.

REFERENCIAS

Abdelrahman, A., et al. (2025). Exploring zero-shot SLM ensembles as an alternative to LLMs for sentiment analysis. Information Fusion. https://www.sciencedirect.com/science/article/pii/S1566253525007389

Adak, A., Pradhan, B., & Shukla, N. (2022). Sentiment Analysis of Customer Reviews of Food Delivery Services Using Deep Learning and Explainable Artificial Intelligence: Systematic Review. Foods, 11(10), 1500. https://doi.org/10.3390/foods11101500

Alahmadi, K., Alharbi, S., Chen, J., & Wang, X. (2025). Generalizing sentiment analysis: a review of progress, challenges, and emerging directions. Social Network Analysis and Mining, 15(1), 45. https://doi.org/10.1007/s13278-025-01461-8

Baberwal, S. K., Shelke, N. A., & Anwar, K. (2025). Systematic review of recent advances in multimodal sentiment analysis. Discover Computing, 28(1), 270. https://doi.org/10.1007/s10791-025-09727-7

Bernal-Beltrán, T., Paredes-Valverde, M. A., Salas-Zárate, M. d. P., García-Díaz, J. A., & Valencia-García, R. (2025). Sentiment Analysis in Mexican Spanish: A Comparison Between Fine-Tuning and In-Context Learning with Large Language Models. Future Internet, 17(10), 445. https://doi.org/10.3390/fi17100445

Chae, Y., & Davidson, T. (2025). Large language models for text classification: From zero-shot learning to instruction-tuning. Sociological Methods & Research. https://journals.sagepub.com/doi/10.1177/00491241251325243

Chandan, M. K., & Mandal, S. (2025). A comprehensive survey on sentiment analysis: Framework, techniques, and applications. Computer Science Review, 58, 100777. https://doi.org/10.1016/j.cosrev.2025.100777

Chauhan, S. B., & Mavani, J. M. (2025). Transformer-Based Sentiment Analysis on Amazon Reviews Using Kaggle Dataset. International Journal of Innovative Science and Research Technology, 10(12). https://doi.org/10.38124/ijisrt/25dec103

Cui, J., Wang, Z., Ho, S.-B., & Cambria, E. (2023). Survey on sentiment analysis: evolution of research methods and topics. Artificial Intelligence Review, 56(8), 8469–8510. https://doi.org/10.1007/s10462-022-10386-z

Daza, A., González Rueda, N. D., Aguilar Sánchez, M. S., Robles Espíritu, W. F., & Chauca Quiñones, M. E. (2024). Sentiment Analysis on E-Commerce Product Reviews Using Machine Learning and Deep Learning Algorithms: A Bibliometric Analysis, Systematic Literature Review, Challenges and Future Works. International Journal of Information Management Data Insights, 4(2), 100267. https://doi.org/10.1016/j.jjimei.2024.100267

Hasan, M. A., et al. (2024). Zero- and few-shot prompting with LLMs: A comparative study with fine-tuned models for Bangla sentiment analysis. arXiv. https://arxiv.org/abs/2308.10783

Indriyatmoko, T., Rahardi, M., Utama, H., & Frobenius, A. C. (2026). Comparative Analysis of BERT and LSTM Models for Sentiment Classification of Mobile Game User Reviews. Journal of Applied Informatics and Computing, 10(1), 1093-1100. https://doi.org/10.30871/jaic.v10i1.12149

Jahin, M. A., Shovon, M. S. H., Mridha, M. F., Islam, M. R., & Watanobe, Y. (2024). A hybrid transformer and attention based recurrent neural network for robust and interpretable sentiment analysis of tweets. Scientific Reports, 14(1), 24882. https://doi.org/10.1038/s41598-024-76079-5

Lakshmi, L., Ali, B. M., Sree Devi, K. D., Rafiq, M., Shernazarov, I., Othman, N. A., & Khan, M. I. (2025). Opinion mining in e-commerce: Evaluating machine learning approaches for sentiment analysis. Results in Control and Optimization, 19(7), 100575. https://doi.org/10.1016/j.rico.2025.100575

Lowmanstone, L., Wan, R., Owan, R., Kim, J., & Kang, D. (2023). Annotation Imputation to Individualize Predictions: Initial Studies on Distribution Dynamics and Model Predictions. arXiv. https://arxiv.org/abs/2305.15070v3

Mabokela, K. R., Primus, M., & Celik, T. (2024). Explainable Pre-Trained Language Models for Sentiment Analysis in Low-Resourced Languages. Big Data and Cognitive Computing, 8(11), 160. https://doi.org/10.3390/bdcc8110160

Mabrouk, A., Redondo, R. P. D., & Kayed, M. (2020). Deep Learning-Based Sentiment Classification: A Comparative Survey. IEEE Access, 8, 85616-85638. https://doi.org/10.1109/ACCESS.2020.2992013

Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. Journal of King Saud University - Computer and Information Sciences, 36(4), 102048. https://doi.org/10.1016/j.jksuci.2024.102048

Merayo, N., Vegas, J., Llamas, C., & Fernández, P. (2023). Social Network Sentiment Analysis Using Hybrid Deep Learning Models. Applied Sciences, 13(20), 11608. https://doi.org/10.3390/app132011608

Meshkin, H., et al. (2024). Harnessing large language models' zero-shot and few-shot learning capabilities for regulatory research. Briefings in Bioinformatics, 25(5), bbae354. https://academic.oup.com/bib/article/25/5/bbae354/7739674

Mohammadi, H., Giachanou, A., & Bagheri, A. (2025). Explainability in Practice: A Survey of Explainable NLP Across Various Domains. arXiv preprint, arXiv:2502.00837. https://arxiv.org/abs/2502.00837

Osman, A., et al. (2025). A review on NLP zero-shot and few-shot learning: Methods and applications. Discover Applied Sciences. https://link.springer.com/article/10.1007/s42452-025-07225-5

Rodríguez-Ibáñez, M., Gimeno-Blanes, F. J., Cuenca-Asensi, A., Alonso-García, J. I., &... (2023). A review on sentiment analysis from social media platforms. Expert Systems with Applications, 223, 119862. https://doi.org/10.1016/j.eswa.2023.119862

Venkit, P. N., Srinath, M., Gautam, S., Venkatraman, S., Gupta, V., Passonneau, R. J., & Wilson, S. (2023). The Sentiment Problem: A Critical Survey towards Deconstructing Sentiment Analysis. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 13743–13763). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.848

Wang, G., & Jaber, M. M. (2025). A Deep Learning Approach to Sentiment Analysis of Hotel Reviews: Comparing BERT and LSTM Models. International Journal of Advances in Artificial Intelligence and Machine Learning, 2(2), 67-74. https://doi.org/10.58723/ijaaiml.v2i2.403

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55(7), 5731–5780. https://doi.org/10.1007/s10462-022-10144-1