

Procesamiento de Lenguaje Natural Aplicado a la Selección de Personal

Natural Language Processing Applied to Recruitment

Luisa María Carabali-Sanchez ^{1*}

¹ Universidad Distrital Francisco José De Caldas, Colombia;

Orcid: <https://orcid.org/0009-0000-2322-4048>

* Correo para correspondencia: lmcarabalis@udistrital.edu.co

Resumen

En la actualidad, la información crece exponencialmente cada día y con esto las técnicas tradicionales de selección de personal se han transformado ya que las organizaciones están comenzando a adoptar y aprovechar la funcionalidad de la inteligencia artificial en sus procesos de contratación. Este estudio examina la aplicación del Procesamiento de Lenguaje Natural mediante una metodología de procesamiento de texto diseñada para identificar relaciones entre perfiles de candidatos y ofertas laborales. El desarrollo del proyecto fue compuesto por la captura de datos de la red social LinkedIn utilizando Web Scraping (raspado web), una limpieza, adecuación y transformación de la información, para utilizar diferentes modelos de Word Embeddings y Transformes en el contexto de Procesamiento de Lenguaje Natural a fin de clasificar los candidatos más compatibles con la oferta laboral y así generar un nuevo conjunto de herramientas que faciliten la toma de decisiones en la selección de personal.

Palabras clave: procesamiento del lenguaje natural, reclutamiento, empleabilidad, coincidencia de empleo, automatización.

Abstract

Information is growing exponentially, and with this, traditional personnel selection techniques have been transformed as organizations have begun to adopt and take advantage of artificial intelligence in their hiring processes. This work addresses the use of Natural Language Processing by proposing a text-processing methodology that helps determine relationships between candidate profiles and job offers. The project involves capturing data from the LinkedIn social network using Web Scraping, then cleaning, adapting, and transforming the information to utilize different Word embedding and Transformer models in Natural Language Processing. The goal is to classify the candidates most compatible with the job offers, thus generating a new set of tools to facilitate decision-making in personnel selection.

Keywords: natural language processing, recruitment, employability, job matching, automation.

Introducción

En esta nueva era de información, se han llevado a cabo pluralidad de desarrollos, creando una amplia gama de herramientas e infraestructura para procesar información y con ello gran cantidad de organizaciones utilizan procesos de selección de personal totalmente digitalizados, que permitan una fácil clasificación, comunicación y evaluación de sus candidatos. Existen diferentes alternativas para realizar este tipo de tareas como: una hoja de vida, un formulario web, un vídeo, o un perfil en redes sociales, etc. La capacidad de acceder a grandes cantidades de información se ha vuelto mucho más fácil.

En la actualidad el análisis de texto ayuda a extraer significados, patrones y estructuras ocultas en los datos no estructurados (Sabry, 2023). Entre algunos ejemplos tenemos el análisis de sentimientos en las redes sociales, las herramientas de traducción en línea, el reconocimiento de voz, la indexación de documentos en la web, entre otros como la clasificación del correo spam.

Este artículo presenta el resumen del desarrollo de una herramienta de apoyo en la selección de personal junto con su implementación en un caso particular obtenido con el trabajo de grado realizado para la maestría en Ingeniería Industrial en la Universidad Distrital Francisco José de Caldas.

El procesamiento del lenguaje natural (PLN) ha evolucionado significativamente, integrando técnicas de Deep Learning para extraer valor de los textos y encontrar relaciones entre ellos, lo que permite generar conocimiento implícito que no estaba explícitamente presente en las colecciones textuales originales (Cesar & Daniel, 2008). Estas técnicas se han aplicado ampliamente en el campo del reclutamiento y la selección de personal, donde la automatización y el análisis semántico juegan un papel crucial en la identificación y evaluación de candidatos.

Por ejemplo, el estudio de Adugna, Ansari, Fathi, y Tijdens (2018) profundiza en el análisis semántico de descripciones de empleo para mejorar la coincidencia de candidatos en los procesos de reclutamiento. Su enfoque se centra en aprovechar las técnicas de PLN para vincular las habilidades y requisitos mencionados en las ofertas de trabajo con los perfiles de los candidatos. Esto es coherente con investigaciones actuales que han utilizado PLN para extraer información clave y optimizar el proceso de selección, como en el caso de Resume Parser with Natural Language Processing (Sroison & Chan, 2021), donde se presenta un sistema automatizado que convierte currículums a formato de texto y compara la similitud con descripciones de puestos.

De manera similar, en “Automated Resume Screener using Natural Language Processing (NLP)” (Harsha, 2022), se desarrolla un sistema automatizado que analiza currículums y ofertas laborales utilizando técnicas de PLN para extraer información relevante y evaluar las habilidades de los solicitantes. Esta investigación sigue la línea de estudios anteriores que buscan optimizar los procesos de selección automatizados mediante la clasificación y evaluación de candidatos, mejorando la eficiencia y la precisión del reclutamiento. Estos enfoques se alinean con el trabajo de Odili et al. (2024), quienes examinan el impacto de la IA en los procesos de selección dentro de la industria del petróleo y gas, donde las tecnologías avanzadas están transformando las metodologías tradicionales de reclutamiento y eliminando sesgos.

Además, estudios como “Natural Language Processing and Text Mining to Identify Knowledge Profiles for Software Engineering Positions” (Rogelio et al., 2017) y “Data Mining Approach to Monitoring The Requirements of the Job Market” (Ioannis, Barla, & Aristides) muestran cómo la minería de texto y el PLN se pueden aplicar para identificar perfiles técnicos y tendencias laborales, respectivamente, ayudando a las empresas a encontrar candidatos adecuados y a anticiparse a las demandas del mercado. Estas técnicas también son fundamentales en la recomendación de empleos, como se observa en “Machine Learned Job Recommendation” (Ioannis, Barla, & Aristides), donde el aprendizaje automático se utiliza para predecir las transiciones laborales basadas en datos históricos, lo que mejora la precisión en la coincidencia de candidatos.

Finalmente, estos enfoques están complementados por soluciones más sofisticadas como las presentadas en “Text Analysis for Job Matching Quality Improvement” (Yasunobu et al., 2017), que se centran en extraer palabras claves que mejoran la calidad de la coincidencia entre candidatos y ofertas laborales, y en “Matching People and Jobs: A Bilateral Recommendation Approach” (Jochen et al., 2006), donde se aplican sistemas de recomendación para optimizar tanto el lado del candidato como el del empleador, cerrando el ciclo de innovación en el uso del PLN y la IA en el campo del reclutamiento.

En conjunto, estas investigaciones ilustran cómo el procesamiento del lenguaje natural y las técnicas de aprendizaje profundo están transformando los procesos de reclutamiento, desde la automatización de la selección de currículums hasta la predicción de transiciones laborales y la mejora de la calidad en la coincidencia entre candidatos y vacantes.

Este trabajo aborda el uso de técnicas de Procesamiento de Lenguaje Natural (PLN) y Deep Learning para mejorar los procesos de selección de personal. A continuación, describimos las principales contribuciones del estudio:

- **Extracción de datos mediante Web Scraping:** Se desarrolló un proceso automatizado para capturar información desde LinkedIn, que permite acceder a un vasto conjunto de perfiles profesionales. Este método proporciona un enfoque actualizado para obtener datos en tiempo real sobre candidatos potenciales, facilitando la actualización y expansión constante de los datos disponibles para el estudio.
- **Limpieza, adecuación y transformación de la información:** El estudio incluye un preprocesamiento de datos, que abarca la limpieza y estructuración de la información obtenida a través del Web Scraping. Estas técnicas aseguran la calidad de los datos, eliminando ruido y preparando los textos para el análisis mediante algoritmos de PLN y modelos de aprendizaje profundo.

- **Aplicación de modelos de Word Embeddings y Transformers:** Se emplearon modelos Word Embeddings (como FastText) y Transformers (como BERT y Universal Sentence Encoder) para representar semánticamente los textos extraídos de los perfiles de candidatos y las ofertas laborales. Estas representaciones contextuales mejoran la capacidad del sistema para captar relaciones más complejas entre candidatos y vacantes, más allá de las coincidencias léxicas.
- **Generación de herramientas para la toma de decisiones:** El resultado de este proyecto incluye el desarrollo de una opción práctica integrando el PLN para proporcionando recomendaciones sobre los mejores currículos para cada oferta laboral, lo que optimiza el proceso de contratación.

Este estudio contribuye a la evolución del reclutamiento automatizado, integrando metodologías innovadoras que aprovechan los avances recientes en inteligencia artificial para transformar el proceso de selección de personal en las organizaciones.

Algunas limitaciones que deben ser consideradas para una adecuada interpretación de los resultados. En primer lugar, el tamaño y la representatividad de los datos extraídos de LinkedIn mediante raspado web pueden no reflejar la totalidad del mercado laboral.

Debido al alto uso de recursos computacionales necesarios para entrenar modelos de Deep Learning y Word Embeddings, el alcance de las pruebas y ajustes del modelo presentado en este artículo se ha limitado a la comparación de perfiles y ofertas laborales en el ámbito de la ingeniería, utilizando un subconjunto de 10 ofertas laborales y 10 currículums de candidatos. Esta selección específica permite una evaluación detallada y efectiva en el contexto del estudio.

Materiales y Métodos

Se utilizó el lenguaje de programación Python en su versión 3.10 a través de Google Colab debido a su compatibilidad con una serie de bibliotecas esenciales para el procesamiento y análisis de datos textuales. El Natural Language Toolkit (NLTK) se utiliza para el procesamiento del lenguaje en español, abarcando tareas cruciales como la tokenización, la eliminación de stopwords y el stemming, que son fundamentales para preparar el texto para el análisis.

La biblioteca Gensim desempeña un papel crucial al facilitar la creación y el manejo de modelos de Word Embeddings, tales como FastText. Convierte palabras en vectores numéricos que capturan las relaciones semánticas entre ellas, mientras que FastText mejora esta representación al considerar subpalabras, permitiendo así el tratamiento de palabras raras o desconocidas.

Para la generación de representaciones contextuales más avanzadas del texto, se emplean Transformers y BETO (Bidirectional Encoder Representations BERT en Español). BETO permite una comprensión más profunda del contexto en el que las palabras aparecen dentro de una oración, mejorando significativamente la precisión en tareas como la clasificación de texto y el análisis de sentimientos. Además, el Universal Sentence Encoder proporciona representaciones vectoriales de oraciones completas, facilitando la comparación semántica entre textos y contribuyendo a una evaluación más precisa de la similitud entre documentos.

Finalmente, scikit-learn (sklearn) se emplea para el pre procesamiento de datos y la implementación de algoritmos de aprendizaje automático, incluyendo el cálculo de similitudes de coseno entre vectores de texto, lo cual es esencial para la comparación y clasificación de documentos.

En conjunto, estas herramientas permiten un análisis exhaustivo de textos, desde el procesamiento inicial hasta la comparación y clasificación de datos, optimizando la capacidad de extraer y evaluar patrones y relaciones en los datos textuales.

Diseño de la Investigación

Este artículo presenta una investigación que combina enfoques aplicados y cuantitativos para abordar un problema práctico en el ámbito del procesamiento de lenguaje natural (PLN). La investigación aplicada se centra en desarrollar soluciones concretas a problemas reales. En este caso, se busca mejorar el proceso de selección de personal mediante la aplicación de técnicas avanzadas de PLN. La investigación aplicada se orienta hacia la creación de un sistema que optimice la correspondencia entre currículos y ofertas laborales, lo cual tiene una relevancia directa y práctica en el ámbito de la contratación (Bai & Li, 2021).

Por otro lado, la naturaleza cuantitativa de esta investigación implica el uso de datos numéricos y métodos estadísticos para analizar y evaluar los resultados del sistema propuesto. En este estudio, se emplean técnicas de análisis cuantitativo para clasificar a los candidatos desde los más adecuados hasta los menos adecuados para un puesto, basándose en medidas de similitud semántica. Se comparan los resultados obtenidos mediante modelos de Word Embeddings y Transformers, estos se comparan con la clasificación realizada por un reclutador humano, utilizando métricas estadísticas para evaluar la precisión y eficacia del modelo.

Así, la combinación de investigación aplicada y cuantitativa permite abordar el problema práctico de manera efectiva y respaldar las soluciones propuestas con análisis numérico riguroso.

Población y Muestra

- **Población:** La población de estudio está compuesta por descripciones de currículos y ofertas laborales provenientes de archivos en PDF del Servicio Público de Empleo en Colombia y ofertas laborales recolectadas de LinkedIn. Estas ofertas están enfocadas en cargos de ingenierías.
- **Muestra:** La muestra incluye un subconjunto de 10 currículos y 10 ofertas laborales que se utilizan para entrenar y evaluar los modelos de Embeddings.

Técnicas de recolección

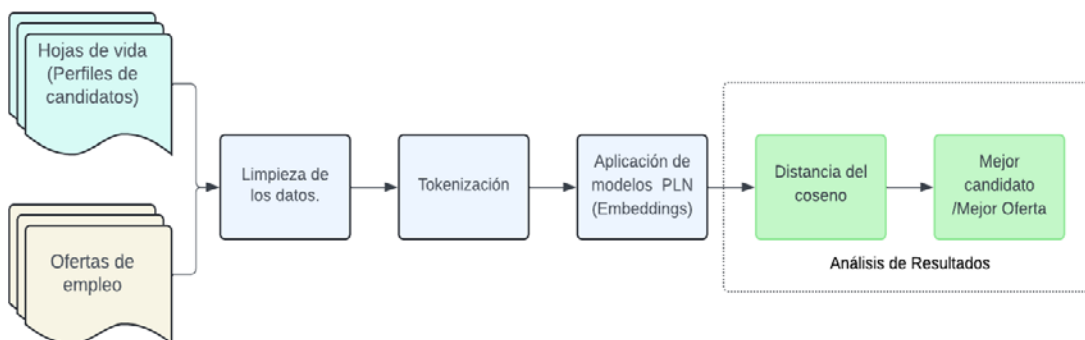
- **Web Scraping:** Las ofertas laborales se recolectan mediante un Bot de Web Scraping que utiliza la biblioteca Selenium para extraer datos de LinkedIn. El Bot automatiza el proceso de navegación en LinkedIn, la búsqueda de ofertas laborales, y la recopilación de descripciones de empleo.
- **Bases de datos:** Los currículos fueron obtenidos mediante la colaboración con reclutadores que tienen acceso a la base de datos del Servicio Público de Empleo en Colombia.

Esta metodología de recolección de datos garantiza que la investigación se base en información real, relevante y significativa, al mismo tiempo que asegura la protección de la privacidad de los candidatos involucrados.

Metodología

Nuestra metodología se estructura en varias etapas clave que garantizan una evaluación precisa y contextualizada de las similitudes entre currículums y ofertas de trabajo como se detalla en el diagrama de flujo de la figura 1.

Figura 1



Pre-procesamiento de Datos

- **Tokenización y Limpieza:** Los textos de los currículums y las ofertas de trabajo son tokenizados y limpiados para eliminar puntuaciones y caracteres no deseados. Esto asegura que el texto esté en una forma que facilita el análisis.
- **Eliminación de stopwords y lematización:** Se eliminan las stopwords (palabras comunes que no aportan significado relevante) y se realiza la lematización para normalizar las palabras, lo que mejora la calidad de los Embeddings al reducir la variabilidad lingüística.

Aplicación de Modelos de Embeddings

- **FastText:** Utilizamos este modelo para generar representaciones vectoriales densas de palabras. FastText captura la semántica de las palabras y sus relaciones contextuales, proporcionando vectores que reflejan significados similares en contextos similares. Esta técnica supera las limitaciones de los métodos tradicionales basados en frecuencia de término.
- **BETO:** Incorporamos BETO, un modelo BERT pre entrenado en español, que ofrece Embeddings contextuales profundos. BETO entiende el contexto completo de cada palabra dentro de una oración, mejorando significativamente la representación semántica y la precisión en la correspondencia de texto.
- **Universal Sentence Encoder:** Utilizamos este modelo para obtener Embeddings de oraciones completas. Esta técnica permite comparar frases y párrafos en lugar de solo palabras, facilitando una evaluación más holística de las similitudes entre currículums y ofertas de trabajo.

Análisis de resultados

Para evaluar la precisión y eficacia de los modelos de procesamiento de lenguaje natural en la correspondencia entre currículums y ofertas de trabajo, se calcularon las similitudes entre las representaciones vectoriales generadas por los modelos utilizando la similitud de coseno. Este método mide la proximidad entre vectores en

el espacio de Embeddings, proporcionando una la similitud semántica entre los textos.

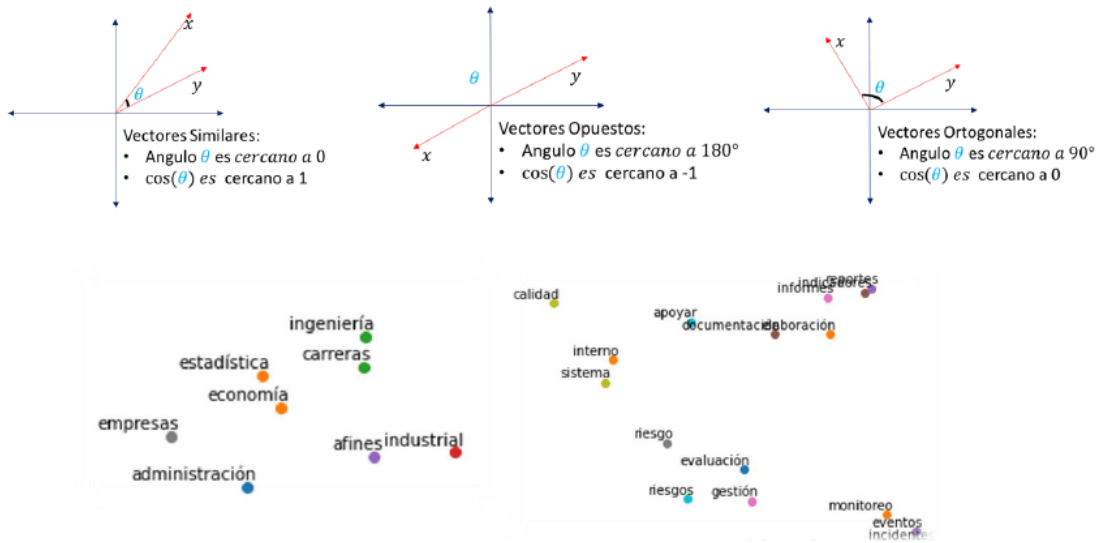
En paralelo, se llevó a cabo un análisis adicional utilizando el coeficiente de Spearman para comparar los rankings de similitud generados por los distintos modelos con el ranking proporcionado por un reclutador humano. Un alto coeficiente de Spearman sugiere que los algoritmos están alineados con el juicio humano, mientras que un bajo coeficiente indica posibles áreas de mejora en los modelos.

Resultados

El proceso de evaluación de currículums comienza con el cálculo de la similitud semántica entre las descripciones de los currículums y las ofertas de trabajo utilizando la similitud de coseno. Esta técnica permite cuantificar el grado de alineación entre los vectores que representan las descripciones textuales, asignando valores dentro de un rango de 0 a 1. Un coeficiente próximo a 1 sugiere una alta correspondencia entre el perfil del candidato y los requisitos del puesto, mientras que un valor cercano a 0 refleja una menor afinidad. A partir de esta información, se genera un ranking de candidatos para cada oferta, priorizando aquellos perfiles cuya representación vectorial presente un coeficiente de similitud del coseno más próximo a uno. Con estos valores, los currículums se clasifican en función de su correspondencia con los requisitos del puesto.

Figura 2

Representaciones vectoriales en base a la Similitud del Coseno.



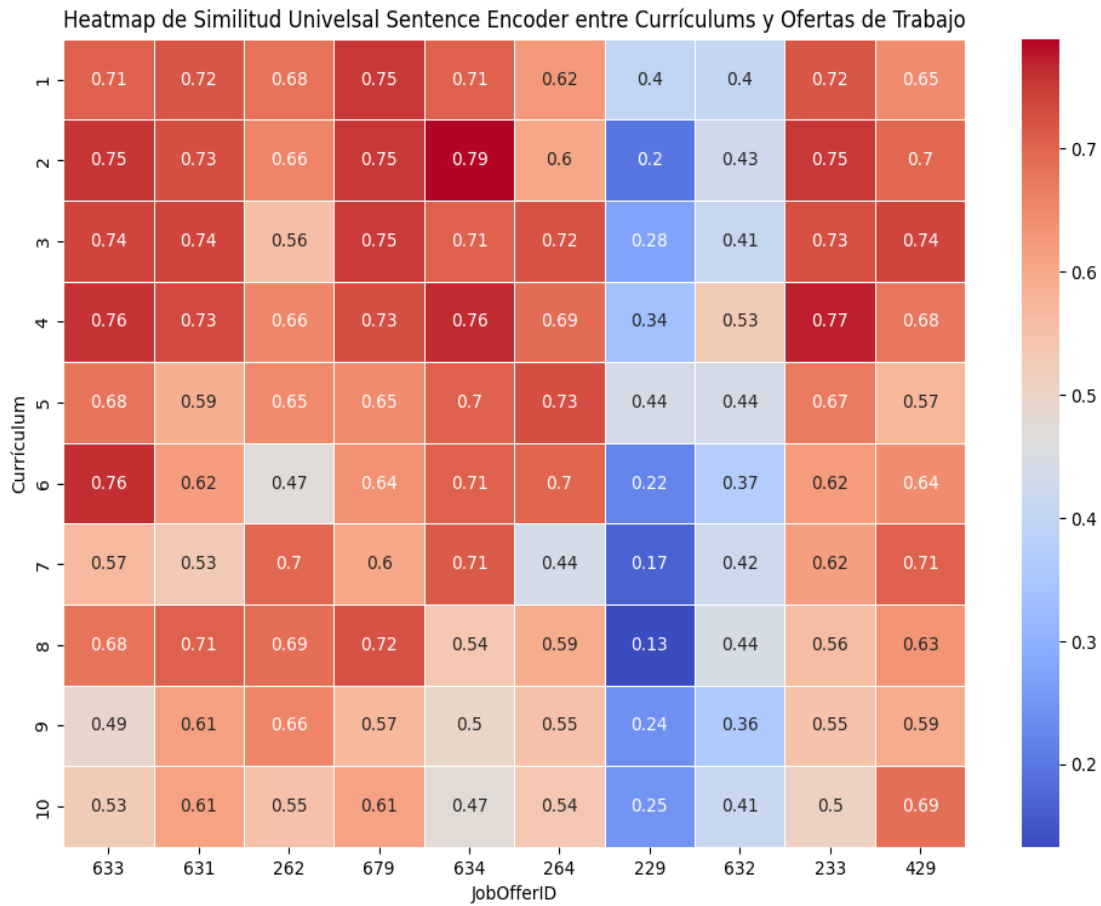
Nota: Esta métrica cuantifica la proximidad entre los vectores que representan los textos, produciendo valores que van de 0 a 1. Un valor cercano a 1 indica una alta similitud semántica, mientras que un valor cercano a 0 indica una baja similitud

En la figura 3, presentamos un mapa de calor correspondiente al resultado encontrado al evaluar el coseno de las distancias entre los 10 currículos de candidatos y las 10 ofertas laborales. Se selecciona el modelo Universal Sentence Encoder (USE) ya que es de los que presenta mejores resultados encontrando así una mayor similitud entre el contenido y contexto semántico en el texto de los currículums y las ofertas laborales.

La figura 4, presenta la matriz de calor de los resultados, donde se evidencia que el modelo que tiene más cercanía con el ranking realizado por un reclutador humano. En el análisis de correlaciones de Spearman entre currículums y ofertas de trabajo, el modelo Universal Sentence Encoder (USE) demostró ser el más efectivo, alcanzando correlaciones significativas y altas en múltiples ofertas, lo que sugiere una mayor capacidad para capturar similitudes semánticas relevantes. Por otro lado, FastText y BERT (BETO) mostraron resultados menos consistentes y algunas correlaciones negativas, indicando posibles desajustes entre los currículums y las ofertas de trabajo. En general, USE se destacó en el análisis, sin embargo, se debe tener en cuenta que el Modelo de BETO contiene resultados similares.

Figura 3

Mapa de Calor del Coseno de las Distancias para el modelo USE.



Al comprar los p-values el modelo de USE tiene un p-value menor que 0.05 (Representado en verde) en varios de los casos, y de ahí podemos concluir que hay una correlación significativa entre los rankings generados por el modelo y los rankings humanos. Es decir que existe evidencia de que el modelo captura la semántica y el contexto completo de las frases, facilitando así una comprensión más precisa y útil del lenguaje en similitud de como lo hace un humano.

Figura 4

Matriz de calor Coeficiente de Spearman y P-Value para combinaciones de currículos de candidatos Vs ofertas.

	Fast Text		USE		BETO	
	Correlación	pValue	Correlación	pValue	Correlación	pValue
CM1 Vs Ofertas	0.418	0.229	0.648	0.043	0.394	0.260
CM2 Vs Ofertas	0.115	0.751	0.636	0.048	0.842	0.002
CM3 Vs Ofertas	0.139	0.701	0.115	0.751	0.261	0.467
CM4 Vs Ofertas	-0.297	0.405	0.867	0.001	0.891	0.001
CM5 Vs Ofertas	0.673	0.033	0.770	0.009	0.612	0.060
CM6 Vs Ofertas	0.539	0.108	0.576	0.082	0.564	0.090
CM7 Vs Ofertas	0.879	0.001	0.406	0.244	0.709	0.022
CM8 Vs Ofertas	-0.370	0.293	0.103	0.777	-0.006	0.987
CM9 Vs Ofertas	-0.127	0.726	0.891	0.001	0.612	0.060
CM10 Vs Ofertas	-0.382	0.276	0.285	0.425	0.067	0.855

Nota: Un p-Value < 0.05 sugiere que la probabilidad de que la diferencia observada sea producto del azar es menor al 5 %, lo que se considera estadísticamente significativo.

Figura 5

Matriz de calor Coeficiente de Spearman y P-Value en el sentido contrario para combinaciones de ofertas laborales Vs currículos de candidatos.

	Fast Text		USE		BETO	
	Correlación	pValue	Correlación	pValue	Correlación	pValue
Oferta1 Vs CMs	0.404	0.247	0.564	0.090	0.164	0.651
Oferta2 Vs CMs	0.398	0.254	0.685	0.029	0.564	0.090
Oferta3 Vs CMs	0.603	0.065	0.721	0.019	0.818	0.004
Oferta4 Vs CMs	0.426	0.220	0.370	0.293	0.467	0.174
Oferta5 Vs CMs	0.267	0.457	0.200	0.580	0.297	0.405
Oferta6 Vs CMs	-0.423	0.223	0.758	0.011	0.685	0.029
Oferta7 Vs CMs	-0.042	0.909	0.297	0.405	-0.042	0.907
Oferta8 Vs CMs	-0.528	0.117	0.285	0.425	0.067	0.855
Oferta9 Vs CMs	0.200	0.580	0.685	0.029	0.624	0.054
Oferta 10 Vs CMs	-0.332	0.349	0.248	0.489	0.358	0.310

Discusión

Los resultados encontrados son consistentes con investigaciones previas que exploran el uso de técnicas de procesamiento de lenguaje natural (PLN) en el ámbito del reclutamiento. El trabajo de Adugna, Ansari, Fathi, y Tijdens (2018) destaca la importancia del análisis semántico en las descripciones de empleo, similar a nuestro enfoque con USE que identifica el contexto y logra realizar coincidencias clave entre currículums y ofertas laborales.

De manera similar, los estudios de Sroison y Chan (2021) y Harsha (2022) presentan sistemas automatizados que, como nuestro trabajo, emplean técnicas de PLN para extraer información relevante de currículums y compararlos con las descripciones de las ofertas. Ambos estudios coinciden en que la automatización de estos procesos puede mejorar tanto el tiempo como la eficiencia en la selección de candidatos.

Por otro lado, Odili et al. (2024) exploran el impacto de la inteligencia artificial (IA) en el reclutamiento dentro de la industria del petróleo y gas, donde se observa una transformación en los métodos tradicionales de selección. Este enfoque es coherente con nuestros resultados, ya que el uso de USE puede ser una herramienta poderosa para reducir sesgos y mejorar la toma de decisiones en los procesos de reclutamiento.

Finalmente, los trabajos de Yasunobu et al. (2017) y Jochen et al. (2006) introducen el uso de sistemas de recomendación y análisis de texto para mejorar la calidad de las coincidencias en el proceso de selección. Estos enfoques, alineados con nuestro experimento del modelo Universal Sentence Encoder (USE), pueden ser relevantes como una herramienta PLN e IA en la modernización del reclutamiento, optimizando tanto la experiencia del candidato como la del empleador.

En conclusión, el estudio destaca la efectividad de los modelos de embeddings como FastText, BERT y Universal Sentence Encoder (USE) para transformar texto no estructurado en representaciones vectoriales. Se observó que algunos de estos modelos lograron mayor precisión en la identificación de candidatos, aunque su

desempeño depende del contexto y los datos utilizados. La elección del modelo más adecuado debe ajustarse a los objetivos y necesidades específicas de la tarea.

Los resultados obtenidos en este estudio muestran que el modelo Universal Sentence Encoder (USE) se destaca como una herramienta útil para capturar la semántica y el contexto de los currículums en comparación con las ofertas laborales, presentando una alta concordancia con los rankings humanos. El análisis del coseno de las distancias y el coeficiente de Spearman revelan una correlación significativa, lo que indica que el modelo puede emular de manera efectiva la evaluación humana en procesos de selección. Este hallazgo respalda el uso de este modelo embedding como un recurso clave en la mejora de los sistemas automatizados de reclutamiento, ofreciendo una evaluación útil para mejorar la calidad y eficiencia en la selección de candidatos.

La evaluación del modelo mediante estadística inferencial y distancias del coseno permite analizar su eficacia incluso sin valores esperados. Conjuntamente, la validación manual de las ofertas de trabajo y su comparación con los resultados del modelo mejoran la exactitud. En este contexto la metodología propuesta facilita la identificación de similitudes entre descripciones de puestos y habilidades de los postulantes, mejorando el proceso de selección de personal.

Los hallazgos obtenidos representan un avance en la automatización de los procesos de reclutamiento y destacan la necesidad de futuras investigaciones sobre la adaptabilidad de los modelos de Deep Learning en distintos sectores laborales e industriales. Dado que la estructura y el contenido de las ofertas varían, es fundamental evaluar su capacidad de generalización en contextos con datos heterogéneos. Dado que las ofertas de trabajo pueden variar en su forma y nivel de detalle según la empresa o la industria, es importante evaluar si los modelos pueden identificar correctamente la relación entre candidatos y puestos en diferentes contextos, sin necesidad de ajustes específicos para cada caso. Además, se sugiere la integración de enfoques híbridos que combinen embeddings tradicionales con arquitecturas avanzadas, como transformers especializados en emparejamiento

semántico.

Agradecimientos

Agradecemos a la Universidad Distrital Francisco José de Caldas, mis profesores de la maestría en especial Gerardo Muñoz por su apoyo en este proyecto y a mi profesor al profesor de NPL Carlos Isaac Zainea por compartir su conocimiento y hacerme amar esta materia. A mis tutores en especial al profesor Sergio Rojas por su apoyo resolviendo dudas y a los reclutadores anónimos que me apoyaron en este estudio.

Referencias bibliográficas

- Alberto, M. C. (2019). Utilización del machine learning en la industria 4.0. Universidad de Valladolid.
- Alicia García-García, X. G.-M.-M. (2014). Text mining versus redes neuronales. Obtenido de <http://eprints.rclis.org/23152/1/pesetcrecs14.pdf>
- Andrés, H., & Eugenio, M. (2021). Gestión De Personas En Organizaciones Innovadoras: Manual De Teoría Y Practica. Argentina: Ediciones Granica.
- A dugna Chala, S., Ansari, F., Fathi, M., & Tijdens, K. (2018). Semantic matching of job seeker to vacancy: A bidirectional approach. International Journal of Manpower, 1047-2063. London: Emerald Publishing Limited.
- Apaza Alanoca, H., Rubin de Celis Vidal, A., & Chire Saire, J. E. (2021). Curriculum Vitae Recommendation Based on Text Mining.
- Ayub Zubeda, J. A., Ayyas Shaheen, M. A., Narsayya Godavari , G. R., & Sadiq Naseem, S. Z. (2019). Resume Ranking using NLP and Machine Learning.
- Bai, X., & Li, J. (2021). Applied Research of Knowledge in the Field of Artificial Intelligence in the Intelligent Retrieval of Teaching Resources. Scientific Programming.
-

-
- Camacho-Collados, J., & Taher Pilehvar, M. (2020). Embeddings in natural language processing. En Proceedings of the 28th International Conference on Computational Linguistics: Tutorial Abstracts (pp. 10-15). <https://doi.org/10.18653/v1/2020.coling-tutorials.2>
- Cesar, P. L., & Daniel, S. G. (2008). Minería de datos. Técnicas y herramientas: técnicas y herramientas. Madrid: Editorial Paraninfo.
- Chavan, P. R., Chandurkar, Y., Tidake, A., Lavankar, G., Gaikwad, S., & Chavan, R. (2024). Enhancing recruitment efficiency: An advanced Applicant Tracking System (ATS). Pune, India: Arts and Science Press Pte. Ltd.
- Cifuentes Cortés, J. D., & Martínez Naizaque, J. M. (2018). Revisión sistemática para las técnicas de minería web de. Bogota D.C: Universidad Católica de Colombia.
- Devlin, J., Chang, M.-W., Kenton, L., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. En Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.
- EQUIPOS Y TALENTO . (2020, 08 06). Obtenido de EQUIPOS Y TALENTO : <https://www.equiposytalento.com/noticias/2020/06/08/un-36-de-las-empresas-aumentara-los-procesos-de-seleccion-digital>
- Garcia de Sousa Neto, M., & Saraiva, F. (2023). Resume analysis in Portuguese using word embeddings: Development of a decision support system for candidate selection. En Anais do XX Encontro Nacional de Inteligência Artificial e Computacional (pp. 1074-1088). <https://doi.org/10.5753/eniac.2023.234609>
- Gary, M. J., Andrew, F., Thomas, H., Robert, N., & Dursun, D. (2012). Practical Text Mining and Statistical Analysis for Non-structured Text Data. Waltham: Elsevier.
- Godoy Viera, A. F. (2017, Enero). Obtenido de www.scielo.org: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-358X2017000100103
- Goutam, C., Murali, P., & Satish, G. (2013). Text Mining and Analysis: Practical Methods, Examples, and Case Studies. USA: S.A.S Institute Inc.
- Ibáñez, A. (2015). .Machine Learning in Scientometrics. M. Madrid: Universidad Politecnica de Madrid.
-

-
- Lim, Z., Pushpanathan, K., Min Er Yew, S., Lai, Y., Sun, C.-H., Lam, J., ... Tham, Y.-C. (2023). Benchmarking large language models performances for myopia care: A comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and Google Bard. *EBioMedicine*, 1-11.
- Luis, J. A. (2016). *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. Mexico: Alfaomega.
- Myoung-Jong Kim*, I. H. (2003). The discovery of experts decision rules from qualitative bankruptcy. *Expert Systems with Applications* 25, 637–646.
- Odili, P. O., Daudu, C. D., Adefemi, A., Ekemezie, I. O., & Usiagu, G. S. (2024, February) The impact of artificial intelligence on recruitment and selection processes in the oil and gas industry: A Review. *Engineering Science & Technology Journal*, pp. 612-638.
- Power Data. (n.d.). Obtenido de Power Data: <https://www.powerdata.es/big-data#:~:text=Solo%20el%2020%25%20de%20informaci%C3%B3n,proyecto%20de%20calidad%20de%20datos>
- Rojas-Galeano, S., Posada, J., & Ordoñez, E. (2022). A bibliometric perspective on AI research for job resume matching. *The Scientific World Journal*. Obtenido de <https://onlinelibrary.wiley.com/doi/10.1155/2022/8002363>
- Sabry, F. (2023). *Text Mining. USA: One Billion Knowledgeable*.
- Sippy, M., Khandelwal, J., Jain, A., & Mathew, K. (2021). *ResumeScan: Application Tracking and Career Prediction Model*.
- Sroison, P., & Chan, J. H. (2021, 12 29). *Resume Parser with Natural Language*. Obtenido de <https://www.techrxiv.org/doi/full/10.36227/techrxiv.17641604.v1>
- T. M. Harsha, G. S. (2022). Automated Resume Screener using Natural Language Processing(NLP). En *International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 1772-1777). Tirunelveli, India: IEEE.
- Winasti, W. E. (2018). Inpatient fow management: A systematic review. *Health Care Qual.*, 718–734.
-