

Análisis de preguntas abiertas en la evaluación del desempeño docente utilizando Minería de Texto en R y Chat GPT

Analysis of Open-Ended Questions in Teacher Performance Evaluation Using Text Mining in R and Chat GPT

Byron Jiménez Oviedo ^{1,*}  ; Katalina Oviedo Rodríguez ¹  ;
Jorge Arroyo Hernández ¹  ; Federico Mora Mora ¹  ;
Giselle Ñurinda Montoya ²  ; Rita Hernández Gómez ²  ;
Kendall Ruíz Benavides ² 

¹ Facultad de Ciencias Exactas y Naturales, Escuela de Matemática, Universidad Nacional, Costa Rica.

² Vicerrectoría de Docencia, Universidad Nacional, Costa Rica.

* Autor correspondiente: byron.jimenez.oviedo@una.cr (B. Jiménez)

DOI: [10.17268/scien.inge.2025.03.03](https://doi.org/10.17268/scien.inge.2025.03.03)

RESUMEN

Este estudio analiza comparativamente las aplicaciones de Minería de Texto en R y el modelo de lenguaje ChatGPT para el procesamiento de respuestas abiertas en evaluaciones del desempeño docente en la Universidad Nacional (UNA) de Costa Rica. Se emplearon datos simulados basados en evaluaciones históricas del estudiantado, garantizando la confidencialidad. El análisis en R incluyó técnicas de tokenización, frecuencia de términos, asociación de palabras y análisis de sentimientos, mientras que ChatGPT se utilizó para la interpretación semántica, la generación de resúmenes y la categorización automática de fortalezas y áreas de mejora. Los resultados evidencian que R ofrece precisión estadística y capacidad para explorar patrones textuales, mientras que ChatGPT destaca por su flexibilidad interpretativa y su potencial para automatizar la generación de reportes analíticos. La correlación alta entre ambos enfoques en la detección de emociones y temas clave demuestra la complementariedad de sus resultados. Se concluye que la integración de herramientas de minería de texto y modelos de lenguaje puede optimizar el análisis cualitativo en contextos educativos, promoviendo procesos de evaluación más objetivos, eficientes y basados en evidencia. Este enfoque ofrece un marco replicable para fortalecer la gestión académica y la mejora continua de la docencia universitaria.

Palabras clave: Minería de Texto; chat GPT; Desempeño docente; Preguntas abiertas; Análisis automatizado de texto.

ABSTRACT

This study presents a comparative analysis of Text Mining applications in R and the ChatGPT language model for processing open-ended responses in teaching performance evaluations at the National University (UNA) of Costa Rica. Simulated data based on historical student evaluations were used to ensure confidentiality. The analysis in R included techniques such as tokenization, term frequency analysis, word association, and sentiment analysis, while ChatGPT was employed for semantic interpretation, summary generation, and automatic categorization of strengths and areas for improvement. The results show that R provides statistical precision and the ability to explore textual patterns, whereas ChatGPT excels in interpretative flexibility and its potential to automate analytical report generation. The high correlation between both approaches in the detection of emotions and key topics demonstrates the complementarity of their outcomes. It is concluded that integrating text mining tools and language models can optimize qualitative analysis in educational contexts, promoting more objective, efficient, and evidence-based evaluation processes. This combined approach offers a replicable framework to strengthen academic management and foster continuous improvement in university teaching quality.

Keywords: Text Mining; Chat GPT; Teacher Performance; Open-Ended Questions; Automated Text Analysis.



1. INTRODUCCIÓN

La evaluación del desempeño docente en la Universidad Nacional (UNA, Costa Rica) comenzó en 1999 con la creación del Programa de Evaluación Académica, cuyo objetivo era identificar fortalezas y debilidades del personal docente para mejorar los procesos académicos. Inicialmente, la evaluación se aplicó a solicitud de las direcciones académicas; sin embargo, a partir de 2003, se estableció como un proceso institucional regular. En 2004, se introdujo un sistema de evaluación basado en áreas temáticas para mejorar la interpretación de los datos y fortalecer las competencias docentes, promoviendo la autoevaluación y el desarrollo profesional. En 2006, el Programa de Evaluación y Desarrollo Profesional Docente se consolidó como un sistema integral para la mejora continua de la docencia, reconociendo la complejidad de este proceso. Este programa, ahora conocido como Sistema de Evaluación del Desempeño Académico (SEDA), asesora a las direcciones de las unidades académicas en aspectos de evaluación docente, ofrece criterios y elementos para construir modelos de evaluación en la UNA, diseña y rediseña instrumentos de evaluación, y propone opciones para mejorar las prácticas docentes. Al abarcar aspectos técnicos, académicos y sociales, el SEDA busca promover una evaluación integral que permita identificar fortalezas y debilidades en el desempeño docente.

En la educación superior, la evaluación docente por parte de los estudiantes es crucial para la mejora continua de la calidad educativa. Según Pacheco et al. (2018), la evaluación del desempeño docente es un “ejercicio sistemático que, con fundamento en un conjunto de evidencias, juzga cómo llevan a cabo su labor los profesores” (p. 1). Martínez & García (2017) señalan que el desempeño docente se refiere a las operaciones realizadas por las personas docentes durante el proceso de enseñanza. Rodríguez et al. (2021) enumeran varias de estas operaciones, como la planificación y estructuración de los contenidos, las estrategias didácticas, la comunicación efectiva, el manejo de nuevas tecnologías, el desarrollo de tareas instructivas con retroalimentación, la relación asertiva con los estudiantes para asegurar su maduración y desarrollo personal, el acompañamiento en la formación, la evaluación, la inclinación hacia la reflexión y la investigación, y el trabajo en equipo.

La UNA define la evaluación del desempeño docente como un proceso que recoge evidencias de validez y confiabilidad, permitiendo valorar las funciones, tareas y roles del quehacer docente, y relacionar actitudes, valores, saberes y habilidades desde una perspectiva personal, disciplinar y pedagógica en el proceso de enseñanza-aprendizaje, tanto dentro como fuera del aula.

En este contexto, el SEDA diseña instrumentos de evaluación del desempeño docente con el objetivo de obtener información consistente y sistemática, tanto cuantitativa como cualitativa, para reflexionar y promover cambios en la enseñanza y el aprendizaje. Como parte del diseño de estos instrumentos, se utilizan ítems abiertos, que representan un reto para el análisis debido al tipo de dato (textual) y al volumen de estos por periodo de evaluación.

Dado lo anterior, surge como objetivo para este estudio comparar las aplicaciones de Minería de Texto en R y ChatGPT para el análisis de ítems abiertos en evaluaciones del desempeño docente. Sumado a esto, se pretende mostrar las capacidades de análisis de texto disponibles en R para procesamiento de texto (como tokenización, análisis de sentimientos y extracción de temas) con las capacidades de generación y análisis de texto de ChatGPT y describir la eficiencia y facilidad de implementación de cada una de estas herramientas (A Fuller et al., 2024; Li et al., 2024; Mamo et al., 2024).

Para alcanzar estos objetivos, hay que profundizar en los instrumentos que miden el desempeño docente, diseñados por especialistas en psicometría del SEDA. El instrumento que explora la percepción estudiantil es uno de los recursos más utilizados por las instituciones de educación superior para evaluar la labor del profesorado universitario (Gómez & Valdés, 2019). Este tipo de instrumento se basa en un modelo orientado a los resultados, en el que el estudiantado, como receptor del servicio, evalúa el desempeño del profesorado a partir de su percepción (Montoya et al., 2014). Su popularidad se debe a factores como la facilidad de aplicación y el procesamiento eficiente de la información. Hoy se utilizan cuestionarios en línea, lo que agiliza el procesamiento de los datos, ya que los programas informáticos permiten generar gráficos, realizar comparaciones de puntajes, entre otros (Wellein et al., 2009).

El instrumento incluye una sección cuantitativa, compuesta exclusivamente por ítems cerrados, que permite asignar una calificación final al docente. Sin embargo, también se utilizan ítems abiertos, en los que el estudiantado puede escribir en prosa su percepción sobre aspectos relacionados con la práctica docente. Las respuestas a estos ítems abiertos son difíciles de cuantificar y, normalmente, son revisadas manualmente por un superior jerárquico, quien proporciona retroalimentación al docente (Peña-Torres, 2024).

El proceso de revisión de preguntas abiertas puede ser laborioso, dependiendo del número de docentes en cada escuela o unidad académica. Tradicionalmente, la recopilación y análisis de estas evaluaciones se ha realizado mediante métodos convencionales, que resultan limitados en cuanto a la profundidad del análisis.

No obstante, con los avances en técnicas de minería de texto e inteligencia artificial, se presentan nuevas oportunidades para extraer información significativa de grandes volúmenes de comentarios estudiantiles de forma más eficiente y efectiva (Buenano-Fernandez et al., 2020; Caballero & Enciso, 2022; Meléndez et al., 2023). En este sentido, el análisis de respuestas a preguntas abiertas puede facilitarse mediante técnicas computacionales, como la Minería de Texto.

La Minería de Texto es el proceso de extraer patrones o conocimientos previamente desconocidos, comprensibles y prácticos de colecciones de datos textuales (Zhang et al., 2015). Hotho et al. (2005) la definen como la aplicación de algoritmos y métodos provenientes del aprendizaje automático y la estadística para encontrar patrones útiles en textos. La minería de textos es un campo multidisciplinario que abarca la recuperación de información, el análisis de textos, la extracción de información, la categorización, la agrupación, la visualización, la minería de datos y el aprendizaje automático (Dang & Ahmad, 2014). El uso de técnicas de minería de datos tiene muchos beneficios en el manejo de información no estructurada; Ngai & Lee (2016) mencionan que las aplicaciones de minería de textos ayudan a los responsables de políticas a procesar información y descubrir nuevos conocimientos a bajo costo, además de aumentar la efectividad y eficiencia de estos procedimientos en las tomas de decisiones.

Dentro de la Minería de Texto se encuentra la Minería de Opiniones o Text Opinion, la cual permite analizar los sentimientos expresados de manera escrita. El análisis de los sentimientos consiste en determinar y analizar las emociones que se plasman en un texto en relación con un tema o campo específico. Liu (2010) menciona que la investigación en este campo comenzó con la clasificación de sentimientos y de la subjetividad, en donde se trataban problemas de clasificación de textos. La clasificación de sentimientos determina si un documento con opiniones (por ejemplo, reseñas de productos) expresa una opinión positiva o negativa. Actualmente, este campo de estudio analiza las opiniones de las personas, los sentimientos, las evaluaciones (Liu, 2022), actitudes y emociones hacia entidades, como productos, servicios, organizaciones, individuos, entre otros. Zhang et al. (2022) presentan una revisión completa sobre el avance de investigaciones en esta área, exponiendo los conjuntos de datos comúnmente utilizados, las métricas de evaluación, el proceso de desarrollo y las tendencias de investigación que incluyen métodos tradicionales de machine learning y métodos de deep learning.

Recientemente, otra herramienta utilizada para el análisis de texto han sido las inteligencias artificiales, una de las más conocidas es ChatGPT, desarrollada por OpenAI. Este chat utiliza lenguaje artificial y computación cognitiva y está basado en el preentrenamiento generativo GPT (Radford et al., 2018). Respecto a su uso para el análisis de sentimiento en textos escritos, Wang et al. (2023) presentan un análisis preliminar de su rendimiento, determinando que ChatGPT tiene una capacidad notable para inferir sentimientos, mostrando un rendimiento similar a los modelos supervisados de última generación (A Fuller et al., 2024).

2. METODOLOGÍA

Se realizó un estudio comparativo, entre R y el ChatGPT, con el fin de conocer las capacidades de análisis de texto con respecto a preguntas abiertas del instrumento de evaluación del desempeño docente. En aras de proteger la privacidad de datos sensibles, las respuestas que se analizaron fueron simuladas a partir de respuestas dadas por estudiantes de la UNA para un profesor específico. Posteriormente, se analizaron las respuestas simuladas utilizando las herramientas ya mencionadas.

2.1 Simulación de datos

La simulación se hizo con patrones históricos de preguntas abiertas de evaluaciones del desempeño de un docente de la UNA, correspondientes al periodo 2018-2024. El reporte para este profesor, en este periodo, daba un total de 770 respuestas, lo que equivale, en este caso, a 13791 palabras, para un aproximado de 40 páginas en una configuración estándar de Word. A manera de ejemplo, se muestran dos de las preguntas del instrumento de evaluación:

1. En su opinión, ¿qué fortalezas reconoce en la persona docente en el desarrollo de este curso específico? (Respuesta abierta)
2. En su opinión, ¿qué áreas de mejora tiene la persona docente para el desarrollo de este curso específico? (Respuesta abierta)

Para la simulación se utilizaron estos datos históricos y el ChatGPT, realizando el siguiente procedimiento: Primero, como la cantidad de palabras es muy grande el ChatGPT no logra procesarlo, por lo cual, se debió realizar una segmentación del texto en cuatro partes. Luego, para cada uno de estos tramos, se le pidió al

ChatGPT simular un nuevo texto, manteniendo las mismas ideas de los comentarios y la misma cantidad de palabras. Posteriormente, se tomaron los datos simulados y se colocaron en un texto plano para su análisis.

2.2 Herramientas y versiones utilizadas

Los análisis estadísticos y de procesamiento de texto se realizaron utilizando R versión 4.4.1. Para la manipulación y visualización de datos se emplearon las librerías [ggplot2](#) y [RColorBrewer](#), que facilitaron la creación de gráficos claros y estéticamente consistentes. El análisis de texto y minería de información se llevó a cabo utilizando [NLP](#) y [tm](#) para la preparación y limpieza de los textos, [SnowballC](#) para la normalización de palabras mediante stemming, [wordcloud](#) para la representación visual de frecuencias y [syuzhet](#) para el análisis de sentimientos. Por su parte, el análisis con ChatGPT se realizó utilizando la plataforma de OpenAI, correspondiente al modelo GPT-5.

2.3 Preprocesamiento para R

Para analizar el texto mediante R (ver código en [GitHub](#)) se procesaron y limpiaron los datos. Seguido, se cargó el corpus y sus metadatos. Una vez hecho este proceso se realizaron algunas transformaciones como, por ejemplo: se quitaron espacios en blanco extras, se substituyeron por espacios en blanco caracteres como ‘|’, ‘/’, ‘@’ o números, se pasaron todas las letras mayúsculas a minúsculas (tolower), se eliminaron palabras vacías (stopwords), se redujeron las palabras a su raíz (stemming), y por último, se removieron signos de puntuación y palabras específicas que no aportaran información importante, por ejemplo: área, toda, hace, estudiante, práctica, profesor, curso, materia, laboratorio, clase, tema.

Cabe destacar que como las preguntas abiertas fueron contestadas en español se utilizó dicho idioma para las transformaciones o eliminaciones. A modo de ejemplo, se presenta la comparación de un texto antes y después de su limpieza.

Tabla 1. Ejemplo respuesta simula y procesada.

Respuesta simulada	Respuesta procesada
Expresa un verdadero entusiasmo por que sus estudiantes comprendan y se adentren en el tema, brindando ejemplos claros y estando siempre dispuesto a responder preguntas de manera eficiente.	expresa verdadero entusiasmo estudiantes comprendan adentren brindando ejemplos claros siempre dispuestos responder preguntas manera eficiente.

Como último paso, se realizó la tokenización, que se refiere al proceso de dividir texto en unidades más pequeñas llamadas tokens. El tamaño de un token puede ser variado, desde una sola letra o carácter hasta palabras completas.

2.3 Preprocesamiento para chatGPT

Por otro lado, el preprocesamiento para modelos como ChatGPT se centra en optimizar el texto de entrada para la generación de respuestas naturales y coherentes. Se debe considerar la coherencia semántica y el estilo del texto, asegurando que las respuestas generadas mantengan fluidez y relevancia en el contexto. Por esta razón, para efecto de este estudio, se utilizaron los datos en minúsculas, sin los signos de puntuación y sin los stopwords. Además, en este caso se debe considerar el número de palabras (más de 3000) pues, como ya se mencionó, el ChatGPT en una sola corrida solo puede procesar hasta un máximo de 2000 palabras, por lo que se debe considerar realizar el análisis en segmentos de texto.

Tabla 2. Ejemplo respuesta simula y procesada.

Respuesta simulada	Respuesta procesada
Expresa un verdadero entusiasmo por que sus estudiantes comprendan y se adentren en el tema, brindando ejemplos claros y estando siempre dispuesto a responder preguntas de manera eficiente.	expresa un verdadero entusiasmo por que sus estudiantes comprendan y se adentren en el tema brindando ejemplos claros y estando siempre dispuesto a responder preguntas de manera eficiente.

Cada enfoque se adapta a las necesidades y objetivos particulares del análisis de datos y a la generación de texto, garantizando que los datos preparados sean efectivos para el análisis estadístico riguroso o en la creación de contenido textual inteligente. A pesar de los diferentes enfoques finales de cada aplicación (análisis

estadístico versus generación de texto), el proceso de preprocesamiento puede ser similar y se centra en optimizar los datos para su uso efectivo en cada contexto específico.

3. RESULTADOS Y DISCUSIÓN

3.1 Resultados usando R

El análisis de frecuencias permitió identificar patrones en las respuestas simuladas respecto a la percepción de los estudiantes hacia sus docentes y el curso. Se contabilizó un total de 3,885 palabras, de las cuales 1,460 eran únicas, lo que refleja una diversidad moderada en el vocabulario utilizado.

Entre las palabras más frecuentes, se destacan: excelente (83; 2,14%), bien (64; 1,65%), explica (59; 1,52%), manera (40; 1,03%), paciencia (36; 0,93%), conocimiento (30; 0,77%), atento (26; 0,67%), interés (26; 0,67%), sabe (23; 0,59%) y comprensivo (23; 0,59%). Estos resultados evidencian una valoración general positiva, centrada en la claridad de las explicaciones, la paciencia y el conocimiento de los docentes. Como parte de la interpretación, estos datos señalarían la importancia que los estudiantes otorgan al dominio de los contenidos por parte de las personas docentes, así como a su habilidad para explicar y la disposición mostrada hacia el aprendizaje. La alta frecuencia de términos como explica y paciencia puede sugerir el compromiso por parte de las personas docentes en proporcionar una experiencia educativa de calidad.

La Figura 1 muestra un gráfico de barras con las nueve palabras más frecuentes, donde *excelente* se posiciona como la de mayor frecuencia, mientras que *sabe* es la menos mencionada.

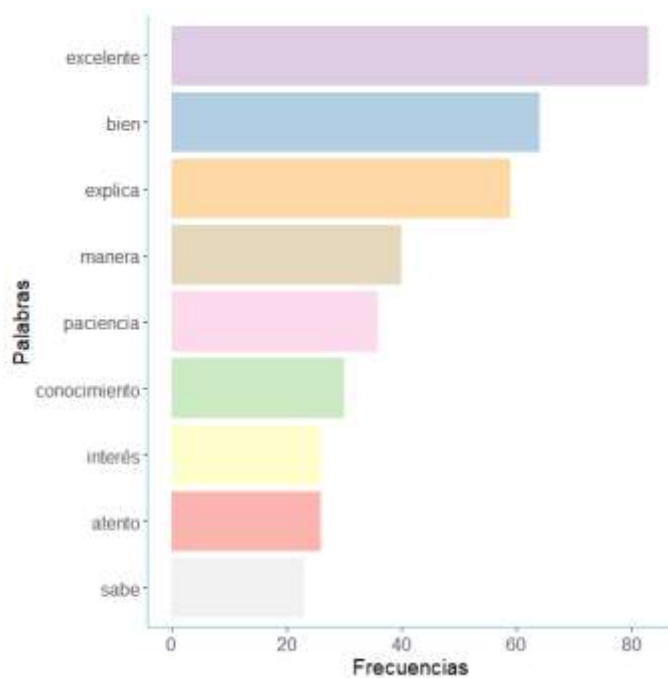


Figura 1. Frecuencia de palabras.

La nube de palabras (Figura 2) refleja visualmente la prominencia de los términos más relevantes. Las palabras más grandes, como *excelente*, *bien*, *explica*, *manera* y *paciencia*, destacan la percepción positiva sobre el desempeño de los docentes y su disposición hacia los estudiantes.

ron promedios similares: -0,009 y -0,129, respectivamente, lo que refuerza la tendencia a una percepción predominantemente neutral-positiva.

La distribución de emociones según el método NRC evidenció que confianza (43,88%), alegría (16,53%) y tristeza (15,17%) fueron las emociones más frecuentes, mientras que ira (1,67%) y disgusto (1,78%) fueron poco mencionadas. Estas frecuencias sugieren que los estudiantes valoran principalmente la seguridad y confianza transmitida por sus docentes (ver Figura 3).

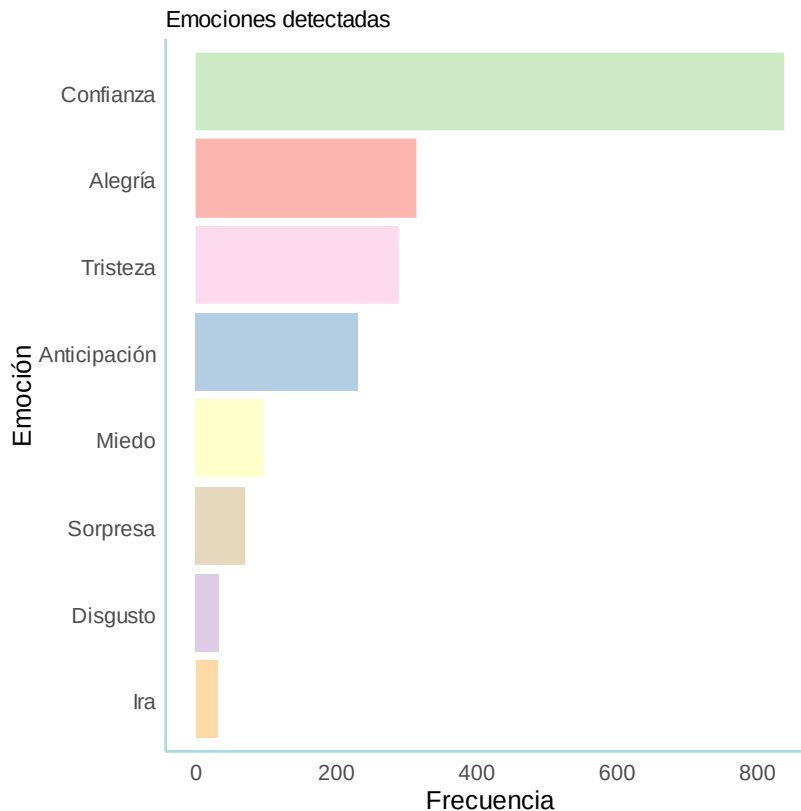


Figura 3. Frecuencia de diferentes sentimientos.

En términos de polaridad, la cantidad de respuestas positivas fue mayor, con 1,487 palabras frente a 371 palabras negativas, mostrando una clara tendencia positiva en el texto.

Los puntajes promedio de sentimiento obtenidos con Syuzhet muestran un valor cercano a cero (-0,031), lo que indica que, aunque hay una mayor cantidad de palabras positivas, el sentimiento global del texto se mantiene cercano a la neutralidad. La desviación estándar de 0,221 refleja que existen variaciones locales de sentimientos positivos y negativos dentro del contenido.

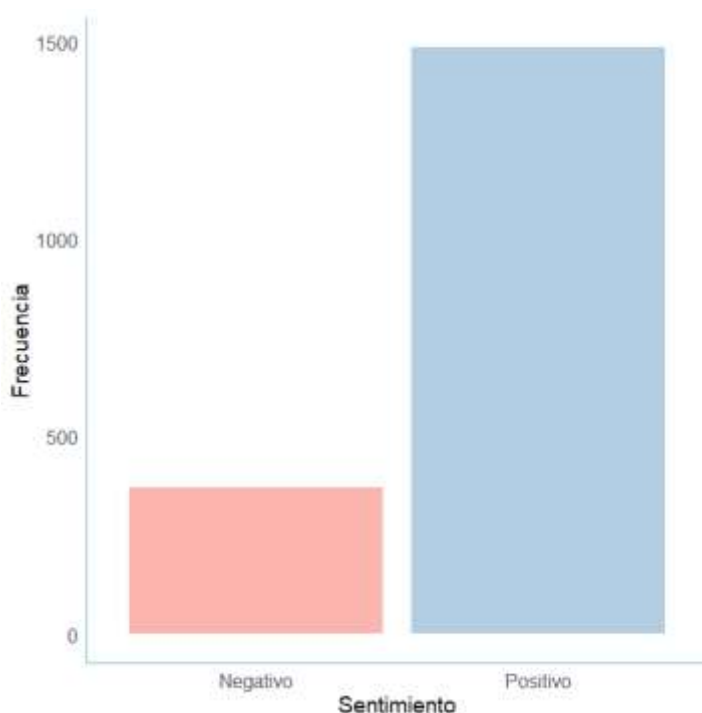


Figura 4. Frecuencia de los sentimientos positivos o negativos.

3.2 Resultados usando el ChatGPT

En primera instancia, se solicitó al ChatGPT un análisis y un reporte del texto, sin embargo, se obtuvo el siguiente mensaje "The message you submitted was too long, please reload the conversation and submit something shorter" lo que refleja una restricción técnica en la plataforma de análisis, probablemente implementada para optimizar el procesamiento eficiente de consultas. Esta limitación puede afectar la capacidad de analizar textos extensos de manera inmediata. Para abordar esta limitación, el ChatGPT en el mismo mensaje sugiere dividir el texto en secciones más manejables o proporcionar un resumen conciso, facilitando así una mejor interacción con la herramienta de análisis.

Como la primera instrucción no se pudo completar, se dio la siguiente instrucción al ChatGPT: "Te proporcionaré extractos de textos de una encuesta de evaluación docente en trozos. Por favor, pregúntame cada vez si he terminado de darte todo el fragmento. Cuando te indique que sí, procede a darme tu percepción objetiva en forma de reporte."

Sin embargo, el ChatGPT no ejecutó la instrucción correctamente. Originalmente se tenían previstos 4 trozos de texto, pero en el segundo ingreso, el ChatGPT proporcionó el reporte anticipadamente. Fue necesario indicarle que aún faltaban dos trozos más. Una vez ingresados estos, el chat arrojó un resumen en forma de reporte. Un extracto del texto es el siguiente:

"El profesor es altamente valorado por su conocimiento, paciencia y actitud positiva. Sin embargo, mejorar la velocidad de la enseñanza, la claridad en la presentación y el enfoque de las evaluaciones podría mejorar significativamente la experiencia de aprendizaje para todos los estudiantes."

Luego de esto, se hizo otro intento sin darle ninguna instrucción específica. Se ingresaron los cuatros extractos de texto, y para cada ingreso el chat generó un reporte de análisis del texto, con un resumen de aspectos positivos, áreas de mejora, sugerencia para la cátedra y una conclusión general. A continuación, se presentan los aspectos positivos que menciona el chat:

1. Dominio del Tema: Los comentarios destacan consistentemente que el profesor tiene un profundo conocimiento de la materia y un excelente manejo del contenido del curso.
2. Paciencia y Disponibilidad: El profesor es descrito como muy paciente y dispuesto a explicar los conceptos tantas veces como sea necesario para asegurar la comprensión.
3. Claridad en la Explicación: Se valora mucho su capacidad para explicar de manera clara y detallada, así como su disposición para aclarar dudas.

4. Interés en los Estudiantes: El profesor muestra un interés genuino en el aprendizaje y el éxito de sus estudiantes, lo que se refleja en su actitud alentadora y su esfuerzo por ayudar.

5. Actitud Positiva: La actitud respetuosa, carismática y comprensiva del profesor es frecuentemente elogiada, creando un ambiente de aprendizaje positivo.

Este último reporte fue mejor que el dado en un inicio, de lo que se concluye que la instrucción para el chat debe ser clara y guiada.

Se le pidió también hacer una tabla para cuantificar las fortalezas y debilidades o puntos de mejora. Los resultados que se obtuvieron fueron los siguientes.

Tabla 4. Cuantificación de las fortalezas.

Categoría	Frecuencia estimada	Porcentaje aproximado
Paciencia / claridad en la explicación	85	30%
Dominio de la materia / conocimiento	50	18%
Actitud positiva / carisma / motivación	60	21%
Empatía / humanismo / comprensión	70	25%
Preparación / organización de la clase	25	9%
Dinamismo / interacción en clase	20	7%

Tabla 5. Cuantificación de áreas a mejorar.

Categoría	Frecuencia estimada	Porcentaje aproximado
Evaluaciones (tipo y cantidad)	40	35%
Tiempo de quices / exámenes / clases largas	20	17%
Organización de materiales y aula virtual	15	13%
Manejo de participación en clase	10	9%
Profundización de contenidos en horas de consulta	8	7%
Puntualidad en subir documentos / recursos	7	6%

Estos resultados son importantes porque se puede visualizar como el estudiantado valora la paciencia, claridad y empatía del docente, así como su dominio de la materia y actitud motivadora. Además, pone en evidencia aspectos a mejorar como la cantidad y tipo de evaluaciones, el tiempo en quices y clases largas, lo que sugiere la necesidad de ajustar la evaluación sin comprometer la calidad pedagógica.

3.3 Comentarios finales

Finalmente, se solicitó al chat realizar un análisis de sentimientos similar al que produce R, utilizando la lista de emociones de la biblioteca NRC, con el objetivo de comparar directamente los resultados. Una vez realizado este procedimiento, se obtuvieron los resultados mostrados en la siguiente tabla, los cuales refuerzan las observaciones preliminares sobre la percepción de los estudiantes:

Tabla 6. Comparación de sentimientos.

Emoción	R	Chat	Diferencia
Ira	1,67	2,0	0,33
Anticipación	12,13	11,5	-0,63
Disgusto	1,78	1,6	-0,18
Miedo	5,13	4,8	-0,33
Alegría	16,53	17,0	0,47
Tristeza	15,17	15,0	-0,17
Sorpresa	3,71	4,0	0,29
Confianza	43,88	44,1	0,22

Con estos resultados se obtuvo una correlación cercana a 1, indicando que el análisis realizado por el chat reproduce de manera muy similar los valores obtenidos por R. Las diferencias observadas son mínimas, lo que refuerza la confiabilidad de los hallazgos y confirma que las emociones predominantes entre los comen-

tarios de los estudiantes son confianza y alegría, mientras que emociones negativas como ira, disgusto y miedo se presentan en menor medida.

El uso de R como del chat representan opciones válidas y efectivas para el análisis de respuestas a preguntas abiertas. Ambas herramientas permiten cuantificar emociones, identificar patrones y generar visualizaciones de manera rápida, lo que facilita la interpretación de grandes volúmenes de comentarios. Además, su uso puede ofrecer múltiples beneficios, como ahorrar tiempo en el procesamiento de datos, estandarizar criterios de análisis y complementar la evaluación cualitativa con resultados más objetivos.

4. CONCLUSIONES

Analizar las respuestas a preguntas abiertas en evaluaciones docentes, proporciona perspectivas valiosas sobre la percepción de los estudiantes hacia sus profesores, asimismo, ofrece una guía práctica para mejorar continuamente la calidad docente en la educación superior. Comúnmente, estas respuestas son analizadas de forma manual por las direcciones de las unidades académicas, lo que representa un trabajo arduo y que puede tomar mucho tiempo. Aunque las respuestas abiertas suelen ser breves, el volumen total de respuestas hace lento y laborioso el proceso de análisis. En este contexto, la automatización del análisis resulta ser una solución eficiente, al liberar tiempo y esfuerzo de la persona responsable del análisis.

El uso de herramientas como R y modelos de lenguaje como ChatGPT para analizar texto, presenta diversas fortalezas y limitaciones. Por un lado, R permite realizar análisis exhaustivos y detallados de grandes volúmenes de texto, utilizando técnicas estadísticas avanzadas y herramientas gráficas para explorar patrones y relaciones. Este lenguaje de programación es adecuado para tareas como el análisis de frecuencia de palabras, la detección de temas y el modelado predictivo basado en texto, permitiendo profundizar en la estructura y el contenido de los datos textuales con un alto grado de precisión y flexibilidad.

Por otro lado, el ChatGPT genera respuestas coherentes en lenguaje natural, lo que facilita la interpretación y comunicación de resultados complejos de manera accesible. Esto puede representar una ventaja en el análisis de datos textuales, para personas que no manejen técnicas avanzadas de análisis de texto. Además, esta capacidad de generar texto natural también puede ser útil en la creación automática de resúmenes o reportes de evaluaciones, como lo es el caso de las evaluaciones a docentes.

A pesar de estas ventajas, el ChatGPT presenta limitaciones en comparación con R, especialmente en términos de capacidad para realizar análisis estadísticos, modelado predictivo avanzado o visualización gráfica detallada. La dependencia del contexto de los datos y la calidad del modelo subyacente pueden afectar la precisión y consistencia de las respuestas generadas por el ChatGPT.

Por lo tanto, la elección entre R y ChatGPT depende del objetivo específico del análisis de texto y las habilidades técnicas del usuario. La combinación de ambos enfoques puede ofrecer un marco integral para el análisis de datos textuales.

Gracias a los insumos obtenidos del análisis comparativo entre R y el ChatGPT, se abre la posibilidad de implementar técnicas que permitan a las direcciones académicas y otras autoridades obtener un panorama justificado estadísticamente de la evaluación en preguntas abiertas del desempeño docente. Esto no solo facilitaría la identificación de patrones y tendencias en los datos textuales, sino que también optimizaría el proceso de evaluación al ahorrar tiempo y recursos significativos. La automatización de estas tareas mediante las herramientas mencionadas no solo mejoraría la eficiencia operativa, sino que también ofrecería perspectivas más profundas y contextualizadas para la toma de decisiones informadas.

El uso de estas herramientas tiene un impacto directo en las personas docentes y principalmente a las direcciones de las unidades académicas. En efecto, permiten identificar fortalezas y áreas de mejora de manera objetiva, optimiza la eficiencia en el procesamiento de datos y contribuye a sistemas de evaluación más transparentes, equitativos y basados en evidencia. En última instancia, la automatización del análisis de respuestas abiertas no solo mejora la eficiencia operativa, sino que también enriquece la comprensión de la experiencia educativa, potenciando decisiones más informadas y estratégicas para la mejora de la enseñanza.

5. AGRADECIMIENTOS

Agradecemos a la Vicerrectoría de Docencia por proporcionarnos los datos utilizados en este estudio y por confiar en que es posible mejorar mediante el uso de la tecnología. Su apoyo fue fundamental para el desarrollo de este trabajo.

REFERENCIAS BIBLIOGRÁFICAS

- A Fuller, K., Morbitzer, K. A., Zeeman, J. M., M Persky, A., C Savage, A., & McLaughlin, J. E. (2024). Exploring the use of ChatGPT to analyze student course evaluation comments. *BMC Medical Education*, 24(1), 423. <https://doi.org/10.1186/s12909-024-05316-2>
- Buenano-Fernandez, D., Gonzalez, M., Gil, D., & Lujan-Mora, S. (2020). Text mining of open-ended questions in self-assessment of university teachers: An LDA topic modeling approach. *IEEE Access: Practical Innovations, Open Solutions*, 8, 35318–35330. <https://doi.org/10.1109/access.2020.2974983>
- Caballero, J. & Enciso, J. E. (2022). Minería de texto aplicado en preguntas abiertas sobre evaluación docente: Enfoque de modelado de tópicos con LDA. Recuperado de: <http://hdl.handle.net/10554/63615>
- Dang, S., & Ahmad, P. H. (2014). Text mining: Techniques and its application. *International Journal of Engineering & Technology Innovation*, 1(4), 22–25.
- Gómez, L. F., & Valdés, M. G. (2019). La evaluación del desempeño docente en la educación superior. *Propósitos y Representaciones*, 7(2), 479–515.
- Hotho, A., Nürnberger, A., & Paaß, G. (2005). A brief survey of text mining. *Journal of Language Technology and Computational Linguistics*, 20(1), 19–62.
- Liu, B. (2010). Sentiment analysis: A multi-faceted problem. *IEEE Intelligent Systems*, 25(3), 76–80.
- Liu, B. (2022). Sentiment analysis and opinion mining. Springer Nature.
- Li, L., Ma, Z., Fan, L., Lee, S., Yu, H., & Hemphill, L. (2024). ChatGPT in education: a discourse analysis of worries and concerns on social media. *Education and Information Technologies*, 29(9), 10729–10762. <https://doi.org/10.1007/s10639-023-12256-9>
- Mamo, Y., Crompton, H., Burke, D., & Nickel, C. (2024). Higher education faculty perceptions of ChatGPT and the influencing factors: A sentiment analysis of X. *TechTrends : For Leaders in Education & Training*, 68(3), 520–534. <https://doi.org/10.1007/s11528-024-00954-1>
- Martínez, S., & García, J. (2017). Aproximación al concepto de desempeño docente, una revisión conceptual sobre su delimitación. Congreso Nacional de Investigación Educativa. <http://www.comie.org.mx/congreso/memoriaelectronica/v14/doc/2657.pdf>
- Meléndez, N., Briceño, M., Giberton, J., & Lucente, R. (2023). MINERÍA DE TEXTO APLICADO AL ANÁLISIS DE SENTIMIENTOS PARA LA BÚSQUEDA DE PATRONES DE MOTIVACIÓN EN PROGRAMA DE ACTUALIZACIÓN DOCENTE. *Revista Académica Creatividad e Innovación En Educación*, 2(1), 32–40. <https://doi.org/10.47300/2953-3015-v2i1-03>
- Montenegro-Rueda, M., Fernández-Cerero, J., Fernández-Batanero, J. M., & López-Meneses, E. (2023). Impact of the implementation of ChatGPT in education: A systematic review. *Computers*, 12(8), 153.
- Montoya, J., Arbesú, I., Contreras, G., & Conzuelo, S. (2014). Evaluación de la docencia universitaria en México, Chile y Colombia: Análisis de experiencias. *Revista Iberoamericana de Evaluación Educativa*, 7(2e), 15–42. <https://revistas.uam.es/riee/article/view/3111>
- Ngai, E. W., & Lee, P. T. Y. (2016). A review of the literature on applications of text mining in policy making.
- Pacheco, M. L. C., Ibarra, I., Iñiguez, M. E., Lee, H., & Sánchez, C. V. (2018). La evaluación del desempeño docente en la educación superior. *Revista Digital Universitaria (RDU)*, 19(6), 1–11. <https://doi.org/10.22201/codeic.16076079e.2018.v19n6.a2>
- Peña-Torres, J. A. (2024). Towards an improved of teaching practice using Sentiment Analysis in Student Evaluation. *Ingeniería y Competitividad*, 26(2). <https://doi.org/10.25100/iyv.v26i2.13759>
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. <https://www.mikecaptain.com/resources/pdf/GPT-1.pdf>
- Rodríguez Siu, J. L., Rodríguez Salazar, R. E., & Fuerte Montaña, L. (2021). Habilidades blandas y el desempeño docente en el nivel superior de la educación. *Propósitos y Representaciones*, 9(1).
- Wang, Z., Xie, Q., Feng, Y., Ding, Z., Yang, Z., & Xia, R. (2023). Is ChatGPT a good sentiment analyzer? A preliminary study. *arXiv*. <https://arxiv.org/abs/2304.04339>
- Wellein, M. G., Ragucci, K. R., & Lapointe, M. (2009). A peer review process for classroom teaching. *American Journal of Pharmaceutical Education*, 73(5), 1–7. <https://doi.org/10.5688/aj730579>

- Zhang, Y., Chen, M., & Liu, L. (2015). A review on text mining. In 2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS) (pp. 681–685).
- Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2022). A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 35(11), 11019–11038.