

Optimización de la generación de informes de auditoría mediante procesamiento de lenguaje natural: un enfoque basado en datos de auditorías de calidad en educación superior

Optimizing audit reporting using natural language processing: a data-driven approach from quality audits in higher education

MSc. Alveiro Rosado Gómez ¹, PhD. Claudia Marcela Duran Chinchilla ²,
MSc. Deccy Arias Rodríguez ³

¹ Universidad Francisco de Paula Santander, Facultad de Ingeniería, Grupo de Investigación en Desarrollo Tecnológico en Ingeniería (GITYD), Ocaña, Norte de Santander, Colombia.

² Universidad Francisco de Paula Santander, Departamento de Humanidades, Grupo de Investigación de la Facultad de Educación, Artes y Humanidades (GIFEAH), Ocaña, Norte de Santander, Colombia.

³ Universidad Francisco de Paula Santander, Facultad de Ciencias Agrarias y del Ambiente, Especialización en Sistemas de Gestión Integral HSEQ, Ocaña, Norte de Santander, Colombia.

Correspondencia: aarosadog@ufpso.edu.co

Recibido: 21 enero 2024. Aceptado: 11 junio 2024. Publicado: 23 julio 2024.

Cómo citar: A. A. Rosado Gómez, C. M. Duran Chinchilla, y D. Arias Rodríguez, «Optimización de la generación de informes de auditoría mediante procesamiento de lenguaje natural: un enfoque basado en datos de auditorías de calidad en educación superior», RCTA, vol. 2, n.º 44, pp. 89–96, jul. 2024.

Recuperado de <https://ojs.unipamplona.edu.co/index.php/rcta/article/view/3018>

Derechos de autor 2024 Revista Colombiana de Tecnologías de Avanzada (RCTA).
Esta obra está bajo una licencia internacional [Creative Commons Atribución-NoComercial 4.0](https://creativecommons.org/licenses/by-nc/4.0/).



Resumen: Esta investigación se centró en la automatización de la comprensión e identificación semántica de hallazgos para su clasificación en auditorías internas, utilizando técnicas de procesamiento de lenguaje natural. Se analizaron informes de auditorías internas para extraer textos vinculados a no conformidades, fortalezas y oportunidades de mejora. Para optimizar la presentación del texto para diversos algoritmos, se examinaron métodos como bolsa de palabras (BoW), frecuencia de término-frecuencia inversa de documento (TF-IDF), así como representaciones de texto a través de vectores de palabras incrustadas como Word2Vec y FastText. Se determinó que la mejor combinación de rendimiento provino de un clasificador lineal, que utiliza datos transformados mediante palabras incrustadas y equilibra las clases con sobre-muestreo. Este modelo fundamenta sus clasificaciones en palabras que capturan adecuadamente el sentido y contexto del hallazgo analizado.

Palabras clave: Aprendizaje automático, auditoría interna, aprendizaje supervisado, inteligencia artificial, procesamiento del lenguaje natural.

Abstract: This research focused on automating the understanding and semantic identification of findings for classification in internal audits using natural language processing techniques. Internal audit reports were analyzed to extract texts linked to non-conformities, strengths, and opportunities for improvement. To optimize text presentation

for various algorithms, methods such as bag-of-words (BoW), term frequency-inverse document frequency (TF-IDF), and text representations via embedded word vectors such as Word2Vec and FastText. The best combination of performance was determined to come from a linear classifier, which uses data transformed by word embeddings and balances oversampled classes. This model bases its classifications on words that adequately capture the meaning and context of the analyzed finding.

Keywords: Machine learning, internal audit, supervised learning, artificial intelligence, natural language processing.

1. INTRODUCCIÓN

Las organizaciones constituyen agrupaciones de individuos con identidad legal propia, que colaboran bajo reglas específicas para lograr objetivos compartidos, centrados principalmente en brindar servicios y productos a la sociedad [1]. En el ámbito de la gestión, estas entidades enfrentan retos continuos, incluyendo la calidad de sus productos o servicios, el cumplimiento normativo, la satisfacción del cliente y la búsqueda constante de mejoras. Para abordar estos desafíos, las organizaciones definen elementos fundamentales como directrices, metas y procedimientos. No obstante, la complejidad de estructurar y llevar a cabo estos componentes a menudo las lleva a apoyarse en sistemas de gestión [2]. Un sistema de gestión de la calidad (SGC) representa la forma en que una organización guía y supervisa sus actividades relacionadas con la consecución de los resultados deseados [1]. Este sistema engloba varios aspectos organizativos, planificación, procesos y recursos, enfocados hacia la calidad enmarcados dentro de la política de calidad que mantiene estándares internos de conformidad del producto [3]. En este contexto, la norma ISO 9001 es una alternativa importante porque se centra en criterios de calidad en la gestión, promoviendo la implementación requisitos y estándares [4].

La norma ISO 9001, requiere de herramientas que le permitan determinar si el sistema de gestión de calidad cumple con los requisitos que ella enmarca. Para lograrlo, se realizan evaluaciones sistemáticas e independientes llamadas auditorías, que pueden ser de naturaleza interna, llevadas a cabo por el personal de la propia organización, o externas, realizadas por entidades independientes como organismos de certificación [2] [4]. Los resultados derivados de la valoración de la evidencia recopilada durante la auditoría en comparación con los estándares establecidos se denominan hallazgos de auditoría, los cuales abarcan tanto la conformidad

como la no conformidad con los criterios de auditoría, además de posibles áreas de mejora [5]. Los hallazgos de la auditoría se constituyen en la base de los planes de mejora que realizan los procesos que incumplieron con algunos de los requisitos de la norma, por lo tanto, la forma en que se redacten guarda relación con la importancia y alcance de darle solución [1] [3]. Estos textos deben expresar con claridad los problemas y sus implicaciones, buscando un equilibrio entre proporcionar suficientes detalles para respaldar los hallazgos de manera lógica y coherente con síntesis que permitan una lectura efectiva [6].

Dentro del proceso de identificación y comunicación de hallazgos en un informe de auditoría, se requiere una clasificación en tres categorías: cuando el hallazgo es positivo, cuando es negativo o cuando existe espacio para mejorar. Al resaltar fortalezas, se describen los logros y prácticas eficaces que contribuyen tanto a los objetivos organizacionales como al sistema de gestión. En relación a las no conformidades, se propone una estructura que comienza con una clara identificación del incumplimiento, seguida por detalles y evidencia específicos, y se analiza cómo este afecta a procesos, calidad y objetivos. Al explorar oportunidades de mejora, se enfatiza la descripción de áreas con potencial para un cambio positivo, se presentan recomendaciones prácticas y los beneficios previstos para la organización en términos de desempeño y logro de metas [7].

Los auditores deben entregar el informe de auditoría a la persona o rol responsable de consolidar y revisar la información, con el propósito de evaluar su calidad, claridad y relevancia. Esta etapa implica una inversión significativa de tiempo y esfuerzo administrativo, ya que además de la duración del proceso de evaluación, se suma el tiempo necesario para que el auditor implemente las correcciones requeridas en el informe de auditoría. Lo que genera una mayor demora en la conclusión del proceso de auditoría y en el inicio del correspondiente plan de

mejora [1] [2] [5]. Esta situación invita a adelantar acciones encaminadas a reducir el tiempo y los recursos involucrados en el aseguramiento de la calidad en la gestión de la organización, por lo tanto, una opción que se plantea es la aplicación de tecnología que automatice el proceso de comprensión de los resultados de la auditoría [8] [7].

Desde una perspectiva de automatización, resulta necesario contar con soluciones tecnológicas inteligentes capaces de manejar datos no estructurados para la identificación de información pertinente en auditorías [8]. Mediante el procesamiento del lenguaje natural (PLN), es posible extraer datos de estos formatos, automatizando en gran medida la evaluación y validación de la calidad textual [9]. Por esta razón, el objetivo de esta investigación fue incorporar técnicas de PLN en la valoración de los hallazgos consignados en los informes de auditoría, permitiendo reducir el tiempo, la incidencia de errores humanos en la redacción y categorización de los hallazgos.

2. METODOLOGÍA

El conjunto de datos que fue utilizado, corresponde a los resultados de las auditorías que se han realizado en una institución de educación superior certificada en la norma ISO 9001: 2015. El dataset contenía dos columnas una con descripción del hallazgo y la otra con su etiqueta de hallazgo, esta última columna contenía tres valores; No conformidad, que se utilizó cuando se incumplía un requisito de la norma. Fortaleza, cuando el proceso mostraba un grado de madurez superior al solicitado por la norma y en pro de la gestión. El tercer valor es oportunidad de mejora, la cual hace referencia al cumplimiento de la norma de forma aceptable o a un incumplimiento de las normas internas de la institución. La tabla 1, muestra la distribución de los 864 registros utilizados.

Tabla 1: Distribución de clases

Clase	Cantidad
FORTALEZA	294
NO CONFORMIDAD	78
OPORTUNIDAD DE MEJORA	492

Fuente: elaboración propia

En el ámbito del Machine Learning (ML), el enfoque del aprendizaje supervisado se inicia con un conjunto de datos de entrada al cual se le asignan etiquetas de salida. Estas etiquetas representan los resultados deseados y permiten al modelo generalizar y aprender a través de las relaciones

establecidas entre las entradas, las etiquetas originales y las clasificaciones resultantes [10] [11]. La naturaleza no estructurada de los atributos de tipo texto, demanda hacer algún tipo de transformación a una forma que sea adecuada para ML como son formatos numéricos y categóricos [12] [13].

En esta investigación se utilizaron varias formas de procesar texto, con el fin de seleccionar la presentación que mejor rendimiento produjera para diferentes algoritmos. El enfoque bolsa de palabras (en inglés Bag of Words, BoW) se refiere a una técnica de representación textual que convierte el contenido en un vector numérico. Cada palabra en el texto se convierte en un "token", y el texto se transforma en un vector en el cual cada valor refleja la frecuencia de aparición de una palabra específica [12] [14]. También el texto fue procesado mediante, Frecuencia de Término-Frecuencia Inversa de Documento (en inglés Term frequency – Inverse document frequency, TF-IDF) que es una estrategia que otorga peso a las palabras en un documento según su relevancia.

Las palabras que se repiten con mayor frecuencia en un documento, pero son escasas en el conjunto total, son consideradas más significativas [12] [15] [16]. Adicionalmente, se utilizaron representaciones de texto mediante vectores de palabras con contexto, uno de ellos es Word2Vec, que es una técnica de "embedding" de palabras que emplea redes neuronales para aprender representaciones vectoriales para capturar la semántica y las relaciones entre las palabras al observar cómo se distribuyen en un corpus de texto [17] [18]. El otro tipo de representación utilizado fue FastText, el cual permite representar palabras como conjuntos de subcadenas de caracteres, lo cual le permite interpretar palabras que no están en el vocabulario original [19].

Cada conjunto de datos que se generó por la transformación del texto en formato numérico fue dividido en conjunto de entrenamiento (80%) y pruebas (20%), para luego ser utilizado con diferentes clasificadores con el fin de seleccionar la combinación con mejor rendimiento entre datos y algoritmo que produjera. Fueron utilizados, clasificadores familias como Support Vector Machines (SVM), que se caracterizan por buscar hiperplanos óptimos para separar clases en espacios de alta dimensión. Fueron utilizados Ensemble Methods, que combinan múltiples clasificadores para mejorar el rendimiento general. Otra familia que se trabajó, fue Generalized Linear Models (GLM) que incluyen modelos como la regresión

logística y Ridge para ajustar relaciones entre características y objetivos. Los K-Nearest Neighbors (K-NN) asignan clases en función de los vecinos más cercanos, mientras que los Decision Trees dividen los espacios según las características. El Discriminant Analysis, busca una proyección de las características originales en un nuevo espacio en el que las clases estén mejor separadas, y XGBoost emplea el boosting combinados con residuos para entrenar nuevos árboles y regularización para evitar sobreajuste [14] [20] [21] [22].

Luego se seleccionó el mejor algoritmo por tipo de transformación, estos algoritmos y conjunto de datos se les hicieron intervenciones como el balanceo de clases, utilizando el submuestreo aleatorio (Random undersampling, RUS) reduce ejemplos de la clase mayoritaria, afectando la generalización del modelo. El sobre-muestreo aleatorio (Random oversampling, ROS) replica instancias de la clase minoritaria. La técnica Synthetic minority oversampling technique (SMOTE) crea instancias sintéticas de la clase minoritaria a partir de vecinos cercanos, mejorando el equilibrio y la generalización del modelo [21]. Adicionalmente, se realizó ajuste paralelo de hiper-parámetros, mediante la librería Mango [23].

A partir de estas intervenciones se seleccionó el modelo con mayor valor de F1, como criterio de mejor combinación de precisión (porcentaje de predicciones correctas entre las predicciones positivas) y sensibilidad (porcentaje de instancias positivas correctamente identificadas) [24]. Al tratarse de un problema multiclase la selección inicial de los mejores modelos se realizó mediante el cálculo promedio del valor de cada métrica [25].

3. RESULTADOS

En la tabla 2, se muestran la distribución de los resultados, para cada uno de los métodos utilizados en la transformación del texto a una representación numérica. Tanto BoW, como TF-IDF, coinciden en el mismo número de atributos, esto se debe a la forma de presentar las palabras. Pero el contenido de cada valor dentro de la matriz resultado es diferentes porque depende de cómo se calculan esos valores dentro de la matriz y cómo estos valores se ponderan en función de la frecuencia y la rareza de las palabras en el corpus [26]. La dimensión del espacio vectorial en el que se representarán las palabras se definió en 500 para Word2Vec y FastText. Se puede observar cómo los métodos basados en frecuencia de aparición proporcionan una mayor dimensionalidad al conjunto de datos,

mientras que los métodos basados en representación con vectores, este valor es configurable y puede ser menor [27].

Tabla 2: Datos por representación

Clase	Atributos
BoW	1432
TF-IDF	1432
Word2Vec	500
FastText	500

Fuente: elaboración propia

La tabla 3, presenta los resultados sobresalientes para cada técnica de procesamiento de texto. Los resultados que se muestran a la implementación de los clasificadores sin balancear los datos. Se puede ver como los conjuntos de datos basados en vocabulario, tienen un mayor rendimiento con algoritmos lineales que los basados en vectores. Adicionalmente se destaca que las incrustaciones de palabras logran un mejor rendimiento en conjunto con el algoritmo XGBClassifier.

Tabla 3: Métricas por modelo

Algoritmo	F1	Técnica
LogisticRegressionCV	0.905	tfidf
PassiveAggressiveClassifier	0.900	bow
XGBClassifier	0.846	word2vec
XGBClassifier	0.802	fasttext

Fuente: elaboración propia

Haciendo el análisis del valor de F1 para cada una de las clases, como se muestra en la tabla 4, todos los algoritmos tienen un mejor comportamiento identificando no conformidades (NC) y un desempeño menor en las fortalezas (FO). Para las oportunidades de mejora (OP), los resultados tienen ligeras variaciones entre los algoritmos. Al igual que los resultados promediados, al revisar la salida por clase, se sigue teniendo en cuenta el mejor rendimiento de los datos basados en vocabularios de palabras.

Tabla 4: F1 por modelo

Algoritmo	FO	NC	OM
LogisticRegressionCV	0.824	1	0.891
PassiveAggressiveClassifier	0.833	0.960	0.893
XGBClassifier	0.755	0.916	0.867
XGBClassifier	0.716	0.846	0.843

Fuente: elaboración propia

La tabla 5, presenta los resultados sobresalientes para cada técnica de procesamiento de texto. Se entrenaron los clasificadores con los conjuntos de datos mediante diferentes técnicas de balanceo de clases. Con el ánimo de lograr el mejor modelo, se ejecutaron las representaciones de las familias de clasificadores. Para todos los clasificadores la combinación con ROS fue la que produjo la mejor

salida y se lograron mejores resultados de F1, sin ninguna intervención en el conjunto de datos. Sin embargo, cambiaron los clasificadores para tres de las técnicas, mientras que se mantiene las técnicas basadas en vocabularios permiten generalizar mejor los textos, especialmente TF-IDF.

Tabla 5. Clases balanceadas

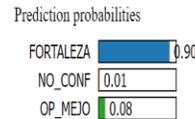
Algoritmo	F1	Técnica
LogisticRegressionCV	0.917	tfidf
RandomForestClassifier	0.911	bow
LogisticRegressionCV	0.909	word2vec
GradientBoostingClassifier	0.818	fasttext

Fuente: elaboración propia

Luego, se procedió a entrenar los clasificadores que mostraron el mejor rendimiento para cada variante de procesamiento de texto. Estos modelos se ajustaron utilizando datos balanceados mediante las técnicas identificadas previamente, además se realizó optimización de hiper-parámetros. Destacando, únicamente el algoritmo LogisticRegressionCV, al trabajar con el conjunto de datos procesados mediante word2vec y aplicando el balanceo de clases con ROS, logró una mejora en su puntuación de F1, aumentando a 0.918. En contraste, el rendimiento de los demás modelos disminuyó al implementar los nuevos hiper-parámetros.

Para explicar por qué un modelo hizo una predicción particular, se utilizó LimeTextExplainer es una parte de la biblioteca lime que permite entender cómo funcionan los modelos de clasificación de texto al introducir pequeñas modificaciones en el texto original y crear un modelo más simple que facilita generar explicaciones que resaltan las palabras clave en una instancia de texto [28]. Las instancias que fueron utilizadas, están relacionadas con las palabras frecuentes cada una de las clases. Por ejemplo, en las fortalezas palabras como permitido se repite 63 veces, para oportunidad de mejora la palabra revisión se repite 61 veces y en no conformidad numeral se repite 30 veces. Otras palabras relacionadas con palabras vacías no fueron contadas, dado que estas no fueron incluidas en los conjuntos de datos basados en vocabularios.

La figura 1, se ilustra la explicación de una instancia específica del conjunto de datos de pruebas que fue clasificada como "fortaleza". En esta explicación, se destacan las palabras clave "porque" y "permitido", las cuales tuvieron una influencia significativa en la predicción resultante.



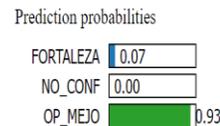
Text with highlighted words

La realización y presentación de los informes sobre las actividades que se le desarrollan a los procesos y/o dependencias, **porque** **se** **permitido** evidenciar las estrategias diseñadas y **garantizar** el cumplimiento de las solicitudes requeridas.

Fig. 1. Clasificación fortaleza.

Fuente: elaboración propia.

La figura 2 presenta la salida del modelo para una instancia asociada a una oportunidad de mejora. En esta visualización, se evidencian palabras como "revisión", "actualización" e "interesadas" las cuales tuvieron una contribución positiva en la clasificación de esta instancia.



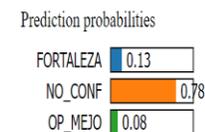
Text with highlighted words

La **revisión**, **actualización** y estandarización de los documentos del proceso **para** que la información sea controlada **y** de fácil utilización por las partes **interesadas**

Fig. 2. Clasificación oportunidad de mejora.

Fuente: elaboración propia.

Relacionado con las no conformidades, se muestra como en la figura 3, las palabras No, es muy significativa por su intensidad de color, también se tiene la palabra evidencia y numeral como contribuciones positivas a la clasificación.



Text with highlighted words

No se encuentra **evidencia** que respalde el monitoreo y la evaluación **de** **los** proveedores responsables **de** **los** servicios gestionados bajo órdenes generales, con el fin **de** determinar su desempeño. Este incumplimiento está en contradicción con el **Numeral** 9.1.3 literal F contemplado en la Norma Técnica Colombiana NTC:ISO:9001.

Fig. 2. Clasificación no conformidad.

Fuente: elaboración propia.

Las apariciones que hacen en la imagen 2 y 3 de palabras vacías, sugiere la utilización de estas en la argumentación de fortalezas y así es como el modelo lo entiende en donde el modelo explicativo local, utilizo dichas palabras presentes en las instancias perturbadas, para generar sus respectivas clasificaciones del modelo original.

4. DISCUSIÓN

Esta investigación siguió las recomendaciones propuestas en la literatura, relacionada con el procesamiento de texto, especialmente durante el preprocesamiento, donde enfoques como Bag-of-Words (BoW) y TF-IDF requieren de la eliminación de palabras poco informativas, como stopwords comunes o poco frecuentes, con el propósito de mejorar la eficacia de estos métodos [12]. El enfoque BoW y TF-IDF tal y como lo señalan Gasparetto, Marcuzzo, Zangari, & Albarelli [29], resulto vectores de alta dimensionalidad (1432 atributos). En contraste, word2vec y FastText generan vectores con dimensiones fijas independientemente del tamaño del vocabulario, preservando el orden original y la captura de patrones semánticos y relaciones entre palabras con mayor precisión [29]. A pesar de los problemas que puede generar una alta dimensionalidad en la generalización de la función objetivo que presente los datos [30], en esta investigación se mostró que, para todos los métodos de transformación de la representación de texto utilizados, los modelos entrenados generaban métricas de F1 por encima del 80% [31].

La combinación utilizada con LogisticRegressionCV, entrenado utilizando Word2Vec en el procesamiento de texto y aplicando la técnica de Random Oversampling (ROS) para balancear clases, demostró su eficacia en la clasificación de hallazgos identificados en auditorías internas. Las características del modelo elegido se centran en el uso del clasificador LogisticRegressionCV, que emplea validación cruzada para optimizar la regularización. Este enfoque al entrenarse con datos transformados mediante Word2Vec, permitió capturar las relaciones semánticas presentes en el texto. Además, la inclusión de la técnica de sobre muestreo de clases con ROS, facilitó un mejor rendimiento en la clasificación de distintas clases [32] [33].

Con la salida del modelo construido, se encontró que las palabras adquieren significado dependiendo del hallazgo que se esté reportando [34]. Para el caso del término “permitido”, el cual es usado frecuentemente en la redacción de una fortaleza, en un contexto de calidad, denota acciones encaminadas a cumplir de la mejor forma posible con requisitos, normativas o estándares establecidos [35].

Desde la clasificación de las oportunidades de mejora, se destaca el término “actualización”, que

denota hacer cambios o mejoras en algo para que esté en línea con información o tecnología más reciente. En calidad, puede implicar mantener los métodos y procesos al día. Connota, adaptación, innovación y progreso. También puede sugerir la exploración de la excelencia y la voluntad de mantenerse relevante en un entorno en constante cambio. Complementariamente, la palabra revisión, se refiere denotativamente a un análisis exhaustivo y crítico de algo para evaluar su calidad, precisión y adecuación. En el contexto de calidad, se trata de examinar procesos y resultados en busca de posibles mejoras. Connota, introspección, análisis detallado y búsqueda de posibles áreas de fortaleza y debilidad [1].

Los términos usados en un contexto de no conformidades: “evidencia” y “numeral”, palabras que son clave esencial para comprender completamente la naturaleza y los alcances de cada término. Respecto a la palabra evidencia denotativamente, se refiere a los datos, hechos o información que respaldan o corroboran una afirmación, hipótesis o conclusión. En el contexto de una no conformidad, la evidencia puede ser crucial para identificar y comprender el problema. Connota objetividad, comprobación y fundamento sólido. El término numeral, denota una sección o cláusula en una política o regulación. Connota, especificidad, detalle y formalidad en la presentación de información [36].

Finalmente, la investigación aprovechó los recursos que proporciona el procesamiento del lenguaje natural para lograr una identificación semántica que permitiera entender el significado de la redacción de los hallazgos dentro de un contexto lingüístico y cultural particular [37] [38]. Los resultados demostraron la eficacia de la combinación de las palabras incrustadas, con equilibrio de clases mediante sobre muestreo, para entrenar un modelo para clasificación de hallazgos detectados en auditorías internas, que permitirá optimizar tiempos y recursos en los procesos de mejora continua de la organización.

5. CONCLUSIONES

En esta investigación se resaltó la importancia de la automatización en la comprensión y clasificación de hallazgos en auditorías internas. Mediante técnicas de procesamiento de lenguaje natural, se logra analizar y categorizar de manera efectiva los resultados de las auditorías, lo que facilita la toma de decisiones y abre las posibilidades para evaluar

la calidad de un texto de acuerdo a como lo clasifica el modelo.

Si bien, el conjunto de datos que fue seleccionado como el que mejor generalizaba los datos fue Word2Vec, las diferentes técnicas de procesamiento de texto, como Bag-of-Words (BoW), TF-IDF, y FastText, también generan resultados cercanos al seleccionado, esto sugiere que no existe un dominio entre los métodos basados en vocabulario con relación a los métodos basados en vectores de palabras incrustadas. Adicionalmente, una vez tratado el texto y convertido a valores numéricos se pueden utilizar otras técnicas propias del aprendizaje automático como el balanceo de clases, y la optimización de hiper-parámetros para mejorar el rendimiento del modelo.

Los resultados indican como la aplicación de procesamiento de lenguaje natural puede ayudar a mejorar la comprensión de los informes de auditoría, dado que las palabras utilizadas en los hallazgos adquieren significados específicos dependiendo de la categoría (fortalezas, no conformidades u oportunidades de mejora). Esto indica que el modelo identificó adecuadamente la semántica y el contexto lingüístico utilizado por los auditores, al utilizar palabras propias de cada tipo de hallazgo para hacer la clasificación.

REFERENCIAS

- [1] N. Calso, Guía práctica para la integración de sistemas de gestión. ISO 9001, ISO 14001 e ISO 45001, Madrid: AENOR - Asociación Española de Normalización y Certificación, 2018.
- [2] M. Espino, Fundamentos de auditoría, México: Grupo Editorial Patria, 2015.
- [3] AENOR, ISO 9001: 2015 para la pequeña empresa: recomendaciones del ISO/TC 176, Madrid: AENOR Internacional, 2016.
- [4] J. Cortés, Sistemas de gestión de calidad (ISO 9001:2015), Málaga: Interconsulting Bureau, 2017.
- [5] T. Sevilla, Auditoría de los sistemas integrados de gestión ISO 9001:2015, ISO 14001:2015, ISO 45001:2018, Madrid: FC Editorial, 2019.
- [6] M. Vásquez, 6 pecados con la ISO 9001, Santa Cruz de la Sierra: El Cid Editor, 2020.
- [7] T. Xiao, C. Geng y C. Yuan, «How audit effort affects audit quality: An audit process and audit output perspective,» *China Journal of Accounting Research*, pp. 109-127, 2020.
- [8] G. Boskou, E. Kirkos y C. Spathis, «Classifying internal audit quality using textual analysis: the case of auditor selection,» *Managerial Auditing Journal*, pp. 925-950, 2019.
- [9] D. Khurana, A. Koli, K. Khatter y S. Singh, «Natural language processing: state of the art, current trends and challenges,» *Multimedia Tools and Applications*, p. 3713–3744, 2023.
- [10] R. Stuart y P. Norvig, Artificial Intelligence: A Modern Approach, Englewood Cliffs: Prentice Hall, 1995.
- [11] J. Han, M. Kamber y J. Pei, Data Mining Concepts and Techniques, Tercera ed., Waltham: Morgan Kaufmann, 2012.
- [12] V. Lakshmanan, S. Robinson y M. Munn, Machine Learning Design Patterns, Sebastopol: O'Reilly Media, 2020.
- [13] F. K. Khattak, S. Jebleea, C. Pou-Proma, M. Abdalla, C. Meaney y F. Rudzicz, «A survey of word embeddings for clinical text,» *Journal of Biomedical Informatics*, 2019.
- [14] A. Müller y S. Guido, Introduction to Machine Learning with Python, Sebastopol: O'Reilly, 2017.
- [15] T. Verdonck, B. Baesens, M. Óskarsdóttir y S. Broucke, «Special issue on feature engineering editorial,» *Machine Learning*, 2021.
- [16] S. Raschka y V. Mirjalili, Python Machine Learning Third Edition, Birmingham: Packt, 2019.
- [17] T. Mikolov, K. Chen, G. Corrado y J. Dean, «Efficient Estimation of Word Representations in Vector Space,» *arXiv*, pp. 1-12, 2013.
- [18] T. Mikolov, I. Sutskever, K. Chen, G. Corrado y J. Dean, «Distributed Representations of Words and Phrases and their Compositionality,» *arXiv*, pp. 1-9, 2013.
- [19] P. Bojanowski, E. Grave, A. Joulin y T. Mikolov, «Enriching Word Vectors with Subword Information,» *arXiv*, 2016.
- [20] A. Géron, Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, Sebastopol: O'Reilly, 2019.
- [21] M. Galar, A. Fernández, E. Barrenechea, H. Bustince y F. Herrera, «A Review on

- Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches.» *IEEE Trans Syst Man Cybern Part C*, p. 463–484, 2012.
- [22] M. Lango y J. Stefanowski, «Multi-class and feature selection extensions of Roughly Balanced Bagging for imbalanced data.» *Journal of Intelligent Information Systems*, p. 97–127, 2018.
- [23] S. Sandha, M. Aggarwal, I. Fedorov y M. Srivastava, «Mango: A Python Library for Parallel Hyperparameter Tuning.» de *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, 2020.
- [24] A. Zheng, *Evaluating Machine Learning Models*, Sebastopol: O'Reilly Media, 2015.
- [25] I. Witten, E. Frank, M. Hall y C. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, Burlington: Morgan Kaufmann, 2017.
- [26] S. Ahmed, M. Singh, B. Doherty, E. Ramlan, K. Harkin, M. Bucholc y D. Coyle, «An Empirical Analysis of State-of-Art Classification Models in an IT Incident Severity Prediction Framework.» *Applied Sciences*, pp. 1-27, 2023.
- [27] W. Zhou, H. Wang, H. Sun y T. Sun, «A Method of Short Text Representation Based on the Feature Probability Embedded Vector.» *Sensor*, 2019.
- [28] A. Bhattacharya, *Applied Machine Learning Explainability Techniques: Make ML models explainable and trustworthy for practical applications using LIME, SHAP, and more*, Birmingham: Packt, 2022.
- [29] A. Gasparetto, M. Marcuzzo, A. Zangari y A. Albarelli, «A Survey on Text Classification Algorithms: From Text to Predictions.» *Information*, pp. 1-39, 2022.
- [30] S. Galli, *Python Feature Engineering Cookbook*, Birmingham: Packt Publishing, 2020.
- [31] Z. Zhao, G. Feng, J. Zhu y Q. Shen, «Manifold learning: Dimensionality reduction and high dimensional data reconstruction via dictionary learning.» *Neurocomputing*, p. 268–285, 2016.
- [32] A. Akkasi y M.-F. Moens, «Causal relationship extraction from biomedical text using deep neural models: A comprehensive survey.» *Journal of Biomedical Informatics*, pp. 1-12, 2021.
- [33] K. Ghosh, A. Banerjee, S. Chatterjee y S. Sen, «Imbalanced Twitter Sentiment Analysis using Minority Oversampling.» de *International Conference on Awareness Science and Technology (iCAST)*, Morioka, 2019.
- [34] M. García , «La polisemia en el lenguaje cotidiano.» *Revista de Linguística Moderna* 7(2) , pp. 45-58 <https://doi.org/10.12345/rlm.2015.7.2.45> , 2015.
- [35] P. Robayo, «La innovación como proceso y su gestión en la organización: una aplicación para el sector gráfico colombiano.» *Suma de Negocios*, pp. 125-140, 2016.
- [36] C. Zheng, B. Huang, A. Agazaryan, B. Creekmur, T. Osuj y M. Gould, «Natural Language Processing to Identify Pulmonary Nodules and Extract Nodule Characteristics From Radiology Reports.» *Chest*, pp. 1902-1914, 2021.
- [37] J. Smith, *Semántica y significado*, Editorial Lingua , 2010.
- [38] R. García y M. Huerta , «Significado y sociedad.» *Sincronía*, núm. 77. Disponible en: <https://www.redalyc.org/articulo.oa?id=513862147026>, pp. 530-544, 2020.
- [39] M. Schonlau y R. Y. Zou, «The random forest algorithm for statistical learning.» *The Stata Journal*, pp. 3-29, 2020.