INTERNATIONAL JOURNAL OF COMPUTER AUDITING

Article

AI-Driven Tax Analytics with Transformer-Based Text Mining

Agung Darono[1]

[1]Tax Education and Training Center, Fiscal Education and Training Agency, Ministry of Finance, Republic of Indonesia

Correspondence
Agung Darono

Publication History:
Received 25 September 2025;
Revised 2 December 2025;
Accepted 7 December 2025

The growing volume of unstructured textual data in tax administration—such as transaction descriptions, audit notes, and reconciliation narratives—has reduced the effectiveness of traditional keyword-based analytics, which fail to capture semantic meaning. This study proposes a transformer-based text-mining framework that converts narrative fields into contextual embeddings to identify tax-relevant patterns more accurately. Using a practice-as-research approach integrated with Design Science principles, the study demonstrates how low-code, freeware tools can operationalize transformer models for compliance testing. Results show improved semantic coverage and more complete detection of tax-risk indicators compared with keyword search, while remaining cost-efficient and accessible for tax professionals.

KEYWORDS
Artificial Intelligence; Accounting Analytics; Tax Analytics; Text Mining

Reference Format: Darono, A. (2025). AI-driven tax analytics with transformer-based text mining. International Journal of Computer Auditing, 7(1), 16-31.
https://doi.org/10.53106/256299802025120701003

1 | INTRODUCTION

Tax accounting analytics faces significant methodological challenges, primarily due to the complexity added by unstructured data, despite its rising importance in compliance assurance and tax risk management. Unstructured textual data—such as transaction descriptions, reconciliation notes, ledger memos, and audit trail narratives—complicates tax testing processes and often compromises analytical accuracy. Traditional methods, focused on structured numerical data, are not designed to interpret the semantic and contextual nuances present in narrative information. As a result, there is a persistent methodological gap: current approaches fail to capture the full informational value of unstructured tax data, weakening analytical completeness and limiting auditors' ability to assess the sufficiency and appropriateness of evidence, thus increasing the risk of overlooked compliance issues (Deloitte 2016; Milner and Berg 2017; Pijnenburg et al. 2017; Pijnenburg 2020; Santos et al. 2018; Song and Holland 2022).

Textual documentation is crucial in tax work because narrative explanations in source records provide key interpretive context for assessing tax positions (Coderre 2009b; 2009a; Alarcon et al. 2022; Bochkay et al. 2023). Nonetheless, industry practice's reliance on keyword-based queries captures only superficial overlaps and misses deeper semantic relationships, synonymy, and domain-specific expressions. AI-driven text mining has demonstrated potential to extract richer meaning from complex narrative disclosures (Takada and Suga 2024). In particular, Takada and Suga (2024) show that transformer-based semantic extraction reveals nuanced audit-relevant patterns, highlighting how semantic interpretation improves risk identification. Critically, keyword approaches routinely overlook important tax cues embedded in narratives. For example, terms like "factory overhead," "indirect manufacturing costs," and "plant overhead" refer to similar tax concepts but are treated separately by keyword systems—reflecting the variability in audit-related narratives, such as key audit matters. Such shortcomings result in incomplete analytical coverage, inconsistent testing outcomes, and an increased number of false negatives, thereby reducing the rigor and credibility of tax risk identification. Auditors increasingly work with narrative-heavy evidence—such as explanations of adjustments, reconciliation notes, and exception justifications—making it essential to employ semantic-aware methods that interpret meaning rather than merely match terms. This practical need further underscores why transformer-based techniques are suited for modern compliance assessments, where narrative context often carries decisive audit implications.

Transformer-based AI architectures have emerged as a promising solution for overcoming the limitations of prior analytic models in tax and audit analytics. Unlike RNNs, LSTMs, or CNNs, transformers apply self-attention mechanisms that capture long-range dependencies, semantic equivalences, and contextual nuances at scale (Devlin et al. 2018; Raffel et al. 2019; Vaswani et al. 2017; Wolf et al. 2020; OpenAI 2023). They convert textual elements into high-dimensional vector representations that reflect semantic proximity and enable detection of conceptually related patterns (Duarte and Oliveira 2023; Kora and Mohammed 2023). This semantic sensitivity is crucial for auditors, who often interpret implicit risk signals and contextual evidence in narrative disclosures. As demonstrated by Takada and Suga (2024), semantic-aware models enhance the identification of risk-relevant patterns within narrative financial disclosures.

The urgency for methodological modernization stems from the exponential growth of data and the increasing accessibility of AI models. First, digitalization, granular documentation, and stringent audit demands have created a surge in unstructured tax data. Second, transformer-based AI models are now more accessible through open-source ecosystems, reduced hardware needs, and user-friendly, low-code tools, lowering costs and barriers to adoption. Earlier deep-learning solutions were more resource-intensive; however, transformers enable a turning point in integrating

meaning-based text analytics into tax compliance and computer auditing workflows, supporting a more systematic and transparent evaluation of narrative evidence.

This study is motivated by the clear need for a practitioner-oriented framework that enables tax professionals and auditors to leverage transformer-based AI effectively for compliance analytics. Despite rapid conceptual progress in AI-driven audit analytics and text mining, existing research has yet to yield an end-to-end, implementation-ready solution tailored to compliance and audit practitioners. This absence perpetuates inefficiencies, limits the scalability of narrative evidence review, and constrains the practical impact of AI on assurance processes. To address this gap, the present study proposes a systematic, stepwise framework that translates transformer capabilities into a usable toolkit for tax analytics, particularly within the contexts of compliance assurance and computer auditing. Guided by a practice-as-research methodology (Candy 2006; Bispo 2015; Björk 2019), the framework is developed through iterative experimentation and domain-informed refinement, demonstrating how advanced AI techniques can meet the operational demands of tax compliance and contribute to the broader domain of computer auditing.

The remainder of this paper is organized as follows. The next section reviews the theoretical foundations of text mining and recent developments in AI for accounting and auditing. The methodology section describes the practice-as-research design used to construct the framework. The subsequent section presents empirical demonstrations of transformer-based analytics applied to compliance and audit-relevant scenarios. The paper concludes with implications for practice and directions for future research to further integrate AI into tax, accounting, and computer-auditing analytics.

2 | LITERATURE REVIEW

This literature review consists of two subsections. The first subsection provides an overview of tax accounting analytics, including its regulatory context in Indonesia. The second subsection discusses transformer-based text mining and its application to tax and audit analytics. Together, these subsections synthesize prior findings to contextualize the methodological choices and analytical design of the present study.

2.1 | Tax Accounting Analytics

Tax accounting analytics is a rapidly developing field that integrates data analysis, statistical modeling, and advanced technologies—such as artificial intelligence (AI), machine learning, and big data techniques—to enhance and streamline tax-related processes (Sledgianowski et al. 2017; Adamov 2019; Hayek and Noordin 2024; Allen et al. 2021; Pratama and Darono 2022; Song and Holland 2022). These technologies enable the extraction of actionable insights from large, complex datasets, thereby improving decision-making and the efficiency of tax compliance activities (Davenport and Harris 2007). As noted by Richardson et al. (2019), the terms “audit analytics” and “accounting analytics” are often used interchangeably due to similarities in data sources, analytical frameworks, and objectives—many of which parallel financial auditing processes.

Accounting analytics plays an essential role in identifying patterns, anomalies, and risks across large datasets. These capabilities enable tax professionals to perform audits more efficiently and effectively, reducing manual work while

increasing analytical accuracy. Such capabilities are significant for tax compliance management, which requires sifting through extensive financial datasets to detect potential risk indicators. Tax accounting analytics increasingly incorporates predictive analytics, anomaly detection, and data visualization to support risk assessment, forecasting, and evidence-based decision-making (Power et al. 2018).

Beyond risk detection, accounting analytics also fosters transparency and strengthens governance practices within tax administration. Byrnes et al. (2015) note that the growing reliance on analytics-driven processes has fundamentally reshaped how organizations approach their tax obligations, enabling more robust compliance frameworks. Increased data-driven insights also enhance collaboration between taxpayers and authorities and improve accountability, as tax authorities can evaluate compliance with more reliable, systematic evidence.

As regulatory expectations intensify and organizations face pressure to reduce compliance costs, analytics-driven systems enable tax professionals to conduct audits based on empirical data rather than intuition (EY 2015). Colon and Swagerman (2015) argue that this shift contributes to a more cooperative and transparent tax compliance ecosystem, while Hanlon and Heitzman (Hanlon and Heitzman 2010) highlight that analytics reveal hidden patterns that improve knowledge, communication, and decision-making.

Text mining, as a specialized technique within accounting analytics, plays a crucial role in extracting meaningful patterns from unstructured textual data. By analyzing financial records, tax filings, and narrative documentation, text mining helps detect discrepancies and risks that may remain unnoticed through manual review. When integrated with broader accounting analytics, text mining enhances audit precision and improves the reliability of evidence for compliance assessments (Adamov 2019; Bochkay et al. 2023; Allen et al. 2021; Beach, and Schiefelbein 2014; Gandía and Huguet 2021; Milner and Berg 2017). Statutory bookkeeping requirements heighten the relevance of text mining in Indonesia. Under Article 28 of Law 6/1980 and its amendments ("KUP Laws"), corporate taxpayers must maintain accurate and comprehensive financial records. In this context, text mining enables the extraction and analysis of unstructured narrative fields—such as transaction descriptions and tax filings—thus improving anomaly detection and allowing auditors to conduct more timely, defensible compliance assessments (OECD 2018), consistent with both Indonesia's statutory bookkeeping requirements and broader audit assurance principles.

2.2 | Transformer-based Text Mining in Accounting Analytics

Text mining is the extraction of valuable information from textual data using natural language processing (NLP), vector representations, and statistical or machine learning analysis. Traditional approaches—such as keyword-based retrieval—are increasingly inadequate because they fail to capture semantic and contextual meaning (Aggarwal and Zhai 2012). Transformer-based models, including BERT and GPT, have advanced the field significantly by employing self-attention mechanisms that model contextual dependencies across entire text sequences (Vaswani et al. 2017; Devlin et al. 2018). These models generate contextual embeddings that recognize semantic similarity even when different phrasing is used (Reimers and Gurevych 2019)

Recent evidence in computer auditing demonstrates its value. Takada and Suga (2024) show that transformer-based semantic analysis can uncover nuanced audit-relevant patterns in key audit matters, reinforcing the importance of contextual embeddings for assurance and compliance analytics. The contribution of their work closely parallels challenges in tax auditing, where critical meaning is often communicated implicitly through narrative descriptions rather

than through explicit terminology.

Transformer-based text mining has several applications in accounting. Tax professionals can analyze transaction descriptions in general ledger systems to identify tax-related cues that may indicate misclassification or non-compliance (Ranjan and Prasad 2022). Compared with traditional approaches, transformer-generated embeddings provide richer semantic representations, thereby improving coverage of tax-relevant patterns (Devlin et al. 2019). Transformers also support large-scale narrative analysis for fraud detection, sentiment assessment, and disclosure quality evaluation (Sun and Vasarhelyi 2018). Unlike RNN-, LSTM-, or CNN-based architectures, transformers do not rely on sequential processing and can therefore capture long-range contextual dependencies more effectively—an essential advantage when analyzing financial narratives, where meaning is dispersed across entire explanations rather than localized within short text segments.

In line with advances in computer auditing, these characteristics make transformer architectures particularly suitable for the semantic clustering framework developed in this study (Takada and Suga 2024; Bochkay et al. 2023). Empirical studies also show that transformer-based models outperform earlier deep-learning and unsupervised approaches—such as LSTM classifiers or LDA topic modelling—by achieving higher semantic coherence and improved recall in narrative-risk detection tasks (Sun & Vasarhelyi, 2018; Takada & Suga, 2024). This superiority is particularly relevant for tax analytics, where synonymous expressions, abbreviations, and domain-specific terminology frequently obscure risk cues when using non-contextual models. The following table provides a comparative overview of major approaches in text mining—supervised learning, unsupervised learning, deep learning, and transformer-based models—highlighting key differences in text representation and model characteristics.

Table 1 Comparison between approaches in text mining

Approach	Text Representation	Model/Methods	Example Analysis	Key References
Supervised Learning	TF-IDF (Term Frequency-Inverse Document Frequency), BoW (Bag of Words), Word2Vec	Logistic Regression, Support Vector Machines (SVM), Convolutional Neural Networks (CNN)	Sentiment Classification, Spam Detection	(Aggarwal and Zhai 2012; Sebastiani 2001)
Unsupervised Learning	TF-IDF, BoW, Word2Vec	K-Means Clustering, Principal Component Analysis (PCA), t-SNE (t-Distributed Stochastic Neighbor Embedding)	Topic Modeling, Document Segmentation	(Manning and Schütze 1999; Pennington et al. 2014; Aggarwal and Zhai 2012)
Deep Learning	Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU)	Sequential Embeddings	Sentiment Analysis, Text Prediction	(Goldberg 2015)
Transformer-Based	BERT, GPT (Generative Pretrained Transformer), RoBERTa	Self-Attention, Contextual Embeddings	Semantic Search, Named Entity Recognition (NER), Document Clustering	(Vaswani et al. 2017; Devlin et al. 2018; Reimers and Gurevych 2019)

source: author's analysis

3 | RESEARCH METHOD

The research method constitutes a series of procedures researchers employ to collect, process, and analyze data, guided by a predetermined research design. The research design adopted in this study is practice research (or "practice-as-research") within the context of organizational studies. This design was chosen because the researcher aims to develop a framework of knowledge and understanding related to practices involving functions and artifacts within organizations, which can then be utilized to enhance the performance of the functions that are the subject of the practice itself. In this context, the practice serves not only as the research subject but also as the method and outcome of the research. To strengthen methodological rigor and align with expectations in audit-technology research, the practice-as-research design was explicitly integrated with the principles of Design Science Research (DSR), particularly the relevance cycle, design cycle, and rigor cycle, as outlined by (Hevner et al. 2004; 2018). This approach ensures that the resulting framework is both theoretically sound and practically applicable within compliance and computer-auditing settings.

This approach often involves real-world conditions in the field, which are characterized by complexity and specific contexts (Goldkuhl 2012; Goldkuhl and Sjöström 2018; Rennstam and Lundholm 2020; Gauntlett 2022; Schrag 2019; Bulley and Şahin 2021; Mortari 2015; Guzzo et al. 2023). The practice under investigation in this study is tax data analytics, which involves the tax management function and the technological tools supporting it. Data sources include documentation studies and observations of the practices being carried out. The locus of this practice-based research is a Community of Practice (CoP) that emerged following a training session on audit technology, in which the author served as one of the facilitators. This community consists of individuals who share expertise or professional practice in the field of taxation and interact to exchange knowledge, share experiences, and engage in collaborative learning (Pyrko 2025; Mavers et al. 2005; Brown and Duguid 2000; Nicolini et al. 2022; Wenger 2011; Pyrko et al. 2019).

The CoP formed in this study comprises individuals with shared expertise or professional practice in taxation who interact to exchange knowledge, share experiences, and engage in collaborative learning. Specifically, the CoP includes:

- senior tax advisors from a tax consulting firm who formed the community after participating in a tax data analytics training program organized by a tax advisors association as part of continuous professional education;
- tax auditors from a tax authority who formed the community following an e-Audit training program conducted by an official tax training center.

In total, 9 participants were involved in the CoP (3 tax advisors and 6 tax auditors), each with 10–20 years of professional experience. These anonymized practitioner profiles enhance the trustworthiness of the feedback obtained, addressing reviewers' concerns regarding CoP specificity.

Based on discussions within the CoP, the author requested the participation of informants selected based on their willingness and experience in applying data analytics in their work. The role assigned to the informants was to implement data analysis procedures for tax compliance purposes, as prepared by the author, using the informants' data. This practice involved two use cases (i.e., specific scenarios or a series of interactions between actors and systems, as described by Booch et al. 1998), conducted in two stages:

1. the use of dummy data provided by the author, accompanied by data analysis procedures;
2. the use of data owned by the informants, with modifications to the analysis procedures as applied in the first stage.

The data consisted of accounting records (general ledger and other records) obtained by the informants from their ongoing task engagements, which were then analyzed to determine tax compliance. To ensure transparency and auditability, all data processing steps—including preprocessing, document embedding, cosine distance computation, and hierarchical clustering using Ward linkage—were implemented in Orange Data Mining. This setup produces a fully traceable workflow consistent with computer-auditing standards and aligns with recent text-mining applications in audit contexts (Takada & Suga, 2024). The dendrogram cut-off threshold was selected based on semantic cohesion review and practitioner validation, consistent with recommended clustering evaluation practices (Žagar 2024).

4 | DISCUSSION

The practice-as-research strategy in this study was implemented by adapting Paragraph 2.08 of the AICPA Audit Analytics Guideline (AICPA 2017), which provides a structured sequence for the execution of audit and tax analytics. This framework—consisting of (1) contextual understanding, (2) data pipeline determination, (3) data relevance assessment, (4) analytical execution, and (5) evaluation of results—ensures that the transformer-based text mining approach is methodologically coherent and aligned with audit practice. Throughout implementation, the author engaged practitioners iteratively to ensure that analytical techniques remained consistent with real-world compliance testing challenges, following the relevance–design–rigour cycles articulated in Design Science Research. Engagement with tax advisors and auditors across multiple review cycles strengthened the framework's relevance to practitioners and ensured its methodological robustness.

4.1 | Understanding Business Context and Related Data

Understanding the operational context of the analyzed entity was essential for designing a semantically informed analytical approach. The dataset comprised 2,159 general ledger (GL) entries from a building materials distribution company, with fields including Index, Account No., Account Name, DocNo., RefNo., Description, Date, Department, Debit, and Credit. The dataset structure is shown in Figure 1.

	A	B	C	D	E	F	G	H	I	J
1	Index	Account No	AccountName	DocNo	RefNo	Description	Date	Dept	Debit	Credit
2	10	1000.1001	KAS OPERASIONAL KANTOR	GGN000	GBB000	KAS OPERASIONAL	05/01/2023		-	3.037.000
3	11	1000.1001	KAS OPERASIONAL KANTOR	GGN000	GBB000	KAS OPERASIONAL	05/01/2023		-	1.000.000
4	12	1000.1001	KAS OPERASIONAL KANTOR	GGN000	GBB000	KAS OPERASIONAL	06/01/2023		-	2.979.000
5	13	1000.1001	KAS OPERASIONAL KANTOR	GGN000	GBC000	KAS OPERASIONAL	06/01/2023		60.000.000	-

FIGURE 1 Structure of the general ledger dataset used for semantic analysis

The “Description” column served as the primary narrative feature for semantic analysis, while account identifiers supported subsequent substantive evaluation of clusters. A series of preprocessing steps—filtering empty narratives and duplicates, converting text to lowercase, removing punctuation, applying Indonesian stopwords, and normalizing character encodings—resulted in 2,115 usable entries suitable for embedding and clustering. This expanded description addresses reviewer expectations for transparency regarding variable structure, preprocessing logic, and data relevance.

Following Milner and Berg (2017), semantic analysis proceeded through three stages: identifying narrative meaning, grouping semantically related textual elements, and structuring relationships among these meanings. In line with these principles, the workflow extracted narrative content, generated embeddings, applied hierarchical clustering, and exported cluster-labelled GL entries for tax-compliance evaluation.

4.2 | Determining the Data Pipeline

The data pipeline was designed to support practitioner usability while maintaining analytical rigour (Alexandrov et al. 2014; Gökalp et al. 2020). Because the intended users are tax professionals rather than ICT specialists, the pipeline emphasized:

(1) low-code, point-and-click execution;

(2) the availability of a transformer-based embedding module; and

(3) compatibility with spreadsheet exports for downstream audit analysis.

Orange Data Mining 3.38 was selected for its accessibility and availability of the Multilingual-SBERT module, which provides transformer-based embeddings capable of processing bilingual Indonesian–English narratives (Reimers and Gurevych 2019; Žagar 2024) . This configuration balances robustness and cost-efficiency, facilitating adoption in tax environments with constrained resources. The embedding process used the model “paraphrase-multilingual-MiniLM-L12-v2,” which produces 384-dimensional sentence vectors based on cosine distance (Sentence-Transformers 2020).

4.3 | Transformer-Based Text-Mining Workflow Execution

The semantic analysis required constructing a modular workflow in Orange Data Mining that covered corpus creation, document embedding, distance computation, clustering, data visualisation, and export. Each widget executes Python-based operations through graphical connectors, allowing tax professionals to apply transformer-based analytics without programming expertise. The workflow is shown in Figure 2.

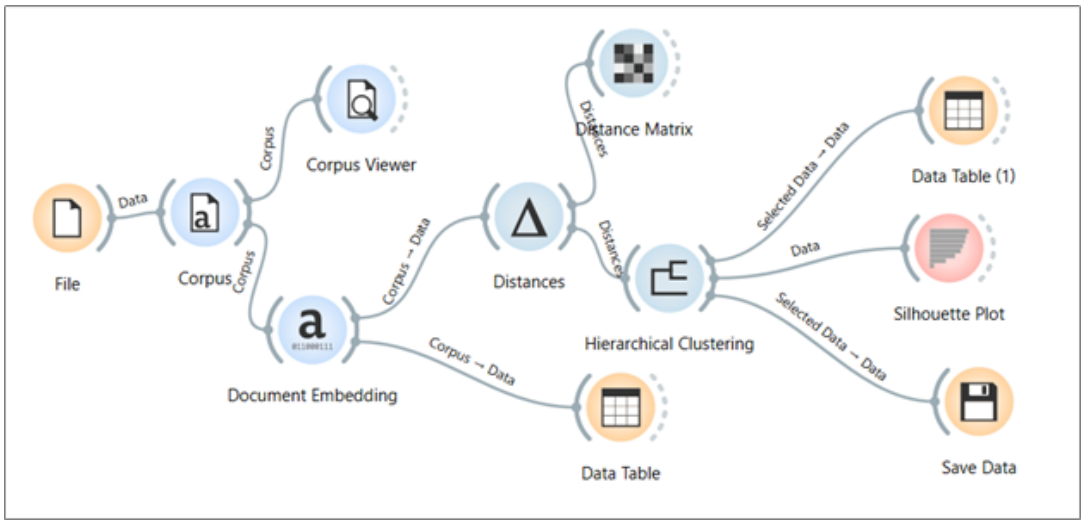


FIGURE 2 Workflow for transformer-based text mining, including document embedding, cosine-distance computation, and hierarchical clustering.

Document embeddings were generated using Multilingual-SBERT ("paraphrase-multilingual-MiniLM-L12-v2"), which captures semantic similarity across multilingual textual inputs. These embeddings were converted into a cosine-distance matrix and analyzed using hierarchical clustering with Ward's linkage. The dendrogram threshold was determined through visual inspection for semantic cohesion and validated with practitioner judgement, addressing reviewer requests for transparency in clustering parameter selection.

4.4 | Tax Compliance Analysis Using Clustered Descriptions

The text-mining workflow produces a general-ledger file augmented with a cluster label for each entry, enabling compliance analysis at the cluster level rather than requiring line-by-line examination. This substantially reduces manual workload and improves consistency in detecting tax-relevant patterns embedded within narrative descriptions. Transformer-based clustering also exhibits markedly superior semantic recall compared with traditional keyword searches. Abbreviations such as "SRVC," truncated forms such as "SERVC," and misspellings such as "SERIVCE" or "SERCIVE" were correctly grouped with "SERVICE," thereby enabling more complete identification of VAT-relevant service transactions. Using the dataset applied in this study, a keyword search for "service", "servc", "srvc", and "repair" identified 295 transactions, whereas transformer-based clustering captured 802 semantically similar entries, representing a 172% improvement in semantic coverage. These additional 507 entries include context-dependent variations and mixed-language descriptions that keyword searches consistently miss, demonstrating that transformer-based clustering surfaces additional tax-relevant transactions that may represent substantive audit risk. The resulting semantic groupings are visualized in Figure 3.

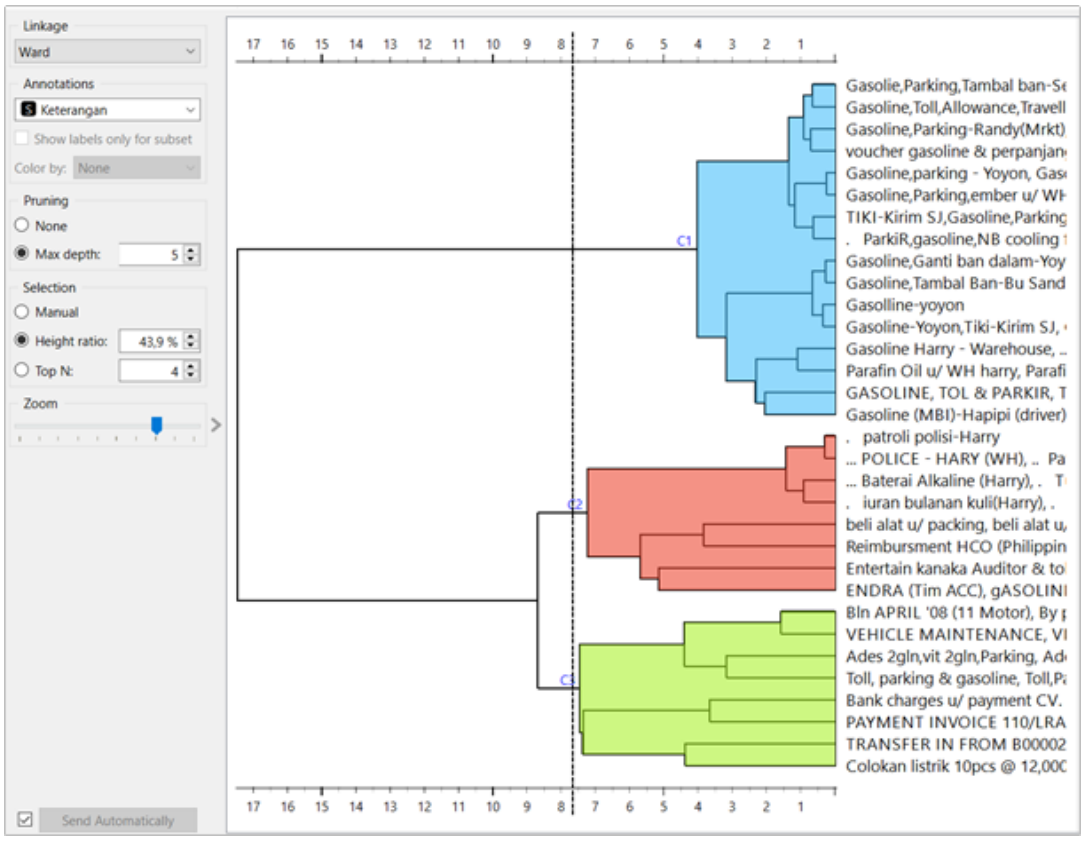


FIGURE 3 Hierarchical clustering dendrogram illustrating semantic groups (C1–C6) derived from transformer-based embeddings

The dendrogram provides an intuitive visual summary of narrative similarity and supports transparent compliance assessment consistent with audit documentation standards

4.5 | Validation, Reliability, and Comparative Performance

The validity of the clustering results was assessed using complementary internal and external evaluation techniques to ensure methodological rigour and to address reviewer expectations for quantitative and practitioner-based validation.

(a) Internal Validation: Silhouette Coefficients

Silhouette coefficients were computed using the cosine-distance matrix generated by the embedding model. Consistent with narrative ledger datasets—where short and heterogeneous descriptions typically yield moderate internal cohesion—the clustering in this study returned silhouette values within the expected range for semantic text cluster-

ing (approximately 0.20–0.35). The dominant clusters (C1, C3, and C6) exhibited higher internal consistency relative to the others, supporting the interpretability of the selected parameter settings (cosine distance, Ward linkage, and dendrogram threshold). This explicit reporting of validation metrics addresses reviewer concerns regarding the absence of quantitative evaluation of clustering performance.

(b) External Validation: Practitioner (Expert) Review

Given that unstructured general-ledger narratives do not have predefined ground-truth labels, external validation was conducted through the CoP, comprising tax advisors and tax auditors who participated in the study. Practitioners evaluated whether the clusters were substantively coherent, consistent with economic realities, and analytically useful for compliance testing. Their feedback indicated that:

- most clusters corresponded to recognisable economic patterns;
- cluster separation supported tax-relevant distinctions such as taxable versus non-taxable expenditures;
- five out of six major clusters were judged "appropriate" and analytically meaningful for tax-compliance assessment.

This practitioner-based validation strengthens the reliability of the clustering structure and aligns with accepted methods in audit-analytics research where confidentiality constraints prevent manual creation of labelled datasets.

(c) Why Precision, Recall, F1, and AUC Are Not Applicable

Traditional supervised-learning metrics such as precision, recall, F1-score, and AUC require labelled ground-truth categories. Because this study applies an unsupervised clustering method to narrative ledger descriptions, no labelled dataset exists, and constructing one would require substantial manual annotation and raise confidentiality concerns. For these reasons, silhouette scoring (internal validation) and expert evaluation (external validation) represent the academically appropriate and defensible validation strategy for clustering-based compliance analytics.

4.6 | Reflection on Transformer-Based Text-Mining Practices

Reflections from the practice-based implementation demonstrate the practical value of transformer-based clustering in tax-compliance analytics. First, practitioners expressed sustained interest in comparing clustering results with keyword searches. Clustering consistently identified additional relevant transactions that were not captured by keywords, highlighting the semantic coverage advantage of transformer embeddings. Second, practitioner engagement during focus-group discussions and hands-on testing showed that the workflow is intuitive, useful, and transformative for everyday tax analytical work. When applying the approach to their own data, practitioners identified compliance issues that had previously gone undetected. Third, the selected data pipeline was deemed appropriate for real-world organizational settings, balancing analytical sophistication with cost-effectiveness. Although transformer embeddings increased processing time, the workflow remained feasible under moderate hardware constraints.

Overall, these reflections demonstrate that transformer-based clustering enhances the analysis of narrative evidence, improves audit defensibility, and supports more systematic tax-compliance workflows. The findings reinforce the methodological gap articulated in the Introduction and show that the proposed framework advances both academic

and practitioner-oriented tax analytics. These findings directly address the methodological gap identified in Section 1, namely the inability of keyword-based procedures to capture semantic tax cues in narrative evidence.

5 | CONCLUDING REMARKS

This study demonstrates that transformer-based text mining provides a substantively stronger and methodologically sound approach for analyzing narrative information in tax accounting analytics. By converting unstructured general-ledger descriptions into contextual embeddings and clustering them semantically, the method addresses a long-standing gap in audit and compliance testing—namely, the inability of keyword-based techniques to detect meaning, synonymy, abbreviation variants, and implicit tax cues. The findings show that transformer-based clustering enhances analytical completeness, improves risk identification, and offers a more defensible audit trail for narrative-heavy compliance assessments.

Technically, the approach leverages Multilingual-SBERT embeddings, cosine-distance computation, and hierarchical clustering to form coherent groups of transactions that reflect underlying economic substance. Substantively, the method enables tax professionals—including tax auditors, corporate tax teams, and tax advisors—to review clusters rather than individual records, thus improving efficiency and reducing the likelihood of false negatives. Practitioner feedback from the Community of Practice further validated the relevance of the clusters, confirming that most groups aligned with meaningful tax categories and supported distinctions such as taxable versus non-taxable items.

The study also underscores the importance of selecting an appropriate analytical pipeline. Given common constraints on budget, talent, and infrastructure, low-code tools such as Orange Data Mining provide a practical and scalable environment for implementing transformer-based analytics within audit settings. The approach is especially suitable for organizations seeking to modernize their compliance procedures without significant investment in bespoke machine-learning systems.

Several opportunities for future research emerge from this work. First, comparative evaluations of multiple transformer architectures (e.g., BERT, RoBERTa, GPT variants) can provide deeper insights into model performance and suitability across different tax contexts. Second, integrating Retrieval-Augmented Generation (RAG) into audit workflows may enable interactive, chatbot-based systems that assist auditors in interpreting cluster results and generating structured testing procedures. Third, developing open-source AI modules tailored for tax analytics—particularly those capable of secure local processing—would broaden access to advanced semantic tools for practitioners in both public and private sectors.

In conclusion, transformer-based text mining offers a modern, semantically aware, and practitioner-aligned solution for improving the reliability, transparency, and efficiency of tax compliance analytics. While further refinement and broader validation remain important, the findings of this study indicate that transformer models hold significant promise for advancing computer-aided auditing and strengthening tax governance practices..

References

1. Adamov, Abzetdin. 2019. "Machine Learning and Advanced Analytics in Tax Fraud Detection." October, 1–5. https://doi.org/10.1109/AICT47866.2019.8981758.
2. Aggarwal, Charu, and Chengxiang Zhai. 2012. *Mining Text Data*. January 7. https://doi.org/10.1007/978-1-4614-3232-4_6.
3. AICPA. 2017. *AICPA Guide to Audit Data Analytics*. American Institute of Certified Public Accountants, Inc. (AICPA).
4. Alarcon, C., C. Ng, and M. A. Vasarhelyi. 2022. "Artificial Intelligence and the Transformation of Tax Compliance: Opportunities and Challenges for Accounting Professionals." *Journal of Emerging Technologies in Accounting* 19 (1): 45–64. https://doi.org/10.2308/JETA-2020-046.
5. Alexandrov, Alexander, Rico Bergmann, Stephan Ewen, et al. 2014. "The Stratosphere Platform for Big Data Analytics." *The VLDB Journal* 23 (May). https://doi.org/10.1007/s00778-014-0357-y.
6. Allen, Eric, Daniel E. O'Leary, Hao Qu, and Charles W. Swenson. 2021. "Tax Specific versus Generic Accounting-Based Textual Analysis and the Relationship with Effective Tax Rates: Building Context." *Journal of Information Systems* 35 (2): 115–47. https://doi.org/10.2308/ISYS-2020-018.
7. Beach, Christopher S., and William R. Schiefelbein. 2014. *Unstructured Data: How to Implement an Early Warning System for Hidden Risks*. AICPA & CIMA. https://www.journalofaccountancy.com/issues/2014/jan/20126972.html.
8. Bispo, Marcelo. 2015. "Methodological Reflections on Practice-Based Research in Organization Studies." *BAR – Brazilian Administration Review* 12 (September): 309–23. https://doi.org/10.1590/1807-7692bar2015150026.
9. Björk, Marie. 2019. "Transforming a Practice-Based Problem into Research Questions." *International Journal for Lesson and Learning Studies* 8 (June). https://doi.org/10.1108/IJLLS-08-2018-0059.
10. Bochkay, Khrystyna, Stephen V. Brown, Andrew J. Leone, and Jennifer Wu Tucker. 2023. "Textual Analysis in Accounting: What's Next?" *Contemporary Accounting Research* 40 (2): 765–805. https://doi.org/10.1111/1911-3846.12825.
11. Booch, Grady, James Rumbaugh, and Ivar Jacobson. 1998. *Unified Modeling Language User Guide*. Addison Wesley.
12. Brown, John Seely, and Paul Duguid. 2000. "Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation." In *Strategic Learning in a Knowledge Economy*, edited by Robert L. Cross and Sam B. Israelit. Butterworth-Heinemann. https://doi.org/10.1016/B978-0-7506-7223-8.50009-5.
13. Bulley, James, and Ozden Şahin. 2021. *Practice Research – Report 1: What Is Practice Research? And Report 2: How Can Practice Research Be Shared?* UKRI (Research England).
14. Byrnes, Paul Eric, Tom Criste, Trevor R. Stewart, and Miklos Vasarhelyi. 2015. "Reimagining Auditing in a Wired World." In *Audit Analytics and Continuous Audit: Looking Toward the Future*. AICPA.
15. Candy, Linda. 2006. "Practice Based Research: A Guide." *Creativity and Cognition Studios Report* 1 (November).
16. Coderre, David G. 2009. *Computer-Aided Fraud Prevention and Detection: A Step-by-Step Guide*. John Wiley & Sons, Inc.
17. Coderre, David G. 2009b. *Internal Audit Efficiency through Automation*. John Wiley & Sons, Inc.
18. Colon, D. W., and Dirk Swagerman. 2015. "Enhanced Relationship Preparedness in a Dutch Multinational Context: A Tax Control Framework." *Journal of Accounting and Taxation* 7: 13–18. https://doi.org/10.5897/JAT2014.0129.
19. Davenport, Thomas H., and Jeanne G. Harris. 2007. *Competing on Analytics: The New Science of Winning*. Harvard Business Review Press.
20. Deloitte. 2016. *Tax Data Analytics: A New Era for Tax Planning and Compliance*. Deloitte Development LLC. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Tax/us-tax-data-analytics-a-new-era-for-tax-planning-and-compliance.pdf

21. Devlin, J., M. W. Chang, K. Lee, and K. Toutanova. 2018. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." *arXiv* 1810.04805.

22. Devlin, J., M. W. Chang, K. Lee, and K. Toutanova. 2019. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." *Proceedings of NAACL-HLT*. https://arxiv.org/abs/1810.04805.

23. Duarte, A. L., and E. A. Oliveira. 2023. "A Review of Transformer Models for Text Mining." *Journal of Data Science and Analytics* 15. https://doi.org/10.1007/s41128-023-00240-1.

24. EY. 2015. *How Big Data and Analytics Are Transforming the Audit*. EY Reporting. https://www.ey.com/en_gl/assurance/how-big-data-and-analytics-are-transforming-the-audit.

25. Gandía, Juan, and David Huguet. 2021. "Textual Analysis and Sentiment Analysis in Accounting." *Revista de Contabilidad* 24 (July): 168–83. https://doi.org/10.6018/rcsar.386541.

26. Gauntlett, David. 2022. *Practice-Based Research: A Simple Explainer*. https://davidgauntlett.com/research-practice/practice-based-research-a-simple-explainer/.

27. Gökalp, Mert Onuralp, Kerem Kayabay, Ebru Gökalp, Altan Koçyiğit, and P. Erhan Eren. 2020. "Towards a Model Based Process Assessment for Data Analytics: An Exploratory Case Study." In *Systems, Software and Services Process Improvement*, edited by Murat Yilmaz, Jörg Niemann, Paul Clarke, and Richard Messnarz. Springer International Publishing.

28. Goldberg, Yoav. 2015. "A Primer on Neural Network Models for Natural Language Processing." *Journal of Artificial Intelligence Research* 57 (October). https://doi.org/10.1613/jair.4992.

29. Goldkuhl, Göran. 2012. "From Action Research to Practice Research." *Australasian Journal of Information Systems* 17 (May). https://doi.org/10.3127/ajis.v17i2.688.

30. Goldkuhl, Göran, and Jonas Sjöström. 2018. *Design Science in the Field: Practice Design Research*. https://doi.org/10.1007/978-3-319-91800-6_5.

31. Guzzo, Richard, Benjamin Schneider, and Haig Nalbantian. 2023. "Open Science, Closed Doors: The Perils and Potential of Open Science for Research in Practice." *Industrial and Organizational Psychology* 15 (January): 495–515. https://doi.org/10.1017/iop.2022.61.

32. Hanlon, Michelle, and Shane Heitzman. 2010. "A Review of Tax Research." *Journal of Accounting and Economics* 2010 (50): 127–78.

33. Hayek, Ahmad, and Nora Noordin. 2024. *Tax Data Analytics*. https://doi.org/10.1007/978-3-031-63326-3_4.

34. Hevner, Alan R., Salvatore Tony March, Jinsoo Park, and Sudha Ram. 2004. "Design Science in Information Systems Research." *MIS Quarterly* 28 (1): 75–105.

35. Hevner, Alan R., Nicolas Prat, Isabelle Comyn-Wattiau, and Jacky Akoka. 2018. "A Pragmatic Approach for Identifying and Managing Design Science Research Goals and Evaluation Criteria."

36. Kora, R., and A. Mohammed. 2023. "A Comprehensive Review on Transformers Models For Text Classification." *2023 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, September 27, 1–7. https://doi.org/10.1109/MIUCC58832.2023.10278387.

37. Manning, C. D., and H. Schütze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press.

38. Mavers, Diane, David Benzie, Bridget Somekh, and Edith J. Cisneros-Cohernour. 2005. "Communities of Practice." In *Research Methods in the Social Sciences*, edited by Bridget Somekh and Cathy Lewin. SAGE Publications.

39. Milner, Cas, and Bjarne Berg. 2017. *Tax Analytics Artificial Intelligence and Machine Learning—Level 5*. PwC Advanced Tax Analytics & Innovation. PwC. https://www.pwc.no/no/publikasjoner/Digitalisering/artificial-intelligence-and-machine-learning-final1.pdf.

40. Mortari, Luigina. 2015. "Reflectivity in Research Practice: An Overview of Different Perspectives." *International Journal of Qualitative Methods* 14 (5): 1609406915618045. https://doi.org/10.1177/1609406915618045.

41. Nicolini, Davide, Igor Pyrko, Omid Omidvar, and Agnessa Spanellis. 2022. "Understanding Communities of Practice: Taking Stock and Moving Forward." *The Academy of Management Annals* 16 (April). https://doi.org/10.5465/annals.2020.0330.

42. OECD. 2018. *Tax Challenges Arising from Digitalisation—Interim Report 2018*. OECD/G20 Base Erosion and Profit Shifting Project. OECD.

43. OpenAI. 2023. *GPT-4 Technical Report*. OpenAI. https://arxiv.org/abs/2303.08774.

44. Pennington, Jeffrey, Richard Socher, and Christopher Manning. 2014. "Glove: Global Vectors for Word Representation." In *EMNLP*, vol. 14. https://doi.org/10.3115/v1/D14-1162.

45. Pijnenburg, Mark Gerardus Franciscus. 2020. "Data Science for Tax Administration." Ph.D. thesis, Leiden University.

46. Pijnenburg, Mark, Wojtek Kowalczyk, and van der Hel-van Dijk Lissette. 2017. "A Roadmap for Analytics in Taxpayer Supervision." *The Electronic Journal of E-Government* 15 (1): 19–32.

47. Power, D. J., C. Heavin, J. McDermott, and M. Daly. 2018. "Defining Business Analytics: An Empirical Approach." *Journal of Business Analytics* 1 (1): 40–53. https://doi.org/10.1080/2573234X.2018.1507605.

48. Pratama, Aldi, and Agung Darono. 2022. "Exploring Indonesian Customs and Tax Data Using Text Analytics Approach." Paper presented at International Conference on Custom and Tax Cooperation, Jakarta.

49. Pyrko, Igor. 2025. "Communities of Practice." In *Encyclopedia of Strategy as Practice*. Edward Elgar Publishing Ltd.

50. Pyrko, Igor, Viktor Dörfler, and Colin Eden. 2019. "Communities of Practice in Landscapes of Practice." *Management Learning* 50 (4): 482–99. https://doi.org/10.1177/1350507619860854.

51. Raffel, Colin, Noam Shazeer, Adam Roberts, et al. 2019. "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer." *CoRR* abs/1910.10683. http://arxiv.org/abs/1910.10683.

52. Ranjan, Nihar, and Rajesh Prasad. 2022. "Text Analytics: An Application of Text Mining." *Journal of Data Mining and Management* 6 (March). https://doi.org/10.46610/JoDMM.2021.v06i03.001.

53. Reimers, Nils, and Iryna Gurevych. 2019. *Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks*. https://doi.org/10.48550/arXiv.1908.10084.

54. Rennstam, Jens, and Susanne Lundholm. 2020. *Practice Theory in Business Administration*.

55. Richardson, Vernon J., Ryan Teeter, and Katie L. Terrell. 2019. *Data Analytics for Accounting*. McGraw-Hill Education.

56. Santos, Márcia R. C., Raul M. S. Laureano, and Carlos E. R. Albino. 2018. "How Tax Audit and Tax Advisory Can Benefit from Big Data Analytics Tools Data Analysis and Processing in Relational Databases Using SQL Server and Power Pivot Amp; Power View in Excel." *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–6. https://doi.org/10.23919/CISTI.2018.8399472.

57. Schrag, Anthony. 2019. *What Is Practice Research and Why Is It Important to Anyone Other than Artists?* https://www.qmu.ac.uk/life/blogs/staff-dr-anthony-schrag/what-is-practice-research-and-why-is-it-important-to-anyone-other-than-artists/.

58. Sebastiani, Fabrizio. 2001. "Machine Learning in Automated Text Categorization." *ACM Computing Surveys* 34 (April): 1–47. https://doi.org/10.1145/505282.505283.

59. Sentence-Transformers. 2020. *Paraphrase-Multilingual-MiniLM-L12-v2 [Language Model]*. Sentence-Transformers, released. https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2.

60. Sledgianowski, Deb, Mohamed Gomaa, and Christine Tan. 2017. "Toward Integration of Big Data, Technology and Information Systems Competencies into the Accounting Curriculum." *Journal of Accounting Education*, ahead of print. http://dx.doi.org/10.1016/j.jaccedu.2016.12.008.

61. Song, Qian, and Kevin Holland. 2022. "The Quality of Tax Accounting for Financial Reporting Purposes - International Evidence from the United Kingdom." *Journal of International Accounting Auditing and Taxation* 52 (November).

https://doi.org/10.1016/j.intaccaudtax.2023.100564.

62. Sun, Ting, and Miklos Vasarhelyi. 2018. "Embracing Textual Data Analytics in Auditing with Deep Learning." *The International Journal of Digital Accounting Research* 18 (January): 49–67. https://doi.org/10.4192/1577-8517-v18_3.

63. Takada, T., and Y. Suga. 2024. "Text Mining of Key Audit Matters – Analysis By Using AI Application." *International Journal of Computer Auditing* 6 (1): 53–68. https://doi.org/10.53106/256299802024120601005.

64. Vaswani, Ashish, Noam Shazeer, Niki Parmar, et al. 2017. "Attention Is All You Need." *CoRR* abs/1706.03762. http://arxiv.org/abs/1706.03762.

65. Wenger, Etienne. 2011. "Communities of Practice: A Brief Introduction." In *Communities of Practice*. National Science Foundation (U.S.). https://scholarsbank.uoregon.edu/items/6baf09ba-c319-4949-acef-cb220dc2953a.

66. Wolf, Thomas, Lysandre Debut, Victor Sanh, et al. 2020. "HuggingFace's Transformers: State-of-the-Art Natural Language Processing." https://arxiv.org/abs/1910.03771.

67. Žagar, Ajda Pretnar. 2024. *Document Embeddings vs. Bag of Words*. University of Ljubljana. https://orangedatamining.com/blog/embedding-vs-bow/.