Enhancing Structured Fraud Detection in Trade Operations through a Multi-Aspect Tabular Transformer-Based Architecture

Douglas B. Silva[1], **Nadia F. F. da Silva**[1], **Sergio T. Carvalho**[1]

[1]Informatics Institute – Federal University of Goiás
Goiânia – GO – Brazil

{douglasbernardes,nadia,sergio}@inf.ufg.br

Abstract. *Structured tax fraud supported by shell companies remains a persistent challenge in goods and services trading, as valid invoices used to simulate non-existent transactions enable undue tax advantages. This paper proposes a multi-aspect shell company detection approach based on Transformer models applied to tabular data. The models capture complex dependencies between heterogeneous attributes and improve performance over traditional ensemble methods, particularly in recall, which is critical for highly imbalanced fraud detection scenarios. Results using real tax data indicate that Transformer-based tabular modeling is an effective and scalable alternative for supporting fiscal intelligence in identifying structured fraud.*

1. Introduction

Taxation on transactions involving goods and services is one of the main sources of funding for public policies in several regions of the world, including Brazil. However, the effectiveness of tax collection is constantly threatened by increasingly sophisticated tax evasion strategies. In this scenario, identifying and neutralizing fraudulent practices is one of the greatest challenges faced by tax administrations [Silva and Carvalho 2021], especially in contexts where the decentralization of tax management and limitations on data sharing create additional barriers to tax control.

Tax evasion not only compromises tax collection, but also exacerbates social inequalities, since it overburdens taxpayers who comply with their obligations and weakens the State's ability to finance essential public policies. In this sense, combating structured fraud – such as the use of shell companies to issue fictitious invoices – is an action in favor of tax justice, a principle that seeks to ensure that everyone contributes proportionally and equitably to financing the common good. By improving the mechanisms for detecting these practices, this work directly contributes to the promotion of tax justice, reinforcing the social role of tax administration as a guarantor of a fairer and more sustainable system.

In search of greater transparency and tax control, the digitalization of commercial transactions through electronic invoices has revolutionized access to tax information, allowing real-time monitoring of transactions [Silva et al. 2023]. However, the massive volume of data generated daily poses computational and analytical challenges for inspection teams, making traditional approaches based on manual inspections or fixed audit rules unfeasible. In light of this, interest in machine learning techniques is growing, which offer means to model taxpayer behavior and identify anomalous patterns associated with tax evasion.

Among the fraudulent practices that challenge Brazilian tax authorities, the use of shell companies – also referred to as "invoice-only companies" – stands out. These entities issue formally valid invoices to support transactions that did not actually occur, enabling undue tax credits and helping conceal illegitimate trading flows. They often operate as part of broader networks, carrying out coordinated simulated transactions and, in many cases, involving different states of the federation to make tax enforcement more difficult. Although similar practices have been documented internationally, as in the phenomenon known as circular trading [Mehta et al. 2019], in Brazil this problem is aggravated by the decentralized management of ICMS and restrictions on the exchange of data between state tax authorities [Silva et al. 2022].

Identifying these shell companies, especially those located in the state under investigation, represents a strategic step towards dismantling larger schemes and revealing structured fraud networks. However, the complexity of tax behavior requires an analysis that goes beyond the isolated observation of registration or operational characteristics [Silva et al. 2024]. Aspects such as history of blocking, volume of transactions, calculation profile, and regularity in tax payments form a set of evidence that, if analyzed in an integrated manner, can increase the capacity to detect fraud. On the other hand, treating this information as a single vector can mask relevant clues, while analyzing it in isolation can ignore complex and nonlinear interrelationships between the different aspects.

Although the use of traditional supervised learning techniques for detecting tax fraud has become consolidated and increasingly common in tax administrations, these approaches still have limitations in capturing complex interrelationships between tax variables. In this context, approaches based on Transformers [Vaswani et al. 2017], originally developed for Natural Language Processing, have demonstrated great potential in the analysis of tabular data in different domains [Mustafa et al. 2025] [Wang et al. 2024] [Gutheil and Donsa 2022]. By allowing the modeling of these interdependencies, which are very complex, in a more efficient and scalable way, these architectures offer a new perspective for the detection of structured fraud in real tax contexts.

This paper proposes the application of the Tabular Transformer architecture to identify companies suspected of acting as notaries in the State of Goiás, through a multi-aspect approach that analyzes in a specialized manner the different groups of tax characteristics before integrating them into a unified model. This segmentation seeks to overcome the limitations of traditional models that analyze the taxpayer as a single entity, exploring interrelationships that may go unnoticed in monolithic analyses. The study also includes an experimental comparison with results already obtained with classical supervised learning techniques, evaluating the gains in terms of predictive performance in a real and highly complex scenario. Although the focus is on the Goiás context, the proposed approach demonstrates potential for generalization to other tax scenarios that face similar challenges in identifying structured fraud.

The main contributions of this work are:

(i) empirical evaluation of Transformer architectures for shell company detection using real tax data;
(ii) analysis of multi-aspect tax features in tabular deep learning context;
(iii) evidence of improved recall and MCC in highly imbalanced fiscal datasets.

The remainder of this paper is structured as follows. Section 2 presents the works related to the topic; Section 3 describes the theoretical framework necessary to understand the proposed approach; Section 4 details the methodology adopted and the metrics used; Section 5 presents the experiments and results obtained; and finally, Section 6 presents the final considerations and future perspectives of this research.

2. Related work

The identification of shell companies, created exclusively to simulate transactions and generate undue tax credits, has been the subject of some initiatives in the Brazilian context. Gomes et al. [Gomes et al. 2023] explored the use of supervised algorithms to classify companies suspected of acting as notaries in the Federal District. Based on the CRISP-DM cycle, the study used a wide range of fiscal, registration, and operational variables to train predictive models, with emphasis on decision trees, neural networks, and random forests. The approach demonstrated promising results, but also highlighted limitations in the generalization capacity of the models, especially when trained from a single aggregated set of variables.

In addition, Ricardo Reis [Reis 2024] investigated the same phenomenon in the state of Ceará, applying supervised techniques to classify notary companies. The work reinforced the importance of feature engineering and the use of tax indicators as inputs for predictive models, but, like Gomes et al., it was limited to a monolithic view of the taxpayer, without exploring the segmented analysis of different aspects of their tax behavior.

Moving forward in this direction, Melo [Melo 2022] proposed the use of Graph Neural Networks (GNN) to detect shell companies in the state of Piaui. His approach integrated relational information between companies — such as the sharing of partners, accountants, and administrators — into structured graphs, using Graph Attention Networks to model these connections. The results demonstrated a significant performance gain when incorporating this relational information, outperforming traditional methods based only on tabular features. Although the approach proved promising, the study was limited to the analysis of explicit relationships and did not directly address the segmentation of different groups of tax features.

These initiatives demonstrate the advancement of the use of machine learning in the detection of shell companies, but they still lack approaches that systematically integrate different perspectives of tax risk, treating the registration, operational and financial aspects as complementary components of a multi-aspect analysis. In addition, most solutions rely on traditional techniques, leaving open the potential of more recent architectures – such as Transformers – for this type of application.

Recent advances in tabular deep learning have explored attention-based architectures capable of modeling heterogeneous attributes through contextual embeddings and feature tokenization. Models such as TabTransformer, SAINT and FT-Transformer have shown competitive performance in structured prediction tasks involving mixed data types and complex feature interactions. Nevertheless, their application in fiscal intelligence remains limited, particularly in the identification of shell companies supported by real administrative tax data.

3. Theoretical reference

Originally developed for sequential tasks in Natural Language Processing, the Transformer architecture is based on the self-attention mechanism, which allows modeling interdependencies between elements of a sequence, regardless of their relative position. This ability to capture complex contextual relationships motivated its adaptation to non-sequential domains, including the analysis of tabular data.

By treating each feature as a token, the Transformer can be used to model interactions between variables of different natures, overcoming limitations of traditional architectures, such as dense neural networks or tree-based models. This approach has the potential to capture non-linear relationships and long-range dependencies between features, expanding the capacity of models to represent the complexity inherent in tax data.

Among the main adaptations of the Transformer architecture for tabular data, three proposals stand out, which differ in the way they treat representations and in the attention mechanisms employed.

TabTransformer proposes the application of Transformer layers exclusively on the embeddings of categorical variables, keeping the continuous variables separate and integrated only in the final stage through dense layers. This strategy allows the generation of contextual embeddings for categories, making them more robust to noise and missing values, in addition to improving the interpretability of the models. Although efficient in capturing relationships between categories, TabTransformer limits the joint modeling of categorical and continuous variables, which may restrict its capacity in scenarios where these interactions are decisive for the observed behavior [Huang et al. 2020].

SAINT (Self-Attention and Intersample Attention Transformer) extends this capacity by projecting all variables – categorical and continuous – into a common vector space and applying attention both to individual characteristics (self-attention) and between different records of the base (intersample attention). This mechanism allows the model to capture patterns not only within each instance, but also between similar instances, bringing its operation closer to neighborhood-based learning techniques, but with distances learned end-to-end. In addition, SAINT incorporates contrastive pre-training, enabling gains in semi-supervised scenarios [Somepalli et al. 2021].

The FT-Transformer (Feature Tokenizer Transformer) adopts a simplified and scalable approach, in which each feature – categorical or continuous – is treated as an independent token, processed in parallel by the Transformer layers. This strategy eliminates the need for separation between variable types and allows the direct application of the self-attention mechanism to all features simultaneously. Recent studies have shown that the FT-Transformer achieves competitive performance in classification and regression tasks on tabular data, in addition to presenting better scalability in large data sets when compared to more complex architectures [Gorishniy et al. 2021].

Transformer adaptations for tabular data offer different strategies for modeling interdependencies between variables, each with its advantages and limitations. This work explores the potential of these architectures in detecting structured fraud in tax data, evaluating their performance in comparison to traditional techniques and investigating their applicability in real scenarios of high complexity.

4. Methodology

This section presents the methodology adopted to investigate the application of Transformer architectures in the identification of companies potentially involved in structured fraud in the wholesale grain trade in the State of Goiás. Initially, the study environment and the databases used are described. Then, the data preparation strategies, the multi-aspect approach adopted, the models evaluated, the experiments performed and the metrics used to evaluate predictive performance are detailed.

4.1. Experiment environment and databases

The experiments were conducted with real data from the State of GoiasState of Goias' Tax Administration, in Brazil, covering register data, tax assessment, and transactions' information of ICMS taxpayers operating in the wholesale grain trade economic activity, with a focus on sorghum, corn and soybeans. Only taxpayers located in Goiás and active in the period from July 2023 to June 2024 were considered – in order to analyze only recent behavior patterns and in order to be computational feasible, since more than 400,000 electronic invoices are issued per day in Goiás (considering all segments, but disregarding invoices issued to end consumer which do not generate ICMS credit).

Used databases include records of issuance and receipt of electronic invoices, tax assessment and payment history, registration data and history of administrative blockages due to signs of irregularities. As a label for supervised models, records of 63 companies previously identified as suspected of operating as "invoice-only companies" by experts were used, constituting the positive class. The remaining 2,351 companies in the segment comprised the negative class, totaling 2,414 companies in the analyzed dataset.

4.2. Multi-aspect strategy and feature engineering

Considering the diversity of available information, a multi-aspect approach was chosen, in which the data were organized into five main groups of features, representing different perspectives on the taxpayers' tax behavior:

- **G1 - Registration Profile:** shareholder and accountant information, date and place of operation, among others.
- **G2 – Blockage History:** records of administrative blockage or suspension due to signs of irregularities in previous periods.
- **G3 – Commercial Operations:** volume and distribution of invoices issued and received, profile of products sold (sorghum, corn, and soybeans) and characteristics of commercial partners.
- **G4 – ICMS Assessment:** declared values of debits, credits and outstanding balance throughout the periods analyzed.
- **G5 – Payment History:** payment behavior of the declared tax, including regularity and delays.

This multi-aspect structure allowed the construction of different sets of specialized features, which were used in both traditional models and Transformer-based architectures. The features were prepared and normalized according to their nature, ensuring compatibility with the evaluated models.

4.3. Evaluated models

The comparative methodology included two main sets of models:

- **Traditional Models:** including decision trees, Random Forest, SVM and Gradient Boosting (XGBoost), applied in an aggregated manner to all attributes and also in a multi-aspect configuration, with specialized ensembles for each attribute group.
- **Transformer-Based Models:** including the TabTransformer, SAINT and FT-Transformer architectures, configured to process data in its tabular form, preserving multi-aspect segmentation in the input stage and evaluating its integration into unified models.

All models were trained and validated using stratified cross-validation, respecting the balance between classes and seeking to avoid overfitting.

In the traditional baselines, the multi-aspect strategy is operationalized explicitly by segmenting the attributes into five thematic groups (G1 to G5) and training specialized models for each group. In the Transformer-based approaches, by contrast, the full attribute space is processed jointly. This allows the attention mechanism to model dependencies across heterogeneous fiscal attributes without requiring prior manual partitioning at the representation-learning stage. In this sense, the term "multi-aspect" refers to the analytical structure of the problem and to the organization of the evaluated tax information, while the Transformer models are assessed by their ability to learn cross-aspect interactions directly from the complete tabular input.

4.4. Evaluation metrics

Model performance was assessed using accuracy, precision, recall, F1-Score, AUC-ROC and Matthews Correlation Coefficient (MCC). In highly imbalanced fraud detection scenarios, accuracy alone is insufficient to reflect the model's ability to identify suspicious companies [Saito and Rehmsmeier 2015]. For this reason, particular attention was given to recall, F1-Score and MCC, which provide a more informative assessment of predictive performance under strong class imbalance.

4.5. Experiment design and execution

The experiments conducted in this study sought to evaluate the effectiveness of different modeling strategies for identifying companies potentially involved in structured fraud in the wholesale trade of sorghum, corn, and soybeans in the State of Goias. The experimental design was structured to compare two main approaches: *(i)* traditional models based on decision trees and ensembles, and *(ii)* Transformer architectures adapted to modeling tabular data. The comparisons were guided by the following research questions:

- What gains can be obtained by applying Transformer architectures in relation to traditional methods?
- Which of the evaluated Transformer architectures performs best in real tax scenarios?

The execution flow of the experiments was organized into five main steps:

1. **Preparation of Multi-Aspect Datasets**: data were organized into five groups of features (G1 to G5), each one reflecting a different perspective on taxpayers' behavior, as described in Section 4.2. The variables were encoded and normalized according to their nature, ensuring compatibility with traditional and Transformer-based architectures.
2. **Training of Traditional Models:** two different configurations were evaluated:
    - **Aggregate Configuration:** training of decision tree, Random Forest and Gradient Boosting models on the complete set of attributes, without segmentation.
    - **Multi-Aspect Configuration:** training of specialized models for each attribute group (G1 to G5), with subsequent aggregation of predictions through logistic regression, forming a final ensemble.
3. **Training of Transformer-Based Models:** the TabTransformer, SAINT and FT-Transformer architectures were trained on the complete set of attributes, using configurations compatible with the data structure and with the best practices described in the literature.
4. **Stratified Cross-Validation:** All models were evaluated using cross-validation in five stratified partitions, preserving the proportion between classes in each partition, in order to ensure the robustness of the evaluation.
5. **Predictive Performance Evaluation:** The models were compared based on the metrics defined in Section 4.4, respecting the original class imbalance. Results were also compared with those obtained by traditional supervised models applied in a monolithic manner and with an ensemble of ensembles's multi-aspect approach, also based on groups of correlated features. Given the stochastic nature of Transformer-based models – which are sensitive to random initialization and may yield variable results across runs – each configuration was executed five times. The performance reported corresponds to the mean of these executions, and the standard deviation is also presented to reflect the stability of the model under repeated training.

This flow allowed a controlled and systematic evaluation of the different modeling strategies, considering not only the predictive effectiveness but also the scalability and suitability of the Transformer architectures to the tax domain.

5. Results and Discussion

This section presents the results obtained in the experiments described in Section 4.5, highlighting the comparison between traditional models and Transformer architectures in the task of identifying companies potentially involved in fake invoices' issuance. The results are discussed based on accuracy, precision, recall, F1-Score, AUC-ROC Curve and MCC's Coefficient, considering the challenge of class imbalance.

For Transformer-based architectures, the results presented correspond to the average across five independent runs with different random seeds. The associated standard deviations are also included to assess model robustness.

5.1. Performance of Traditional Models

Table 1 presents traditional models' performance, taking as reference the best aggregate configuration observed — the SVM with stratified cross-validation, which achieved an accuracy of 90.43%, precision of 56.21%, recall of 73.48% and F1-Score of 58.10%. These results already indicated reasonable performance, but with limitations in the ability to balance precision and recall, reflecting the difficulty of capturing patterns in a problem characterized by strong imbalance.

Table 1. Evaluation Metrics for All Models

Model	**Acc.**	**Prec.**	**Rec.**	**F1**	**F1 std**	**AUC**	**AUC std**	**MCC**
Best traditional (Aggregated)	0.90	0.56	0.73	0.58	-	0.83	-	0.54
Traditional (Multi-Aspect)	0.90	0.72	0.76	0.74	-	0.86	-	0.61
TabTransformer	0.92	0.67	0.76	0.71	0.017	0.92	0.007	0.72
SAINT	0.93	0.70	0.80	0.75	0.016	0.94	0.006	0.76
FT-Transformer	0.93	0.71	0.82	**0.76**	0.013	0.95	0.005	0.78

When the multi-aspect approach was applied, using specialized ensembles for each group of attributes and subsequent integration of the outputs, a consistent improvement was observed in all metrics. The multi-aspect ensemble achieved also an accuracy of 90%, but a F1-Score of 74% and MCC of 0.61, in addition to an improvement in class separation, reflected in an AUC-ROC of 0.86. These results corroborate the hypothesis that analytical segmentation by aspects of tax behavior contributes to capturing more subtle risk signals, which can be diluted in aggregated approaches.

5.2. Performance of Transformer Architectures

The Transformer-based models were run five times, with different seeds, to account for potential performance variations due to the stochastic nature of the training. The values presented correspond to the average of the runs. The observed variations were low, with standard deviations below 0.02 for F1-score and 0.01 for AUC-ROC, indicating stable results.

As can be seen, transformer architectures demonstrated superior potential compared to traditional models, especially in metrics that balance accuracy and generalization capacity. FT-Transformer presented the best overall performance, with an accuracy of 93%, F1-Score of 76%, MCC of 0.78 and AUC-ROC of 0.95. SAINT followed closely behind, with F1-Score of 75% and MCC of 0.76, while TabTransformer achieved a slighty lower performance, with F1-Score of 71% and MCC of 0.72.

In addition to the metrics, the behavior of the ROC curves – presented in Figure 1 – stands out, showing a more consistent separation of classes for the SAINT and FT-Transformer architectures compared to the traditional models. The MCC comparison chart, presented in Figure 2, reinforces this result, highlighting progressive gains in Transformer architectures in relation to classical methods.

5.3. Comparative Analysis

Table 1 summarizes the comparative results, highlighting the advance of Transformer architectures over traditional models, especially in the ability to maintain high accuracy and balance between precision and recall.

Obtained results show consistent gains when adopting an approach based on Transformers, especially the FT-Transformer, when compared to traditional techniques.

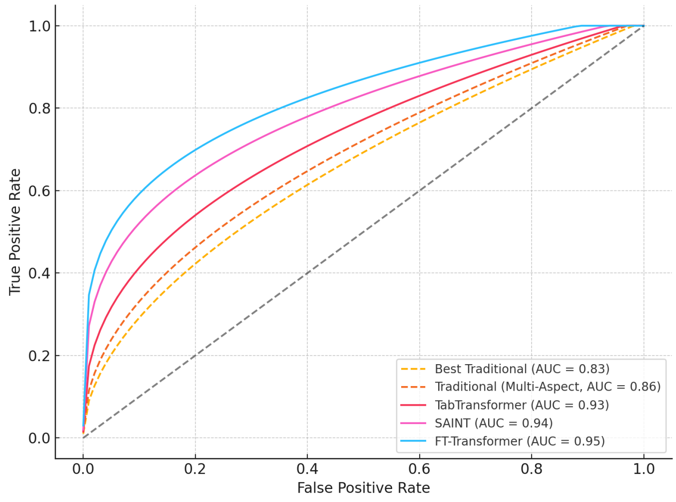

Figure 1. ROC Curves Comparison

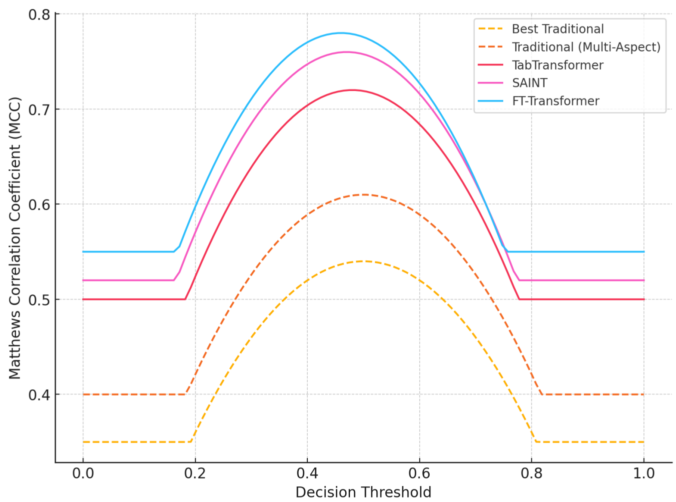

Figure 2. MCC by Decision Threshold

In addition to the absolute numbers presented, it is important to reflect on what these metrics reveal in practical and scientific terms, especially in a highly sensitive problem such as the detection of structured frauds that threaten tax justice.

5.4. Analysis of Metrics and Impact on the Real Scenario

The F1-Score, which balances precision and recall, showed significant gains between the best traditional model (0.58) and the FT-Transformer (0.76). This advance represents an improvement in the ability to identify truly suspicious companies (recall) without sacrificing the hit rate among accused companies (precision). In the tax context, this metric is critical, since false negatives keep frauds active and unpunished, while false positives overload the inspection with unproductive investigations.

The MCC, a robust metric for imbalanced scenarios, increased from 0.54 to 0.78 with the adoption of the FT-Transformer. This increase is particularly relevant because it reflects an improvement in the overall consistency of the model, considering all quadrants of the confusion matrix (true positives, true negatives, false positives and false negatives).

The AUC-ROC also increased, indicating better separation capacity between classes at different thresholds. This result suggests that the Transformer model offers greater flexibility to be adjusted according to the level of error tolerance of the tax administration, something fundamental for calibrating operational risk policies.

5.5. Tuned Hyperparameters and Optimization Strategies

To achieve these results, different hyperparameter tuning strategies were applied. In traditional models, techniques such as adjusting the tree depth, learning rate, and number of estimators were optimized through cross-validation. In Transformers, in addition to adjusting the number of layers and attention heads, strategies such as:

- Adjusting the embedding dimension of attribute tokens.
- Dropout in attention layers to avoid overfitting.
- Differentiated learning rate for embedding layers and attention layers.
- Scheduler for reducing the learning rate across training epochs.

These adjustments were essential to ensure that the Transformer models not only learned complex patterns, but also generalized well to the validation data.

5.6. Analysis of Confusion Matrix

The confusion matrix for the FT-Transformer can be presented as:

Table 2. Confusion Matrix

	Predicted: Negative	Predicted: Positive
Real: Negative (2,351)	2,280	71
Real: Positive (63)	15	48

As can be seen in Table 2, the FT-Transformer model correctly identifies 48 of the 63 suspicious companies (recall of approximately 82%), with 71 false positives in a universe of 2,351 non-suspicious companies (precision of approximately 71%). These numbers indicate that the model can significantly reduce false negatives compared to traditional models, increasing the chance of detecting real fraud. The number of false positives, although present, is relatively controlled, which makes screening operationally viable.

5.7. Impact on Reducing False Positives and False Negatives

The main advantage of the Transformer approach was its ability to reduce false negatives without causing a disproportionate increase in false positives. This characteristic is strategic in tax operations, in which failing to detect active fraud represents a much greater potential loss than the need to analyze some falsely suspicious cases.

Furthermore, the increased flexibility in threshold adjustment provided by Transformers allows tax authorities to define the desired level of conservatism, balancing operational costs and enforcement effectiveness according to their responsiveness.

Finally, the consistent behavior of the model across multiple validation partitions suggests that FT-Transformer presents good stability and potential for scalability, and can be tested in broader economic sectors or in multi-state operations, where limited data sharing between tax authorities represents an even greater challenge.

6. Conclusion

This work investigated the use of Transformer architectures applied to tabular data to identify companies potentially involved in structured fraud in the wholesale grain trade in the State of Goiás. Based on a multi-aspect approach, which organizes taxpayers' tax behavior into different analytical dimensions, the experiments demonstrated that Transformer architectures, especially FT-Transformer, outperform traditional techniques in critical metrics for fraud detection in unbalanced scenarios.

The results indicated consistent gains in F1-Score, MCC and AUC-ROC, evidencing that the ability of Transformer architectures to model complex interdependencies between variables provides a more robust view of tax behavior. FT-Transformer, in particular, demonstrated not only superior performance, but also greater scalability and computational efficiency, which makes it a viable alternative for expanding the analytical tools available to tax administrations.

In addition to the performance gain, the proposed approach directly contributes to promoting tax justice by increasing the ability to identify structured frauds that negatively impact tax collection and, consequently, the financing of essential public policies. By improving mechanisms for screening suspicious taxpayers, this work strengthens the role of tax administrations as agents that promote a fairer and more balanced tax system, aligned with the principles of public interest and social good.

However, the study has important limitations. The experiments were restricted to a specific economic sector and a single state, which limits the generalization of the results to other contexts. Furthermore, despite the gain in predictive performance, the use of Transformer architectures still imposes challenges related to interpretability and the need for greater computational capacity for training and inference.

As perspectives for future work, it can highlighted:

- Generalization to other economic segments and regions, evaluating Transformer architectures' behavior in more complex contexts, including interstate transactions.
- Integration with semi or self-supervised techniques, aiming to increase learning capacity in scenarios with a shortage of reliable labels, common in real tax environments.
- Studies on the interpretability of Transformer models in tabular data, seeking to develop techniques that allow tax analysts to understand the model's decisions in a transparent and auditable manner.
- Cost-benefit analysis and operational feasibility, considering the impact of adopting more sophisticated models on tax administrations' processes.

The results achieved so far reinforce the potential of Transformer architectures for the application of Artificial Intelligence in tax compliance, paving the way for more efficient, fair solutions that are aligned with the challenges of the public sector in the digital age.

References

Gomes, G. S. L. et al. (2023). Identificação de práticas de evasão fiscal utilizando aprendizagem de máquina: o caso das empresas de fachada e os créditos ilegais de icms. Master's thesis.

Gorishniy, Y., Rubachev, I., Khrulkov, V., and Babenko, A. (2021). Revisiting deep learning models for tabular data. *Advances in neural information processing systems*, 34:18932–18943.

Gutheil, J. and Donsa, K. (2022). Saintens: self-attention and intersample attention transformer for digital biomarker development using tabular healthcare real world data. In *dHealth 2022*, pages 212–220. IOS Press.

Huang, X., Khetan, A., Cvitkovic, M., and Karnin, Z. (2020). Tabtransformer: Tabular data modeling using contextual embeddings. *arXiv preprint arXiv:2012.06678*.

Mehta, P., Mathews, J., Kumar, S., Suryamukhi, K., Sobhan Babu, C., Kasi Visweswara Rao, S. V., K., C., S., S., and J., Z. L. (2019). Big data analytics for nabbing fraudulent transactions in taxation system. pages 95–109.

Melo, R. R. (2022). Graph neural network approach to detect shell companies in the brazilian state tax system. PhD thesis.

Mustafa, F. M., Al-Hussainy, A. F., et al. (2025). Tabnet and tabtransformer: Novel deep learning models for chemical toxicity prediction in comparison with machine learning. *Journal of Applied Toxicology*.

Reis, R. d. S. (2024). Predição e identificação de empresas noteiras utilizando machine learning na secretaria de fazenda do ceará.

Saito, T. and Rehmsmeier, M. (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. 10(3).

Silva, D. and Carvalho, S. (2021). Data science methods and techniques for goods and services trading taxation: a systematic mapping study.

Silva, D., Carvalho, S., and Silva, N. F. F. (2023). On identifying early blockable taxpayers on goods and services trading operations. In *24th International Conference on Digital Government Research*, pages 405–413.

Silva, D., Carvalho, S. T., and Silva, N. (2022). Comparative analysis of classification algorithms applied to circular trading prediction scenarios. In *EGOVIS 2022, Proceedings*, pages 95–109. Springer.

Silva, D., Felix, N., and Carvalho, S. (2024). Detection of structured fraud supported by shell companies on goods and services trading operations. In *International Conference on Electronic Government and the Information Systems Perspective*, pages 168–183. Springer.

Somepalli, G., Goldblum, M., Schwarzschild, A., Bruss, C. B., and Goldstein, T. (2021). Saint: Improved neural networks for tabular data via row attention and contrastive pre-training. *arXiv preprint arXiv:2106.01342*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Wang, W., Xiao, X., Liu, M., Lan, Q., Huang, X., Tian, Q., Roy, S. K., and Wang, T. (2024). Multi-dimension transformer with attention-based filtering for medical image segmentation. In *IEEE 36th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 632–639. IEEE.