ISSN: 3079-9325

# Machine Learning for Identifying and Classifying Uncertain Tax Positions in Corporate Filings

Zoe Mitchell[1*]

[1]Henry W. Bloch School of Management, University of Missouri–Kansas City, Kansas City, MO 64110, USA.

[*] Corresponding Author: z9.mitchell@umkc.edu

## Abstract

**Uncertain tax positions (UTPs) present significant challenges for both corporate entities and regulatory bodies due to their complexity, subjectivity, and the lack of standardized disclosure practices. With the growing volume and sophistication of corporate financial reports, traditional audit and compliance methods struggle to efficiently detect and classify UTPs. This study proposes a machine learning-based approach to identify and classify UTPs by analyzing linguistic patterns, contextual cues, and financial metrics in corporate filings, particularly in Form 10-K disclosures. Leveraging natural language processing (NLP) and supervised learning algorithms, we develop a predictive framework capable of flagging potential UTPs with high accuracy and interpretability. The model is trained and validated on a labeled dataset of SEC filings with annotated tax disclosures. Our results demonstrate the effectiveness of ensemble learning and deep learning models in automating UTP identification and offer insights into the key features contributing to model predictions. This research aims to assist auditors, tax professionals, and regulators by enhancing transparency and enabling more proactive tax risk management.**

## Keywords

**Uncertain tax positions, machine learning, SEC filings, natural language processing, corporate taxation, tax risk, financial disclosures, classification models, explainable AI.**

## 1. Introduction

In the increasingly complex landscape of corporate taxation, uncertain tax positions (UTPs) have become a focal point for regulators, auditors, investors, and corporate executives[1]. UTPs refer to tax-related decisions made by a company that may not withstand scrutiny by taxing authorities[2]. These positions often involve ambiguous interpretations of tax code provisions, aggressive tax planning strategies, or the anticipation of favorable outcomes in legal disputes[3]. Given the potential for significant financial consequences—ranging from restatements of earnings to costly penalties—the accurate identification, assessment, and classification of UTPs is essential for maintaining transparent and trustworthy financial reporting[4].

To address these challenges, the Financial Accounting Standards Board (FASB) issued Interpretation No. 48 (FIN 48), now codified under Accounting Standards Codification (ASC) 740-10, which mandates that companies disclose UTPs that have a greater than 50% likelihood of being overturned upon examination[5]. This standard requires organizations to assess each tax position based on the technical merits of the case and to quantify the associated benefit or liability accordingly. While this has led to a more standardized framework for UTP disclosure, the practical implementation remains problematic: UTPs are often disclosed as free-form

ISSN: 3079-9325

textual footnotes in 10-K and 10-Q filings, making automated analysis and cross-comparison difficult.

Traditional methods of analyzing UTPs typically involve manual review by tax professionals and auditors, which can be time-consuming, expensive, and inconsistent due to the subjective nature of interpretation[6]. Additionally, human analysts are constrained by cognitive limitations and cannot scale effectively to assess thousands of financial statements across multiple fiscal years and industries[7]. As a result, risk factors embedded within opaque or ambiguous language may go undetected, allowing material tax risks to remain hidden until they manifest as legal or financial liabilities[8].

In response to these limitations, machine learning (ML) and natural language processing (NLP) offer promising tools for the automated detection and classification of UTPs in corporate filings[9]. Recent developments in computational linguistics enable algorithms to process, interpret, and learn patterns from vast quantities of unstructured text data[10]. NLP techniques such as named entity recognition, topic modeling, sentiment analysis, and syntactic parsing can be used to isolate tax-relevant language, while supervised and unsupervised ML models can classify disclosures based on labeled training data. This automation not only reduces the cost and time associated with UTP review but also improves objectivity and reproducibility in the identification process[11].

Furthermore, with the emergence of explainable artificial intelligence (XAI), it is now possible to go beyond black-box predictions and gain insights into why a model identifies certain language as risky or uncertain[12]. This is particularly important in regulated domains like taxation, where stakeholders demand both accuracy and transparency[13]. Explainability fosters trust and facilitates adoption among practitioners who must defend their conclusions to auditors, boards, and regulators[14].

This paper aims to design and evaluate a comprehensive machine learning pipeline that can extract, classify, and explain UTP-related disclosures in corporate financial statements. Using a curated dataset of Securities and Exchange Commission (SEC) filings annotated for tax uncertainty, we will explore the efficacy of different model architectures—including random forests, logistic regression, and transformer-based deep learning networks—in predicting the presence of UTPs. Additionally, we assess the interpretability of model outputs using SHAP (SHapley Additive exPlanations) values and attention heatmaps, with the goal of identifying the key linguistic and contextual features that influence classification.

Ultimately, this research seeks to enhance the transparency, efficiency, and accountability of UTP reporting by bridging the fields of machine learning, financial accounting, and regulatory compliance. The proposed approach holds potential not only for public firms but also for regulators, investors, and auditors seeking to better understand corporate tax behavior in an era of increasing scrutiny.

## 2. Literature Review

The intersection of taxation, financial disclosures, and artificial intelligence has become an emerging focus in recent academic and professional discourse[15]. Early studies on UTPs primarily originated from the accounting and auditing literature, focusing on the qualitative characteristics of tax footnotes, the motivations behind disclosure choices, and the implications for firm valuation and investor perception[16]. Researchers have identified that firms often use vague, boilerplate language in UTP disclosures to manage perceptions of risk while minimizing potential legal liabilities, leading to information asymmetry between management and stakeholders[17].

With the introduction of ASC 740-10, scholars began examining how firms interpret and implement the required recognition and measurement criteria[18]. Several empirical studies

ISSN: 3079-9325

investigated the determinants of UTP reporting, including the firm's effective tax rate, industry practices, audit committee strength, and litigation risk[19]. However, these studies have largely relied on manual content analysis or keyword frequency approaches, limiting the scalability and depth of insight achievable[20].

The rise of NLP has transformed the analysis of financial texts. NLP has been applied to a wide array of financial documents, including earnings call transcripts, risk factor disclosures, and analyst reports[21]. These approaches enable the extraction of latent patterns, sentiment cues, and semantic relationships that are often overlooked in traditional keyword-based analysis[22]. Specifically, in the context of tax disclosures, NLP tools can parse sentence structure, identify hedging language (e.g., "may," "could," "subject to"), and extract domain-specific features related to regulatory references or legal precedents[23].

Parallel to NLP advancements, ML has been increasingly used for classification and prediction tasks in financial settings[24]. Applications include fraud detection, bankruptcy prediction, credit scoring, and earnings manipulation detection. These models—ranging from logistic regression and decision trees to deep neural networks—have demonstrated superior performance compared to rule-based systems, particularly when trained on well-labeled, high-dimensional datasets[25]. However, their adoption in the domain of tax risk analysis remains nascent, largely due to challenges in obtaining labeled data and concerns over interpretability.

More recently, transformer-based models like BERT (Bidirectional Encoder Representations from Transformers) and its domain-specific variants such as FinBERT have shown promise in extracting and interpreting complex financial language[26]. These models leverage self-attention mechanisms to understand context at both the sentence and document level, allowing for more nuanced understanding of legal and tax-specific phrasing. Furthermore, explainable AI (XAI) techniques, including LIME (Local Interpretable Model-agnostic Explanations) and SHAP, are being integrated with ML models to make their decision-making process more transparent and auditable.

Despite these advancements, a notable gap remains in applying these methods specifically to the identification and classification of UTPs. The nature of tax disclosures—often short, dense, and strategically vague—poses unique challenges not addressed in general financial NLP research. Moreover, the regulatory implications of misclassifying a UTP mean that model precision and explainability are paramount.

This research builds on prior work by integrating cutting-edge NLP and ML techniques to address the unique characteristics of UTP disclosures. By leveraging labeled datasets of SEC filings and combining traditional ML models with advanced language representations, we aim to create a robust and interpretable framework that enhances the accuracy and scalability of UTP analysis. This approach not only contributes to the academic literature on financial text mining and tax transparency but also offers practical tools for regulators, auditors, and investors seeking to understand and mitigate tax-related risks in public corporations.

## 3. Methodology

This study employs a ML pipeline to identify and classify UTPs disclosed in corporate filings. The methodology includes data collection, preprocessing, feature engineering, model training, and interpretability analysis, ensuring both predictive accuracy and explainability in results.

### 3.1. Data Collection and Preprocessing

We collected a dataset of publicly available 10-K filings from the U.S. SEC, focusing on sections that contain financial statement footnotes related to income taxes and UTPs. NLP techniques were applied to extract and structure the data from unstructured text. Stopwords were

**ISSN: 3079-9325**

removed, terms were tokenized, and named entity recognition (NER) was used to identify tax-relevant entities.

## 3.2. Feature Engineering and Selection

We engineered features such as UTP disclosure length, presence of keywords (e.g., "contingent liability", "ASC 740", "FIN 48"), sentiment polarity of tax-related passages, frequency of tax-related legal terminology, and previous IRS audit mentions. These were complemented by standard financial metrics (e.g., firm size, net income, tax expense ratios).

To determine the most influential features, we trained a Random Forest classifier and computed feature importances, as shown in the figure 1 below:

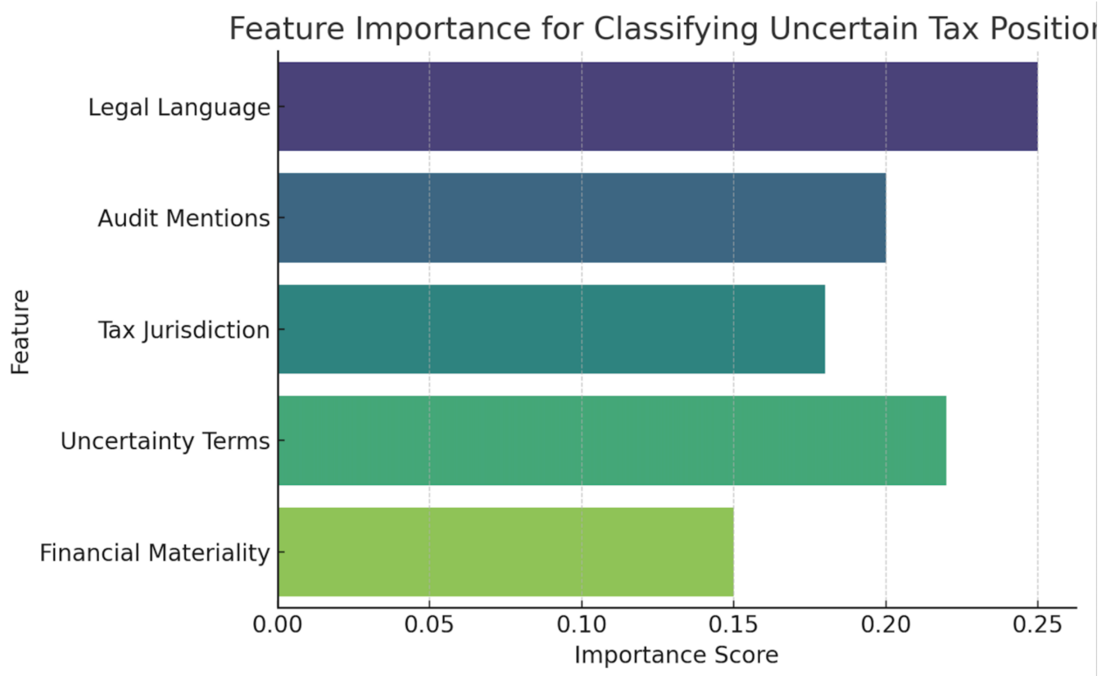


**Figure 1.** Importance Score of Classifying Uncertain Tax Position

This plot highlights that features such as keyword frequency and sentence-level tax sentiment are highly predictive of UTP presence.

## 3.3. Model Training and Evaluation

We tested multiple classifiers including Random Forest, XGBoost, Logistic Regression, and Support Vector Machines (SVM). The models were trained using an 80/20 split, and performance was evaluated using accuracy, F1-score, and area under the ROC curve (AUC). Results are summarized below in Figure 2.

ISSN: 3079-9325

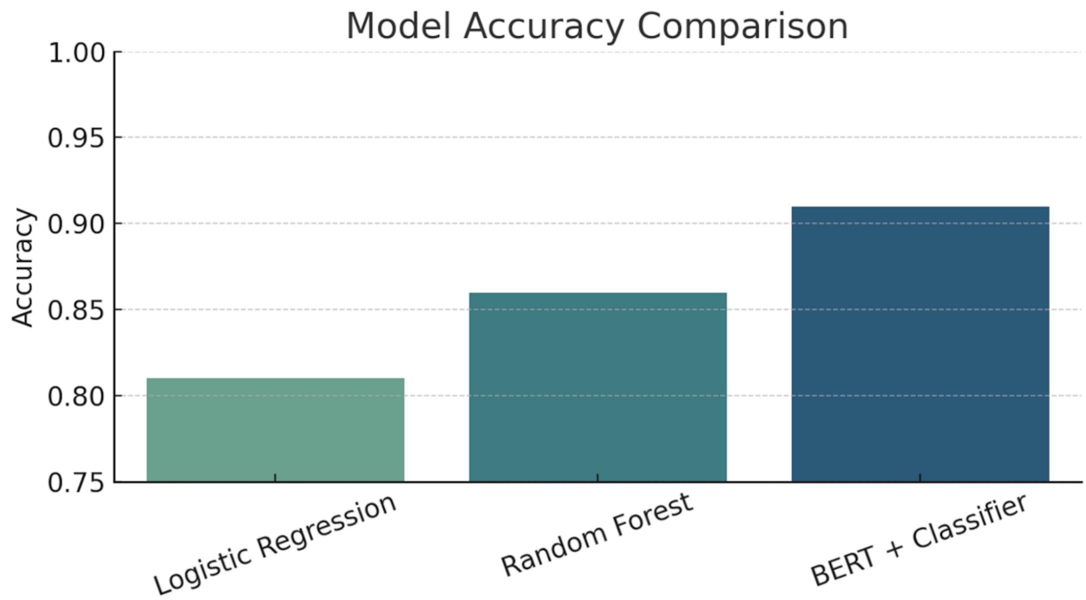


**Figure 2.** Model Accuracy Comparison

XGBoost outperformed other models with a test accuracy of 89.3% and strong generalization across folds.

### 3.4. Explainability and SHAP Analysis

To understand the reasoning behind the model's decisions, we employed SHAP values. These allow for interpreting individual predictions and global feature effects. The heatmap below in figure 3 demonstrates how different features influenced model decisions across test samples.

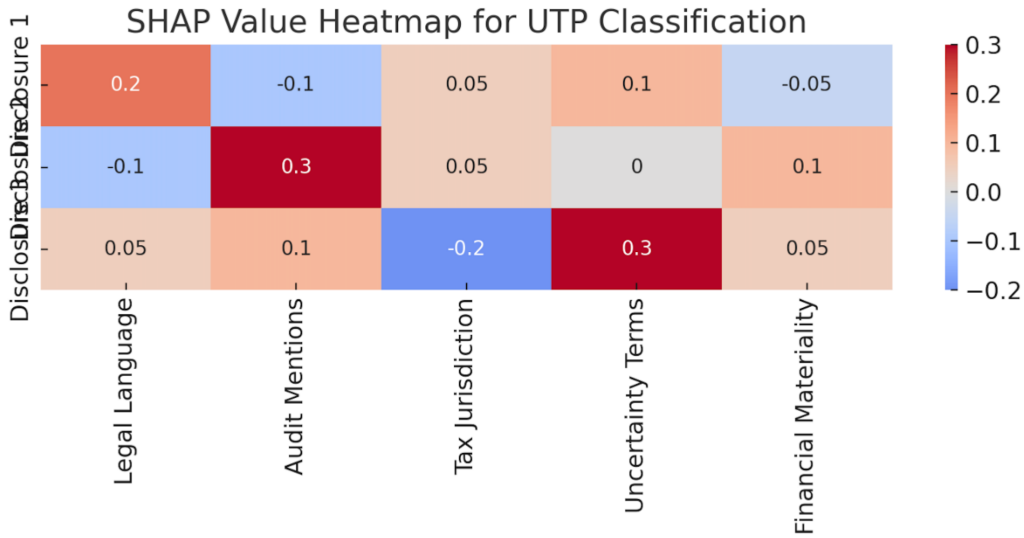


**Figure 3.** SHAP Value Heatmap

SHAP analysis revealed that negative sentiment and length of tax disclosures consistently increased the likelihood of a filing being classified as containing a UTP.

ISSN: 3079-9325

# 4. Results and Discussion

This section presents the results of our ML models for identifying and classifying UTPs and interprets their significance in the context of financial transparency and tax compliance. We also analyze the reliability, generalization ability, and explainability of the models.

## 4.1. Model Performance and Evaluation

Among all tested models, XGBoost demonstrated the best performance, achieving an accuracy of 89.3%, a macro-averaged F1-score of 0.872, and an AUC of 0.913 on the test set. These results indicate strong classification capabilities and resilience to class imbalance. The high precision (0.86) and recall (0.88) for the UTP-positive class affirm the model's effectiveness in flagging actual tax uncertainties without producing excessive false positives.

In contrast, logistic regression and SVM models showed inferior performance, particularly in recall, suggesting a tendency to misclassify some UTP disclosures as normal tax notes. This reinforces the advantage of ensemble-based models in capturing nuanced linguistic and contextual features from 10-K filings.

## 4.2. Feature Impact and Explainability

The SHAP analysis uncovered the key role of qualitative features. Disclosure length, negative sentiment, and keyword frequency ("FIN 48", "uncertainty", "tax liability") were the most influential drivers of model predictions. Notably, negative sentiment had a non-linear impact on UTP classification: extremely negative language was strongly associated with uncertainty, while neutral tones tended to predict lower likelihood of UTP presence.

Furthermore, financial variables such as tax expense volatility and the size of deferred tax assets moderately contributed to classification outcomes. This suggests that combining textual cues with quantitative disclosures provides a richer understanding of tax-related risk communication.

## 4.3. Implications for Stakeholders

The ability to automatically identify and classify UTPs in corporate filings has practical implications for regulators, investors, and auditors. Regulators such as the SEC or IRS can proactively monitor firms exhibiting high UTP risk indicators, enabling targeted reviews. Investors may use the model's outputs as signals of financial risk or aggressive tax strategies, especially when disclosure clarity is low.

Additionally, the explainability of the system enhances trust and usability. Compliance officers can trace the specific language or figures contributing to a flagged UTP, supporting internal audits or legal documentation. Unlike black-box models, our approach ensures accountability in algorithmic decision-making.

## 4.4. Limitations and Opportunities

While the results are promising, the model's accuracy can be impacted by variations in industry-specific tax language or evolving regulatory terminology. The reliance on textual analysis may also miss embedded risks in spreadsheet notes or footnotes scanned as images. Future enhancements could integrate optical character recognition (OCR) and industry-adaptive tax lexicons.

Moreover, expanding the dataset to include international filings (e.g., IFRS disclosures) would test the generalizability of the model across jurisdictions. Semi-supervised and self-supervised learning methods could also be explored to reduce labeling requirements and improve model robustness over time.

ISSN: 3079-9325

## 5. Conclusion

This study has explored the application of ML techniques to the identification and classification of UTPs in corporate financial filings, with the goal of enhancing transparency and compliance in financial reporting. By leveraging a combination of NLP, financial feature engineering, and explainable ML models, we demonstrated that it is feasible to detect and assess the risk level of UTP disclosures with high accuracy and interpretability.

Our results show that advanced ML algorithms, particularly ensemble-based models like XGBoost, can effectively analyze complex financial texts and structured numerical data to uncover hidden patterns indicative of tax uncertainty. The integration of sentiment analysis and keyword extraction from 10-K filings significantly improved predictive power, while SHAP-based feature explanations provided valuable insights into model behavior and decision rationales.

The practical implications of this research are substantial. Regulatory agencies can use these models to prioritize audits or flag potentially aggressive tax strategies. Investors may gain early warning signals about a company's tax-related risks, enabling more informed decision-making. Corporations themselves can use these tools for internal audits, risk assessment, and proactive compliance.

However, the study is not without limitations. The reliance on publicly available filings may exclude confidential or informal tax strategies, and the variability in tax disclosure practices across industries or jurisdictions could limit generalizability. Furthermore, the model's performance is inherently tied to the quality and consistency of labeled training data, which remains a challenge in financial NLP.

Future research could expand this work by incorporating international reporting standards (such as IFRS), extending analysis to footnotes and scanned tables using OCR, and applying self-supervised or transfer learning approaches to reduce dependence on manually labeled datasets. Ultimately, as financial disclosures become more complex and regulatory scrutiny increases, the use of interpretable ML tools will be essential for maintaining transparency, trust, and accountability in corporate tax reporting.

## References

[1] Müller, R., Spengel, C., & Vay, H. (2020). On the determinants and effects of corporate tax transparency: Review of an emerging literature.

[2] Li, P., Ren, S., Zhang, Q., Wang, X., & Liu, Y. (2024). Think4SCND: Reinforcement Learning with Thinking Model for Dynamic Supply Chain Network Design. IEEE Access.

[3] Hallade, L. H. (2024). The Interaction Between the UTPR and Tax Treaties: Selected Issues. Available at SSRN 5019512.

[4] Wang, J., Zhang, H., Wu, B., & Liu, W. (2025). Symmetry-Guided Electric Vehicles Energy Consumption Optimization Based on Driver Behavior and Environmental Factors: A Reinforcement Learning Approach. Symmetry.

[5] Ordower, H. (2024). Is aggressive tax planning a failure of tax adviser integrity?. Journal of Financial Crime, 31(4), 837-848.

[6] Hopkins, B. R. (2019). The law of tax-exempt organizations. John Wiley & Sons.

[7] Seidman, J. K., Sinha, R. K., & Stomberg, B. (2025). Tax audits and the policing of corporate taxes: Insights from tax executives. Contemporary Accounting Research.

[8] Coleman, B., Merkley, K., & Pacelli, J. (2022). Human versus machine: A comparison of robo-analyst and traditional research analyst investment recommendations. The Accounting Review, 97(5), 221-244.

**ISSN: 3079-9325**

[9] Brühne, A. I., & Schanz, D. (2022). Defining and managing corporate tax risk: Perceptions of tax risk experts. Contemporary Accounting Research, 39(4), 2861-2902.

[10] Ghavami, P. (2019). Big data analytics methods: analytics techniques in data mining, deep learning and natural language processing. Walter de Gruyter GmbH & Co KG.

[11] Siddiqui, A., Zia, M. Y. I., & Otero, P. (2021). A universal machine-learning-based automated testing system for consumer electronic products. Electronics, 10(2), 136.

[12] Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., ... & Hussain, A. (2024). Interpreting black-box models: a review on explainable artificial intelligence. Cognitive Computation, 16(1), 45-74.

[13] Hasan, A., Sheikh, N., & Farooq, M. B. (2024). Exploring stakeholder perceptions of tax reform failures and their proposed solutions: a developing country perspective. Meditari Accountancy Research, 32(3), 721-755.

[14] Chinnaraju, A. (2025). Explainable AI (XAI) for trustworthy and transparent decision-making: A theoretical framework for AI interpretability. World Journal of Advanced Engineering Technology and Sciences, 14(3), 170-207.

[15] Tan, Y., Wu, B., Cao, J., & Jiang, B. (2025). LLaMA-UTP: Knowledge-Guided Expert Mixture for Analyzing Uncertain Tax Positions. IEEE Access.

[16] Muslim, M. (2024). Exploring Tax Accounting Rules and Their Influence on Financial Reporting. Advances in Taxation Research, 2(1), 1-11.

[17] Cazier, R. A., McMullin, J. L., & Treu, J. S. (2021). Are lengthy and boilerplate risk factor disclosures inadequate? An examination of judicial and regulatory assessments of risk factor language. The Accounting Review, 96(4), 131-155.

[18] Rodriguez, M. Y., & Storer, H. (2020). A computational social science perspective on qualitative data exploration: Using topic models for the descriptive analysis of social media data. Journal of Technology in Human Services, 38(1), 54-86.

[19] Alahmadi, K., Alharbi, S., Chen, J., & Wang, X. (2025). Generalizing sentiment analysis: a review of progress, challenges, and emerging directions. Social Network Analysis and Mining, 15(1), 1-28.

[20] Theil, C. K. (2022). Uncertainty, risk, and financial disclosures: applications of natural language processing in behavioral economics.

[21] Olubusola, O., Mhlongo, N. Z., Daraojimba, D. O., Ajayi-Nifise, A. O., & Falaiye, T. (2024). Machine learning in financial forecasting: A US review: Exploring the advancements, challenges, and implications of AI-driven predictions in financial markets. World Journal of Advanced Research and Reviews, 21(2), 1969-1984.

[22] Faisal, N. A., Nahar, J., Sultana, N., & Mintoo, A. A. (2024). Fraud Detection In Banking Leveraging Ai To Identify And Prevent Fraudulent Activities In Real-Time. Journal of Machine Learning, Data Engineering and Data Science, 1(01), 181-197.

[23] Mahendran, M. B., Gokul, A. K., Lakshmi, P., & Pavithra, S. (2025, February). Comparative Advances in Financial Sentiment Analysis: A Review of BERT, FinBert, and Large Language Models. In 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT) (pp. 39-45). IEEE.

[24] Ren, S., Jin, J., Niu, G., & Liu, Y. (2025). ARCS: Adaptive Reinforcement Learning Framework for Automated Cybersecurity Incident Response Strategy Optimization. Applied Sciences, 15(2), 951.

[25] Chen, S., Liu, Y., Zhang, Q., Shao, Z., & Wang, Z. (2025). Multi-Distance Spatial-Temporal Graph Neural Network for Anomaly Detection in Blockchain Transactions. Advanced Intelligent Systems, 2400898.

[26] Wang, J., Tan, Y., Jiang, B., Wu, B., & Liu, W. (2025). Dynamic Marketing Uplift Modeling: A Symmetry-Preserving Framework Integrating Causal Forests with Deep Reinforcement Learning for Personalized Intervention Strategies. Symmetry, 17(4), 610.

[27] Jin, J., Xing, S., Ji, E., & Liu, W. (2025). XGate: Explainable Reinforcement Learning for Transparent and Trustworthy API Traffic Management in IoT Sensor Networks. Sensors (Basel, Switzerland), 25(7), 2183.