Vol.1, No.4, Winter 2023

Determination of Audit Fees Using Support Vector Machine: Evidence from the Tehran Stock Exchange

Arezoo Memarimoghadam
Department of Accounting, Islamic Azad University, North Tehran Branch, Tehran, Iran.
a.memarimoghadam@gmail.com

Mohammadhamed Khanmohammadi
Associate Professor, Department of Accounting, Islamic Azad University, Damavand Branch, Damavand, Iran.
(Corresponding Author)
Dr.khanmohammadi@damavand.iau.ac.ir

Mohammad Hassani
Assistant Professor, Department of Accounting and Auditing, Faculty of Management, North Tehran Branch, Islamic Azad University, Tehran, Iran.
m_hassani@iau-tnb.ac.ir

Submit: 2024/02/02 Accept: 2024/04/20

Abstract

Objective: This study explores the determination of audit fees (AF) using Support Vector Regression (SVR) among companies listed on the Iranian stock market from 2017 to 2021. It investigates the relationship between financial variables like financial leverage (DA), current asset ratio (CA), quick ratio (QUICK), ASSETS, current ratio to current liabilities (CR), and long-term debt (DE), with AF as the target.

Design/methodology/approach: Data from 60 listed companies during this period, totaling 279 year-observations, are employed. SVR models are trained on this dataset using Google Colab.

Results: The SVR model achieves a 90.5% R^2 value and a 3.7 Mean Squared Error (MSE) on training data, indicating high explained variance and reasonable error levels. However, on new data, the model's performance diminishes, with an R^2 of 67% and an MSE of 8.1, implying reduced accuracy and intermediate predictive accuracy.

Innovation: This study advances the understanding of AF determination using SVR, highlighting the importance of considering various financial variables.

Keywords: Audit fee, Determination, Iranian Stock Market, SVR

1. Introduction

An audit fee (AF) is a fee charged by a certified public accountant (CPA) or auditing firm for conducting an audit of a company's financial statements. The AF is paid by the company being audited and covers the cost of the auditor's time, expertise, and resources required to thoroughly review the company's financial records, transactions, and internal controls (Kanapathippillai et al., 2024). The fee can vary depending on factors such as the size and complexity of the company, the scope of the audit, and the reputation and experience of the auditing firm (Labri et al., 2024; Subedi, 2024; Ani et al., 2024; Kim et al., 2024).

The determination of AF holds considerable importance within the auditing domain, exerting notable influence on both organizations and auditors (Prabhawa & Harymawan, 2022; Akinola & Olagunju, 2023; Azizkhani et al., 2023). As part of financial auditing procedures, companies engage external auditors are engaged by companies to scrutinize and validate their financial statements, ensuring accuracy and adherence to pertinent accounting standards and regulations. However, the complexity and extent of auditing procedures can vary substantially across companies, leading to disparities in AF (Boynton & Johnson, 2005). Consequently, the determination of AF has garnered considerable scholarly attention over the past four decades (Causholli et al., 2011; Hay, 2017; Xue & O'Sullivan, 2023).

1.1 Literature review

Determining AF involves various methodologies and considerations crucial for defining the financial dynamics between auditors and the entities being audited. These approaches aim to ensure that auditors receive equitable compensation for their services while upholding objectivity and independence in their assessments. For instance, in the investigation conducted by Rusmanto and Waworuntu (2015), which examined the determinants of AF in firms listed on the Indonesia Stock Exchange, findings indicated that factors such as profitability, business intricacy, and the number of subsidiaries did not significantly impact AF. In sum, these diverse elements collectively shape the ultimate cost of auditing services, ensuring a fair and suitable remuneration that reflects the distinctive attributes of each audit engagement.

Determination of AF has traditionally relied on conventional methods, which may now appear insufficient to meet the evolving requirements of contemporary enterprises. With companies encountering escalating complexities and continuously changing financial landscapes, there is a pressing need to explore more efficient and effective approaches to accurately ascertain AF. Within this context, the recent application of Artificial Intelligence (AI), particularly machine learning (ML) techniques, shows considerable potential in reshaping the process of determining various accounting parameters (Ghonji et al., 2020; Moradi et al., 2021; Mohammadi et al., 2021; Alibabae and Khanmohammadi 2022; Duan et al., 2023; Ramzan and Lokanan, 2024; Sun et al., 2024) and especially AF (Hunt et al., 2022; Fedyk et al., 2022; Subedi, 2024; Subedi, 2024; Pham, 2024).

For example, Bao et al. (2020) conducted a study where they discussed the development of an advanced fraud prediction model using a machine learning (ML) approach. The study highlighted the significance of integrating domain knowledge and ML techniques in the construction of the model. In contrast to previous accounting research that relied on financial ratios, Bao et al. (2020) opted for raw accounting numbers as the basis for selecting the model input. Instead of employing the commonly used logistic regression method, they utilized ensemble learning, considered one of the most powerful ML methods. To evaluate the performance of the fraud prediction models, they introduced a novel performance evaluation metric commonly utilized in ranking problems, which proved to be more suitable for the task at hand. By commencing with an identical set of theory-driven raw accounting numbers, Bao et al. (2020) demonstrated a substantial superiority of their new fraud prediction model over two benchmark models.

In this topic, the exploration of incorporating machine learning (ML) methods and novel data sources in the realm of management accounting (MA) research was undertaken by Ranta et al. (2022). A comprehensive examination of existing accounting and related research demonstrated that ML methods were still at an early stage of development in MA. However, several promising opportunities for leveraging ML in MA research were uncovered through an analysis of recently published ML research in related fields. It was proposed by the authors that the most favorable areas for employing ML methods in MA research included: (1) the harnessing of the abundant potential of diverse textual data sources; (2) the quantification of qualitative and unstructured data to generate novel metrics; (3) the enhancement of estimation and prediction capabilities; and (4) the employment of explainable AI techniques for the detailed interpretation of ML models. By utilizing ML methods, MA research could significantly contribute to the creation, advancement, and refinement of theories through induction and abduction, while also providing interventionist study tools. This highlighted the crucial role ML methods could play in the field of MA research.

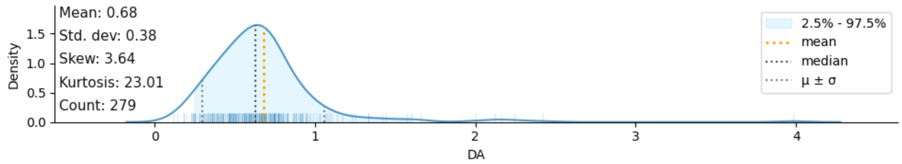
Fedyk et al.'s (2022) research conducted an extensive examination of the utilization of AI methodologies in auditing, presenting notable findings regarding the favorable effects of AI on audit quality and efficiency. The study underscored the potential advantages of AI integration in streamlining and optimizing audit procedures, especially in bolstering precision through advanced data analytics for detecting financial irregularities and fraud. Crucially, the research acknowledged AI as a potent tool to supplement and enhance the capabilities of human auditors rather than supplanting them, leading to more robust and dependable audit outcomes.

1.2 Objectives

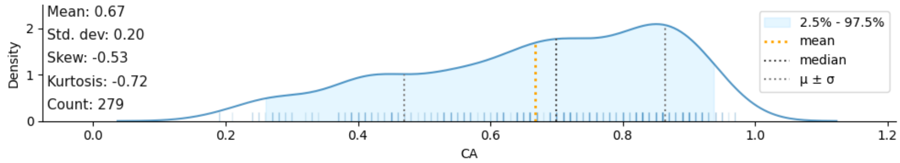
In this study, we aim to employ machine learning (ML) techniques, specifically support vector regression (SVR), utilizing Python libraries, to determine the audit fees (AF) from the Iranian Stock Market. We will leverage the Google Colab environment for modeling the SVR. To assess the effectiveness of the developed models, we will use new data points from various industries to evaluate the accuracy of the most successful model in predicting AF.

2. Data collection for estimating the AF

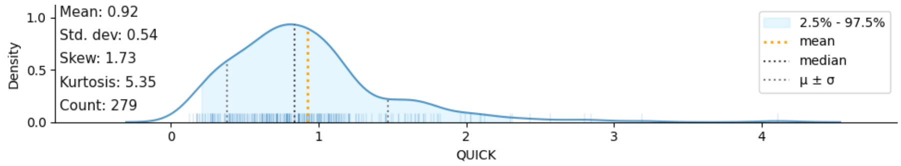
This study concentrates on companies listed on the Iranian Stock Market over a span of four years, spanning from 2017 to 2021. A total of 40 companies and 279 company-year observations were analyzed. The research utilized several input variables, including financial leverage (DA), current asset ratio (CA), quick ratio (QUICK), ASSETS, current ratio to current liabilities (CR), and long-term debt (DE), with AF as the target variable. To ensure the precision of the models, certain large companies with notably high audit fees were omitted from the dataset. This precautionary measure was implemented to prevent potential bias in the analysis and to uphold the relevance of the results within the targeted range of audit fees. Statistical parameters for all variables considered in the study are presented in Figs 1(a) through 1(f), with the account number indicating the total observations at 279. Fig. 1(f) specifically illustrates the AF of the companies, which are represented in logarithmic format. By meticulously controlling the data range, the study endeavors to offer a more focused and accurate exploration of the relationships between the chosen variables and the audit fees of companies within the specified range.



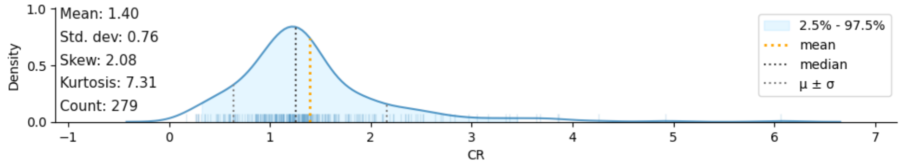
(a)



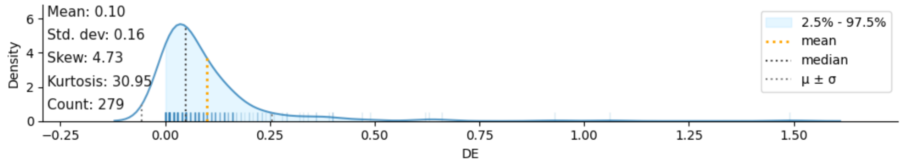
(b)



(c)



(d)



(e)

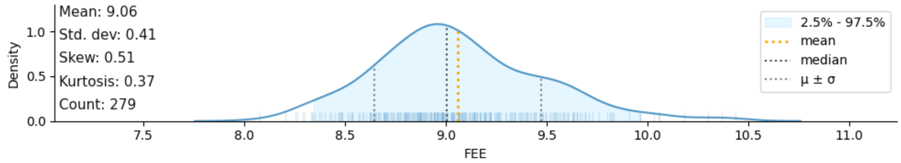


Fig. 1: Statistical summary of the dataset utilized

(f)

3. Development of the SVR model

Support Vector Regression (SVR) is a supervised learning algorithm utilized for regression tasks, functioning akin to Support Vector Machines (SVM) employed in classification but adapted for predicting continuous outcomes. SVR identifies the optimal hyperplane within a high-dimensional space to maximize the margin while minimizing the error between predicted and actual values. Implementation of SVR was conducted using the `SVR` class from the `SVMsvm` module in scikit-learn, specifying hyperparameters including kernel type, regularization parameter (C), and epsilon. The model was trained on the training data using the `fit` method and employed for making predictions on new data via the `predict` method. Notably, the SVR model utilized in this study was developed using the Sklearn Python library. The coding process took place in the Google Colab IDE environment, favored by machine learning and data science professionals for its seamless accessibility and cloud-based execution capabilities. Google Colab offers convenient access to high-performance GPUs and supports parallel code execution, eliminating the need for local development setup management. Fig. 2 depicts the developed code for SVR.

```
# Importing necessary libraries
from sklearn.svm import SVR
import pandas as pd
df = pd.DataFrame(data)
# Separating input features (x) and target values (y)
x = df['inputs']
y = df['target']
# Instantiating SVR model
svr_model = SVR(kernel='rbf', C=1.0, epsilon=0.1)
# Fitting the model to the data
svr_model.fit(x, y)
# Generating sample test data
X_test = df['newdata']# Sample features for testing
# Making predictions on the test data
y_pred = svr_model.predict(X_test)
# Printing the predictions
print("Predictions:", y_pred)
```

Fig 2: Developed code for the SVR

4. Results and discussion

4.1 Correlation matrix

Fig 2 displays the correlation matrix of the parameters. Typically, this matrix illustrates the influence of each parameter on the target variable (AF). However, it does not account for the effects of other parameters. It is evident that the most significant parameter in estimating the audit fee is the company's assets, which is logically consistent.

4.2 SVR results on train data

SVR employs a kernel function to transform the input data into a higher-dimensional space, making it linearly separable. During training, SVR optimizes hyperplane parameters to minimize error and maximize margin from the closest data points, known as support vectors. This optimization involves solving a constrained optimization problem, iteratively adjusting parameters until convergence is reached.

Additionally, hyperparameters like kernel type and regularization parameters are tuned for optimal performance using techniques like cross-validation. The objective of SVR training is to simultaneously minimize prediction error and maximize the margin between the hyperplane and support vectors.

This is achieved through an iterative optimization process, where the model adjusts parameters to fit the training data while maintaining a balance between model complexity and generalization ability. SVR's success relies on its ability to find the optimal trade-off between fitting the training data closely and maintaining flexibility to generalize well to unseen data.

Fig. 4 depicts the results obtained from the trained SVR model, illustrating the comparison between the actual AF and the estimated AF. The plot reveals a close agreement between these parameters, indicating the promising performance of the SVR model in capturing the underlying patterns in the data. However, it's crucial to note that these findings are based solely on the training data, which may not necessarily generalize well to unseen data. It's widely acknowledged in machine learning that model performance metrics, such as the R-squared (R^2) and Mean Squared Error (MSE), tend to be favorable to the training data. In this context, the calculated R^2 value for the SVR model on the training data was approximately 90.5%, suggesting a high degree of explained variance. Additionally, the MSE of the SVR model on the training data was found to be 3.7, indicating a reasonable level of error. While these results are encouraging, it's imperative to evaluate the model's performance on unseen data to assess its true predictive capability and generalization ability. Further validation and testing on independent datasets are essential steps to confirm the reliability and robustness of the SVR model in real-world applications.

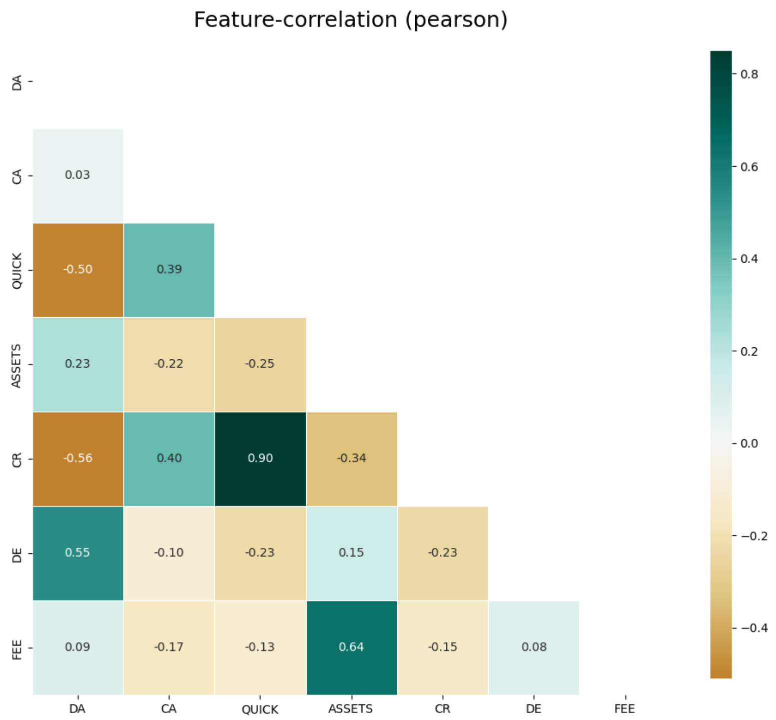


Fig. 3: Correlation matrix of the parameters

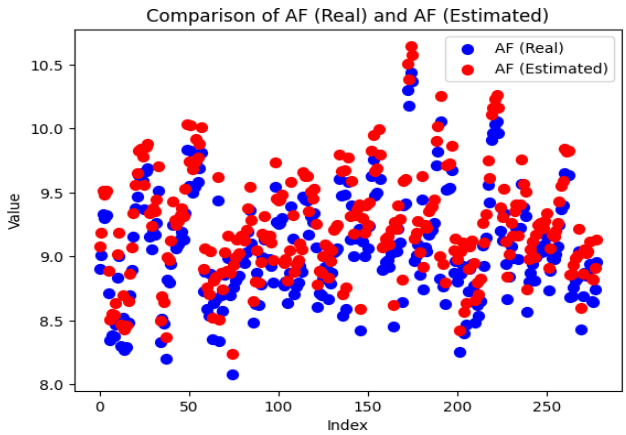


Fig. 4: SVR results

4.2 SVR results on test data

To rigorously validate and ascertain the generalization capabilities of our top-performing model, we opted to scrutinize its efficacy using fresh data. To accomplish this, we sourced data from 10 previously unobserved companies across diverse industries (30 new data), spanning the financial years from 2021 to 2022. Leveraging the Support Vector Regression (SVR) model, we projected the Audit Fees (AF) for these companies and juxtaposed them against the actual fees, as delineated in Fig. 5. Encouragingly, the SVR model showcased notable precision in estimating the new dataset, showcasing a robust alignment with the actual values. The computed R-squared (R^2) value for the SVR model on the training data hovered around 67%, indicating a decrease in the accuracy of the model with new data. Furthermore, the Mean Squared Error (MSE) of the SVR model on the training data was 8.1, signifying an intermediate level of predictive accuracy.

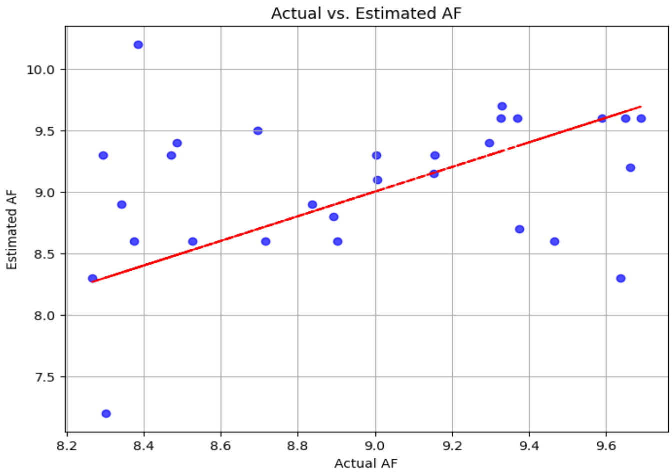


Fig. 5: Validation of the GBR model with new data

5. Conclusions, Limitations, and Future Prospects

In conclusion, our study undertook a comprehensive analysis of audit fee (AF) determination using Support Vector Regression (SVR) on a dataset comprising audit fee data from 60 companies listed on the exchange between 2017 and 2021, resulting in 279 year-observations. Leveraging input variables such as financial leverage (DA), current asset ratio (CA), quick ratio (QUICK), ASSETS, current ratio to current liabilities (CR), and long-term debt (DE), we trained the SVR model on this dataset within the Google Colab environment. Our training results revealed a commendable performance, with an R^2 value of approximately 90.5% and a Mean Squared Error (MSE) of 3.7, indicating a high degree of explained variance and reasonable predictive accuracy, respectively. However, upon testing the model on new data, its performance diminished, as evidenced by a reduced R^2 value of around 67% and an MSE of 8.1. This decline suggests a decrease in accuracy and intermediate predictive performance when confronted with unseen data. Moving forward, addressing the limitations identified, such as data scope and variable selection, presents avenues for future research to enhance the robustness and generalizability of AF determination models.

Although this study has yielded promising outcomes, it is important to acknowledge certain constraints. Notably, the dataset utilized for training and testing the models was limited to firms listed on the Tehran Stock Market from 2017 to 2021. While this dataset offered valuable insights, the ability of the models to generalize to companies beyond this specific timeframe or exchange may be limited. Incorporating data from alternative exchanges or regions could enhance the models' applicability across a wider spectrum of companies. Additionally, the selection of input variables was based on a specific set of parameters, including DA, CA, QUICK, ASSETS, RE, LOSS, CR, and DE. Although these variables were considered relevant and produced satisfactory results, there may exist other factors influencing AF. Future investigations could explore additional variables and employ feature engineering techniques to refine the models and achieve a more comprehensive understanding of the AF determination process.

References

Akinola, A. O., & Olagunju, A. (2023). Automated Accounting System and External Audit Fees: Empirical Evidence from Nigeria. KIU Interdisciplinary Journal of Humanities and Social Sciences, 4(1), 55–81.

Al Ani, M. K., ALshubiri, F., & Al-Shaer, H. (2024). Sustainable products and audit fees: empirical evidence from western European countries. Sustainability Accounting, Management and Policy Journal.

Alibabaee, G., & Khanmohammadi, M. (2022). The Study of the Predictive Power of Meta-Heuristic Algorithms to Provide a Model for Bankruptcy Prediction. International Journal of Finance & Managerial Accounting, 7(26), 33–51.

Azizkhani, M., Hossain, S., & Nguyen, M. (2023). Effects of audit committee chair characteristics on auditor choice, audit fee, and audit quality. Accounting & Finance.

Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting accounting fraud in publicly traded US firms using a machine learning approach. Journal of Accounting Research, 58(1), 199–235.

Bisong E. (2019), Building Machine Learning and Deep Learning Models on theGoogle Cloud Platform (pp. 59–64). Berkeley, CA: Apress.

Boynton WC and Johnson RN (2005) Modern auditing: Assurance services and the integrity of Financial Reporting. John Wiley & Sons.

Causholli M., De Martinis M., Hay D., and Knechel W. (2011). Audit markets, fees and production: Towards an integrated view of empirical audit research. Journal of Accounting Literature, 29, 167–215.

Choi, S. U., Lee, K. C., & Na, H. J. (2022). Exploring the deep neural network model's potential to estimate abnormal audit fees. Management Decision, 60(12), 3304–3323.

Duan, H. K., Vasarhelyi, M. A., Codesso, M., & Alzamil, Z. (2023). Enhancing the government accounting information systems using social media information: An application of text mining and machine learning. International Journal of Accounting Information Systems, 48, 100600.

Fedyk, A., Hodson, J., Khimich, N., and Fedyk, T. (2022). Is artificial intelligence improving the audit process? Review of Accounting Studies, 27(3), 938–985.

Ghonji Feshki, A., Khanmohammadi, M. H., & Yazdani, S. (2020). Political connection and earnings management methods: evidence from the Tehran Stock Exchange, InternationalJournal of Finance & Managerial Accounting, 5(17), 1–17.

Greiner A, Kohlbeck MJ, Smith, TJ (2017). The relationship between aggressive real earnings management and current and future audit fees. Auditing: A Journal of Practice & Theory, 36(1), 85–107.

Hay, D. (2017). Audit fee research on issues related to ethics. Current Issues in Auditing, 11(2), A1–A22.

Hunt, E., Hunt, J., Richardson, V. J., & Rosser, D. (2022). Auditor response to estimated misstatement risk: A machine learning approach. Accounting Horizons, 36(1), 111-130.

Kanapathippillai, S., Yaftian, A., Mirshekary, S., Sami, H., & Gul, F. A. (2024). Director turnover, board monitoring, and audit fees: Some Australian evidence. Pacific-Basin Finance Journal, 83, 102246.

Kim, I., Kong, J. H., & Yang, R. (2024). The impact of board reforms on audit fees: International evidence. Journal of Business Finance & Accounting, 51(1-2), 45–83.

Kramer O. and Kramer O. (2016) Scikit-learn. Machine learning for evolution strategies, pp. 45–53.

Larbi, S. B., Mandzila, E. E. W., Meniaoui, J., & Moor, E. T. (2024). The influence of the auditor and auditee on mandatory audit fees in France. GECONTEC: Revista Internacional de Gestión del Conocimiento y la Tecnología, 12(1), 77–102.

Ngoc Hung D, Thuy Van VT, and Archer L (2023). Factors affecting the quality of financial statements from an audit point of view: A machine learning approach. Cogent Business & Management, 10(1), 2184225.

Mohammadi, M., Yazdani, S., & Khanmohammadi, M. (2021). Presenting a Model for Financial Reporting Fraud Detection Using Genetic Algorithms. Advances in Mathematical Finance and Applications, 6(2), 377–392.

Moradi, Z., Ghilavi, M., Khan Mohammadi, M. H., & Hajiha, Z. (2021). Application of resource-based view theory in assessing the efficiency of Firm listed on the Tehran stock exchange by data envelopment analysis. Advances in Mathematical Finance and Applications, 6(3), 607–629.

Pham, O. T. T. (2024). Forecasting audit opinions on financial statements: statistical algorithm or machine learning? Electronic Journal of Applied Statistical Analysis, 17(1), 133–152.

Prabhawa, A. A., & Harymawan, I. (2022). Readability of Financial Footnotes, Audit Fees, and Risk Management Committee. Risks, 10(9), 170.

Ramzan, S., & Lokanan, M. (2024). The application of machine learning to study fraud in the accounting literature. Journal of Accounting Literature.

Ranta, M., Ylinen, M., & Järvenpää, M. (2023). Machine learning in management accounting research: Literature review and pathways for the future. European Accounting Review, 32(3), 607–636.

Rusmanto T., Waworuntu SR (2015) Factors Influencing Audit Fees in Indonesian Publicly Listed Companies Applying GCG. Procedia-Social and Behavioral Sciences, 172, 63–67.

Saleh MA, Ragab YM (2023). Determining audit fees: evidence from the Egyptian stock market. International Journal of Accounting & Information Management, 31(2), 355–375.

Subedi, M. (2024). Principles-based accounting standards, audit fees, and going concern: evidence using advanced machine learning. International Journal of Accounting & Information Management, 32(2), 308–344.

Subedi, M. (2024). Principles-based accounting standards, audit fees, and going concern: evidence using advanced machine learning. International Journal of Accounting & Information Management, 32(2), 308–344.

Sun, Y., Li, J., Lu, M., & Guo, Z. (2024). Study of the Impact of the Big Data Era on Accounting and Auditing. arXiv preprint arXiv:2403.07180.

Xue B, O'Sullivan N (2023), The determinants of audit fees in the alternative investment market (Aim) in the UK: Evidence on the impact of risk, corporate governance, and auditor size, Journal of International Accounting, Auditing and Taxation, 50, 100523.