International Journal of Multidisciplinary Research and Publications
ISSN (Online): 2581-6187

Machine Learning and Substantive Analytical Procedure in Financial Audit

Pakorn Kotepuchai[1], Yachai Limpiyakorn[2]
[1, 2]Department of Computer Engineering, Chulalongkorn University, Bangkok, Thailand
Corresponding Author: 6570137521@student.chula.ac.th

Abstract**—***In the technology-driven era, the integration of machine learning (ML) into financial auditing offers substantial potential for improving audit quality. This study explores the application of ML to enhance the substantive analytical procedure (SAP), a critical component of financial auditing. Traditional auditing methods, often limited by tools like Excel, struggle to handle extensive and complex datasets, leading to inefficiencies and potential errors. This research aims to develop an advanced substantive analytical procedure using ML to provide deeper insights and more accurate results. Data was acquired from an Enterprise Resource Planning (ERP) system, by merging sales and cost data using invoice numbers. The three models including Linear Regression, Random Forest, and Multi-Layer Perceptron were trained and evaluated. The metrics including MAE, MSE, RMSE, and R-squared are used for model performance evaluation. The Random Forest regressor outperformed the other two models to predict the total sale amount from cost data. Moreover, when compared to the traditional SAP, the Random Forest model also achieved the least deviation between the actual and predicted sale amount. These findings highlight the potential of advanced ML models to enhance the accuracy and reliability of substantive analytical procedures in financial auditing.*

Keywords**—** *Financial audit: substantive analytical procedure: substantive tests: linear regression: random forest: neural network.*

I. INTRODUCTION

In today technology-driven era, machine learning (ML) is transforming various sectors by offering innovative solutions to complex problems. One significant area where the impact can be profound is financial auditing. Traditional financial audits often depend on tools like Excel, which, despite their usefulness, have limitations, particularly when handling extensive datasets. Excel is capable of processing up to 1,048,576 rows, but contemporary financial datasets frequently surpass this threshold, requiring more powerful solutions. Recent studies have highlighted that traditional auditing methods struggle with the increasing volume and complexity of financial data, leading to inefficiencies and potential errors [1, 2]. Machine learning can efficiently process and analyze large volumes of data, presenting a promising alternative compared to traditional methods. This paper aims to study the application of ML to enhance the quality of financial audits. Specifically, the focus is on developing an advanced substantive analytical procedure (SAP) that leverages ML's capabilities. Compared to the previously employed simple methods in SAP, the proposed ML-driven approach is expected to provide deeper insights and more accurate results, thus improving the effectiveness and reliability of financial audits. As noted in recent literature, integrating advanced technologies like ML into auditing practices can significantly improve accuracy and efficiency, addressing many limitations of traditional methods [3].

The objective of this research is to demonstrate the advantages of ML-based SAP over the traditional method, addressing the current challenges and limitations auditors face. By integrating machine learning into the auditing process, we aim to not only improve the quality and accuracy of audits but also ensure they are more adaptable to the complexities of modern financial data. The need for such advancements is underscored by the growing complexity of financial transactions and the limitations of traditional auditing methods in managing these challenges [1, 2, 3].

II. FINANCIAL AUDIT

A. Audit

In order to determine the extent to which claims concerning economic activities and occurrences comply with predetermined criteria, auditing is a methodical procedure that gathers, objectively assesses, and communicates data to interested parties. Audits are typically performed by external auditors who are independent of the organization being audited, providing an unbiased assessment of the financial statements.

The primary purpose of an audit is to enhance the degree of confidence of intended users in the financial statements. This is accomplished by the auditor providing an opinion on whether or not the financial statements are prepared in compliance with the relevant financial reporting framework in all material respects [4].

B. Audit Phases

The audit process generally follows several key phases [4], example as shown in Fig. 1 which contains four major phases critical to ensuring comprehensive and effective audits.

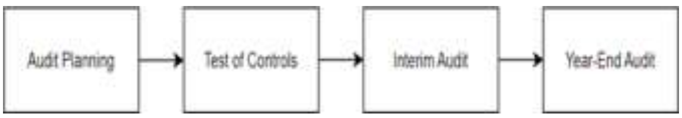

Fig. 1. Example of major audit phases.

- Audit Planning

The process starts with audit planning, where the auditor develops an overall strategy for the ongoing audit. This phase involves understanding the client's business, industry, and internal control environment. The auditor assesses risks and identifies areas that may require special attention. The effective planning ensures that the audit is conducted efficiently and effectively, focusing on areas with higher risks of material misstatement.

- Test of Controls

During this phase, auditors evaluate the effectiveness of an organization's internal controls over financial reporting. Internal controls are processes and procedures implemented by management to ensure the reliability of financial reporting, compliance with laws and regulations, and the effectiveness and efficiency of operations. Testing controls helps auditors determine the nature, timing, and extent of substantive testing needed.

- Interim Audit

Typically, the interim audit occurs before the year-end audit. It involves preliminary audit work: reviewing interim financial statements, performing analytical procedures, and testing transactions that occurred during the first part of the fiscal year. The interim audit helps identify potential issues early, allowing for timely resolution before the final audit.

- Year-End Audit (Substantive Testing)

The year-end audit phase is where substantive testing is performed. Substantive tests are designed to detect material misstatements in the financial statements. This phase involves detailed testing of account balances, transactions, and disclosures to obtain sufficient appropriate audit evidence to form an opinion on the financial statements.

C. *Substantive Analytical Procedures*

Substantive analytical procedures are audit tests that involve evaluating financial information through analysis of plausible relationships among both financial and non-financial data. These procedures are used to identify inconsistencies or unexpected trends that may indicate potential misstatements.

Analytical procedures can be divided into different categories, including:

- *Trend Analysis*: Examining changes in account balances over time to identify unusual fluctuations
- *Ratio Analysis*: Comparing financial ratios from one period to another or against industry benchmarks to detect anomalies
- *Reasonableness Tests*: Assessing whether financial data appears reasonable based on the auditor's knowledge of the business and industry

Substantive analytical procedures are typically more effective when there are predictable relationships among data, such as stable business operations. These procedures can provide strong evidence when used in combination with other substantive tests, particularly in areas where the auditor expects consistent and predictable relationships [4].

III. LITERATURE REVIEW

In recent years, the integration of machine learning and artificial intelligence into the field of accounting and auditing has garnered significant attention. These technologies are promising for transforming traditional practices and providing new insights to enhance the efficiency, accuracy, and reliability of financial processes. In literature, various research has been conducted to highlight the advancements, applications, and challenges of applying machine learning in accounting as summarized in Table 1.

TABLE I. Related work.

Citation/Year of Publishing	Name of Journal/Conference	Technique/Tools	Summary
[5]/2015	Using Classification Algorithms for Smart Suggestions in Accounting Systems	SVM, Feed-Forward Neural Networks	Investigated the performance of SVM and feed-forward neural networks for classifying financial transactions. The authors suggested the use of clustering algorithms to improve the automation process as future work.
[6]/2018	Machine Learning and Rule Induction in Invoice Processing	Rule Induction, SVM, Random Forest	Compared rule induction methods with the selected ML methods for invoice classification. The models of Support Vector Machines (SVM) and Random Forest were used to benchmark against rule induction.
[7]/2020	Automatic Electronic Invoice Classification Using Machine Learning Models	Random Forest, AdaBoost, MLP	Developed a system to automate the classification of invoices into account and VAT codes using multiclass classification algorithms. The study used text mining techniques to extract relevant information from structured electronic invoices in a predefined XML template.
[8]/2021	Machine Learning for Financial Transaction Classification Using Character-Level Word Embeddings of Text Fields	Random Forest, k-Nearest Neighbor, Logistic Regression, Word Embeddings	Developed ML systems for mapping financial transactions to accounts using character-level word embeddings of text fields. The study focused on creating models that could generalize across different companies and sectors.
[9]/2022	Automated Accounting Using Machine Learning	Logistic Regression, Random Forest, RNN, SVM	Explored the automation of invoice classification using various ML models. They experimented with Support Vector Machines (SVM), Logistic Regression, Recurrent Neural Networks (RNN), and Random Forest. They also used text encoding techniques such as TF-IDF and Count Vector.
[10]/2022	Hierarchical Classification for Account Code Suggestion	Logistic Regression, Neural Networks, Random Forest, Hierarchical Classification	Proposed a hierarchical single-label classifier for account code suggestion. They introduced the Top-K Parent Boosting strategy to improve recommendation performance.
[11]/2024	Tree-based Classifiers for Smart General Ledger Code Suggestion	Random Forest, CART, NLP	Utilized tree-based classifiers for general ledger code suggestion. NLP was applied for data preprocessing. The study compared CART and Random Forest models using TF-IDF and Bag of Words vectorization
[12]/2024	Multilabel Classification of Account Code in Double-Entry Bookkeeping	Random Forest, MLP, CART, NLP	Multilabel classification was adopted to suggest account codes in double-entry bookkeeping. The research compared the model performance among CART, Random Forest, and Multilayer Perceptron using alternative text vectorization methods like Bag of Words and TF-IDF.

IV. MODEL CONSTRUCTION

Traditional SAP methods rely on manual analysis and simple statistical techniques, which can be prone to human error and limited in their predictive power. The AI-based Substantive Analytical Procedure (SAP) was proposed to enhance the effectiveness and accuracy of predicting sales based on the important relevant variables.

Initially, the data were extracted from the Enterprise Resource Planning (ERP) system containing the sales data and the cost data (inventory report). The dataset was created by joining the sales and cost information using the invoice number as the key. It contains 1,531,871 transactions of the sale report dated from January 1, 2021 to December 31, 2023. Table 2 describes the six relevant attributes essential for suggesting the target variable *sale* denoting the actual sale amount that the models aim to predict. Steps of data preprocessing and model construction are described in the following subsections.

A. Data Preprocessing

Data type conversion is required for fitting the data into the regressor models, while Normalization is essential for improving the model performance.

- *Label Encoding*− Since the model training required the numeric input, the categorical features such as *Category Name, Channel, Group Name, and Product Type* were then transformed to number using label encoding in this work.
- *Standard Scaler*− The numeric *cost* column was standardized to ensure that the data fits well within the range required for regression analysis. Standardization helps improving the convergence of gradient-based algorithms, leading to the better model. The Z-score normalization was used to adjust feature values to follow a specific distribution, making comparisons between features more meaningful by eliminating disparity in scale.

TABLE II. Feature description of dataset.

Column Name	Description	Data type
Qty.	Quantity of sale	Number
Cat_name	Sale category	Text
Channel	Distribution channel	Text
Group_name	Sale group name (online/offline)	Text
Type	Types of products	Text
Cost	Total cost of goods sold for transaction	Number
Sale	Total sale (actual)	Number

B. Data Splitting

The dataset was split into the training and testing sets based on temporal order, simulating the practical use of historical data for predicting current outcomes. The data was split into training and testing sets with an 80:20 ratio. This split was performed based on specific dates, using the data of previous years to predict the current year, that aligns with the auditor's practice of using historical data to predict and compare with the actual recorded amounts.

C. Model Training

The three machine learning models were selected consisting of Linear Regression, Random Forest, and Multi- Layer Perceptron.

- *Linear Regression (LR)* was chosen as a baseline model due to its simplicity and interpretability. Since linear regression does not have hyperparameters to fine-tune, it was trained directly on the data.
- *Random Forest (RF)* is an ensemble learning method that operates by constructing multiple decision trees during training and outputting the mode of the classes (classification) or mean prediction (regression) of the individual trees [13]. The hyperparameter tuning was performed using Randomized Search CV() with 5-fold cross-validation. It is effective in finding the best hyperparameters by randomly sampling a wide range of values, which can be more efficient than grid search when dealing with large datasets or complex models. The primary hyperparameters tuned for Random Forest include the number of trees in the forest, maximum depth of the trees, and the minimum number of samples required to split an internal node.
- *Multi-Layer Perceptron (MLP)* is a class of feedforward Artificial Neural Network. MLP models consist of one or more hidden layers positioned between the input and output layers. The input layer transmits data to the hidden layers without performing any computations. The hidden layers then execute the necessary computations and pass the information to the output layer, which returns the predicted result [14]. RandomizedSearchCV() with 5-fold cross-validation was also performed for hyperparameter tuning of MLP.

The hyperparameters tuned for the MLP model include learning method, max iteration, learning rate, number of hidden layers, number of neurons in hidden layer, regularization parameter, and activation function.

The environment for model construction includes:

- Processor: 11th Gen Intel Core I i7-1185G7 @ 3.00GHz
- System type: 64-bit operating system, x64-based processor; RAM: 16 GB

Hyperparameters were fine-tuned as below:

- *Linear Regression* criterion: default
- *Random Forest Regressor* criterion: n_estimators:50, min_samples_split: 10, max_depth: 10
- *MLP Regressor* criterion: solver: adam, max_iter: 1000, learning_rate: constant, hidden_layer_sizes: (100, 100), alpha: 0.001, activation: Relu

V. MODEL EVALUATION

The models were evaluated on the test data. Several evaluation metrics were considered to determine the best-performing model including:

- *Mean Absolute Error (MAE)*: Calculates the average error size in a group of forecasts without taking the direction of the mistakes into account. It is useful for understanding the average error in the units of the variable of interest.
- *Mean Squared Error (MSE)*: Measures the average of the squares of the errors, providing a way to identify larger errors due to its squaring effect.
- *Root Mean Squared Error (RMSE)*: The square root of

MSE, which also gives an idea of the error magnitude but in the same units as the original data.

- *R-Squared (R^2)*: Represents the proportion of the variance for the dependent variable that's explained by the independent variables in the model. It is useful for understanding the goodness-of-fit of the model.

Table III summarizes the model performance among the three chosen models. This comparative analysis provides insights into the suitability of each model for financial auditing purposes.

TABLE III. Comparisons of model performance of chosen algorithms.

Model	MAE	MSE	RMSE	R^2
LR	696.13	6,210,464.27	2,492.08	0.72
RF	**307.24**	**1,350,852.82**	**1,162.26**	**0.94**
MLP	460.91	2,136,993.19	1,461.85	0.90

Linear Regression, as expected, served as a baseline model. While it provided a decent R-squared value of 0.72, its error metrics (MAE, MSE, RMSE) were significantly higher compared to the other models, indicating that it may not be as effective in capturing the complex relationships within the data.

The Multi-Layer Perceptron model also performed well, with an R-squared value of 0.94. While its performance was slightly lower than the Random Forest model, it still significantly outperformed the Linear Regression model. The MLP's ability to model non-linear relationships helped it achieve better performance compared to Linear Regression, although it did not surpass the Random Forest model.

The performance MLP model demonstrated the potential of neural networks in this domain. However, the slightly higher error metrics compared to Random Forest indicate that further tuning and possibly more sophisticated neural network architectures could be explored to improve its performance.

The experimental results clearly indicate that the Random Forest is the best performer among the three models evaluated. The Random Forest model, with its ensemble learning approach, was particularly effective, achieving the best performance metrics across all evaluation criteria. This suggests that Random Forest is highly capable of capturing the complex interactions and patterns in the data.

Table 4 summarizes the performance of the three machine learning models compared to that of the traditional SAP method. The deviation between the predicted and the actual sale amount are reported in local currency. It is evident that the ML models significantly outperformed the traditional method in predicting sales. The Random Forest model was the most accurate, highlighting the potential of machine learning to enhance audit accuracy and reliability. This demonstrates the substantial benefits of integrating machine learning into auditing practices.

TABLE IV. Comparisons of model performance with traditional SAP.

Method	Actual Sale Amount	Predicted Sale Amount	Deviation
Traditional SAP	*485,211,461.32*	*455,195,693.03*	*30,015,771.28*
Linear Regression	485,211,461.32	458,140,686.75	27,070,776.56
Random Forest	**485,211,461.32**	**478,098,495.98**	**7,112,965.33**
MLP	485,211,461.32	464,712,095.50	20,499,366.82

The overall results indicate that both Random Forest and Multi-Layer Perceptron models substantially improve the accuracy of substantive analytical procedures in financial auditing compared to the traditional SAP method.

VI. Conclusion

Traditional Substantive Analytical Procedures rely on manual analysis and simple statistical techniques that prone to human error and limitation in predictive power. This study thus presents an application of machine learning to enhance the effectiveness and accuracy of the substantive analytical procedure in financial auditing. The experimental results demonstrated significant improvements over the traditional method. The findings reported the promising results of the ML model performance evaluated by MAE, MSE, RMSE, and R^2. The Random Forest model outperformed the others, followed by the Multi-Layer Perceptron, and the Linear Regression. Compared to the Linear Regression as the baseline, the Random Forest and Multi-Layer Perceptron better handle large datasets and complex relationships, providing more accurate and precise predictions of sales figures. This reduces the risk of misstatements and enhances the reliability of audit results.

Compared to the traditional method of SAP, the Random Forest model achieved the lowest error rates, indicating its superior capability to capture complex patterns in the data. The MLP model also showed the strong performance, suggesting that neural networks can effectively model non- linear relationships in financial data. These findings underscore the potential of machine learning models to enhance the accuracy and reliability of financial audits, providing auditors with powerful tools to detect anomalies and assess financial information more precisely.

Further direction would be exploring the application of machine learning techniques to other SAP areas, such as:

Accounts Receivable vs. Allowance for Doubtful Debt: Implement machine learning models to predict accounts receivable and compare them with the allowance for doubtful debts. This could enhance the accuracy of provisioning and improve the reliability of financial statements.

Selling Expenses Analysis: Analyze the appropriateness of the percentage of each type of selling expense relative to sales. Machine learning can help identify unusual expense patterns and potential areas of concern.

The continuous evolution of AI audit holds the promise of transforming financial auditing into a more precise and insightful practice, ultimately contributing to greater financial transparency and trust.

References

[1] Y. Sun, J. Li, M. Lu, and Z. Guo, "Study of the Impact of the Big Data Era on Accounting and Auditing," *Frontiers in Business, Economics and Management*. 2024.

[2] T. M. Muneerali, A. E. Ries, and K. V. Thomas, "Emerging technology and auditing practice: analysis for future directions," *European Journal of Management Studies,* 27, 1, 99-119, doi: 10.1108/EJMS-06-2021-0058, 2022.

[3] A. Liew, P. Boxall, and D. Setiawan, "The transformation to data analytics in Big-Four financial audit: what, why and how?," *Pacific Accounting Review,* 34, 4, 569-584, doi: 10.1108/PAR-06-2021-0105, 2022.

[4] T. J. Louwers, A. D. Blay, D. H. Sinason, J. R. Strawser, and J. C. Thibodeau. *Auditing & Assurance services*, 7th ed. New York: McGraw-Hill Education, 2018.
[5] B. Hampus and J. Johannes. Using classification algorithms for smart suggestions in accounting systems Master of Science, Department of Computer Science & Engineering, Chalmers University of Technology, 2015.
[6] J. Bergdorf, "Machine learning and rule induction in invoice processing: Comparing machine learning methods in their ability to assign account codes in the bookkeeping process," ed, 2018.
[7] C. Bardelli, A. Rondinelli, R. Vecchio, and S. Figini, "Automatic electronic invoice classification using machine learning models," *Machine Learning and Knowledge Extraction,* 2, 4, 617-629. 2020.
[8] J. R. Kær and I. Christian, "Machine learning for financial transaction classification across companies using character-level word embeddings of text fields," *Intelligent Systems in Accounting, Finance and Management,* 28, 3, 159-172, 2021/07/01. 2021.
[9] P. Meenal. Automated Accounting Using Machine Machine Learning Independent thesis Advanced level (degree of Master (Two Years)) Student thesis, IT, 22028, 2022.
[10] J. Munoz, M. Jalili, and L. Tafakori, "Hierarchical classification for account code suggestion," *Knowledge-Based Systems,* 251, 109302. 2022.
[11] P. Kotepuchai and Y. Limpiyakorn, "Tree-based Classifiers for Smart General Ledger Code Suggestion," presented at the *Proceedings of the 2024 13th International Conference on Informatics, Environment, Energy and Applications*, Tokyo, Japan, pp. 17–22, 2024.
[12] P. Kotepuchai and Y. Limpiyakorn, "Multilabel Classification of Account Code in Double-Entry Bookkeeping," presented at the *2024 10th International Conference on Computer Technology Applications (ICCTA)*, Vienna, Austria, 2024.
[13] L. Breiman, "Random forests," *Machine Learning,* 45, 1, 5-32, doi:10.1023/A:1010933404324, 2001.
[14] R. Vang-Mata. *Multilayer Perceptrons: Theory and Applications* (Computer Science, Technology and Applications). Hauppauge, New York: Nova (in English), 2020.