Vol. 8, No. 2, April 2026

Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Integrating ML with Electronic Fiscal Devices for Real-Time Underpricing Detection in Tanzania

Benitho Alphonce Chengula[1], Judith Leo[2], Cyril Chimilila[3]

[1,2] School of Computational and Communication Science and Engineering, Nelson Mandela African Institution of Science and Technology, Arusha, Tanzania

[3]Institute of Tax Administration, Dar es Salaam, Tanzania

Received:
January 12, 2026
Revised:
March 26, 2026
Accepted:
April 26, 2026
Published:
May 26, 2026

Corresponding Author:
Author Name*:
Benitho Alphonce Chengula
Email*: chengulab@nm-aist.ac.tz

DOI:
10.63158/journalisi.v8i2.1582

© 2026 Journal of Information Systems and Informatics. This open access article is distributed under a (CC-BY License)

Abstract. This study aims to develop a machine learning-based tool integrated into Electronic Fiscal Devices (EFDs) to detect underpricing fraud in real time in Tanzania. The motivation for this research arises from the limitations of existing EFD systems, which rely on manual and post-audit mechanisms that are ineffective in identifying fraudulent pricing during transactions. A mixed-methods approach was employed, combining qualitative insights from tax officers with quantitative data collected from traders and buyers. A dataset of 5,000 mobile phone sales transactions collected from Arusha, Dar es Salaam, and Iringa in Tanzania, was pre-processed and used to train and evaluate multiple machine learning models, including Logistic Regression, Support Vector Machine, XGBoost, and Random Forest, using 5-fold cross-validation. The experimental results show that the Random Forest model outperformed other models, achieving an accuracy of 99.6% along with strong precision, recall, and F1-score values. To demonstrate practical applicability, the trained model was further integrated into a prototype EFD environment, where it enabled near real-time fraud detection and generated automated alerts for traders and tax authorities, with geolocation features supporting targeted enforcement. However, the dataset is limited to mobile phone transactions within selected regions of Tanzania, which may affect the generalizability of the findings. The novelty of this study lies in integrating machine learning–based price validation into EFD systems to support proactive detection of underpricing fraud at the point of transaction, thereby enhancing tax compliance and revenue protection.

Keywords: Electronic Fiscal Devices (EFDs), Machine Learning, Underpricing Fraud, Real-time Detection, Random Forest.

1. INTRODUCTION

Value Added Tax (VAT) is one of the most widely adopted indirect tax systems globally, first introduced in Germany in 1958 and now implemented in more than 160 countries. It serves as a major source of government revenue, collected by traders on behalf of the state from the final consumer. In Tanzania, VAT administration has been strengthened through the introduction of Electronic Fiscal Devices (EFDs) in 2010 under the Value Added Tax (Electronic Fiscal Device) Regulations, 2010, through Government Notice No. 192 issued on May 8, 2010 [1]–[3]. These devices are designed to record sales transactions, support business operations such as stock control and reporting, and ensure that taxes are computed based on declared sales values. As a result, EFDs play a critical role in enhancing transparency, accountability, and compliance in tax collection.

To ensure effective use of EFDs, the Tanzania Revenue Authority (TRA) has implemented various compliance monitoring and enforcement mechanisms. These include desk examinations, field visits, online monitoring of transactions, surveillance, spot checks, and periodic inspections of EFD systems, along with legal sanctions for non-compliance [4]. While these approaches contribute to improving compliance, they are largely manual, resource-intensive, and reactive. Their implementation requires significant human effort, time, and financial resources, and they often detect irregularities only after transactions have already occurred. Consequently, these methods are insufficient in fully addressing evolving forms of tax evasion.

Numerous studies have proposed technological solutions to enhance tax compliance and combat fraud. For example, SMS-based notification systems have been developed to improve transparency in tax reporting by notifying authorities after transactions occur [5]. Mobile-based applications have also been introduced to verify the authenticity of EFD receipts and reduce cases of receipt forgery [6]. Other approaches include integrating stock tracking mechanisms within EFD systems to validate sales against inventory movement [7], as well as applying machine learning techniques to detect anomalies in submitted tax reports [8]. In addition, advanced frameworks incorporating Machine Learning (AI), Robotic Process Automation (RPA), and interconnected financial systems have been suggested to identify inconsistencies in tax-related data and improve fraud detection processes [9]–[11].

Despite these advancements, existing solutions exhibit several limitations. Many approaches focus on post-transaction analysis, meaning fraud is detected only after reports have been submitted, which delays corrective action. Others concentrate on verifying the authenticity of receipts without validating the correctness of transaction values, leaving underpricing fraud undetected. Furthermore, machine learning-based approaches often rely heavily on historical data and lack the capability to operate in real time during transaction processing. As a result, these systems are unable to prevent fraud at the point of sale. These limitations contribute to persistent challenges in Tanzania, including underreporting of sales, manipulation of transaction data, and declining or fluctuating VAT contributions to GDP as it contribution is less than 3% which is lower than East African countries [12] , dropping of VAT eligibles as decreased by 38.6%, dropping from 73,882 to 45,363 in the 2020/2021 and also decrease by 35.3% in 2022/2023 financial year [13] and fluctuation of increase of VAT collection [14].

While previous studies have contributed to tax fraud detection, most existing approaches focus on post-transaction analysis or static rule-based systems, limiting their effectiveness in detecting underpricing fraud during the transaction process. In particular, these solutions provide limited support for real-time validation and immediate feedback within Electronic Fiscal Device (EFD) environments, and they often underutilize contextual and product-specific attributes such as location, specifications, and historical pricing patterns. Consequently, there remains a gap in enabling proactive, data-driven detection of underpricing at the point of sale. To address this limitation, this study proposes a machine learning–based approach integrated into the EFD framework to support real-time price validation and alerting during transaction processing. By analyzing transaction data as it is entered and leveraging learned patterns from historical data, the proposed system provides immediate feedback on suspicious pricing, thereby enhancing tax compliance, reducing opportunities for manipulation, and improving the efficiency of tax administration in Tanzania.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

2. METHODS

2.1. Research Design

The study adopted the framework of Design Science Research (DSR) methodology. This DSR methodology was chosen because it enables a holistic consideration of factors relating to the problem, the design of a solution (model), and evaluation of the model. Figure 1 below depicts the schematic of a research design and flow. The implementation of the DSR followed six main stages. The first involved problem identification which was done through review of literature and various reports including reports of the Office of the Control and Auditor General (CAG) which indicated the severity of the problem of low taxpayers' compliance in using EFD receipts and associated underpricing of invoices. This stage included also validation of practices and utilization of the system to identify gaps in current Electronic Fiscal Device Systems (EFDS). Completion of this stage enabled identification of key challenges identified included lack of real-time validation, limited intelligence in detecting price anomalies, and vulnerability to intentional underpricing.

The second stage involved data collection both qualitative and quantitative. Quantitative data were gathered from traders using questionnaires and qualitative interviews administered to tax officers. Quantitative data included sales data. These methods of data collection enables gather detailed insight on the intensity of underpricing fraud in Tanzania. After preparation of the dataset, the third stage involved preparation of functional and non-functional requirements, including real-time price validation, high detection accuracy, low false-positive rates, and the ability to handle both categorical and numerical product attributes such as product name, brand, model, RAM, storage, location, and unit price. This was concluded by designing objectives of the system based of system requirements.

The fourth stage was system development. System development was conducted iteratively using the Scrum framework, chosen for its flexibility and suitability in handling evolving requirements in machine learning-based applications. Development was organized into multiple sprints, each focusing on specific components such as data preprocessing, feature engineering, model training, and mobile application integration. Sprint reviews were guided by predefined acceptance criteria, including correctness of data processing, model performance, and usability of the mobile application interface.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

The fifth stage was system evaluation based on pre-determined criteria applied at each iteration. Four iterations were performed. In the first iteration, acceptance was based on successful mapping of research objectives to system requirements and dataset features without inconsistencies. The second iteration was accepted when data ingestion, preprocessing, and feature engineering processes produced consistent and accurate outputs across multiple runs. The third iteration focused on machine learning model development, where acceptance required successful training, validation, and generation of evaluation metrics such as accuracy, precision, recall, and F1-score using unseen data. In the final iteration, the system was evaluated based on its integration into a mobile application, ensuring that real-time predictions, user inputs, and system outputs functioned correctly and aligned with expected behaviour. The last stage was system integration and User Acceptance Testing (UAT). Preliminary testing was conducted with potential users to assess usability and practicality of the EFD system.

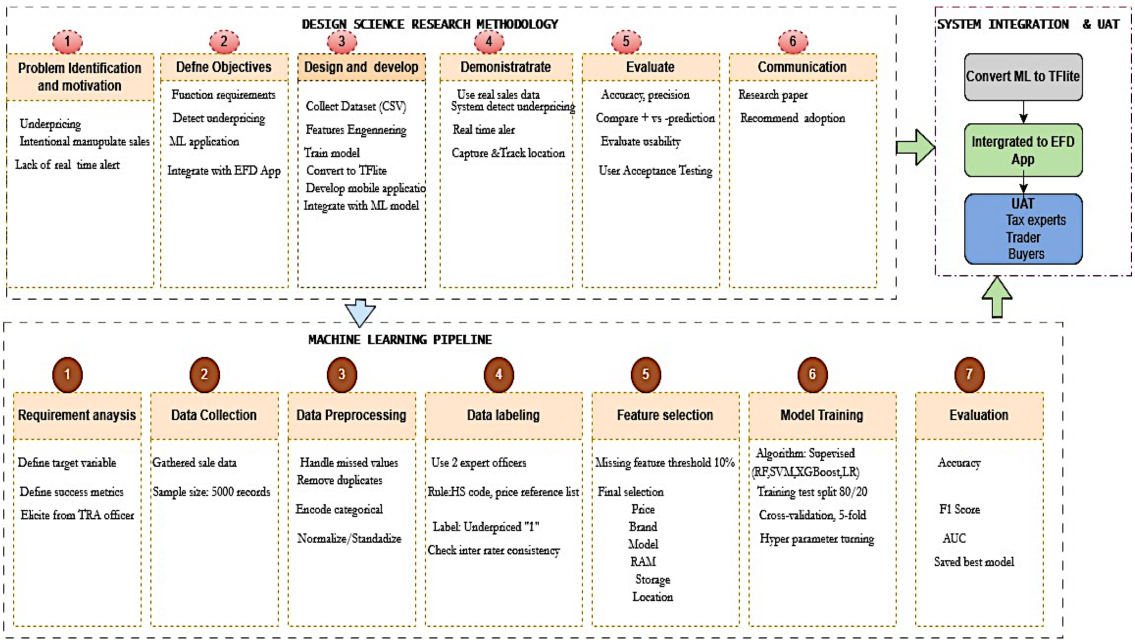


Figure 1. Research design and Flow

2.2. Participants and data collection

Participants were solicited from both public sector entities and the commercial sector. Participants from the public sector were the officials from the Tanzania Revenue Authority (TRA), who hold responsibilities for tax administration and the oversight of EFD systems, while for commercial sector participants were mobile phone retailers, who engage directly in pricing strategies and tax compliance at the operational level. the use of tax officers as participants enables elicit of information regarding regulatory enforcement, while traders were essential to reveal the actual business practices. These participants were sourced from two distinct regions; Dar es Salaam region in the East of Tanzania (tax officials at TRA headquarters) and Arusha in the North (mobile phones traders).

Primary data were collecting using semi-structured and in-person interviews. with six TRA officials (two system administrators and four representatives from the EFD department). Before data collection at TRA, consent was sought and approval granted through formal permission via official email in order to abide to ethical considerations of research involving human subjects as well as to ensure authenticity of data and the information provided. Purposive sampling based on their direct expertise in tax operations was used to select TRA officials for the interviews. These interviews were conducted through an open-ended questionnaire.

In addition, the study gathered data from a sample of 78 randomly selected mobile phone traders and 200 randomly sampled buyers in the Arusha region using a structured questionnaire administered through Google Forms. Ethical consideration was also ensured through the use of digital consent filter questions to ensure that participants are only those who explicitly agreed to the study's terms. Apart from qualitative data, a total of 5,000 sales records were also collected from the traders. Other data collected from the traders include multiple features such as Brand, Model, Storage, RAM, location, unit_price, underpriced, Quantity, Discount_Flag, Discount_Percent, Payment_Method, Stock_Level, Previous Stock, Tax_Rate, and Customs_Value.

2.3. Data Analysis Techniques

Data analysis involved qualitative analysis and descriptive analysis. Qualitative analysis of interview responses and text literature was performed using Atlas. ti version 9.2 software. This analysis enabled identification of key themes and patterns. Descriptive data analysis of the structured questionnaires was performed using SPSS version 29. The analysis involved computation of descriptive statistics, frequency distribution and statistical charts and graphs. This analysis facilitated summarization of the findings and identify trends, hence provided a detailed understanding of the mechanisms and challenges of VAT compliance [15]. Analysis and detection of tax evasion (underpricing) was based on comparison of the variability of the feature data from one trader to another using price database and profit margin analysis using harmonized system codes. Thus, differing invoice prices for same features indicates potential evasion. Furthermore, all preprocessing and modeling steps were implemented using Python (pandas and scikit-learn) to ensure reproducibility.

2.4. Data preparation and feature Engineering

Data preparation involved cleaning and transformation of the 5,000 sales records collected from various shops selling mobile phones. Data processing and feature selection procedure was applied to prepare the dataset for machine learning modeling. Features with more than 40% missing values were removed to ensure data quality and reduce bias. These included payment method, stock level, discount flag, discount percentage, custom ratio, and previous stock. For the remaining features with missing values below the threshold, regression imputation using a linear regression model was applied to estimate missing entries based on other available variables. Duplicate records were removed using pandas' drop duplicates() function, and data cleaning was performed to ensure consistency. The final dataset consisted of six (6) features: unit_price, brand, model, RAM, storage capacity, and location. These were selected based on data availability, preprocessing results, and domain relevance to mobile phone pricing in the Tanzanian context. As shown in Table 1, the most common feature customer asking are those documented by traders, except for those using computerized systems. Where the custom ratio is zero (0), meaning it is not included in retail sales, as no documentation for that. Features engineering helps to improve model accuracy, reduces noise, improves generalization, and helps algorithms learn more efficiently [16].

Table 1. Feature contributing to tax computation

Feature	Available data	Feature	Available data
1. Brand	5000	8. Location	5000
2. Model	5000	9. Unit Price	5000
3. Storage	5000	10. Quantity	5000
4. RAM	5000	11. Discount Flag	14

Feature	Available data	Feature	Available data
5. Location	5000	12. Discount percent	14
6. Payment method	50	13. Custom ratio	0
7. Stock level	89	14. Previous Stock	89

Further, the target variable, "underpriced," was labeled in collaboration with tax experts based on established assessment practices, using (i) HS code classification to align products with standard import valuation categories and (ii) a reference price list allowing up to 20% variation for discounts and promotions. Where labelling by the two independent tax experts could no tally, disagreements were resolved through consensus. Following that, transactions below acceptable price thresholds were labeled as "underpriced." Feature importance analysis presented in Figure 2 show that unit_price, model, and location contributed most to underpricing prediction, while storage capacity had relatively low influence.

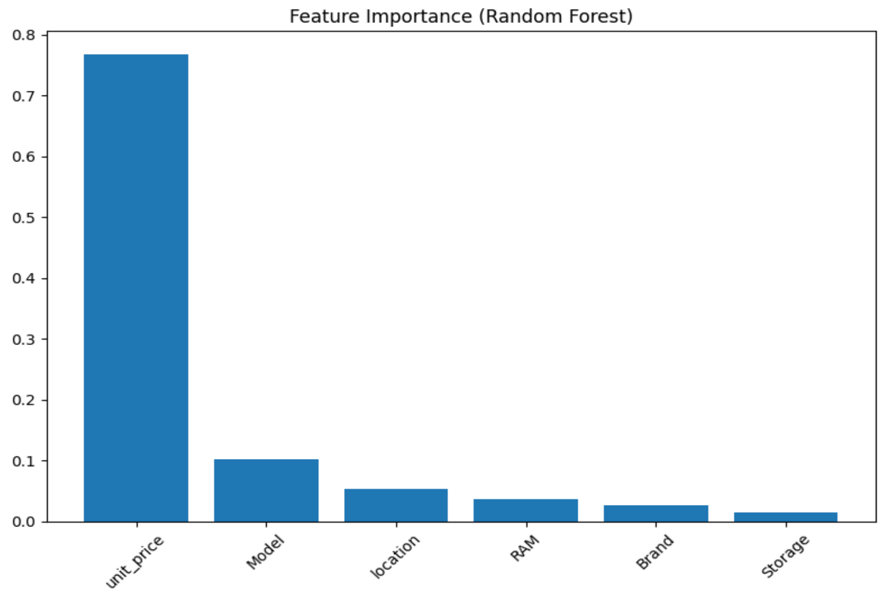


Figure 2. Features importance

2.5. Model Development Method

The model development process was formulated as a supervised machine learning classification task aimed at detecting underpriced transactions within Electronic Fiscal Device (EFD) systems, where each transaction was classified as either "underpriced" or "normal" based on historical patterns and expert-defined criteria. The dataset was split

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

into 80 % training and 20% testing sets, and a 5-fold cross-validation approach was applied to enhance model reliability and reduce overfitting. Several supervised machine learning algorithms, including Logistic Regression, Support Vector Machine (SVM), XGBoost, and Random Forest, were trained using the same feature set to ensure fair comparison, and hyperparameter tuning using grid search was performed to optimize model performance. Evaluation was conducted using accuracy, precision, recall, and F1-score.

2.6 System Development Methodology

Scrum methodology was applied in the development process, emphasizing iterative and incremental delivery. Agile (Scrum) software development was chosen because it allows flexibility in handling changing requirements throughout the project [17]–[19]. The process began with requirements gathering, where user needs were collected from key stakeholders, including traders and tax officers, and prioritized in the product backlog. Development was carried out in short iterations known as sprints, during which a simple user interface and system features were designed, implemented, and tested. At the end of each sprint, a working prototype was presented to stakeholders during sprint review meetings. Stakeholders evaluated the system and provided feedback, which was used to refine and reprioritize requirements in the backlog. This ensured continuous improvement and alignment with user needs. Additionally, user acceptance testing (UAT) was conducted iteratively, allowing stakeholders to validate system functionality and suggest enhancements. Through continuous collaboration between stakeholders and the development team, the system evolved incrementally, in order to ensure that the final product effectively addressed user requirements.

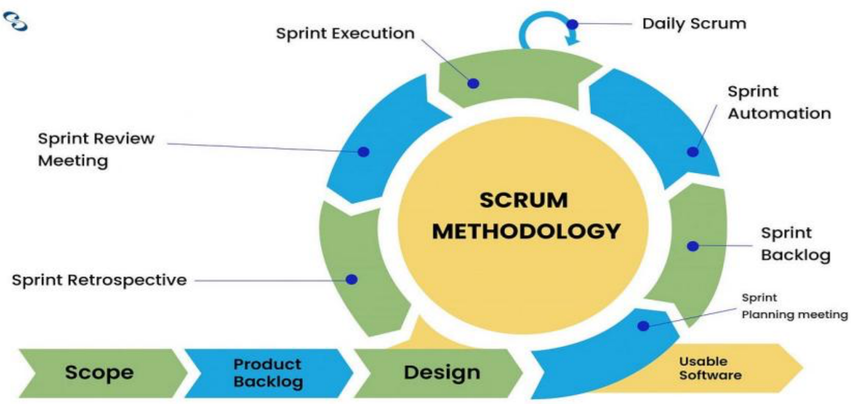


Figure 3. Scrum System development methodology

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

2.6. System Integration

The selected Random Forest model was converted and deployed using TensorFlow Lite (TFLite) and integrated directly into the mobile-based Electronic Fiscal Device (EFD) application to enable real-time, on-device detection of underpricing fraud. Transaction data entered at the point of sale is preprocessed within the application and passed to the embedded TFLite model for inference, which classifies each transaction as either "underpriced" or "normal." This on-device integration ensures low latency, fast response time, and reliable operation even in environments with limited internet connectivity.

After classification, transaction data together with the model output is stored in Firebase, where it is managed and prepared for further use. An Application Programming Interface (API) is then used to fetch and synchronize the stored data from Firebase to the Electronic Fiscal Device Management System (EFDMS). This integration enables tax authorities to access detailed transaction records, including flagged underpricing cases, through a centralized platform. Consequently, tax officers can monitor, analyze, and take appropriate actions based on timely and data-driven insights, thereby enhancing tax compliance and enforcement efficiency Figure 4 components are integrated where a mobile application embedded with machine Learning using Tflite captures inputs, including location data, and submits to the Firebase database so that it can be retrieved by EFMS via API.

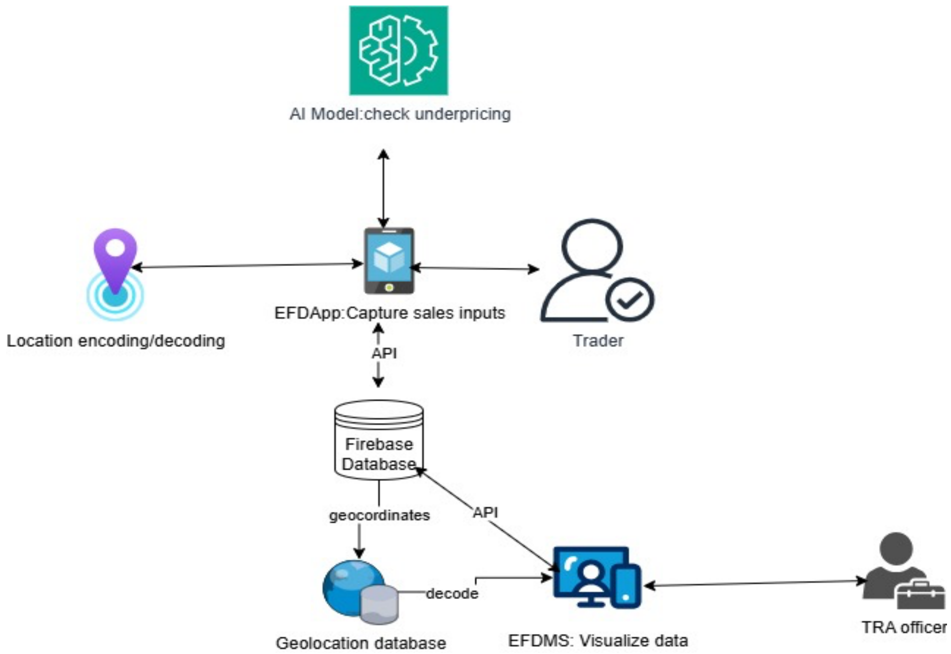


Figure 4. System architecture

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

The system functions based on the flowchart outlined in Figure 5 Below that, it requires the user to turn on device location before proceeding with the transaction, but also validates the data inputs, and at the end, if detected underpricing, it alert trader to cancel the transaction or to proceed with the transaction. The detection of underpricing fraud in real time reduces auditing costs, inspection costs, and the overall manpower required to ensure tax compliance.

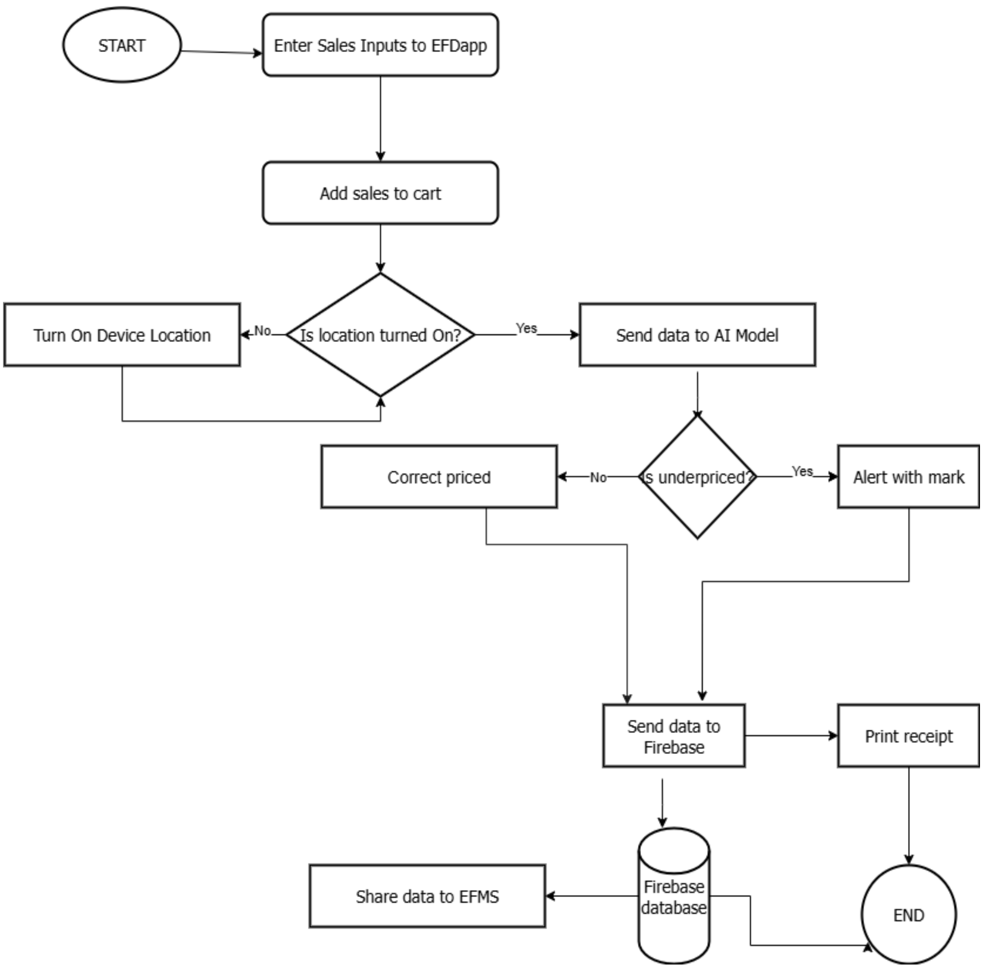


Figure 5. System Data Flow Diagram

2.7 User Acceptance Testing Methodology

User Acceptance Testing (UAT) was conducted to evaluate the usability, functionality, and effectiveness of the developed system in a real-world context. The testing involved selected end users, including 10 traders, 4 buyers, and 10 Tanzania Revenue Authority (TRA) officers, who interacted with the mobile-based Electronic Fiscal Device (EFD) application during transaction processing. The objective was to assess whether the system meets user requirements, particularly in detecting underpricing fraud in real time and providing appropriate feedback to both sellers and buyers.

A structured questionnaire was used to collect user feedback after interacting with the system. The questionnaire focused on key aspects such as system usability, ease of use, accuracy of fraud detection, response time, and overall user satisfaction. Participants performed transaction-related tasks and then responded based on their experience. The collected data were analyzed to evaluate system acceptance and identify areas for improvement.

2.7 Limitations of the study

The study used traders from only one region. This may limit generalizability to alternative regions owing to disparities in market dynamics and enforcement rigor. As such. subsequent investigations may broaden the framework to encompass other regions or nations to substantiate its resilience across varied conditions.

3. RESULTS AND DISCUSSION

3.1. Descriptive Analysis

As shown in Figure *6*, the findings revealed that 62.8% of the respondents were male, while 37.2% were female, highlighting that males are more involved in the mobile phone selling business compared to females. This suggests a higher likelihood of males engaging in fraudulent activities, as indicated by the study of Wang et al. (2022). The findings suggest that women in corporate leadership roles are generally more risk-averse and exhibit a stronger commitment to ethical practices compared to their male counterparts. Additionally, the age distribution of the respondents shows that 62 participants (79.5%) were under the age of 45, an age group often associated with the development of criminal careers, indicating a potential link between age and engagement in unethical or fraudulent activities. (Basto-Pereira & Farrington, 2020)

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

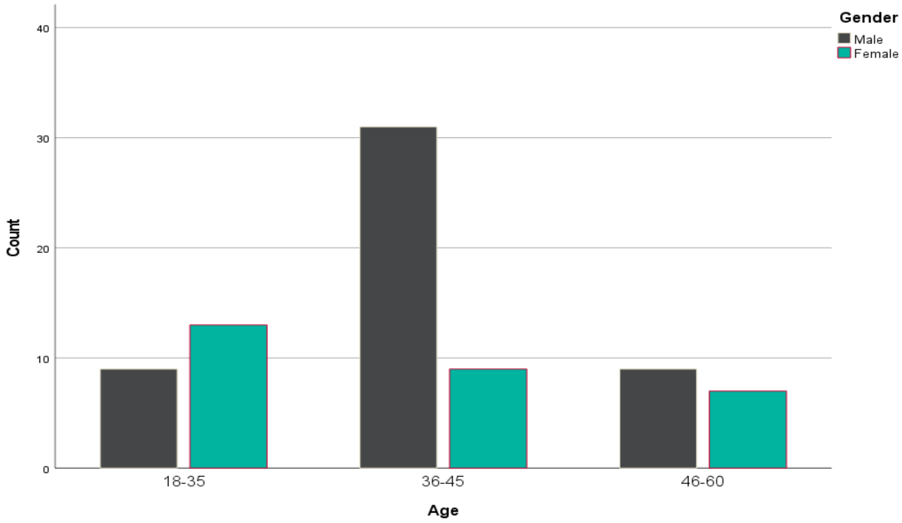


Figure 6. Traders age and gender distribution

Among the respondents, a quantitative measure was used to determine the methods they use to report tax information to the tax authority (Tanzania Revenue Authority). From **Error! Reference source not found.**, The results showed that 75.6% use the yearly tax estimate method, 10% use the EFD system, and 10.3% use the E-filing system and monthly returns. This indicates that only a small percentage of traders allow the tax authority to track their sales in real time. Tax collection is not based on actual income generated but rather on estimations, which creates opportunities for traders to underreport their sales. This is due to the absence of mechanisms capable of capturing and validating sales income, particularly in terms of specified prices, despite the method proposed by [7] can only detect receipt not issued, but the correctness of the amount stated in the receipt it will be impossible, together with [6] suggested to scan the receipt to check the validity, but it can fail to identify the relationship between the price and the product stated. As a study of [20] It was insisted that Artificial Intelligence be used to create a decision-making system that can detect tax evasion fraud as soon as it happens, but AI mechanism developed by [8] can only detect fraud from submitted reports, that create the necessity of have real time detection of fraud.

Table 2. Tax information submitting methods

Reporting method	Frequency	Percentage
EFD	10	12.8%

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Reporting method	Frequency	Percentage
E-Filing and Monthly	8	10.3%
Yearly tax estimate	59	75.6
Not paying Tax at all	1	1.3%
Total	**78**	**100%**

Most traders are aware that there are techniques they use to evade tax, such as pricing strategies, where they underprice products to reduce tax liability. In fact, 76% of the respondents agreed that underpricing is one of the techniques used in tax evasion. Additionally, it was found that the Tanzania Revenue Authority (TRA) very rarely conducts investigations to check for tax compliance among traders, either by visiting their premises or using other sophisticated mechanisms. The majority of respondents 81 % indicated that TRA officers visit business premises only once yearly, as shown in Figure 7, this indicates that while authorities regulation require premises visit once monthly ensure compliance through visits to business premises, detecting tax evasion through underpricing becomes more challenging, as only 5% of the trader visited monthly and 13% never visited. According to [21] Argues that face-to-face supervision can be ineffective due to resource constraints and trader manipulation. Regardless of study of [5] guaranteed the online submission of the data from EFD machine to tax authority web system (EFDMS) but EFD have no mechanisms to detect the price invalidity and [22] proposed to detect fraudulent of submitted data using AI, but cannot check price validity and no instant alert to tax authority with location to ensure focused auditing/visiting. This demands the implementation AI with the current used online data submission to ensure focused auditing/visiting.

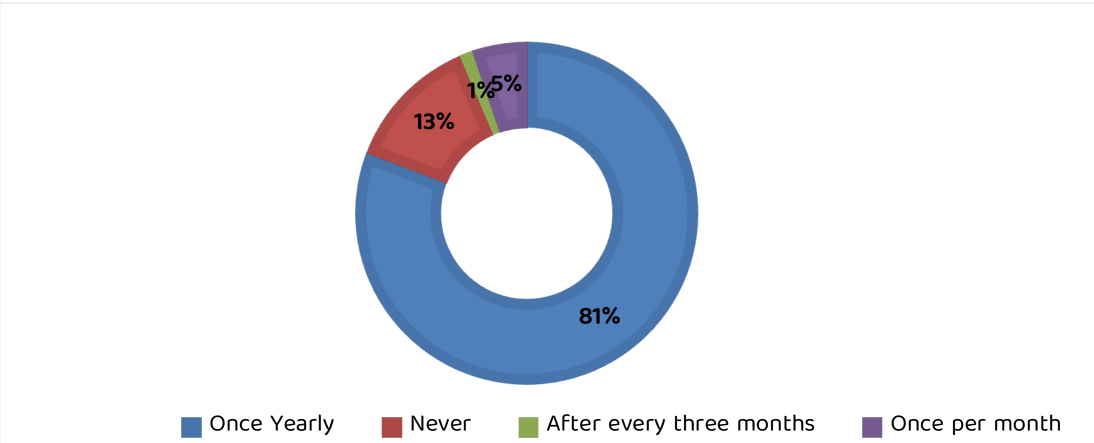


Figure 7. TRA Visit business premises

Additionally, the questionnaire was distributed among buyers to gain a broader understanding of how traders might exploit the weaknesses in the tax reporting system particularly given the infrequency of visits by tax officers to their business locations, as pointed out by [23] customers, being central to the market as key stakeholders, should be actively involved in strategies to combat tax evasion in order to achieve more effective and favourable outcomes. As indicated in Figure *8* customers reported experiencing price discrepancies on receipts whenever official receipts were issued, with 73% regularly receiving receipts showing different prices for the same product, and 19% received correctly priced receipts. So, according to the study of [6]–[9], [11], [24] Regardless of ensuring the issuing of a receipt, checking the validity of the receipt by scanning, using machine learning to check the fraudulent nature of the transaction, or linking systems to check discrepancies between transactions using machine learning, it will not help because the receipt will be issued from a valid machine, but to validate the correctness of the price stated in that receipt, you need to take the relationship between the features of the product and the stated price, which is not implemented by any of the researchers.

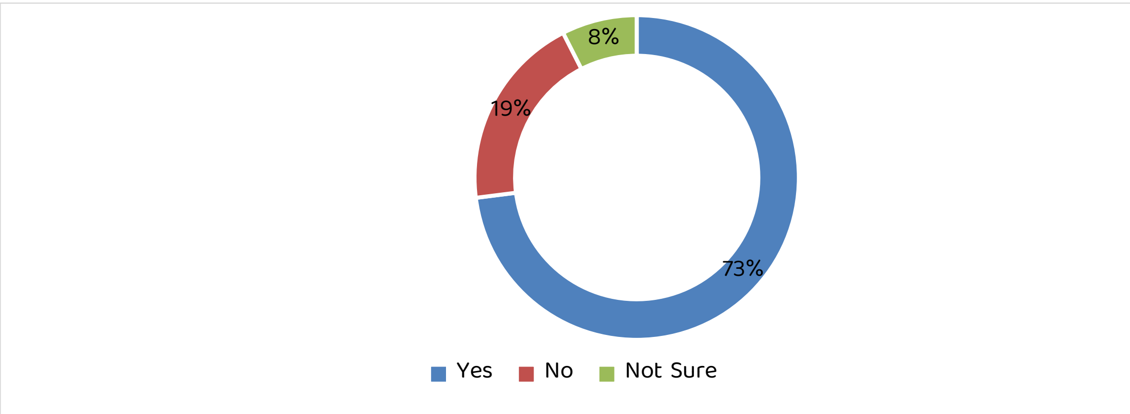


Figure 8. Issue receipt with price discrepancies

Furthermore, customers are often encouraged by business owners to accept receipts showing a lower price than what was actually paid, under the promise of receiving a discount on the product. According to the data presented in Table 3, 78s% of customers confirmed that they were urged by traders to accept such receipts, which ultimately results in underreporting of sales and a reduction in the trader's tax liability. Due to lack of constraints of the current EFD trader can write any price on the product, where the previous researches of [8], [9] Waiting for the data input from the EFD/EBM to process in machine learning to detect tax fraud, it will change to "garbage in, garbage out." Most important is to check the validity of data at the input stage and that the proposed solution performs.

Table 3. The business owner encourages underpricing in the receipt

Responses	Frequency	Percentage
Yes	156	78%
No	44	22%
Total	200	100%

It was also revealed that some traders deliberately lower the price of a product if the customer does not request a receipt, as a tactic to conceal sales and reduce their tax liability, given that tax is calculated based on reported sales. According to the findings, 58% of customers indicated that this practice occurs frequently, with traders offering lower prices when receipts are not requested; meanwhile, 29% of respondents reported that this situation happens occasionally, as shown in Figure 9 (b), In this case this study will be efficient once working together with the study by [7], stock tracking is embedded, and to ensure the issuing of the receipt, machine learning will work to validate the accuracy of the price stated upon processing of the transaction via EFD. Otherwise, this study will be limited to performing functionality designed for the receipt if it is not issued.

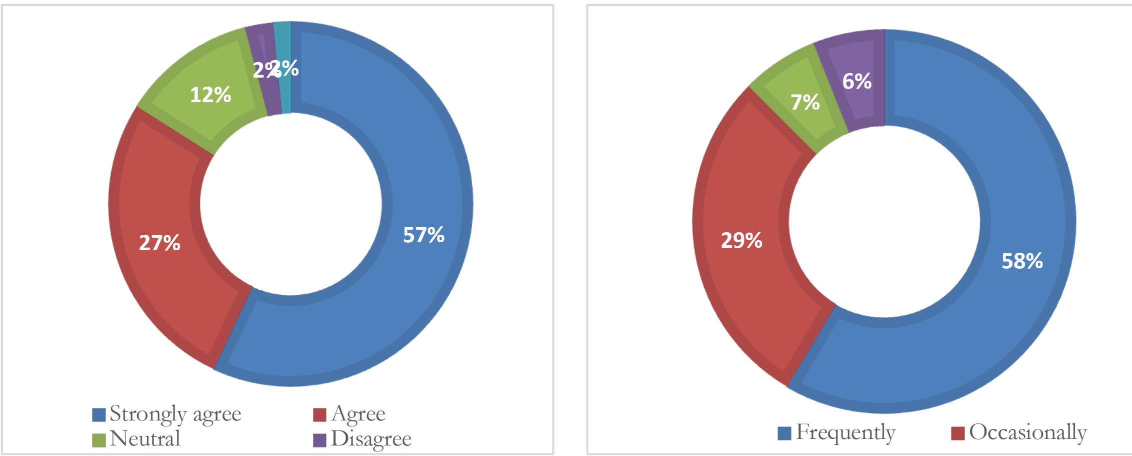
This indicates that customers are aware that the underpricing of products is intentionally done by traders to minimize their tax obligations to the tax authority. Respondents are aware that traders deliberately record a lower price on the EFD receipt to reduce the amount of tax payable, an act that constitutes one of the recognized EFD offences as described by [25] involves presenting false information on the receipt, which highlights the need for sophisticated technology to effectively detect and address such fraudulent practices.

There is also a common practice among business owners to issue the receipts that show a lower price than the actual amount paid in order to reduce their tax liability. About 57%

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

of the 200 respondents as shown in Figure 9(a) confirmed that traders often issue receipts indicating a lower price than what was truly paid. These receipts are mainly used as a precautionary measure to avoid penalties from tax compliance task forces, which punish transactions conducted without official receipts. However, it remains difficult to verify the accuracy of the amounts stated. According to [26], Firms often engage in tax evasion by generating fake or fraudulent receipts that understate the actual sales value of products.

Moreover, customers have observed inconsistencies between the prices indicated on receipts and the actual amounts they paid. According to [27], instances where products with identical features and characteristics are recorded in the system at different prices, leading to discrepancies that affect sales tax liabilities. Therefore, there is a need for an automated system capable of standardizing price ranges for products with similar



features and ensuring that the same amount is accurately reflected on the generated receipts. This would help guarantee the correct computation of Value Added Tax (VAT), which is 18% of the sales price in Tanzania.

(a) (b)

Figure 9. Customer aware trader underprices products to avoid tax liability

3.2. Qualitative Data Analysis Results

To gain a detailed understanding of tax issues, methods used to ensure compliance, and existing loopholes, seven (8% of the respondents) TRA officers from different sections were interviewed. These included system administrators (2 respondents), officers from the EFD department (4 respondents), and the head of the EFD department from the TRA

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

headquarters. The one-on-one interview method was used due to the respondents' availability. As suggested by [28], video, telephone, and online interviews are valid and trustworthy alternatives to traditional face-to-face interviews. The data collected were analyzed using Atlas.ti version 9. Additionally, the document analysis method was adopted, where various sources such as videos, magazines, government reports, images, and online documents were analyzed using the same software. The following themes were extracted and their corresponding network diagram.

1) Lack automation enables tax evasion.

According to [29] implementation of Automatic Exchange of Information (AEoI) is effective in minimizing tax evasion. **Respondent4** said that "manual audit it play the crucial role in controlling tax evasion…", **Respondent2** added that "…it need much involvement of many manpower and time consuming like having special task force", **Respondent5** '….is challenge due to broad product description, that a make difficult to detect underpriced that "Advanced data analytics and AI can help standardize pricing data". The Figure 10 below shows the current mechanisms used to ensure compliance especially to under declaration of sales, which most of them are manual techniques characterized by too much resource demanding. Previous researcher [5], [6], [8] suggested scanning of receipts to check for validity and processing submitted reports to detect understated sales fraud. Both studies are reactive, not proactive, which are still resource-demanding and are not focused. Real-time detection and alerts with integrated location will be more efficient.

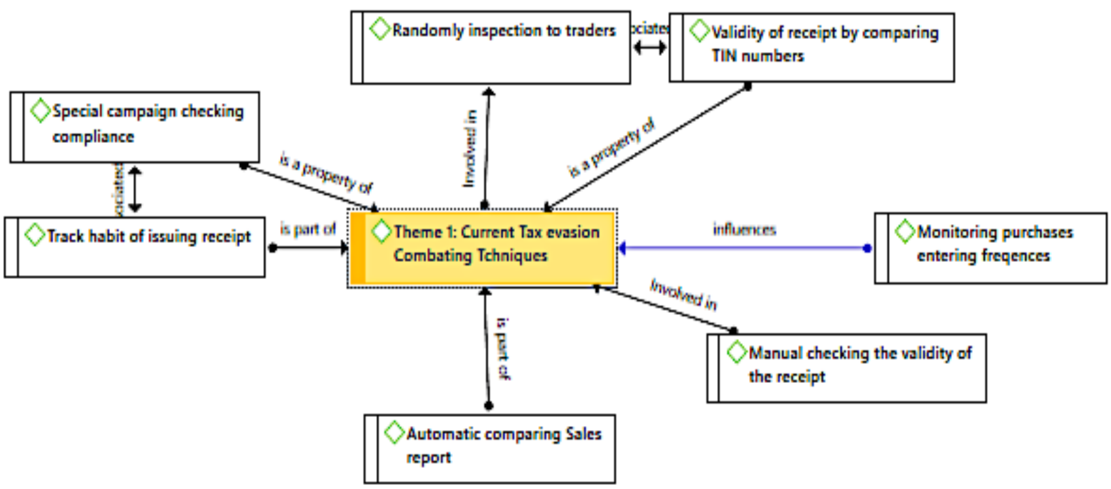


Figure 10. Techniques used to check tax compliance

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

2) Use of sophisticated application with Real time alert.

[9] It is proposed that using Artificial Intelligence can help combat tax evasion involving complex network schemes**. Respondent1** said that "AI can replace the human being activity, may be if possible we can have robots that will be doing inspection to traders instead of us….", **respondent2** added that "Additionally, lack of real time monitoring and immediately feedback…"," sophisticated analytics tools can detect patterns that indicate potential fraud," **respondent4** argued that "Implementing Application Programming Interfaces (APIs) facilitates real-time data integration and synchronization", respondent3 added. The Figure 11 below describes the proposed method that the respondents TRA tax officers believe may be effective in combating the tax evasion especially underpricing fraud, where the use of sophisticated application suggested by the most of the respondents. Meaning leveraging the current technologies is very important to optimize utilization of the resources. Regardless of [9]–[11] integrated the sophisticated application, which is AI, in detecting fraud, still failed to detect underpricing in real time, and sent no alert send to the tax authority in real time, solely dealing with analyzing submitted reports. That foster thinking on implementing the real-time application to detect understated sales.

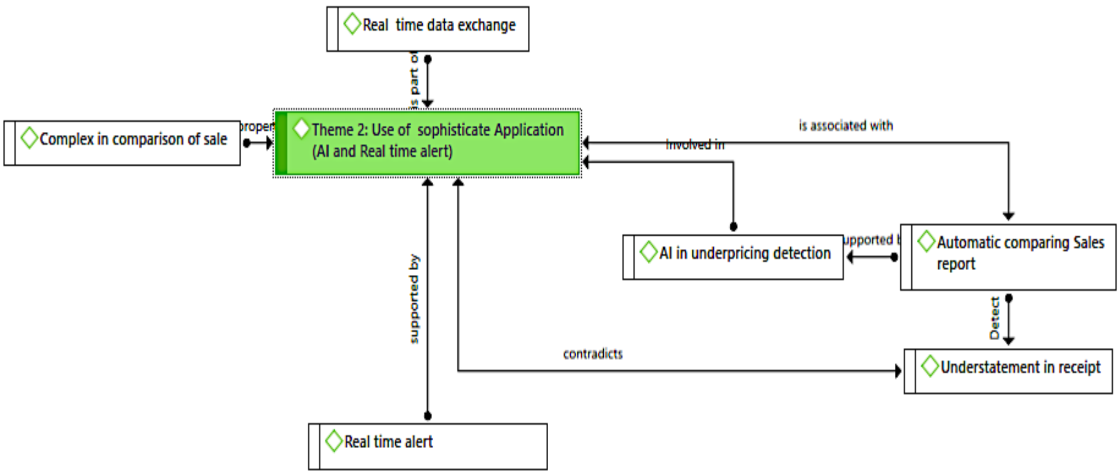


Figure 11. Proposed techniques theme to be implemented

The proposed study aimed to enhance revenue collection by minimizing tax evasion through the understatement of sales caused by underpricing. The system verifies the validity of the stated price against the product's features to detect fraudulent activity at an early stage. When underpricing is suspected, an automated alert containing the seller's location is sent to the tax authority to support focused auditing. This approach promotes compliance even when resources such as time, personnel, and finances are limited. Through regulatory updates and staff training, TRA can operate the system efficiently while ensuring that only tax-related data is processed. All machine-learning analysis should comply with Personal Data Protection Commission (PDPC) and Global Data Protection Alliance (GDPA) requirements, anonymizing data with high data minimization and avoiding personal identifiers to protect taxpayers and customers' privacy. However, the study is constrained by the use of historical sales data, while product prices and specifications frequently change. Since the system relies on EFD-issued receipts, continuous real-time data mining and periodic AI model retraining are necessary to maintain high accuracy in price validation.

3.3. Tool Development

The data collected consisted of phone sales with additional features recommended by experts (tax officers), as these features are considered important for detecting underpriced products. These features include product specifications, the corresponding sale price, and the actual perception of whether the product is underpriced [30], as shown in the Figure 12. Other potential features, such as discount rate allowed, current stock, previous stock, and tax compliance score history, were largely missing and therefore excluded during data preporocessing. The data preprocessing involved examining the correlation of features with the target variable, encoding categorical data, and standardizing numerical features through scaling.

The correlation heatmap shows mild to moderate correlations among numerical features, with higher phone specifications linked to higher prices. Categorical model and location variables also display slight relationships with pricing, reflecting model and market variations. The underpriced label shows weak correlations, indicating it depends on multiple combined factors. Overall, no strong multicollinearity exists, confirming the features are suitable for machine learning models.

Although several factors such as RAM, battery power, pixel dimensions, screen size, touch screen capability, and Wi-Fi availability can influence a phone's price, these features are generally not considered significant determinants of phone pricing in the context of Tanzania.The analysis revealed that specific features influence the price range of mobile

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

phones, with the price range designated as the dependent variable and the phone features as independent variables. As [31] demonstrated, machine learning models can capture the impact of independent variables on the dependent variable which is underpriced.

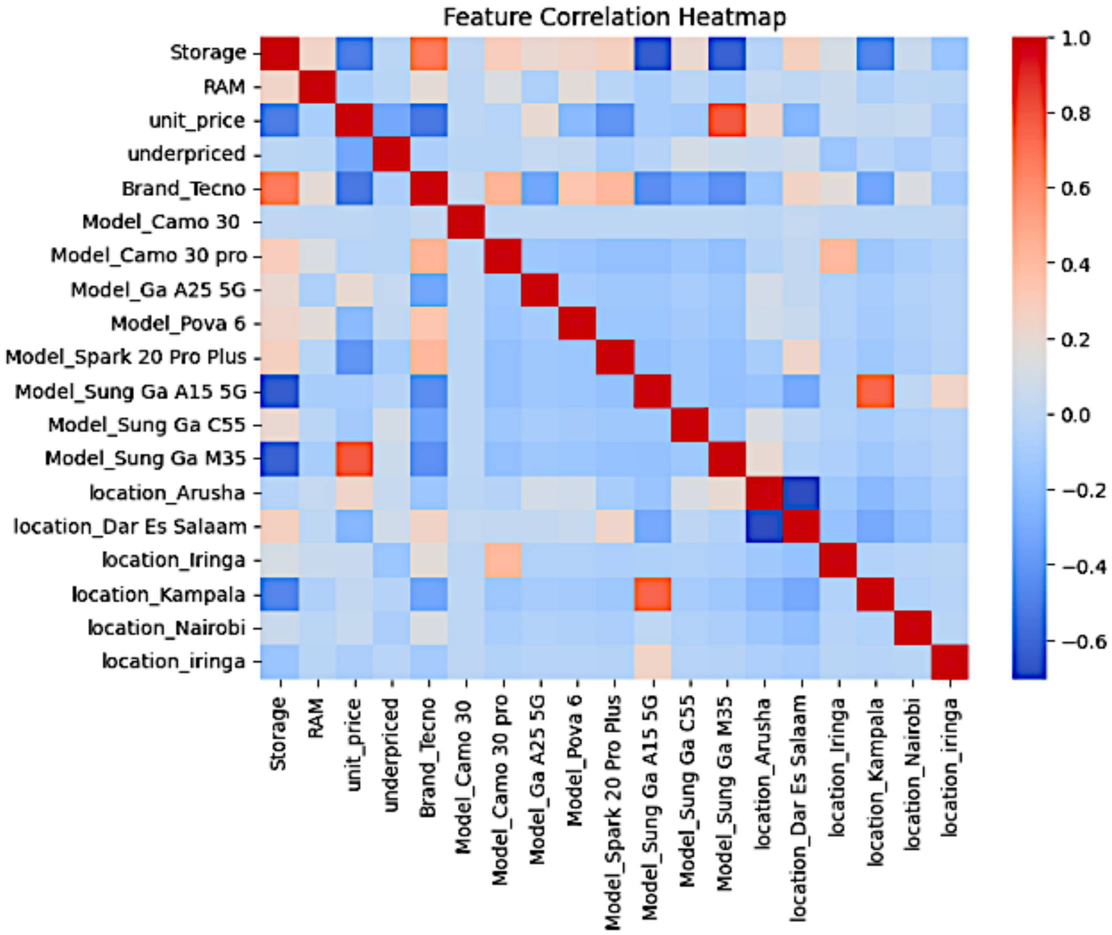


Figure 12. Features correlation

The dataset was further preprocessed using scikit learn and pandas so to clean and encode data so as to fit the purpose and to optimize it performance. A new column, *underpriced*, was introduced to serve as the new dependent variable. The preprocessed dataset allowed for the evaluation of feature importance, highlighting the contribution of each feature to underpricing prediction, as illustrated in the **Error! Reference source not found.** below, features such *as unit_price, Model, location, RAM, Brand and storage* while some features are among the less important factors influencing the model's decision to predict whether a specific sale transaction involves underpricing especially in Tanzania, according to the study [32], feature selection plays a crucial role in datasets with a large number of variables and features. Eliminating irrelevant variables helps improve both the accuracy and effectiveness of the classification process. So, six best features selected to optimize the efficiency of the model.

Then, different machine learning models were then developed, including Logistic Regression (LR), Support Vector Machine (SVM), XGBoost, and Random Forest, trained on a clean dataset of 5,000 sales records. The models were subsequently tested on unseen data to evaluate their performance. The performance comparison, as shown in Table 4, indicates that Random Forest outperformed the other models in terms of training and testing accuracy. Consequently, the Random Forest model was selected for deployment in the EFD application for real-time underpricing detection.

The development of the sales prediction model follows a structured multi-stage machine learning pipeline, as illustrated in Figure 1. The process begins with Dataset Collection, involving the acquisition of 5,000 historical sales records. To ensure data integrity, a Data Cleaning phase is executed to handle missing values and noise, followed by Data Transformation, where categorical variables are converted into numerical formats through encoding techniques to facilitate algorithmic processing.

In the analytical phase, Feature Selection is performed to identify and retain only the most significant variables, thereby reducing dimensionality and computational overhead. The refined dataset is then subjected to a Train/Test Data Split 80/20% to provide a dedicated portion of data for model validation. For the core predictive task, the Random Forest algorithm was selected during the Model Selection stage due to its robustness in handling non-linear relationships within sales data to with the ability of handling both numerical and categorical data. The pipeline concludes with model integration to EFD application, where the optimized model is integrated into a Mobile Application environment for real-time inference and end-user accessibility.

The model's performance was validated using 5-fold cross-validation, where the dataset was divided into five parts. In each round, four folds were used for training while one-fold was used for testing, ensuring every sample was evaluated once. This method provides a reliable estimate of model performance by reducing random bias and variance.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

The Random Forest model showed consistent results across all folds, achieving an average accuracy of 99.6% as shown in Table 4, confirming its strong generalization ability and suitability for detecting underpricing in real-time for the dataset comprising numerical and categorical data.

Table 4. Model performance comparison

Model	Accuracy	Precision	Recall	F1	ROC_AUC
Random Forest	0.9962	0.9933	0.9969	0.9951	0.9982
XGBoost	0.9936	0.9902	0.9932	0.9917	0.9987
SVM	0.9930	0.9922	0.9896	0.9909	0.9935
Logistic Regression	0.9850	0.9904	0.9703	0.9803	0.9910

In this study, the Random Forest (RF) algorithm was implemented to detect underpricing in mobile phone sales data. The model achieved an accuracy of 99.6% and an AUC score of 0.998 as shown in Figure 13 indicating excellent performance in distinguishing underpriced devices. Owing to its robustness and high predictive capability, the Random Forest model was selected for deployment. The trained model was then converted into TensorFlow Lite (TFLite) format and integrated into the EFD application, a mobile application developed using Flutter, to enable real-time detection of underpricing in mobile phone sales [33].

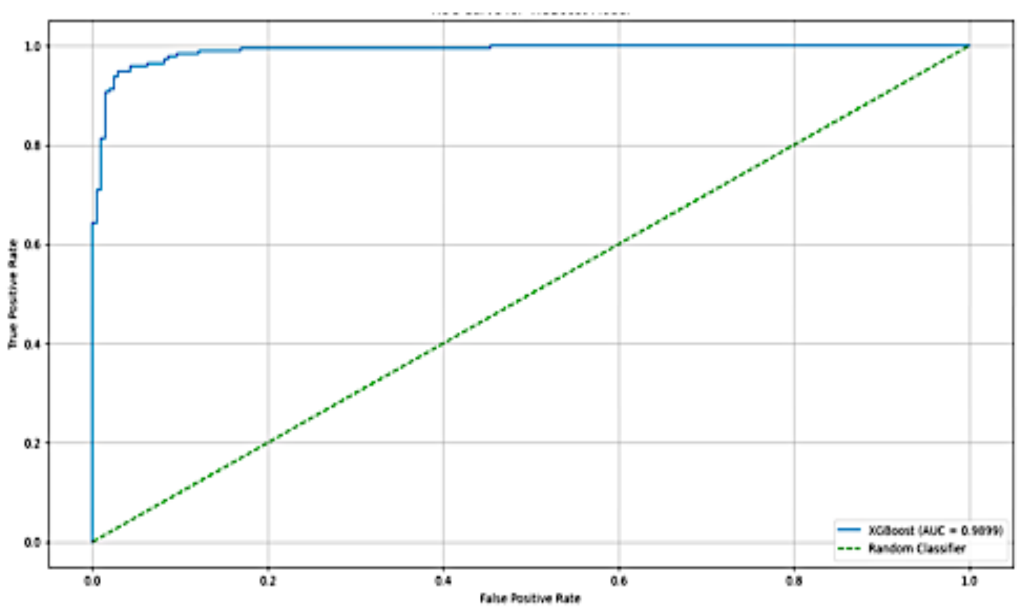


Figure 13. Random Forest model AUC

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

3.4. Tool Integration

In the system implementation, an AI model is integrated into the EFD application to analyze product features and compare the entered product details to determine if it is underpriced. For this research, mobile phones were chosen as the product to test the application's functionality due to their diverse features and the rapidly growing mobile phone market. Additionally, the mobile phone market has a high incidence of tax fraud through underpricing.

The dataset used for this study was collected from a reliable mobile phone sales shop, pre-processed, and trained using various models before being loaded into Streamlit for performance evaluation in a production environment. **Error! Reference source not found.** Figure 14(a) display the application homepage interface for the general EFD application and Figure 14(b) POS interface for entering sales details, the interface implemented with input validation, including phone number validation, RAM input, and storage inputs, is done to improve the quality of data that will affect the model performance. Also, (a) and (b) showing alerts for underpriced and correctly priced products, respectively, that alert the trader that submitting the data will lead to further investigation by the authority, and if the trader does not submit the data will not be subjected to further investigation. After clicking check out, it allows the data to be submitted to the authority data database, which can also synchronize data to EFDM that is monitored by tax officers. During the EFD sales recording process, the product features, such as Brand, RAM, Model, and other important attributes, along with the price range, are entered. The application then alerts the trader whether the product is underpriced based on the correlation between the features and underpricing. The EFD application alerts traders when a product is underpriced and sends the related sales and location data to Firebase. Tax officers from the Tanzania Revenue Authority (TRA) can then access this information to identify and trace potential fraud through Google Maps, improving the efficiency and focus of tax compliance operations, as shown in (b), This give the option for the trader not to checkout for a avoid being caught by the authority, so can opt to recorrect the price to the correct price. As study by [34] insisted the security mechanisms to be more proactive that reactive, by implementing the techniques that can detect the possibility for fraud before it occur and implement corrective measures.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

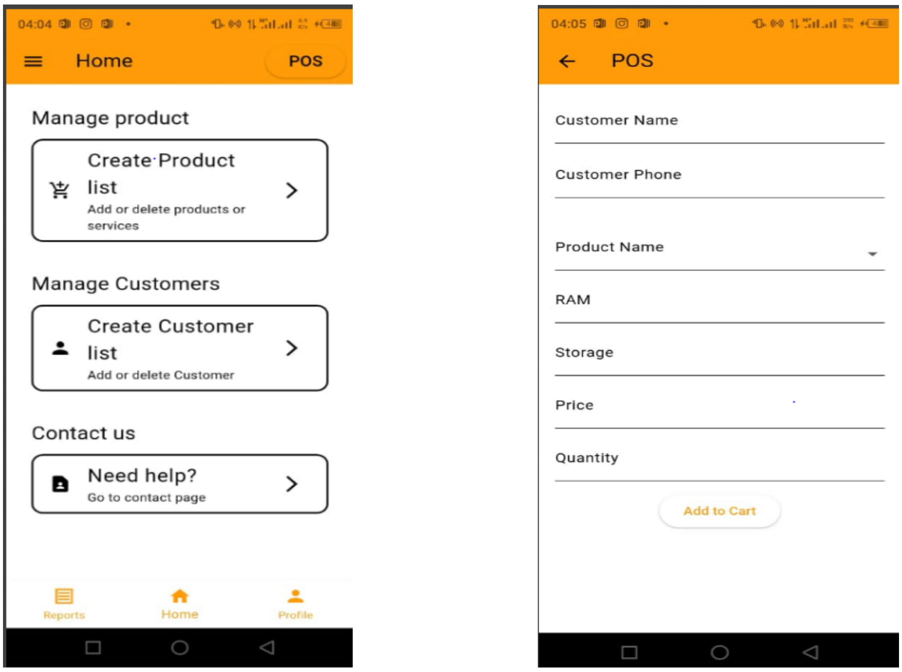


Figure 14. (a) Homepage of the EFD application (b) POS home page

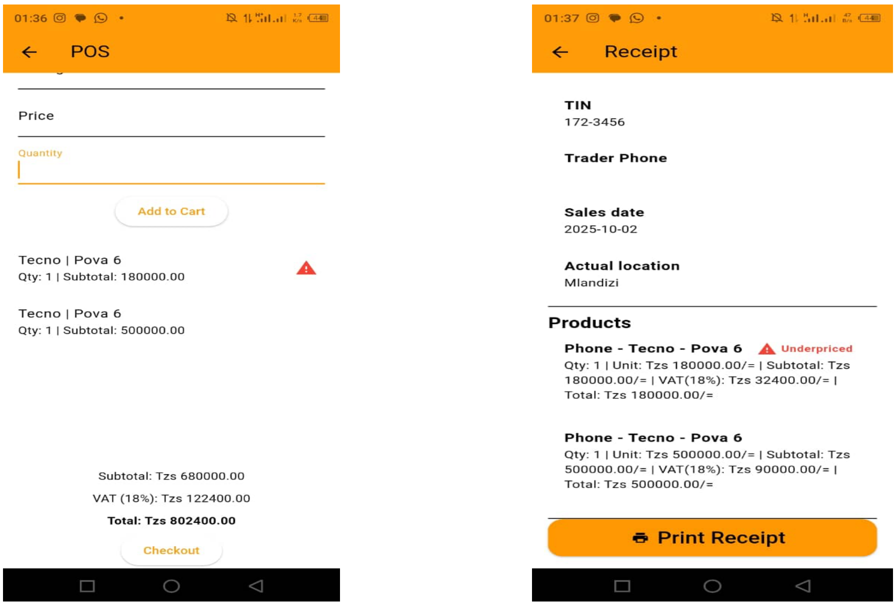


Figure 15. (a): Add to cart underpricing alert (b): Checkout underpricing alert

All data collected by the EFD Application is transmitted to the Firebase Real-time Database, a secure cloud-based platform that ensures real-time data accessibility and synchronization. This setup enhances transparency and supports the timely detection of irregularities such as underpricing or tax evasion [35]. The data is then accessed by the EFD Management System, where tax officers (TRA) can receive real-time alerts indicating the status of each transaction performed by traders, showing "true" if underpricing is

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

detected or "false" if the price is valid. Additionally, the system provides location traceability, enabling officers to identify and locate potential tax evasion sites through an integrated map view, as visualized Figure 16, the tax authority can go direct to that location as shown google map, which will help the compliance monitoring to be mor focused hence it reduce the resource needed to ensure compliance including human, financial and time which current are used without being focused, which limits they efficiency due the fact that, human resource are limited and to ensure compliance need huge investment in human resources [36].

Furthermore, within the EFD Management System, officers can generate sales reports for specific traders based on their TIN numbers and defined date ranges in real time. These reports can then be compared with those generated from the EFD Application to identify any discrepancies or inconsistencies, thereby strengthening oversight and ensuring compliance with tax regulations. Specified date range, then can compare with the report generated by EFD Application to check for discrepancies as shown Figure *17*, Due to the need for data synchronization between the two systems, EFD and EFDMS, comparing reports from both can reveal the correctness of the data synchronization [37].

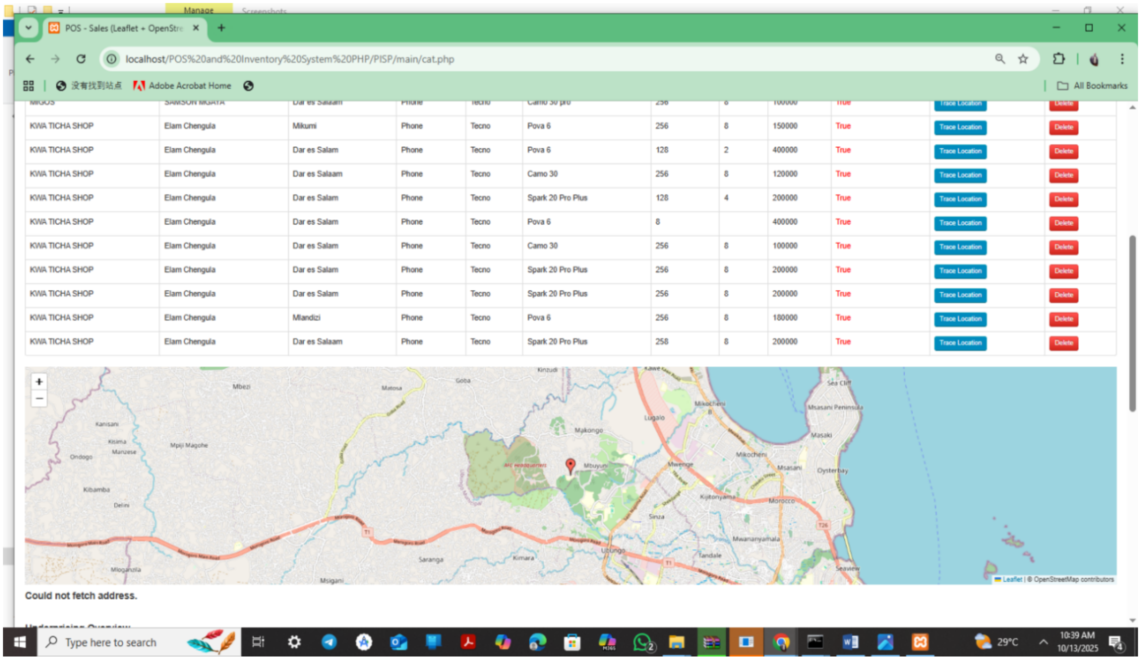


Figure 16. EFD Management system with map location tracing

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

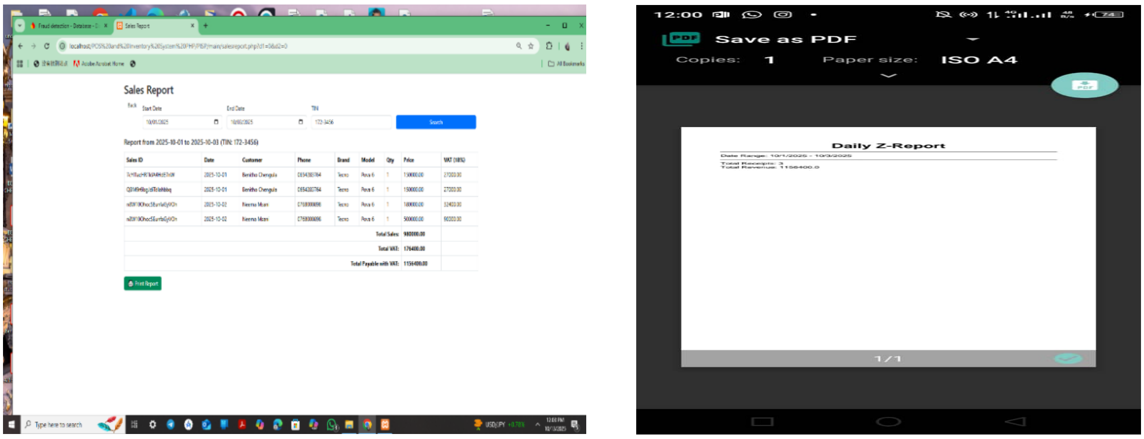


Figure 17. Comparing (a) EFD management system report and (b) EFD application report

In addition, buyers are also integrated into the developed system, which embeds machine learning within the EFD operations. Immediately after transaction data is submitted to Firebase, a receipt is generated and printed with a QR code. Customers can scan this QR code to verify whether the transaction contains any underpricing fraud. As illustrated in Figure 18 (a), the receipt includes standard sales details, while additional fraud-related information is securely encoded within the QR code. Upon scanning, the customer can access this hidden information, including the underpricing status (underpriced or not underpriced). This feature empowers customers to detect suspicious transactions and, if necessary, report them to the TRA authority. Figure 18 (b) further demonstrates how the decoded QR code displays the underpricing status to the user, where the buyer also includes to ensure pricing compliance, creating the chance to report if any underpricing is noticed. Customers need to be involved in ensuring compliance to make it easier; once noticed, underpricing on the receipt can be reported to the authority [38].

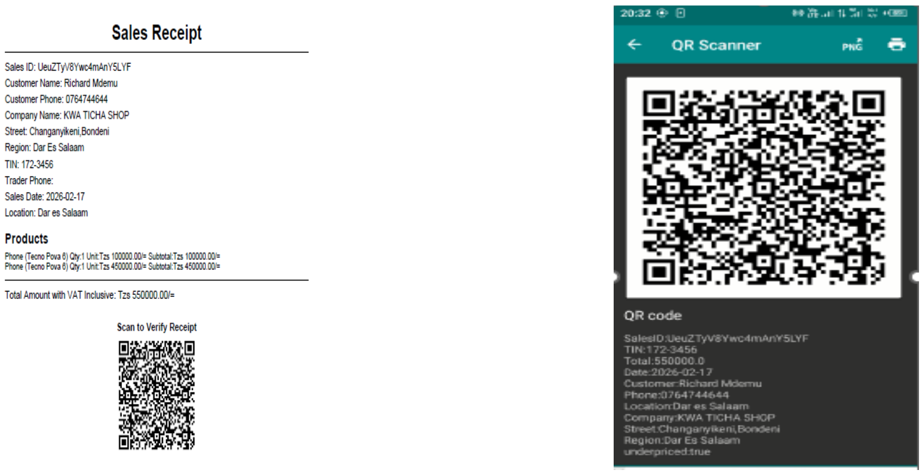


Figure 18. (a) Receipt with QR code (b) QR scan output

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

3.5. Discussion

The machine learning-based underpricing detection model developed in this study provides a practical and modern approach to addressing limitations in traditional Electronic Fiscal Device (EFD) systems. By enabling real-time validation of transaction prices, automating secure data transmission, and reducing opportunities for manual interference, the system enhances transparency, accuracy, and timeliness in tax reporting. These capabilities align with ongoing efforts to modernize tax administration and improve compliance efficiency in Tanzania.

The superior performance of the Random Forest model can be attributed to its ensemble learning capability, which combines multiple decision trees to capture complex, non-linear relationships within transaction data. Unlike models such as Logistic Regression and Support Vector Machine, which may struggle with highly variable pricing patterns, Random Forest is more robust to noise and can effectively handle interactions between features such as product specifications, location, and price deviations. The selected features, including product attributes (e.g., brand, RAM, storage), location, and pricing-related variables, contribute significantly to model performance by providing both technical and contextual information necessary to distinguish normal pricing from underpricing. In particular, price deviation serves as a critical indicator, while location and product specifications help the model adapt to variations in market conditions across regions.

Compared to previous studies, [5]–[9], [11], [24], [39], [40] which primarily rely on post-transaction audits and historical comparisons, this study demonstrates a shift toward proactive fraud detection at the point of sale. Traditional approaches are resource-intensive and reactive, often identifying fraud only after it has occurred. In contrast, the proposed system enables immediate detection and alerting, reducing the reliance on manual audits and improving responsiveness to fraudulent activities. Furthermore, the inclusion of buyers in the verification process introduces an additional layer of transparency, which can enhance trust in the system and encourage compliance.

However, the reliance on historical mobile phone pricing data introduces certain risks and limitations. Market prices for electronic products can change rapidly due to

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

technological advancements, supply chain dynamics, and regional demand variations. As a result, models trained on historical data may become less accurate over time if not regularly updated. This highlights the importance of continuous data updating and model retraining to maintain performance. Additionally, while the system improves detection accuracy, there remains a risk of false positives and false negatives, which could affect user trust and require further refinement of model thresholds and validation mechanisms.

From a broader perspective, the integration of machine learning into EFD systems has important implications for real-time tax enforcement, data quality, and scalability. Real-time detection reduces delays in identifying fraud and enables immediate corrective action, thereby strengthening enforcement mechanisms. However, the effectiveness of such systems depends heavily on the quality, completeness, and consistency of input data. Poor data quality can significantly affect model predictions and reduce reliability. In terms of scalability, while the proposed approach demonstrates strong potential, extending it to other sectors or regions will require adaptation to different pricing structures, data availability, and infrastructure constraints. Overall, the findings suggest that integrating machine learning into tax compliance systems can significantly enhance fraud detection capabilities, but sustained effectiveness will depend on continuous data management, model updating, and system adaptation to evolving market conditions.

Future studies can take this work even further by incorporating real product price data mining, allowing the model to retrain itself regularly using up-to-date market information. The system can also be expanded to include additional features such as tracking sudden changes in business location to address ghost taxpayers and monitoring the time delay between issuing a receipt and sending it to the system, so that unusual delays trigger real-time alerts for tax authorities. With continued development and the ability to scale into other sector areas like detecting repeated large discounts for the aim to avoid tax, this AI-driven approach has strong potential to transform digital tax compliance and improve revenue protection across the country.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

4. CONCLUSION

The research addressed the persistent problem of underpricing, which remains a major contributor to tax evasion in Tanzania's VAT system. Despite the adoption of EFD and EFDMS technologies, challenges such as manual monitoring, limited staff capacity, and vulnerabilities in locally stored sales data continue to create opportunities for manipulation, particularly in underpricing and underreporting. The prototype is tested in pilot not yet in operation, and accepted by more that 70% in reducing manual work, its accuracy in detecting, response type, and secured during User Acceptance Testing included 10 traders, 20 tax officers, and 4 buyers. The proposed AI-based solution enhances compliance by enabling real-time fraud detection, secure cloud-based data storage, and intelligent monitoring within the EFD ecosystem. This approach will support the Tanzania Revenue Authority in efficiently minimizing tax evasion through targeted auditing and improved fraud detection, even with limited resources such as time, personnel, and finances once adopted. Ultimately, the system will contribute to increased revenue collection and aligns with the country's development agenda, particularly the 2030 Sustainable Development Goal 1 (targets 1.a–1. b) on resource mobilization and poverty eradication. Random Forest-based EFD prototype shows promising capability in detecting underpricing fraud within the tested context. However, its effectiveness is limited by a small dataset focused only on mobile phones and a narrow deployment scope. Future improvements should expand the dataset to include diverse products and larger samples, extend the system beyond mobile phones, and integrate it into e-commerce and other sales platforms to improve scalability and real-world applicability, as it currently focuses on utilizing AI in ensuring tax compliance.

ACKNOWLEDGMENTS

I wish to express my deepest gratitude to the HEET project at Ardhi University (ARU) for awarding me the study leave permission that enabled me to conduct this research. Their generous support has made my dream a reality, and I am profoundly appreciative. I would also like to thank all taxpayers from the Arusha region who participated in providing valuable insights into the loopholes in the Electronic Fiscal Device (EFD) system and shared information on how some individuals anonymously evade taxes.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Additionally, I acknowledge the invaluable support of the Tanzania Revenue Authority (TRA) Headquarters. My special thanks go to Annastazia Henry, Senior Economist; Mr. Mkini, Head of EFD; and the TRA Headquarters ICT Department for their comprehensive information about tax collection in Tanzania and their assistance with documentation during the data collection phase.

DATA AVAILABILITY STATEMENT

The data supporting the findings of this study are fully anonymized and were used solely for research purposes. No confidential sales data from Electronic Fiscal Devices (EFDs) were shared or utilized. All qualitative and quantitative data collected during this research are anonymous and comply with ethical research guidelines. For further information, please contact the corresponding author or access to Mendeley data repository with the link: Chengula, Benitho (2025), "Underpricing Fraud Detection using Artificial Intelligence Technology to reduce Tax evasion in Tanzania", Mendeley Data, V1, DOI: https://doi.org/10.17632/ypb66kydvw.1 that support Licence : CC BY 4.0.

NON-COMPETING INTEREST DISCLOSURE STATEMENT

We declare that there are no competing interests associated with this research. The study was conducted independently, and we confirm that no financial, personal, or professional relationships exist that could have influenced the design, execution, or interpretation of the results. Furthermore, there were no affiliations, funding agreements, or external pressures that impacted the objectivity or integrity of this work. This statement reflects our collective commitment to upholding the highest standards of ethical research and academic integrity.

REFERENCES

[1] E. W. Lubua, "Implication of using electronic fiscal devices on audit effectiveness among small business owners of Arusha–Tanzania," *Pan-African J. Bus. Manag. 7*, pp. 41–57, 2023.

[2] J. Longo and N. M. Kenya, "Alignment of Expectations in Electronic Fiscal Devices Adoption: Perspectives of Enforcers and Traders in Tanzania," *sub Sahar. J. Soc. Sci. Humanit.*, vol. 1, no. 1, pp. 1–10, 2025.

[3] J. B. Ndibalema, N. Mayala, and L. Komba, "Influence of Individual Factors on the Usage of Electronic Fiscal Devices Among Smallholder Taxpayers in Moshi Municipality," *Indian J. Econ. Financ.*, vol. 5, no. 1, pp. 18–26, 2025.

[4] A. Kalongola, "Impact of Electronic Fiscal Devices." Institute of Accountancy Arusha (IAA), 2023.

[5] A. B. Bañes, "Tax Management System for Local Government Unit with SMS Notification," 2023.

[6] E. Njogela, J. Ally, B. Lufyagila, and F. Ruambo, "Mobile Application for Enhancing Revenue Collections in Local Government Authorities. A Case of Tanzania," *East African J. Inf. Technol.*, vol. 5, no. 1, pp. 164–179, 2022.

[7] P. E. Shao and M. Dida, "The Implementation of an Enhanced EFD System with an Embedded Tax Evasion Detection Features: A Case of Tanzania," 2020.

[8] B. F. Mururunkwere, J. F. Ihirwe, I. Kayijuka, J. Nzabanita, and D. Haughton, "Comparison of tree-based machine learning algorithms to predict reporting behavior of electronic billing machines," *Information*, vol. 14, no. 3, p. 140, 2023.

[9] N. Nuryani, A. B. Mutiara, I. M. Wiryana, D. Purnamasari, and S. N. W. Putra, "Artificial intelligence model for detecting tax evasion involving complex network schemes," *Aptisi Trans. Technopreneursh.*, vol. 6, no. 3, pp. 339–356, 2024.

[10] S. A. El-Manaseer, J. H. Al-Kayid, A. M. Al Khawatreh, and M. Shamim, "The Impact of Digital Transformation on Combating Tax Evasion (Electronic Billing System as a Model)," in *Artificial Intelligence (AI) and Finance*, Springer, 2023, pp. 679–690.

[11] S. Shaikh, "Balancing Innovation and Oversight: AI in the US Treasury and IRS: A Survey," *arXiv Prepr. arXiv2509.16294*, 2025.

[12] H. E. Nguvava and S. K. Kadenge, "Gross Domestic Product (GDP) determinant of Direct and Indirect Tax Revenue Collection in Tanzania," *Int. J. Econ. Bus. Manag. Res.*, vol. 08, no. 12, pp. 80–90, 2024, doi: 10.51505/ijebmr.2024.81206.

[13] Nbs, "Tax Statistics Report 2020 / 21 Tanzania Mainland National Bureau of Statistics," 2024.

[14] H. E. Nguvava and N. N. Athanas, "Performance of Vat System in Tanzania Since Enactment of The Vat Act in 2014," TIA, 2022.

[15] F. Ramdani, F. Hilmiyah, and V. Indriyani, "The impact of SPSS on research completion," *TOFEDU Futur. Educ. J.*, vol. 4, no. 2, pp. 419–429, 2025.

[16] A. Demir, S. Demir, and E. K. Sahin, "CatGrass: a feature engineering framework to forecast the seismic response of low-rise RC frames using CatBoost algorithm integrated with grasshopper optimization," *Neural Comput. Appl.*, vol. 37, no. 22, pp. 18425–18447, 2025.

[17] L. Williams, "Agile software development methodologies and practices," in *Advances in computers*, Elsevier, 2010, pp. 1–44.

[18] S. Al-Saqqa, S. Sawalha, and H. AbdelNabi, "Agile software development: Methodologies and trends.," *Int. J. Interact. Mob. Technol.*, vol. 14, no. 11, 2020.

[19] D. O. Beerbaum, "Applying Agile Methodology to regulatory compliance projects in the financial industry: A case study research," *Available SSRN 3834205*, 2021.

[20] A. Raikov, "Decreasing tax evasion by artificial intelligence," *IFAC-PapersOnLine*, vol. 54, no. 13, pp. 172–177, 2021.

[21] J. S. Sora, "Determinants of Value Added Tax Performance among Small and Medium Enterprises in Nakuru Central Business District-Kenya," 2021.

[22] S. R. Gwimile, "Value added tax (VAT) compliance through use of electronic fiscal device (EFD) in Tanzania: a case study of Musoma district in Mara." Mzumbe University, 2013.

[23] O.-H. Fjeldstad, C. Kagoma, E. Mdee, I. H. Sjursen, and V. Somville, "The customer is king: Evidence on VAT compliance in Tanzania," *World Dev.*, vol. 128, p. 104841, 2020.

[24] I. Ahmed, G. Jeon, and F. Piccialli, "From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where," *IEEE Trans. Ind. Informatics*, vol. 18, no. 8, pp. 5031–5042, 2022.

[25] J. Martínez-Vázquez, E. Sanz-Arcega, and J. M. T. Martín, "Tax revenue management and reform in the digital era in developing and developed countries," in *Research Handbook on Public Financial Management*, Edward Elgar Publishing, 2023, pp. 202–225.

[26] P. Carrillo, D. Donaldson, D. Pomeranz, and M. Singhal, "Ghosting the tax authority: fake firms and tax fraud in Ecuador," *Am. Econ. Rev. Insights*, vol. 5, no. 4, pp. 427–444, 2023.

[27] B. Bollinger and S. E. Sexton, "Local excise taxes, sticky prices, and spillovers: evidence from Berkeley's soda tax," *Quant. Mark. Econ.*, vol. 21, no. 2, pp. 281–331, 2023.

[28] M. Saarijärvi and E.-L. Bratt, "When face-to-face interviews are not possible: tips and tricks for video, telephone, online chat, and email interviews in qualitative research." Oxford University Press, 2021.

[29] I. Chiaji, C. Kapkiyai, and D. Kirui, "Tax System Automation and Value Added Tax Compliance; The Moderating Role of Obligation Cost," *South Asian J. Soc. Stud. Econ.*, vol. 21, no. 10, pp. 1–13, 2024.

[30] B. Chengula, "Underpricing Fraud Detection using Artificial Intelligence Technology to reduce Tax evasion in Tanzaniaa." Mendeley Data, 2025. doi: 10.17632/YPB66KYDVW.

[31] X. Zhao, R. Lovreglio, and D. Nilsson, "Modelling and interpreting pre-evacuation decision-making using machine learning," *Autom. Constr.*, vol. 113, p. 103140, 2020.

[32] R.-C. Chen, C. Dewi, S.-W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *J. Big Data*, vol. 7, no. 1, p. 52, 2020.

[33] F. Ghafouri-Kesbi, "Comparing the performance of xgboost, Gradient Boosting and GBLUP models under different genomic prediction scenarios.," *J. Livest. Sci. Technol.*, vol. 12, no. 1, 2024.

[34] N. Chhabra Roy and S. P, "Proactive cyber fraud response: a comprehensive framework from detection to mitigation in banks," *Digit. Policy, Regul. Gov.*, vol. 26, no. 6, pp. 678–707, 2024.

[35] M. F. Younis and Z. S. Alwan, "Monitoring the performance of cloud real-time databases: A firebase case study," in *2023 Al-Sadiq International Conference on Communication and Information Technology (AICCIT)*, IEEE, 2023, pp. 240–245.

[36] I. H. Yakin, "Implementation of Human Resources Management in Efforts to Improve the Performance of Tax Instructors," *Atestasi J. Ilm. Akunt.*, vol. 7, no. 1, pp. 287–299, 2024.

[37] K. Bulubaa, O.-H. Fjeldstad, O. M. Nawab, I. H. Sjursen, and V. Somville, "Improving VAT compliance by incentivizing customers: Evidence from Tanzania," *C. Work. Pap.*, 2025.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[38] D. Cipullo, D. G. Degl'Innocenti, M. Le Moglie, and F. Passerini, "Inverting the chain? VAT collection regimes and tax compliance," *VAT Collect. regimes tax compliance*(June 29, 2024)*, 2024.

[39] B. Vyas, "Java in Action: AI for Fraud Detection and Prevention," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, pp. 58–69, 2023.

[40] N. Shafi, F. Atif, A. Atif, and R. Atif, "The Advanced Feature-Driven Smartphone Price Prediction Using Machine Learning," *J. Res. Artif. Intell. Syst. Appl.*, vol. 1, no. 2, pp. 14–20, 2025.