International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 2 (2026)
https://doi.org/10.31436/ijpcc.v12i2.716

Machine Learning–Based Classification of Mobile Money Transactions

Mohamed Abdinur Hussein[1], Hassan Warsame Mohamed[2], Adamu Abubakar[3]
[1,2]Center for Postgraduate Studies, University of Bossaso – Garowe, Puntland, Somalia
[3]Department of Computer Science, International Islamic University Malaysia

Corresponding Author: adamu@iium.edu.my
(Received: 3th June 2026; Accepted: 4th July, 2026; Published on-line: 30th July, 2026)

Abstract— Mobile money has become the backbone of everyday payments across much of the Horn of Africa, and the transaction records it leaves behind are a valuable but under-used source of insight. This paper investigates whether the type of a mobile money transaction can be recovered automatically from its attributes, a capability that would support transaction monitoring and customer analytics. Working within the Knowledge Discovery in Databases framework and using the WEKA platform, we compared three well-established classifiers—Naïve Bayes, J48 (WEKA's implementation of the C4.5 decision tree), and Random Forest—on a real dataset of 3,731 transactions described by ten attributes and labelled with one of ten transaction types. All models were assessed under 10-fold stratified cross-validation. The two tree-based methods clearly outperformed the probabilistic baseline: J48 reached 76.01% accuracy with a Kappa of 0.6147 and the lowest mean absolute error (0.0615), while Random Forest was almost identical at 75.58% and produced the smallest root mean squared error (0.1821); the difference between the two tree-based models was not statistically distinguishable. Naïve Bayes, hampered by its independence assumption, managed only 51.09%. A per-class breakdown showed near-perfect recognition of common categories such as P2P receive (F-measure above 0.93) but a complete failure on the rarest classes, several of which scored zero. This imbalance effect is quantified by the macro-averaged F1 of just 0.40 and balanced accuracy of 38.02% for the best model, far below its 76.01% overall accuracy. This gap is evidence that headline accuracy is inflated by the dominant classes. On balance we recommend J48 for deployment under the present experimental setting, since it pairs competitive accuracy with fast training and human-readable rules that translate directly into monitoring logic.

Keywords— mobile money; machine learning; transaction classification; class imbalance data mining; classification; decision tree; random forest; naïve Bayes

I. INTRODUCTION

Mobile money services have formed the backbone of financial transactions in Puntland and throughout Somalia, where limited access to traditional banks has hastened the growth of digital financial platforms [1]. Individuals and businesses are rapidly using mobile money for daily financial transactions, including person-to-person transfers, merchant payments, salary disbursements, utility bill payments, overseas remittances, among others [2]. Mobile money is a way to get financial services quickly, easily and cheaply via a mobile device, even in the most distant places, and it allows financial inclusion, unlike traditional banking systems. As transaction volumes increase, service providers generate huge amounts of transactional data that mirror financial behaviour of clients, spending patterns, transaction frequency, payment channels and temporal aspects [3]. These digital records are significant information resources that may be studied to reveal hidden patterns of behaviour, boost operational efficiency, reinforce transaction monitoring, aid fraud detection and enhance decision-making through data-driven intelligence.

The fast rise of mobile money services has greatly increased the number and complexity of transaction data generated on a daily basis [4]. The rapidity of the transactions [5] and the heterogeneity of transaction types have made manual processing of these large-scale information increasingly unfeasible. Transaction classification is often inconsistent across different mobile money platforms due to inadequate records, human error, or operational discrepancies. Wrong or delayed classification decreases the effectiveness of transaction monitoring, financial reporting capabilities, and fraud detection methods [6]. As the transaction volumes are increasing, the financial institutions demand automated techniques to precisely identify the transaction types, being efficient and scalable.

https://doi.org/10.31436/ijpcc.v12i2.716

Although the use of machine learning in analysis of financial data is ubiquitous, the selection of a suitable classification algorithm is still a daunting task because the prediction power of different algorithms varies depending on the nature of the dataset [7]. Transactions in mobile money usually possess highly linked features, non-linear relationship and significant class imbalance where common transaction categories are dominant and crucial minority categories appear rarely. These factors diminish the performance of many standard classification methods and complicate model selection. Hence, it is necessary to identify which machine learning method offers the optimal trade-off between predictive accuracy, computational efficiency and model interpretability for mobile money transaction classification in the Puntland setting.

Correct classification of mobile money transactions is critical to provide secure, dependable and effective delivery of digital financial services. Proper classification of transactions helps service providers to improve customer service, prepare correct financial reports, monitor transaction behaviour, detect operational abnormalities and ensure regulatory compliance [8]. Furthermore, good classification leads to improved resource allocation, increased financial transparency, and greater customer confidence in digital monetary systems. As mobile money is replacing conventional banking in many developing countries, efficient transaction classification is becoming a crucial task to provide sustainable financial service provision and to safeguard the integrity of the digital payment ecosystem [9].

Machine learning can be very helpful in assessing massive volumes of transaction data, in that it can automatically find complicated linkages and behaviour patterns that are hard to spot using standard statistical or rule-based approaches. Unlike manual categorization, machine learning algorithms learn continually from past transaction records and identify unseen transactions with minimal human intervention. Decision tree and ensemble learning methods also provide interpretable decision rules that help financial analysts to analyze transaction behaviour. Therefore, machine learning has become an important technology to improve the accuracy of categorization, reduce the operating costs, speed up the transaction process, and enable the intelligent fraud detection of modern monetary systems.

Many studies have been conducted to apply machine learning methods to financial transaction analysis, fraud identification, customer behavior prediction and electronic payment systems. Naïve Bayes, decision tree models, Support Vector Machines, Artificial Neural Networks, Random Forest and Gradient Boosting classification methods have shown promising results on various financial datasets. However, the performance they describe differs significantly, as transaction datasets differ in size, feature qualities, class distributions and domain-specific behaviours. There are no simple generalizations of banking or credit card transaction studies to mobile money contexts, especially in developing nations where the nature of transactions is very different.

Furthermore, many of the prior works are mostly based on binary fraud detection instead of multi-class transaction classification. Others test algorithms on balanced data sets or generated data that is not representative enough of real-world operational contexts. There have been relatively few studies using real mobile money transaction data from Somalia or Puntland and it is not clear what classification technique would perform best given the highly imbalanced transaction distributions, whilst still generating interpretable decision rules which could be used in practice.

The study will compare the performance of Naïve Bayes, J48 Decision Tree and Random Forest classifiers in classifying mobile money transactions based on Knowledge Discovery in Databases (KDD) framework and identify the most accurate, efficient and interpretable model for knowledge discovery and transaction monitoring in the mobile money environment of Puntland.

This study makes several important contributions to the field of mobile money transaction analytics:

- It is one of the few empirical studies that assess machine learning classification approaches using real mobile money transaction data obtained from Puntland, Somalia.
- It applies whole Knowledge Discovery in Databases (KDD) process including data pre-processing, transformation, model creation, evaluation and knowledge deployment to mobile money transaction analysis.
- It offers a comparative analysis of three popular classification algorithms, namely Naïve Bayes, J48 Decision Tree, and Random Forest, under same experimental settings using 10-fold stratified cross-validation.
- It analyzes the performance of the classifiers by means of a number of complementing performance metrics such as classification accuracy, Kappa statistic, precision, recall, F-measure, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and computational time.
- Results indicate that J48 Decision Tree offers the optimum blend of predictive performance, computational efficiency, and model interpretability. It produces transparent decision rules which can enable transaction monitoring, fraud detection, and operational decision making.

https://doi.org/10.31436/ijpcc.v12i2.716

- Finally, the results provide a practical baseline for future research on intelligent mobile money analytics, such as fraud detection, class imbalance management, and sophisticated ensemble learning methods within the expanding digital financial ecosystem.

The remaining sections of the paper are structured as follows: Section 2 presents the relevant work, section 3 outlines the research methods, section 4 details the study results, section 5 comprises the discussion, and section 6 concludes the paper.

II. RELATED WORK

The articles that were studied laid the groundwork for data mining and supervised categorization, both theoretically and in terms of methodology and empirical evidence. They lay out the steps involved in finding new information, provide the most popular classification algorithms, compare how well they forecast, and give statistical metrics for gauging how well two algorithms agree on a classification. Taken together, these research papers back the use of Naïve Bayes, J48, and Random Forest for classifying mobile money transactions, with the caveat that the success of the algorithms relies on the specifics of the dataset.

For quite some time, financial and behavioral data has been subject to classification. When it comes to algorithms and their evaluation, the traditional go-to sources are the standard texts by Han et al. [11] and Witten and Frank [12]. In particular, Quinlan's ID3 and its descendant C4.5 [13], with the latter being implemented in WEKA as J48, are the ancestors of decision trees. In order to create a collection of layered, understandable rules that naturally reflect the interactions between attributes, a tree continually splits the data on the attribute that decreases uncertainty the greatest. Expanding on this concept, Random Forest [14] typically reduces variation and increases stability by averaging the results of constructing several trees on bootstrap samples.

Han et al. [11] proposed a complete architecture of Knowledge Discovery in Databases which includes data cleansing, integration, transformation, data mining, evaluation, and display of knowledge. Using method explanations and application examples, the authors showed how classification, clustering, association analysis, and outlier identification extract useful patterns from enormous data sets. They found that both proper algorithms and thorough data preparation are needed for good knowledge discovery.

T. Witten et al. [12] took a pragmatic approach to experimental machine learning with the WEKA platform. The book provides real-world instances of data preparation, classification, feature selection, cross-validation, and interpretation of performance. Its major finding is that reliable data mining is not only a matter of algorithm choice, but also of good preprocessing, experimental design, validation, and careful interpretation of model outputs.

Quinlan [13] created C4.5 decision-tree approach as an extension of ID3 method. The method chooses splitting attributes, manages continuous and missing data, prunes overfitted branches and turns trees into classification rules. Experiments on different data sets demonstrated that C4.5 could deal successfully with huge instances with many attributes and to provide accurate, interpretable models. It was shown that decision tree models were efficient in performing practical classification.

Breiman [14] proposed Random Forest, an ensemble method that constructs numerous decision tree models from bootstrap samples and a random selection of predictors. Theoretical research shows that the generalization error of its ensemble converges with the number of trees and depends on the strength of the individual trees and their association. Experiments showed that Random Forest was competitive in accuracy, robust to overfitting, able to handle noise and offered important information about variable importance.

Domingos and Pazzani [15] studied both theoretically and empirically the reasons why the simple Bayesian classifier can perform well even when it violates conditional-independence assumptions. They studied zero-one loss and experimented with dependent qualities, showing that accurate probability estimate is not required as long as the proper class still has the highest probability. They showed that Naïve Bayes is best under more general conditions than attribute independence.

Caruana and Niculescu-Mizil [16] performed a comparison of ten supervised learning algorithms on numerous data sets using nine metrics of performance including accuracy, ROC area, precision-recall, squared error and cross-entropy. They found boosted trees, random forests, neural networks, and support vector machines to usually perform well, however the rankings differed by metric. The authors determined that the better model depended on both the data set and the evaluation criterion.

Fernández-Delgado et al. [17] assessed 179 classifiers from 17 method families on 121 real-world datasets. Overall, Random Forest variations performed best, support vector machines and neural networks also performed well. However, no classifier dominated all datasets. The study concludes that substantial empirical comparison is still required as classification success strongly depends on the properties of the application data.

Wolpert [18] compared learning methods with off-training-set error and mathematical analysis without any

https://doi.org/10.31436/ijpcc.v12i2.716

assumptions about the distribution of the problems. The results showed that, averaged uniformly over all conceivable learning tasks, gains of one approach are lost elsewhere. The study proved the No Free Lunch principle, which states that no classifier is better than any other for all possible problems without making assumptions about the target problem.

A statistical method for assessing the agreement between observers for multivariate categorical data was proposed by Landis and Koch [19]. Their technique extended Kappa statistics and agreement functions that account for agreement expected by chance to the observed agreement. An application to categorical reliability data illustrated the effectiveness of the method in assessing consistency in classification. They also offered interpretation ranges for Kappa values from little or poor agreement to practically perfect agreement.

Naive Bayes follows a different path, estimating class probabilities assuming characteristics are independent given the class. The approach can still perform well even when that assumption is broken, as Domingos and Pazzani showed notably [15], but its accuracy does deteriorate when attributes are highly connected. Transaction data is a common example when amount, balance, channel and direction move together. Tree-based and ensemble methods have been shown to be among the strongest general-purpose classifiers in broad empirical surveys [16], [17]. Wolpert's No Free Lunch theorem [18] tells us that no algorithm is the best for all datasets, and this is precisely why an empirical comparison is needed.

Work on transaction and fraud analytics often reports accuracy, precision, recall and the F-measure together with the Kappa statistic interpreted using the benchmarks of Landis and Koch [19], and error terms such as MAE and RMSE validated by cross-validation. One surface gap that keeps coming up is the class imbalance in multi-class transaction data, where the rare categories most relevant for fraud are also the hardest to learn. Our study directly addresses this gap by comparing three classifiers on a real data set from Puntland and studying in detail the effects of imbalance on the outcomes.

III. METHODOLOGY

In this study, we apply the Knowledge Discovery in Databases (KDD) process to build and test a machine learning model to detect fraudulent mobile money transactions. The methodology has successive phases including data collection, pre-processing, transformation, model building, evaluation and knowledge deployment (see Figure 1). Each phase contributes to increasing data quality, establishing robust prediction models, and discovering actionable knowledge to facilitate fraud detection and decision-making in mobile monetary systems.

Figure 1 shows the Knowledge Discovery in Databases (KDD) methodology employed in this study for mobile money transaction classification. The procedure starts with the gathering of a dataset containing 3,731 mobile money transactions. Data pre-processing then follows, such as removing extraneous features, handling missing values, and encoding nominal variables. The cleaned dataset is translated into WEKA ARFF format for analysis. Three classification algorithms including Naïve Bayes, J48 Decision Tree and Random Forest are constructed and tested using the 10-fold stratified cross-validation. Finally, the top performing model is deployed to extract decision rules and to identify critical predictive variables to facilitate transaction classification and knowledge discovery.

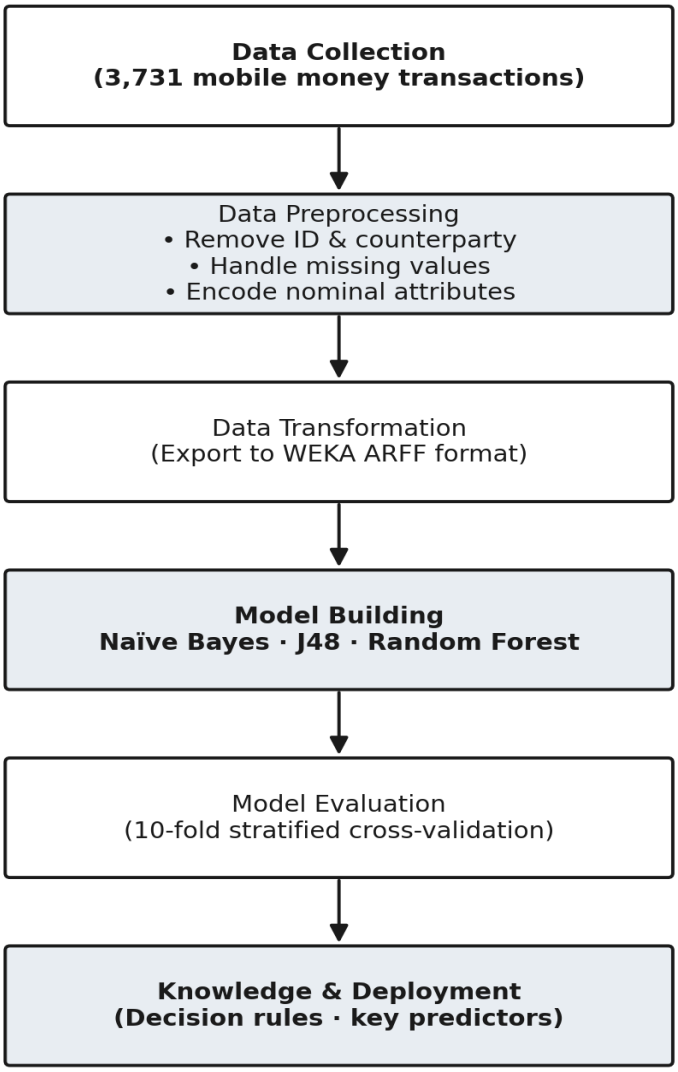

Fig. 1 KDD workflow followed in this study.

https://doi.org/10.31436/ijpcc.v12i2.716

A. Dataset

The data are from mobile money transaction logs gathered over approximately twelve months in a Puntland context. The raw file contains 3,731 transactions. After eliminating a distinctive transaction identification and a high-cardinality counterparty attribute—both of which lack generalizability—ten properties persist, one of which serves as the class label. The label, txn_type, is nominal and encompasses ten values, including merchant payment, two types of person-to-person transfers, overseas remittance, bill payment, and other less common categories. The predictor attributes delineate the transaction's direction, access channel, amount, amount-to-balance ratio, remaining balance, hour of the day, day of the week, whether the counterparty is new, and the time elapsed since the user's last transaction. Table I illustrates a significant skew in class distribution.

TABLE I

CLASS DISTRIBUTION OF TRANSACTION TYPES

Transaction Type	Count	Transaction Type	Count
Merchant payment	1,776	Bill payment	18
P2P send	1,323	Agent deposit	14
P2P receive	503	Reversal	6
Bank withdraw	64	Bank transfer	5
Intl. remittance	19	Airtime/data	3

B. Pre-processing and Attribute Selection

Preparation was carried out entirely in WEKA. The identifier and counterparty fields were removed with the Remove filter; the categorical attributes were encoded as nominal while the numeric ones were kept on their original scale; and the small number of missing entries—chiefly the time-since-previous field for a user's first transaction, together with a few unrecorded channel values—were handled with the ReplaceMissingValues filter where appropriate. The cleaned data were exported to ARFF for modelling. We retained all nine remaining predictors, since each captures a distinct facet of behaviour; an information-gain ranking confirmed that amount, the amount-to-balance ratio, direction, and channel carry the most discriminative power.

C. Experimental Design

Three classifiers were chosen to contrast a probabilistic method with a single tree and with an ensemble of trees. Naïve Bayes served as a fast-probabilistic baseline. J48, WEKA's implementation of C4.5, was run with its default settings (a confidence factor of 0.25 and a minimum of two instances per leaf). Random Forest was likewise left at its defaults, growing one hundred trees. Every model was evaluated using 10-fold stratified cross-validation, which repeatedly trains on nine folds and tests on the tenth while preserving the class proportions in each fold; this uses every record for testing exactly once and gives a more trustworthy estimate than a single train–test split. We recorded the count of correctly and incorrectly classified instances, overall accuracy, the Kappa statistic, precision, recall, the F-measure, MAE, RMSE, and the time taken to build each model.

Fig. 2 depicts the experimental pipeline: the prepared dataset is passed through 10-fold cross-validation to each of the three classifiers in parallel, their performance metrics are collected, and the results are then compared to identify the strongest model.

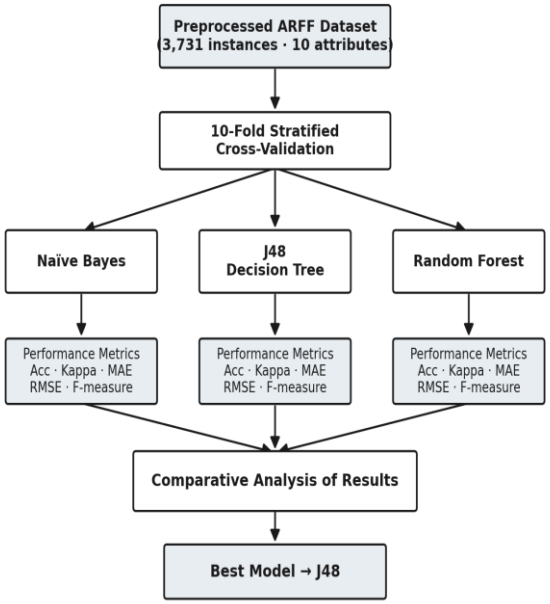

Fig. 2 Model training and comparison pipeline.

IV. RESULTS AND DISCUSSION

Table II summarises how the three classifiers performed across the headline metrics. The gap between the probabilistic baseline and the two tree-based methods is stark. Naïve Bayes classified just over half of the transactions correctly, whereas J48 and Random Forest were both close to three-quarters. J48 edged ahead on most measures the highest accuracy at 76.01%, the highest Kappa at 0.6147, and the lowest mean absolute error while Random Forest returned the smallest root mean squared error, hinting at marginally better calibrated probabilities. In build time the ordering reversed, with Naïve Bayes fastest and Random Forest the most expensive, though none of the models was slow in absolute terms.

https://doi.org/10.31436/ijpcc.v12i2.716

TABLE II

OVERALL PERFORMANCE OF THE THREE CLASSIFIERS

Metric	Naïve Bayes	J48	Random Forest
Correctly classified	1,906	2,836	2,820
Incorrectly classified	1,825	895	911
Accuracy (%)	51.09	76.01	75.58
Kappa	0.2117	0.6147	0.6091
MAE	0.1018	0.0615	0.0651
RMSE	0.2639	0.1935	0.1821
Build time (s)	0.01	0.10	0.57

Read against the Landis and Koch benchmarks [9], the Kappa values tell a clear story: both tree models land in the substantial-agreement band, while Naïve Bayes reaches only fair agreement, confirming that its apparent accuracy owes a good deal to the dominant classes rather than to genuine discrimination.

Averaged figures conceal an important weakness, so Table III reports the F-measure for each transaction type. The common categories were handled well—P2P receive reached an F-measure above 0.93 under both tree models, and merchant payment and P2P send were also recognised reliably. The rare categories were another matter entirely. Classes such as airtime/data, bank transfer, and reversal, each represented by only a handful of records, were effectively invisible to the tree models, which scored zero on them. This is the familiar consequence of class imbalance: with so few examples, there is simply not enough signal for the learner to form a rule.

TABLE III

CLASS-LEVEL F-MEASURE BY CLASSIFIER

Transaction Type	NB	J48	RF
Merchant payment	0.709	0.779	0.769
P2P send	0.244	0.691	0.688
P2P receive	0.243	0.941	0.939
Banks withdraw	0.167	0.455	0.478
Intl. remittance	0.160	0.750	0.788
Bill payment	0.163	0.211	0.095
Agent deposit	0.176	0.000	0.000
Reversal	0.092	0.000	0.000
Airtime/data, Bank transfer	0.000	0.000	0.000

The pattern is easy to explain once the assumptions behind each method are laid out. Naïve Bayes treats the predictors as independent given the class, but in mobile money data the amount, the balance, the channel, and the direction plainly move together, so the assumption is violated and accuracy pays the price [5]. J48 and Random Forest make no such assumption; they carve the feature space into nested regions and thereby capture the very interactions that trip up the Bayesian model. It is worth noting that ensemble theory would lead one to expect Random Forest to beat a single tree [4], yet here the two were separated by less than half a percentage point. That near-tie is not a contradiction so much as a reminder, in the spirit of the No Free Lunch theorem [8], that the best model is a property of the dataset, and on this dataset a single well-pruned tree was enough.

V. DISCUSSION

Mobile money transfers are essential in providing financial services to consumers and businesses with restricted access to conventional banking institutions. They facilitate users in transferring funds, receiving remittances, making merchant payments, purchasing airtime, settling bills, and managing daily financial transactions through their mobile devices. Furthermore, mobile money eliminates geographical, physical, and administrative barriers to financial participation in Somalia and other emerging countries [1], [2]. Its simplicity, cost-effectiveness, speed, and accessibility render it especially vital in rural and underprivileged communities. Robust mobile money platforms are crucial for promoting financial inclusion, enhancing business operations, improving household welfare, and fostering broader economic development.

The swift expansion of mobile money generates extensive transaction records that encompass valuable insights into client behavior, transaction frequency, payment methods, balances, transaction directions, and spending patterns. However, this data holds no use until it is continuously transformed into valuable knowledge. Knowledge Discovery in Databases provides a systematic methodology for cleansing, transforming, analyzing, evaluating, and interpreting transactional data [11]. Knowledge discovery techniques can be utilized to reveal concealed behavioral patterns, aberrant behaviors, trends in service consumption, operational vulnerabilities, and potentially fraudulent transactions. This intelligence assists providers in making superior, evidence-based decisions, enhancing financial reporting, segmenting clients, monitoring transactions, allocating funds, adhering to rules, managing risk, and more.

The increasing volume of transactions is unmanageable through human review, necessitating mobile money providers to seek efficient transaction analysis. Manual classification is labor-intensive, unreliable, prone to human error, and less appropriate for high-volume financial settings. Inaccurate transaction classification can affect financial reporting, client accounts, fraud investigations, and

https://doi.org/10.31436/ijpcc.v12i2.716

regulatory oversight. Automated analysis enables providers to differentiate among merchant payments, person-to-person transfers, international remittances, bank withdrawals, bill payments, deposits, reversals, and other transactions. It can also facilitate the swift identification of anomalous transaction behavior. Consequently, transaction analysis is essential for ensuring service reliability, protecting users, improving financial transparency, and maintaining trust in the digital payment system.

The significance lies in the fact that mobile money variables frequently exhibit complex, non-linear, and interdependent interactions that conventional manual or rule-based methods cannot adequately capture. Machine learning algorithms can identify transaction trends from past data, autonomously categorize new transactions, reduce processing expenses, and produce decision rules that improve monitoring, fraud detection, and operational decision-making [7], [8].

The experiment involved evaluating Naïve Bayes, J48, and Random Forest on 3,731 transactions utilizing 10-fold stratified cross-validation. Naïve Bayes accurately classified 1,906 transactions and erroneously categorized 1,825, resulting in an accuracy rate of 51.09%. The Kappa statistic was 0.2117, the Mean Absolute Error (MAE) was 0.1018, and the Root Mean Square Error (RMSE) was 0.2639. Nonetheless, it was constructed in under 0.01 seconds and exhibited poor accuracy and moderate concordance, signifying insufficient discriminatory power. Thus, computation speed alone is insufficient when the foundational model assumptions do not accurately reflect the linkages present in mobile money data.

J48 accurately identified 2,836 transactions and misclassified 895 transactions, with a maximum accuracy of 76.01%. The highest Kappa value recorded was 0.6147, the lowest MAE was 0.0615, the RMSE was 0.1935, and the model construction time was 0.10 seconds. The Kappa value signifies substantial agreement above random occurrence. The low MAE indicates relatively minor average prediction errors. The findings demonstrate that J48 successfully identified substantial correlations among the transaction amount, balance ratio, direction, and channel. The model features a discernible decision-tree framework, appropriate for operational analysis and practical implementation.

The Random Forest correctly detected 2820 transactions and misclassified 911. The precision was 75.58%. The Kappa statistic was 0.6091 and the Mean Absolute Error (MAE) was 0.0651. It exhibited the lowest RMSE of 0.1821. The model's construction time was 0.57 seconds, slower than Naïve Bayes and J48, however adequate for practical application. The low RMSE signifies that probability estimations are adequately calibrated. Nevertheless, it exhibited an accuracy that was 0.43 percentage points inferior to J48, indicating that the additional complexity of the ensemble did not provide a substantial classification advantage for this dataset.

The J48 and Random Forest models performed effectively for the prevalent transaction types at the class level. The F-measures for merchant payment were 0.709, 0.779, and 0.769 for Naïve Bayes, J48, and Random Forest, respectively. The respective figures for P2P are 0.244, 0.691, and 0.688. P2P reception recorded values of 0.243, 0.941, and 0.939. The bank withdrawal received scores of 0.167, 0.455, and 0.478. International remittance achieved scores of 0.160, 0.750, and 0.788. The findings indicate that tree-based models excel at recognizing prevalent and moderately represented transaction categories.

The performance of minority transaction types was markedly inferior. The F-measures for the bill payment dataset were 0.163, 0.211, and 0.095 for Naïve Bayes, J48, and Random Forest, respectively. The agent deposit was documented as 0.176 for Naïve Bayes and 0.000 for both tree-based models. The reversal achieved a score of 0.092 with Naïve Bayes, and 0.000 with both J48 and Random Forest. All three classifiers received a score of 0.000 for airtime/data and bank transfer. # These outcomes are due to the very low frequency of the classes, comprising merely 18 bill payments, 14 agent deposits, 6 reversals, 5 bank transfers, and 3 airtime/data transactions, which hampers the establishment of dependable patterns.The study demonstrates that mobile money transaction data may be converted into pertinent operational insights via machine learning. J48 provides the optimal balance of classification accuracy, computing efficiency, error minimization, and interpretability. The decision criteria are beneficial for automated transaction classification, customer behavior analysis, financial reporting, transaction monitoring, and preliminary fraud detection. In contexts where probability calibration is crucial, Random Forest proves advantageous, whereas Naïve Bayes serves merely as a first benchmark. The results indicate that overall accuracy should not be evaluated without considering class-level performance, as a model may demonstrate accuracy in most classes while failing to identify infrequent but operationally significant transactions.

The primary concern is the significant imbalance among the classes in the dataset. Merchant payments and P2P transfers recorded 1,776 and 1,323 transactions, respectively. Multiple minority classes contained fewer than twenty observations. The classifiers predominantly identified patterns associated with prevailing transaction types, but they continuously failed to recognize airtime/data, bank transfer, reversal, and agent-deposit transactions. The dataset was obtained in a specific Puntland mobile money context during around twelve months. Transaction patterns,

channel utilization, and class distributions may not accurately reflect different geographical regions, service providers, customer demographics, or temporal contexts.

Subsequent research should compile larger and more diverse datasets from various providers and locations. Enhancing minority classes can be accomplished via additional data acquisition, SMOTE, adaptive synthetic sampling, random undersampling, cost-sensitive learning, or class-weighted models. Moreover, researchers ought to evaluate Gradient Boosting, XGBoost, LightGBM, Support Vector Machines, and deep learning techniques. The models ought to be trained on prior transactions and subsequently evaluated on later periods. This is referred to as temporal validation. Additional behavioral markers, including transaction velocity, device alterations, geographical patterns, recurring counterparties, failed transactions, and account age, may improve classification. Future research should encompass the examination of explainability, concept drift, real-time implementation, and fraud-specific assessment utilizing precision-recall curves and class-sensitive performance metrics.

VI. Conclusions

We applied the KDD workflow to the problem of classifying mobile money transactions, comparing Naïve Bayes, J48, and Random Forest on 3,731 real records under 10-fold cross-validation. The tree-based methods were decisively stronger than the probabilistic baseline. J48 gave the best overall balance 76.01% accuracy, a Kappa of 0.6147, and the lowest MAE—while Random Forest matched it on accuracy and posted the lowest RMSE. Naïve Bayes, at 51.09%, was held back by an independence assumption that the data do not support. We therefore recommend J48 for deployment, not only for its accuracy but for the transparent rules it yields, which are of immediate use in transaction monitoring and fraud detection. The clearest limitation is class imbalance, which left every model unable to learn the rarest transaction types. The obvious next step is to address that imbalance directly through resampling methods such as SMOTE, cost-sensitive learning, or stronger gradient-boosted ensembles—so that the rare but consequential categories can be detected as well as the common ones

Acknowledgment

The author wishes to acknowledge the Center for Postgraduate Studies, University of Bossaso – Garowe, Puntland, Somalia for their support.

Conflict of Interest

The authors declare that there is no conflict of interest.

Authors Contribution Statement

Mohamed Abdinur Hussein: conceptualisation, data collection and pre-processing, WEKA experiments, formal analysis, writing of the original draft. Hassan Warsame Mohamed: methodology design, validation of results, interpretation of findings, review and editing of the manuscript. Adamu Abubakar supervise the work. All authors read and approved the final version.

Data Availability Statement

The dataset used for this study has been cited in the paper.

Ethics Statement

This study did not require ethical approval.

References

[1] P. Chonka, "Displacement stigma, insecurity and inequality in urban Somalia: The limits of digital financial empowerment," Contemporary Politics, pp. 1–21, 2025, doi: 10.1080/13569775.2025.2522448.

[2] T. Suri, J. Aker, C. Batista, M. Callen, T. Ghani, W. Jack, et al., "Mobile money," VoxDevLit, vol. 2, no. 2, p. 3, 2023. [Online]. Available: https://www.voxdev.org/sites/default/files/2023-09/Mobile_Money_2.pdf

[3] S. R. Baker and L. Kueng, "Household financial transaction data," Annual Review of Economics, vol. 14, no. 1, pp. 47–67, 2022, doi: 10.1146/annurev-economics-051520-023646.

[4] D. Namakhwa, B. H. M. Tchereni, W. Masanjala, C. D. Namakhwa, S. L. Kuchande, and W. R. Mgomezulu, "Modelling and forecasting mobile money customer transaction volumes in rural and semi-urban Malawi: An autoregressive integrated moving average spatial decomposition," Scientific African, vol. 26, Art. no. e02430, 2024, doi: 10.1016/j.sciaf.2024.e02430.

[5] M. R. Miah and M. M. Islam, "Data preprocessing and feature engineering strategies for large-scale predictive modeling applications," Review of Applied Science and Technology, vol. 3, no. 1, pp. 263–302, 2024, doi: 10.63125/tqqqed47.

[6] A. Agarwal and V. A. Assenova, "Mobile money as a stepping stone to financial inclusion: How digital multisided platforms fill institutional voids," Organization Science, vol. 35, no. 3, pp. 769–787, 2024, doi: 10.1287/orsc.2022.16562.

[7] H. Gao, G. Kou, H. Liang, H. Zhang, X. Chao, C. C. Li, and Y. Dong, "Machine learning in business and finance: A literature review and research opportunities," Financial Innovation, vol. 10, no. 1, Art. no. 86, 2024, doi: 10.1186/s40854-024-00629-z.

[8] N. Hanbali and A. El-Yahyaoui, "Advanced machine learning and deep learning approaches for fraud detection in mobile money transactions," Innovations in Systems and Software Engineering, vol. 21, no. 2, pp. 333–353, 2025, doi: 10.1007/s11334-025-00605-5.

[9] J. Fan, L. K. Shar, R. Zhang, Z. Liu, W. Yang, D. Niyato, and K. Y. Lam, "Deep learning approaches for anti-money laundering on mobile transactions: Review, framework, and directions," IEEE Internet of Things Journal, early access, 2026, doi: 10.1109/JIOT.2026.3653434.

[10] A. Ali, S. Abd Razak, S. H. Othman, T. A. E. Eisa, A. Al-Dhaqm, M. Nasser, et al., "Financial fraud detection based on machine learning: A systematic literature review," Applied Sciences, vol. 12, no. 19, Art. no. 9637, 2022, doi: 10.3390/app12199637.

[11] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Waltham, MA: Morgan Kaufmann, 2012.

[12] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 4th ed. Cambridge, MA: Morgan Kaufmann, 2016.

[13] J. R. Quinlan, C4.5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann, 1993.
[14] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[15] P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss," Machine Learning, vol. 29, no. 2–3, pp. 103–130, 1997.
[16] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in Proc. 23rd Int. Conf. Machine Learning, 2006, pp. 161–168.
[17] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?" J. Machine Learning Research, vol. 15, no. 1, pp. 3133–3181, 2014.
[18] D. H. Wolpert, "The lack of a priori distinctions between learning algorithms," Neural Computation, vol. 8, no. 7, pp. 1341–1390, 1996.
[19] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," Biometrics, vol. 33, no. 1, pp. 159–174, 1977.