Received: 3 January 2024 | Revised: 29 January 2024 | Accepted: 12 March 2024 | Published online: 19 March 2024

RESEARCH ARTICLE

Archives of Advanced Engineering Science
2024, Vol. 00(00) 1–11
DOI: 10.47852/bonviewAAES42022418

# Comparative Machine Learning Models for Predicting Loan Fructification in a Semi-Urban Area

**Héritier Nsenge Mpia[1,*], Laure Mbambu Syasimwa[1] and Dorcas Masika Muyisa[1]**

[1]*Department of Business and IT, University of the Assumption in Congo, The Democratic Republic of Congo*

**Abstract:** The current research proposes a reliable and robust machine learning (ML) model which outperforms among six other models in predicting loan fructification obtained by entrepreneurs in a semi-urban area. The proposed model predicts if an entrepreneur can make grow a loan from a microfinance firm, a bank, a financial company, or an individual. The proposed model uses primary data collected from entrepreneurs residing in Butembo, a semi-urban town located in eastern Democratic Republic of Congo as dataset. This study uses a dataset that contains 5868 records. Seven ML model performances are compared in the loan fructification prediction: support vector machine (SVM), random forest, extra trees, decision tree, naïve Bayes, k-nearest neighbors, and logistic regression. SVM reveals to be the best model for predicting loan fructification using features such as age, years of working experience of the entrepreneur, entrepreneur loan repayment conviction, used mean by the lender to recover its loan, entrepreneur opinion on the disadvantage of taking out a loan, capacity of the entrepreneur to invest after obtaining loan, entrepreneur position on the possibility of launching a business without a loan, entrepreneur willingness to apply again for loan in the future, and success project after obtaining loan. The study uses accuracy, recall, precision, and *F*1-score as metrics to assess the developed models. The four metrics for SVM scored 95%, 95%, 83%, and 83%, respectively. The proposed model confirms the robustness of SVM in predicting loan fructification.

**Keywords:** quantitative approach, predictive analysis, support vector machine, loan fructification

## 1. Introduction

The authentic foundations of entrepreneurship derive from the economic sciences, whose functional approaches focus a priori on the effects of entrepreneurship and the role of the entrepreneur in the development of the economic system, aiming to understand the personality traits, behavior, origins, motivation, and trajectory of the entrepreneur from a social point of view [1]. Entrepreneurship is also the ability of starting up and developing an activity that gives life to a project, of which the individual him/herself is the initiator. However, since entrepreneurship is a generic term, it exists in several forms [2], including the creation of a business ex nihilo: the creation of a previously nonexistent business; the creation of an existing business; entrepreneurship, which refers to salaried employees of a self-employed person; the takeover of a business, which is the taking over and development of a business in normal operation or in difficulty; social entrepreneurship, which refers to companies that offer solutions in response to society's problems; public entrepreneurship, which refers to public companies that offer services to citizens; and intrapreneurship, which is the creation of a new economic activity within a company that creates an activity external to the company [3].

It is likely to be possible to identify routine factors in entrepreneurship. These factors are, for example, the right age to start a business, the financial resources needed to seize the opportunity to launch a business, family and social background, level of education, employment status and experience of the entrepreneur, which can be measured by the type of job or vocational training he or she has already had before setting up, gender, and perhaps many other factors [4]. In the Democratic Republic of Congo (DRC), despite the various wars and noncompliant infrastructures, the link between the population's mobility and their commitment to entrepreneurship remains strong, particularly for young people aged between 15 and 35 years; attracted by the diversity of opportunities provided by resources across the country, many are turning to entrepreneurial activities [5]. Entrepreneurship is therefore important in the DRC, as innovation is a key element in the DRC's economic growth, and is carried out by an entrepreneur, who will then maintain and nurture it, increasing each household's ability to consume with its own means [5].

The Central Bank of the DRC has shown that the DRC's per capita gross domestic product (GDP) rose from $79.3 in 2002 to $91 in 2006, $100.5 in 2010, $408 in 2014, $466 in 2018, and $703 in 2022, while the population is estimated at 95.89 million over the same period. Poverty affects 6.4 out of 10 households. For every 100 students graduating from Congolese universities, fewer than 30 are employed. The unemployment rate in 2022 is estimated at around 50%, implying a growing number of job

*Corresponding author: Héritier Nsenge Mpia, Department of Business and IT, University of the Assumption in Congo, The Democratic Republic of Congo. Email: nsengempia@uaconline.edu.cd

seekers [6]. However, despite this presented situation, the provinces of the DRC are endowed with more exploitable wealth resources, which can constitute more productive business opportunities [5].

Entrepreneurship is important in the lives of job seekers because, through entrepreneurship, a person is supposed to make his/her dreams come true in what he/she chooses to do for his working life. Responses to questions about feeling capable of entrepreneurship, about the fear of failing in this activity, and about no longer having opportunities to get started, are assimilated to variables that explain the obstacles to entrepreneurship. Income level can also be a motivational constraint to starting a business [7]. By taking out loan, the entrepreneur is able to launch his or her activities. The only real problem is not knowing where to get the loan, but also having an idea of what his or her contribution will be to the activities undertaken, given that it is not at all easy to predict the productivity of any activity [8]. However, for the lender, granting loan means taking on risks of non-repayment, liquidity, interest rates, etc., on the part of the borrower. Risk analysis when taking out loan is largely a banker's job. For companies, the assessment of these risks is based in particular on the borrowing company's vision, its cash flow situation, its operating statement, its shareholders' equity, and the outlook for the general economic situation. It also depends on the type of credit requested and the project to be financed. For individuals, the assessment is based primarily on the borrower's repayment capacity [9].

As a result, some of the loan received by entrepreneurs does not bear fruit. Many researchers in the existing literature have tried to predict whether or not a person can be granted loan based on certain criteria and characteristics of the person using machine learning (ML) [10–12]. Unfortunately, the literature review revealed that there is a lack of research based on the prediction of loan fructification by entrepreneurs. That said, this research used ML to predict the fructification of loan by entrepreneurs in a semi-urban area. The target population of this study was entrepreneurs in the city of Butembo, in particular, those who had already received loan to launch their businesses. Using data from those entrepreneurs, the authors were able to build predictive models to predict whether or not a loan received by an entrepreneur would bear fruit. Butembo is a city in the DRC's North Kivu Province and is known as one of the DRC's commercial cities. Therefore, by studying the entrepreneurial reality in this semi-urban city, the authors felt that the results could be generalized to other DRC cities similar to Butembo.

The main objective of this study was to predict, by applying ML algorithms, whether an entrepreneur residing in a semi-urban area can fructify a loan received from microfinances, banks, financial companies, and individuals. Specifically, this research aimed to: (i) identify the features that predict the fructification of a loan in a semi-urban area; (ii) develop different ML models using the identified features, and (iii) validate the most accurate model. This research on the fructification of loan received by entrepreneurs was motivated by the fact that many entrepreneurs are ready to apply for receiving loan but find it difficult to fructify the obtained loan. The loan granted is often either not remitted in time or not fructified. Several studies have been conducted to predict either loan approval in banks, loan eligibility, or the impact of loan features on bank loan prediction by suggesting random forest (RF) as the best algorithm which outperforms the other methods [12–14]. Using the literature review as a starting point, the authors found that there was a lack of research in the area of loan prediction focused on the prediction of loan fructification. In this paper, the authors examined the effectiveness of various ML classification techniques to predict the loan fructification of an entrepreneur. Contrary to previous work on loan prediction where RF was found to be the best algorithm to predict loan fructification, the findings of this study suggest that support vector machine (SVM) algorithm was more effective in predicting loan fructification than RF. To sum up, this research has contributed to the following three areas: (1) loan fructification prediction, which has not previously been addressed in the context of loan prediction; (2) an in-depth study of the factors that influence the likelihood of loan fructification in semi-urban entrepreneurs; and (3) a comparative analysis of seven ML models designed for optimal selection based on four metrics of evaluation. This study was thus relevant in that the proposed ML model contribute to the identification, through prediction, of entrepreneurs who can potentially grow the loan they can obtain from microfinances, banks, financial companies, and individuals. Managers of banks and entrepreneurs will benefit the most from the current study.

Apart from the introduction, which ends in this paragraph, the rest of the sections of this manuscript are as follows. Section 2 focuses on the literature review, both theoretical and empirical, in order to understand different key concepts referring to this study and to identify the demarcation between previous works similar to this present research. Section 3 presents the methodology and all the tools used to carry out this research. Section 4 briefly illustrates how the analysis was carried out and presents the results obtained in this research. This section concludes with a brief discussion of the results. The final section summarizes the main findings of the study and makes a few recommendations.

## 2. Literature Review

This second section of the paper is based on two main points. First, the essential concepts constituting the object of this research, namely credit, entrepreneurship, and some ML algorithms such as decision tree (DT), RF, extra trees (ET), naïve Bayes (NB), logistic regression (LR), SVM, k-nearest neighbors (KNN), and prediction model are defined. Second, the work related to this research is reviewed in order to place this study in a specific context.

### 2.1. Loan

A loan is an act of trust, esteem, and consideration that takes the form of a credit in cash or in kind granted by a person, in return for a promise to repay within a period of time generally agreed in advance. Loan is therefore a financial operation carried out by a bank or other credit institution, which consists in making resources available to a customer, who in exchange undertakes to repay the sum by a specific date, with interest if necessary [15]. As a means of economic financing, it is used to fund operations on the financial markets, which can be highly profitable thanks to the high leverage provided by bank loans, but also very risky, destabilizing, and whose social usefulness is not always well established. Loan can also enable companies to finance their operating needs while keeping cash at their disposal or to have access to goods produced by others before having produced the equivalent themselves. Conditions of access to loan vary according to the size of the business and the personality of the entrepreneur [16].

### 2.2. Entrepreneur

For economists, an entrepreneur is a person who innovates; for management specialists, an entrepreneur is someone who knows

how to give him/herself common threads, visions around which he/she organizes all his/her activities, someone who knows how to manage and organize him/herself, and someone who shines in the organization and use of the resources around him/her. In the marketers' point of view, an entrepreneur is someone who knows how to identify business opportunities, knows how to do things differently, and knows how to think customer. For those who study business creation, a future entrepreneur is predicted by his or her ability to set and achieve goals, his or her diversity and depth of experience gained in the business sector in which he or she works, and so on [17]. The entrepreneur is someone who must continually learn from what's going on in his environment in order to see business opportunities in what he is doing and adjust his business accordingly [18].

## 2.3. Predictive models

A predictive model is an instance of an ML algorithm that narrows the reality. It is most commonly used to foresee strategic facts and capabilities for companies and society, predicting the evolution of the future in some situation of a certain domain [19]. Predictive models are considered in terms of two factors: system knowledge, which translates a certain reality based on observations of masses of data or by defining a theoretical model of certain actions, and system evolution, which enables predictions to be made based on knowledge. Predictive models are used to predict events, hence the classification model, and to predict quantities, hence the regression model. While the classification model gives the probability of an event by answering the question "will such and such an event occur?", the regression model answers the questions "what quantity will be observed?" based on the observation of data, and "how will the quantity evolve over time?" based on the evolution of data. Predictive models are built using ML algorithms, some of which have been applied in this research [20].

SVM, also known as wide-margin separators, is a powerful ML algorithm based on the linear algorithm, which can learn more than linear models. It is used to predict binary qualitative variables as well as quantitative ones [21]. The aim of SVM is to classify data according to observations and then generalize the different results before building a model. SVM is one of the most widely used algorithms in the field of prediction [19]. SVM is a classification agent that searches the problem-solving space properly and faster [22]. Some SVM hybrid models have been proposed recently such as SVM imperialist competition algorithm (SVM-ICA) and SVM genetic algorithm (SVM-GA) which predict more accurately sell signals in tracking stock price [23]. Loan fructification is challenging to predict and classify. However, SVM has shown to be the most effective classification method due to its unique nature and ability to overcome restrictions. SVM discovers global optimal solutions faster than other algorithms, including neural networks, and it frequently delivers accurate optimal solutions [24]. The authors' decision to use SVM to predict loan fructification was consistent with the advantages indicated above in terms of prediction optimization.

LR is an estimate of the probability of an event occurring. It is often used for classification and predictive analysis problems. It requires a larger representative sample, to have sufficient statistical power to detect a significant effect [25]. LR can be used in fraud detection, disease prediction, and bank failure prediction [26].

KNN belongs to the family of ML algorithms but does not require a learning phase to solve a problem. It is used for both classification and regression problems [27]. It enables companies to use their appropriate data to train algorithms in order to better circumvent their consumption requirements. The KNN algorithm determines the KNN in a mass of objects and then classifies a new object into one or other category, by determining its few characteristics, which are compared with those of any number of object categories, and then assigning the object in question to one of these categories [28]. Recently, KNN has been used to predict breast cancer and its performance in such prediction revealed to be accurate at 100%. In a balanced dataset, KNN defeated SVM, RF, LR, and neural networks [29]. The DT is an algorithmic model used in ML. This model is called a tree, because it is made up of nodes each forming an attribute and each having a certain number of variables [30]. The DT is not the only ML model, but also it is preferred for its ability to analyze a variety of data from different fields of study. Indeed, the main purpose of the DT when developing an ML model is to make a prediction of the value of any target based on the understanding of the indispensable measures of the ends taken from the training data or previously presented data, taken as a reference [31].

The RF is a set-type algorithm used to build individual trees on different samples and using different variables. This algorithm is used to contribute to the resolution of classification and regression problems. It deals with masses of real data in several fields of application such as ecology, biology, economics, population forecasting, and many others [32]. The NB algorithm is one of the algorithms in ML, used naturally in model supervision and for text classification problems. It is characterized by its speed in classification, and by the fact that it handles a small mass of data equally well and easily, on the other hand, NB holds in assumption the independence of variables [33]. NB, also known as a basic method, provides excellent text classification performance and accuracy. NB has recently demonstrated its ability to analyze the sentiments of Indonesians regarding the COVID-19 pandemic [34]. ET is an ensemble method that is built on trees. This algorithm is an extension of the RF algorithm. ET produces an ensemble of a DT that has not been pruned using a traditional top-down process. This current technique goes beyond the RF's usefulness. Each basic estimator is built using a random subset of features [35]. Previous research on stock prediction has primarily concentrated on regression-based models. Despite its ability to manage large amounts of data and avoid overfitting, classification models have received little attention. ET algorithm has recently overcome the drawbacks of RF, owing to faster learning and noise resilience. As a result, ET is now one of the best algorithms for handling complex financial data. An ET classification model has been shown to outperform benchmark methods like RFs and standard regression models in predicting short-term stock market returns [36].

## 2.4. Related works

Viswanatha et al. [13] stated that banks face significant challenges on a daily basis when it comes to assessing loan requests and mitigating the risks associated with potential borrower noncompliance. Since banks have to assess each borrower's eligibility for a loan, this painstaking evaluation process is proving difficult. Thus, they proposed to combine ML models and ensemble learning approaches to determine the probability of accepting individual loan applications. According to Viswanatha et al. [13], the use of ML can increase the accuracy with which qualified candidates are selected from a pool of applicants. The model they have proposed could solve the

problems associated with loan approval processes. Four algorithms were used, namely RF, NB, DT, and KNN. Their results revealed that NB gave a better accuracy of 83.73% [13].

Kumari et al. [14] said that banking is a crucial part of any economy. Banks serve as mediators between savers and borrowers, offering financial services to individuals, corporations, and governments. One of the most significant services offered by banks is lending. Loans enable firms to develop, hire more workers, and contribute to economic growth. Banks contribute significantly to risk management by extending credit. This reduces the risk of default by making loans available to borrowers who are more likely to repay them. Loan applications are examined based on the applicant's circumstances and the lending regulations established by the institutions. However, there is no guarantee that the candidate chosen from among all candidates would be the best. To reduce human mistake, ML can help automate loan eligibility prediction by analyzing vast volumes of data and recognizing patterns and trends. One of the primary benefits of ML is its ability to learn from past data and apply that knowledge to predict future outcomes [14]. To forecast loan eligibility, they created seven models: KNN, DT, SVM, RF, NB, linear regression, and LR. They used a dataset of 614 records that included loan id, property area, applicant gender, dependents in the family, marital status, loan amount and term, qualification, applicant income, employment status, co-applicant income, creditworthiness, and loan status. Their results indicated that RF was the best model, with an accuracy of 90.71% [14].

Dansana et al. [12] estimate that banking is a regulated industry in most nations because of the critical role it plays in a country's economic stability. Most banks' principal business is lending, and loan interest income accounts for a sizable amount of their assets. However, the loan approval process is currently done manually, which takes time and is prone to errors. Defaults can cause huge losses for banks and possibly bank failures, affecting the economy. To that purpose, these authors investigated the application of ML in the lending process, namely the RF, to reliably identify eligible loan applicants while lowering credit risk. The classification model can predict whether or not a loan will be approved, giving the bank a quick and easy way to identify deserving candidates who provide certain benefits such as higher customer satisfaction and lower operating costs. A dataset containing 25 variables was used [12].

From the above-mentioned related work, the authors came up with the comparison illustrated in Table 1.

Ultimately, from Table 1 summary, it can be seen that the current study differs from previous works in terms of the research environment in which the study was carried out, which is the town of Butembo. Moreover, the method used for data collection and the target population of this study make the current research unique. In addition, past research has focused more on predicting loan approval or loan eligibility, whereas this study focuses on predicting the fructification of loan received by entrepreneurs. The validated model was guided by four different metrics results.

**Table 1**
**Highlight and summary of the related works outcomes**

| S/N | Research | Used models | Best model | Accuracy of the best model | Past research weakness | Conclusion |
|---|---|---|---|---|---|---|
| 1 | Viswanatha et al. [13] | RF, NB, DT, and KNN | NB | 83.73% | Use of only one evaluation metric and low accuracy value | NB is the best model in loan approval prediction |
| 2 | Kumari et al. [14] | KNN, DT, SVM, RF, NB, linear regression, and LR | RF | 90.71% | Use of a small dataset (614 records) to conduct an ML project and use of secondary data which sometimes are not suitable for real-world problems and not reliable [37] | RF is the outperformed model in loan eligibility prediction |
| 3 | Dansana et al. [12] | RF | RF | Not reported | Lack of model performance report and use of unique ML model | RF revealed to be the best model in analyzing the impact of loan features on bank loan prediction |
| 4 | Current research | RF, NB, DT, ET, LR, KNN, and SVM | SVM | 95% | | SVM is the best model to predict loan fructification in semi-urban area |

## 3. Methodology

To develop a suitable predictive model that predict the fructification of loans, the authors compared different ML algorithms. Seven models were developed: RF, DT, LR, NB, KNN, ET, and SVM. The ML models developed used contextual variables derived from primary data collected from

entrepreneurs in one of the DRC's semi-urban areas. A questionnaire was designed and administered via Google Forms to collect this data. A total of 5868 records were collected. An exploratory data analysis was carried out to verify the relationships between the various study variables. Exploratory data analysis is an evolution of inquiry involving the use of statistical summaries and graphical tools to instill knowledge of the data and an understanding of what the result might be. Its aims are data exploration, research, and learning [38]. The exploratory data analysis phase was followed by the division of the dataset into two parts: training and test sets. The test set was the 20% of our data, and the training set was 80%. This was followed by the construction of the ML models, whose performance was verified using accuracy, recall, precision, and *F*1-score.

Accuracy provides information on the number of correctly predicted positives, correctly predicted negatives, incorrectly predicted positives, and incorrectly predicted negatives. While recall provides information on the number of positives well predicted and negatives incorrectly predicted by the designed model. Accuracy, somewhat similar to recall, measures the number of well-predicted positives and mispredicted positives. *F*1-score is the combination of recall and accuracy. It assesses the performance of the prediction model [39]. The four used metrics to evaluate the developed models are mathematically expressed by the formulas (1), (2), (3), and (4) as follows:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

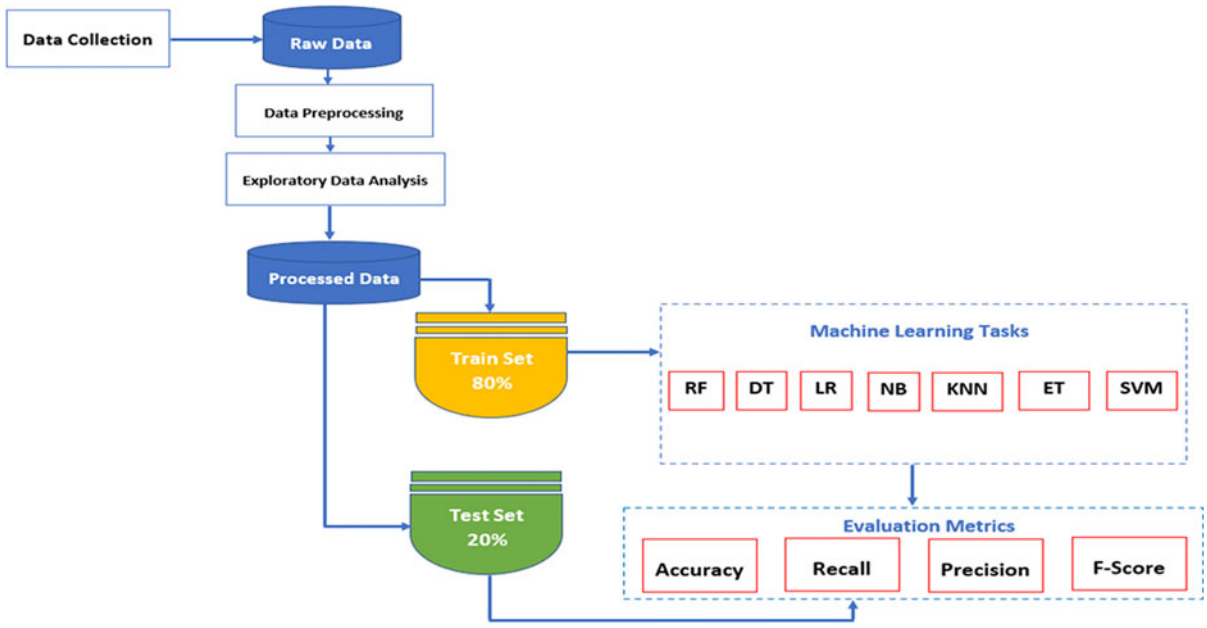
$$F - Score = 2 * \frac{Recall * Precision}{Recall + Precision} \tag{4}$$

The four metrics share the concepts of true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs). TPs correspond to positive data correctly classified by the model. FPs are negative data that the model incorrectly classifies as positive. TNs indicate that the model correctly identifies negative data in the sample. FNs are positive data that the model has incorrectly classified as negative [40]. To comprehend TP, TN, FP, and FN, consider a collection of real values named y from a dataset called D. These are the actual values or the observed data in D's dataset. In binary classification, these actual values are typically represented as 0 or 1. For example, in a loan scenario, "1" may represent the fructification of the obtained loan (positive case) and "0" the loan's no fructification (negative case). By developing a classification model based on D data, a set of results called predicted values is obtained, which can be called $\hat{y}$. Actual values are the true observed values, whereas predicted values are the model's expected results. Unlike actual values (y), predicted values ($\hat{y}$) in a binary classification are probabilities ranging from 0 to 1. These probabilities represent the possibility of a specific outcome. A predicted value of 0.8, for example, in a loan model, indicates that the entrepreneur can fructify the loan 80% of the time. On this approach, the TP in ML is the sum of all $y = 1$ and $\hat{y} = 1$. The TN is the sum of all $y = 0$ and $\hat{y} = 0$. FP is the sum of all $y = 0$ and all $\hat{y} = 1$. FN is the sum of all $y = 1$ and $\hat{y} = 0$ [41].

The proposed methodology for developing the predictive models followed the above-mentioned complementary procedures. These procedures are shown graphically in Figure 1:

To achieve the first objective of this research, features were selected using the feature selection technique. The authors used Python's SelectKBest function for this purpose. SelectKBest is a sklearn function that determines the relevant variables that best predict a supervised problem using the supervised feature

**Figure 1**
**Flowchart of the used methodology of the study**



selection approach. SelectKBest evaluates feature relevance using different scoring systems and eliminates all but the k best characteristics [42].

### 3.1. Sampling and sampling procedure

To collect data related to loan fructification, it was necessary to determine the target population. The considered population was constituted by entrepreneurs residing in the city of Butembo, particularly those who have already received any loan to launch their businesses. Sampling is a technique with a dual meaning: on the one hand, in the strict sense, it designates the exclusive result of an approach aimed at sampling a part of a well-defined whole; on the other hand, in the broad sense, it designates the result of any action aimed at constituting the empirical corpus of any research [43]. Sampling is a method of selecting a sample according to the chosen method. Since the population was not well defined, the authors selected the sample randomly. Simple random sampling is nothing more than one of the methods of probability sampling, which allows each sample element in a population to have an equal chance of being selected from a sample [44]. Hence, the authors resorted to statistical formulas to get an idea of the actual number of sample. Using a hypothesis validation threshold of 95% and a risk of error of 5%, the $Z$ table proposes a confidence level of 1.96. Indeed, when the degree of proportion is not defined, it is advisable to default to 0.5 and $q$ is estimated at 1-p. The following formula (5) was applied:

$$n_0 = \frac{Z^2 * P * q}{e^2} \tag{5}$$

where $n_o$ is the sample, $Z^2$ is the confidence level squared, $p$ is the degree of proportion, and $e^2$ is the error squared. Applying the above formula, the sample for this research was calculated as follows:

$$n_0 = \frac{(1.96)^2(0.5)(1-0.5)}{\left(\frac{5}{100}\right)^2}$$

$$n_0 = \frac{3{,}8416*0{,}5*0{,}5}{0{,}0025}$$

$$n_0 = \frac{0{,}9604}{0{,}0025}$$

$$n_0 = 384{,}16 \cong 384$$

Even though the sample value for this research was estimated at 384 respondents, the authors managed to reach 5868 respondents during the data collection phase, and it was with all the responses received that the analyses and processing were carried out, given that some ML algorithms are sensitive to the number of data.

### 3.2. Research instrument

Quantitative research was used in this study to obtain the primary data. Quantitative research is used to collect pure, concrete data by measuring the opinions or behaviors of a population and describing their characteristics for each type of behavior in order to draw general conclusions about the study. It relates to a purely positive representation of statistical facts, while aiming to test hypotheses and illustrate theories by highlighting correlations between variables, measuring inequalities of distributions between them [45]. The data for this study were collected from entrepreneurs in the city of Butembo, via a survey questionnaire developed using Google Forms. Google Forms is an office tool that enabled the authors to create the form and save the data collected via the link it generates to access the form [6]. The link was sent via social networks such as WhatsApp, Facebook, and WeChat to obtain the data.

Hence, a questionnaire was drawn up, targeting entrepreneurs. The questionnaire consisted of 28 closed questions, with 8 dyadic questions, one likert-3 question, 8 likert-4 questions, 8 likert-5 questions, and 3 likert-6 questions. Table 2 shows the 28 variables as encoded and their descriptions.

### 3.3. Validity and reliability of the instrument

Once the survey questionnaire had been drawn up, it was submitted to a number of financial academics and data science experts for its evaluation, before being submitted to respondents to test its validity. The validity test was carried out in consultation with these experts. The instrument's reliability was tested using Cronbach's alpha coefficient, a statistical practice used to estimate or measure the reliability or internal consistency of answers given to questions in a test or questionnaire on a given subject [46]. Alpha test was computed, using the following formula (6):

$$\alpha = \frac{k * \bar{r}}{1 + (k-1) * \bar{r}} \tag{6}$$

where $\alpha$ stands for Cronbach's alpha coefficient, $k$ the number of items in the submitted questionnaire, and $\bar{r}$ the average correlation between items [47]. The Python code below was therefore adopted from the research of Mpia et al. [6] to compute the alpha coefficient using Python. Figure 2 illustrates the Python sample code used by the authors to compute Cronbach's alpha test

Alpha's result showed that the research tool was reliable, with a score of 0.61.

## 4. Results and Discussion

### 4.1. Research results

Having completed the data analysis and processing stage, it proved essential to present the results obtained in this research. These results are presented according to the objectives of this study.

#### *4.1.1. Results to achieve the research objective one*

The first objective aimed to identify the variables that best predict the fructification of a loan. The authors used the SelectKBest function to compute feature selection. Of the 28 research variables, 9 were selected. The code illustrated in Figure 3 shows how the SelectKBest function was instantiated in this research.

Based on the results illustrated in Figure 3, it was observed that the variables age, anciennete entreprise, jugement remboursement, recouvrement, desavantage credit, reponse motifdemande credit, succes projet, nonrecours credit, and volonteprise credit ulterieur (whose descriptions are illustrated in Table 2) were retained as the contextual variables that best predict the fructification of a loan by entrepreneurs in semi-urban areas.

#### *4.1.2. Results to achieve the research objective two*

The second objective aimed to build seven ML models using the selected best research variables as features. These models were developed using algorithms presented in the research design stage. Table 3 shows the performances of the obtained results.

**Table 2**
**Description of the variables used**

| SN | Variable | Description |
|---|---|---|
| 1 | Commune | Entrepreneur's district |
| 2 | Genre | The gender of the entrepreneur |
| 3 | Age | The age of the entrepreneur |
| 4 | Etude | The education level |
| 5 | Etatcivil | Marital status |
| 6 | Profession | The profession of the entrepreneur |
| 7 | Ancienneteentreprise | Years of working experience |
| 8 | Sourcefinancement | Financial source |
| 9 | Originecredit | Company's main source of financing |
| 10 | Motifprisecredit | Motif of getting the loan |
| 11 | Difficultepourprisecredit | Difficulties that led to obtaining loan |
| 12 | Montantcredit | The amount of the loan granted |
| 13 | Frequencecredit | Number of loan already obtained to start a business |
| 14 | Affectationcredit | The purpose the loan was used |
| 15 | Delairemboursement | Repayment term of loan |
| 16 | Capaciteremboursement | Have you been able to repay the loan after the scheduled repayment period? |
| 17 | Moderemboursement | Interest-free or interest-bearing payment |
| 18 | Jugementremboursement | Do you think it is normal for a loan to be repaid with interest? |
| 19 | Jugementmoderemboursement | Can repaying loan with interest directly influence the income of the amount earned after using the loan? |
| 20 | Recouvrement | The means used by the lender to recover its loan |
| 21 | Desavantagecredit | The disadvantage of taking out a loan |
| 22 | Defisdemandecredit | Challenges faced when applying for a loan |
| 23 | Constatapresusagecredit | Point of view after using the loan |
| 24 | Reponsemotifdemandecredit | After accessing loan, were you able to invest? |
| 25 | Succesprojet | Has the use of loan led your project to success? |
| 26 | Volonteprisecreditulterieur | Are you willing to apply for loan in the future? |
| 27 | Nonrecourscredit | Do you think that you can launch a business without a loan? |
| 28 | Fructificationcredit | Did you fructify the loan you received earlier? |

**Figure 2**
**Cronbach's alpha test for verifying internal consistency of the questionnaire**

```
def test_alpha(dataframe):
    import numpy as np
    data_corr=dataframe.corr()
    N=dataframe.shape[1]

    moyenne_corr=np.array([])
    for i, colonne in enumerate(data_corr.columns):
        somme=data_corr[colonne][i+1:].values
        moyenne_corr=np.append(somme,moyenne_corr)
    moyenne_finale=np.mean(moyenne_corr)

    alpha=(N*moyenne_finale)/(1+(N-1)*moyenne_finale)

    return alpha
```

```
test_alpha(mslrdata)
```

```
0.6154539181536689
```

Based on the above results, the RF model scored 84% for accuracy, 84% for recall, 82% for precision, and 82% for *F*1-score. The DT scored 81% for accuracy, 81% for recall, 81% for precision, and 81% for *F*1-score. The KNN scored 85% for

**Figure 3**
**Best selected features**

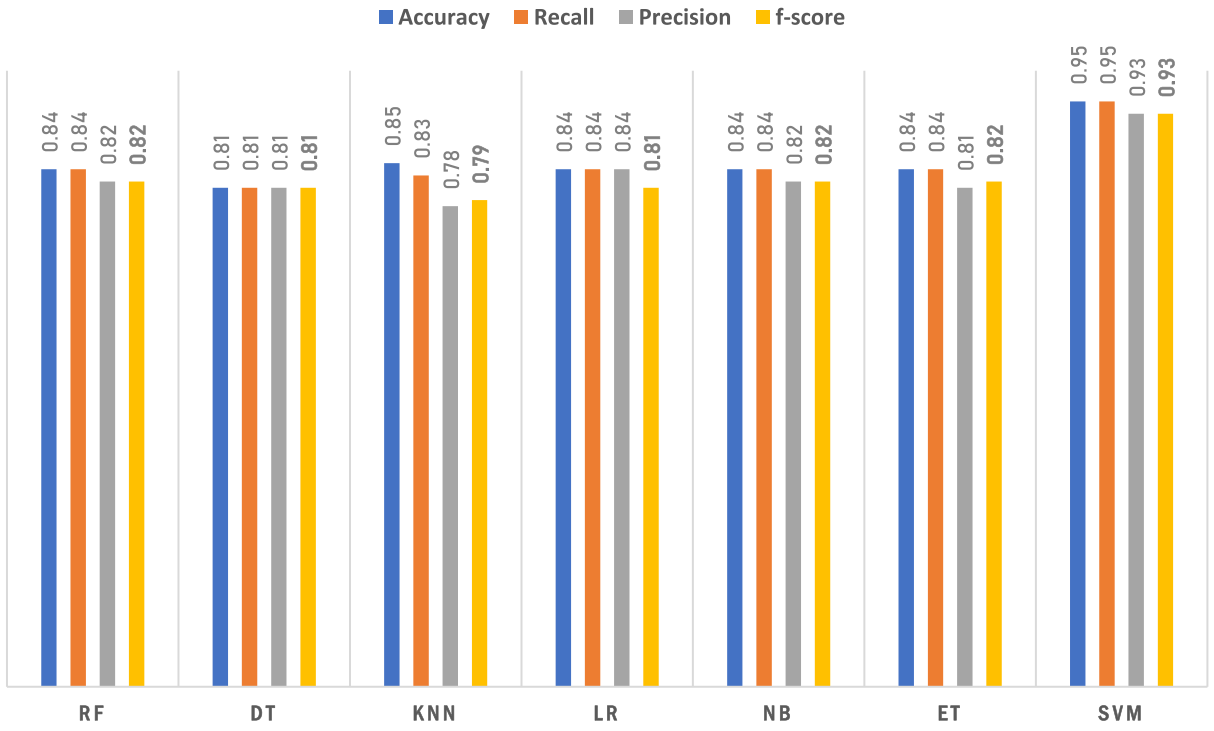
```
from sklearn.feature_selection import SelectKBest, mutual_info_regression
selectMeilleur=SelectKBest(mutual_info_regression,k=9)
selectMeilleur.fit(X,y)
X.columns[selectMeilleur.get_support()]
```

```
ndex(['age', 'ancienneteentreprise', 'jugementremboursement', 'recouvrement',
       'desavantagecredit', 'reponsemotifdemandecredit', 'succesprojet',
       'nonrecourscredit', 'volonteprisecreditulterieur'],
      dtype='object')
```

**Table 3**
**Summary of the evaluation metrics**

| Model | Accuracy | Recall | Precision | *F*1-score |
|---|---|---|---|---|
| Random forest | 84 | 0.84 | 0.82 | 0.82 |
| Decision tree | 81 | 0.81 | 0.81 | 0.81 |
| K-nearest neighbors | 85 | 0.83 | 0.78 | 0.79 |
| Logistic regression | 84 | 0.84 | 0.84 | 0.81 |
| Naïve Bayes | 84 | 0.84 | 0.82 | 0.82 |
| Extra trees | 84 | 0.84 | 0.81 | 0.82 |
| Support vector machine | 95 | 0.95 | 0.93 | 0.93 |

**Figure 4**
**Model performances chart**




accuracy, 83% for recall, 78% for precision, and 79% for *F*1-score. Accuracy for LR reached 84%, with recall at 84%, precision at 84%, and *F*1-score at 81%. NB recorded an accuracy of 84%, a recall of 84%, 82% with precision, and 82% with *F*1-score. The accuracy value for ET was 84%, 84% with recall, 81% with precision, and 82% with *F*1-score. While the SVM scored 95% with accuracy, 95% with recall, 93% with precision, and 93% with *F*1-score.

*4.1.3. Results to achieve the research objective three*

The third research objective was to validate the best-performing model. Figure 4 shows that the SVM model outperformed the six other models in terms of accuracy, *F*1-score, recall, and precision. DT was the least accurate model, with an 81% accuracy. Given that the *F*1-score is a metric that determines the number of correctly classified class 1 objects expressed as a proportion of the number of incorrectly classified class 0 objects [6], the authors observed that the SVM model can make relevant predictions about the loan fructification in the real world, as its *F*1-score value reached 0.93.

Based on the previously presented results, the model that best predicts the fructification of financial credits in a semi-urban area is the SVM model, with 95% for the accuracy metric, 95% for the recall, 93% for the precision, and 93% for the *F*1-score.

## 4.2. Discussion

Based on the past research discussed in the related works section, the authors made a comparison between the results found in the existing literature and those of this study. On this point, most researchers in past research have focused on the prediction of loan approval and eligibility and have used secondary data. They have also predicted the risks of granting credit to individuals. In contrast to these studies, the current study focused on predicting the fructification of loan using ML. Thus, Viswanatha et al. [13] dealt with the prediction of loan approval in banks. They used four ML algorithms, and their results revealed that NB outperformed in terms of accuracy reaching 83.73%. While, the current research NB model achieved an accuracy score of 84%. The study by Kumari et al. [14] focused on predicting the loan eligibility. They developed KNN, DT, SVM, LR, NB, linear regression, and RF models using a dataset of 614 records which yielded, respectively, accuracies of 54.09%, 78.62%, 82.23%, 74.83%, 85.96%, 51.21%, and 90.17%. They validated the RF model. Whereas the RF in this current study, using a dataset of 5856 records, scored an accuracy of 84% and the DT reached 81%, LR and NB scored 84% each, KNN achieved 85%, and SVM outperformed at 95% in terms of accuracy. Dansana et al. [12] proposed a model that analyzes the impact of loan features on bank loan prediction. They developed a RF model. Unfortunately, the authors did not report the accuracy of their model.

The performance of the models developed in this study compared with those of models from past research revealed to be better, demonstrating that the SVM model proposed for the prediction of loan fructification is efficient and more accurate.

## 5. Conclusions and Future Scopes

This research proposed SVM as the best-performing ML model for the prediction of loan fructification obtained by a resident contractor in a semi-urban area. This conclusion was possible after a thorough comparison analysis of seven ML algorithms:

RF, KNN, LR, DT, ET, NB, and SVM. To ensure a fair and reliable comparison, the authors used the same dataset containing 5868 records collected in Butembo, a town in eastern DRC. After conducting the feature selection technique using the Python function SelectKBest, a set of nine relevant factors was extracted which served as features to develop the compared models. The nine selected factors were (1) age of the entrepreneur, (2) years of working experience of the entrepreneur, (3) entrepreneur loan repayment conviction, (4) used mean by the lender to recover its loan, (5) entrepreneur opinion on the disadvantage of taking out a loan, (6) capacity of the entrepreneur to invest after obtaining loan, (7) entrepreneur position on the possibility of launching a business without a loan, (8) entrepreneur willingness to apply again for loan in the future, and (9) success project after obtaining loan. SVM was validated as the best model by applying four different metrics: accuracy, recall, precision, and *F*1-score. The results showed that, in terms of accuracy, DT achieved a score of 81%, RF, LR, NB and ET achieved 84%, KNN achieved 85%, and SVM 95%. Recall results for RF, DT, KNN, LR, NB, ET, and SVM were 0.84, 0.81, 0.83, 0.84, 0.84, 0.84, and 0.95, respectively. In terms of precision, RF scored a value of 0.82, DT 0.81, KNN 0.78, LR 0.84, NB 0.82, ET 0.81, and SVM 0.93. From an *F*1-score point of view, the results for RF, DT, KNN, LR, NB, ET, and SVM were 0.82, 0.81, 0.79, 0.81, 0.82, 0.82, and 0.93, respectively.

The study has three main contribution. First, the authors have conducted a research on loan fructification prediction, while past research on loan prediction has not considered the aspect of fructification of loan. Second, the study has proposed a set of factors which predict loan fructification in a semi-urban area. Therefore, the collected dataset can serve as materials to support data analysts and ML engineers to address several issues related to loan prediction in a semi-urban area. Finally, SVM was proposed as the outperformed model in loan fructification prediction. Indeed, SVM and many of its recent versions proposed by Mahmoodi et al. [48] have been widely used in economics patterns and stock price prediction. This study has certain limitations. The factors used in the research as features to develop the proposed model were not obtained through exploratory factor analysis, but through feature selection technique. As a result, the reliability of these factors remains to be tested through various data analysis techniques. Moreover, the generalizability of the proposed model is still questionable as the used data for testing and evaluating the validated model were still from the same semi-urban zone. With this in mind, the authors recommend in future the application of exploratory factor analysis to reliably capture variables and integrate them as predictors. In addition, the validated model could be deployed in a mobile application for its usability.

## Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

## Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

## Data Availability Statement

Data are available on request from the corresponding author upon reasonable request.

## References

[1] Gomes, S., Ferreira, J., Lopes, J. M., & Farinha, L. (2022). The impacts of the entrepreneurial conditions on economic growth: Evidence from OECD countries. *Economies*, *10*(7), 163. https://doi.org/10.3390/economies10070163

[2] Mets, T., Raudsaar, M., Vahejõe, K., Kaseorg, M., & Vettik-Leemet, P. (2022). Putting entrepreneurial process competence into the focus in entrepreneurship education: Experience from Estonian Universities. *Administrative Sciences*, *12*(2), 67. https://doi.org/10.3390/admsci12020067

[3] Perényi, Á., & Losoncz, M. (2018). A systematic review of international entrepreneurship special issue articles. *Sustainability*, *10*(10), 3476. https://doi.org/10.3390/su10103476

[4] Muhammad, F. S., Kanwal, I. K., Saima, S., & Tayyiba, R. (2021). What factors affect the entrepreneurial intention to start-ups? The role of entrepreneurial skills, propensity to take risks, and innovativeness in open business models. *Journal of Open Innovation: Technology, Market, and Complexity*, *7*(3), 173. https://doi.org/10.3390/joitmc7030173

[5] Kikandi, K. A., Kaghoma, K. C., Mukenyi, J., & Kabamba, A. (2015). *Mobilité Interne Et Entrepreneuriat Des Jeunes En République Démocratique Du Congo (Internal mobility and youth entrepreneurship in the Democratic Republic of Congo)*. Partnership for economic policy (PEP), Cahier de recherche 2015-07. http://doi.org/10.2139/ssrn.2573848

[6] Mpia, H. N, Mwendia, S. N., & Mburu, L. W. (2022). Predicting employability of Congolese information technology graduates using contextual factors: Towards sustainable employability. *Sustainability*, *14*(20), 13001. https://doi.org/10.3390/su142013001

[7] Boubker, O., Arroud, M., & Ouajdouni, A. (2021). Entrepreneurship education versus management students' entrepreneurial intentions. A PLS-SEM approach. *The International Journal of Management Education*, *19*(1), 100450. https://doi.org/10.1016/j.ijme.2020.100450

[8] Chen, Y., Wu, X., Hu, A., & Ju, G. (2021). Social prediction: A new research paradigm based on machine learning. *The Journal of Chinese Sociology*, *8*(15), 1–21. https://doi.org/10.1186/s40711-021-00152-z

[9] Kiros, Y. W. M. (2022). Determinants of loan repayment performance of micro and small enterprises: Empirical evidence from Somali regional state, Ethiopia. *The Journal of Entrepreneurial Finance*, *24*(2), 59–76. https://doi.org/10.57229/2373-1761.1411

[10] Guan, C., Suryanto, H., Mahidadia, A., Bain, M., & Compton, P. (2023). Responsible credit risk assessment with machine learning and knowledge acquisition. *Human-Centric Intelligent Systems*, *3*, 232–243. https://doi.org/10.1007/s44230-023-00035-1

[11] Noriega, J. P., Rivera, L. A., & Herrera, J. A. (2023). Machine learning for credit risk prediction: A systematic literature review. *Data*, *8*(11), 169. https://doi.org/10.3390/data8110169

[12] Dansana, D., Patro, S. G. K., Mishra, B. K., Prasad, V., Razak, A., & Wodajo, A. W. (2023). Analyzing the impact of loan features on bank loan prediction using Random Forest algorithm. *Engineering Reports*, *6*(2), e12707. https://doi.org/10.1002/eng2.12707

[13] Viswanatha, V., Ramachandra, A. C., Vishwas, K. N., & Adithya, G. (2023). Prediction of loan approval in banks using machine learning approach. *International Journal of*

*Engineering and Management Research*, *13*(4), 7–19. https://doi.org/10.31033/ijemr.13.4.2

[14] Kumari, S., Swapnesh, D., & Swarnkar, T. (2023). Loan eligibility prediction using machine learning: A comparative approach. *Global Journal of Modeling and Intelligent Computing*, *3*(1), 48–54.

[15] Moudud-Ul-Huq, S., Halim, M. A., Sobhani, F. A., Karim, Z., & Nesa, Z. (2023). The Nexus of competition, loan quality, and ownership structure for risk-taking behaviour. *Risks*, *11*(4), 68. https://doi.org/10.3390/risks11040068

[16] Pamuk, H., van Asseldonk, M., Ruben, R., Kweka, T., Wattel, C., & Hella, J. P. (2022). Social ties, access to loans, and loan repayments in savings and loan associations: Evidence from rural Tanzania. *Agricultural Finance Review*, *82*(5), 777–796. https://doi.org/10.1108/AFR-03-2021-0036

[17] Landström, H., Harirchi, G., & Åström, F. (2012). Entrepreneurship: Exploring the knowledge base. *Research Policy*, *41*(7), 1154–1181. https://doi.org/10.1016/j.respol.2012.03.009

[18] Méndez-Picazo, M. T., Galindo-Martín, M. A., & Castaño-Martínez, M. S. (2021). Effects of sociocultural and economic factors on social entrepreneurship and sustainable development. *Journal of Innovation & Knowledge*, *6*(2), 69–77. https://doi.org/10.1016/j.jik.2020.06.001

[19] Kalfarisi, R., Chew, A., Cai, J., Xue, M., Pok, J., & Wu, Z. Y. (2022). Predictive modeling framework accelerated by GPU computing for smart water grid data-driven analysis in near real-time. *Advances in Engineering Software*, *173*, 103287. https://doi.org/10.1016/j.advengsoft.2022.103287

[20] Tan, Y. L., Saffari, S. E., & Tan, N. C. K. (2022). A framework for evaluating predictive models. *Journal of Clinical Epidemiology*, *150*, 188–190. https://doi.org/10.1016/j.jclinepi.2022.08.005

[21] Rodríguez-Pérez, R., & Bajorath, J. (2022). Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery. *Journal of Comput-Aided Molecular Design*, *36*, 355–362. https://doi.org/10.1007/s10822-022-00442-9

[22] Mahmoodi, A., Hashemi, L., & Jasemi, M. (2023). Develop an integrated candlestick technical analysis model using meta-heuristic algorithms. *EuroMed Journal of Business*, Ahead-of-print (ahead-of print). https://doi.org/10.1108/EMJB-02-2022-0034

[23] Mahmoodi, A., Hashemi, L., Mahmoodi, A., Mahmoodi, B., & Jasemi, M. (2023a). Novel comparative methodology of hybrid support vector machine with meta-heuristic algorithms to develop an integrated candlestick technical analysis model. *Journal of Capital Markets Studies*, 1–28. https://doi.org/10.1108/JCMS-04-2023-0013

[24] Mahmoodi, A., Hashemi, L., Jasemi, M., Laliberté, J., Millar, R. C., & Noshadi, H. (2023b). A novel approach for candlestick technical analysis using a combination of the support vector machine and particle swarm optimization. *Asian Journal of Economics and Banking*, *7*(1), 2–24. https://doi.org/10.1108/AJEB-11-2021-0131

[25] Boateng, E., & Abaye, D. (2019). A review of the logistic regression model with emphasis on medical research. *Journal of Data Analysis and Information Processing*, *7*, 190–207. https://doi.org/10.4236/jdaip.2019.74012

[26] Borucka, A., & Grzelak, M. (2019). Application of logistic regression for production machinery efficiency evaluation. *Applied Sciences*, *9*(22), 4770. https://doi.org/10.3390/app9224770

[27] Paramasivam, K., Sindha, M. M. R., & Balakrishnan, S. B. (2023). KNN-based machine learning classifier used on deep learned spatial motion features for human action recognition. *Entropy*, *25*(6), 844. https://doi.org/10.3390/e25060844

[28] Tahraoui, H., Toumi, S., Hassein-Bey, A. H., Bousselma, A., Sid, A. N. E. H., Belhadj, A-E., … & Mouni, L. (2023). Advancing water quality research: K-nearest neighbor coupled with the improved grey wolf optimizer algorithm model unveils new possibilities for dry residue prediction. *Water*, *15*(14), 2631. https://doi.org/10.3390/w15142631

[29] Shafique, R., Rustam, F., Choi, G. S., Díez, I. D. L. T., Mahmood, A., Lipari, V., …, & Ashraf, I. (2023). Breast cancer prediction using fine needle aspiration features and upsampling with supervised machine learning. *Cancers*, *15*(3), 1–21. https://doi.org/10.3390/cancers15030681

[30] Chicho, B. T., Abdulazeez, A. M., Zeebaree, D. Q., & Zebari, D. A. (2021). Machine learning classifiers based classification for IRIS recognition. *Qubahan Academic Journal*, *1*(2), 106–118. https://doi.org/10.48161/qaj.v1n2a48

[31] Costa, V. G., & Pedreira, C. E. (2023). Recent advances in decision trees: An updated survey. *Artificial Intelligence Review*, *56*, 4765–4800. https://doi.org/10.1007/s10462-022-10275-5

[32] Vergni, L., & Todisco, F. (2023). A random forest machine learning approach for the identification and quantification of erosive events. *Water*, *15*(12), 2225. https://doi.org/10.3390/w15122225

[33] Zhang, N., Wu, L., Yang, J., & Guan Y. (2018). Naive Bayes bearing fault diagnosis based on enhanced independence of data. *Sensors*, *18*(2), 463. https://doi.org/10.3390/s18020463

[34] Abdillah, L., & Wijaya, D. (2023). Sentiment analysis of Omicron COVID-19 variant using Naïve Bayes classifier and rapidMiner. *Journal of Data Science*, *8*, 1–7. https://ssrn.com/abstract=4598907

[35] Papadopoulos, S., Azar, E., Woon, W. L., & Kontokosta, C. E. (2018). Evaluation of tree-based ensemble learning algorithms for building energy performance estimation. *Journal of Building Performance Simulation*, *11*(3), 322–332. https://doi.org/10.1080/19401493.2017.1354919

[36] Pagliaro, A. (2023). Forecasting significant stock market price changes using machine learning: Extra trees classifier leads. *Electronics*, *12*(21), 1–23. https://doi.org/10.3390/electronics12214551

[37] Mpia, H. N., Waruguru, L. M., & Mwendia, S. N. (2023). Exploratory factor analysis of Congolese information technology graduates' employability: Towards sustainable employment. *SAGE Open*, *13*(4), 1–15. https://doi.org/10.1177/21582440231210109

[38] Nicodemo, C., & Satorra, A. (2022). Exploratory data analysis on large data sets: The example of salary variation in Spanish social security data. *BRQ Business Research Quarterly*, *25*(3), 283–294. https://doi.org/10.1177/2340944420957335

[39] Fränti, P., & Mariescu-Istodor, R. (2023). Soft precision and recall. *Pattern Recognition Letters*, *167*, 115–121. https://doi.org/10.1016/j.patrec.2023.02.005

[40] Xu, Z., Qingyong, C., Xinchang, S., Ping, H., & Lu, P. (2023). Explainable prediction of loan default based on machine learning models. *Data Science and Management*, *6*(3), 123–133. https://doi.org/10.1016/j.dsm.2023.04.003

[41] Picek, S., Heuser, A., Jovic, A., Bhasin, S., & Regazzoni, F. (2018). The curse of class imbalance and conflicting metrics with machine learning for side-channel evaluations. *IACR Transactions on Cryptographic Hardware and Embedded*

*Systems*, *2019*(1), 209–237. https://doi.org/10.13154/tches.v2019.i1.209-237

[42] Mooijman, P., Catal, C., Tekinerdogan, B., Lommen, A., & Blokland, M. (2023). The effects of data balancing approaches: A case study. *Applied Soft Computing*, *132*, 109853. https://doi.org/10.1016/j.asoc.2022.109853

[43] Rahman, M. M., Tabash, M. I., Salamzadeh, A., Selajdin, A., & Rahaman, M. S. (2022). Sampling techniques (Probability) for quantitative social science researchers: A conceptual guidelines with examples. *SEEU Review*, *17*(1), 42–51. https://doi.org/10.2478/seeur-2022-0023

[44] Noor, S., Tajik, O., & Golzar, J. (2022). Simple random sampling. *International Journal of Education & Language Studies*, *1*(2), 78–82. https://doi.org/10.22034/ijels.2022.162982

[45] Franz, D. J. (2023). Quantitative research without measurement. Reinterpreting the better-than-average-effect. *New Ideas in Psychology*, *68*, 100976. https://doi.org/10.1016/j.newideapsych.2022.100976

[46] Béland, S., & Cousineau, D. (2018). Good bye Cronbach's alpha! I found more reliable than you. *Revue de pyschoéducation*, *47*(2), 449–460. https://doi.org/10.7202/1054068ar

[47] Malkewitz, C. P., Schwall, P., Meesters, C., & Hardt, J. (2023). Estimating reliability: A comparison of Cronbach's α, McDonald's ωt and the greatest lower bound. *Social Sciences & Humanities Open*, *7*(1), 100368. https://doi.org/10.1016/j.ssaho.2022.100368

[48] Mahmoodi, A., Hashemi, L., Jasemi, M., Mehraban, S., Laliberté, J., & Millar, R. C. (2023c). A developed stock price forecasting model using support vector machine combined with metaheuristic algorithms. *Opsearch*, *60*, 59–86. https://doi.org/10.1007/s12597-022-00608-x

**How to Cite:** Mpia, H. N., Syasimwa, L. M., & Muyisa, D. M. (2024). Comparative Machine Learning Models for Predicting Loan Fructification in a Semi-Urban Area. *Archives of Advanced Engineering Science*. https://doi.org/10.47852/bonviewAAES42022418