Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 6, December 2025, pp. 4578~4589
ISSN: 2302-9285, DOI: 10.11591/eei.v14i6.9848

Design of a machine learning model for predicting credit risk in microfinance using environmental data

Eladio Alfredo Soto Alvarez[1], Miguel Angel Cano Lengua[1,2]
[1]Unidad de Posgrado de la Facultad de Ingeniería de Sistemas e Informática, Universidad Nacional Mayor de San Marcos, Lima, Perú
[2]Facultad de Ingeniería de Sistemas e Informática, Universidad Tecnológica del Perú, Lima, Perú

Article Info

Article history:

Received Jan 4, 2025
Revised Sep 11, 2025
Accepted Sep 27, 2025

Keywords:

Agriculture
Credit risk
Environmental data
Loan
Prediction

ABSTRACT

Agricultural microfinance is a sector that is significantly impacted by climate-related risks, such as temperature fluctuation, soil degradation, and irregular rainfall. These environmental factors have not only impact on crop yield but also results in influencing borrowers' ability to repay agricultural loans. Traditional credit scoring models lack in predicting due to the complex interplay between environmental and borrower-specific variables. This research study proposes a new predictive machine learning model based on XGBoost for assessing the credit risks in agricultural microfinance. This model utilizes environmental indicators, borrower characteristics, and loan attributes for computing the continuous credit risk score. The model was trained utilizing a real-world dataset of 142,017 loan applications with a 70/30 split. When compared with other traditional models, the results of the model showcases an accuracy of 99%, a recall of 84%, a precision of 89%, and an F1-score of 86%, outperforming traditional algorithms such as logistic regression and decision tree. This model has substantial implications for microfinance organizations. With this model, borrowers can evaluate risk accurately during the loan application stage by utilizing environmental data, resulting in better loan targeting, enhanced financial inclusion, and better risk mitigation for vulnerable farming communities in climate-sensitive regions.

Corresponding Author:

Eladio Alfredo Soto Alvarez
Unidad de Posgrado de la Facultad de Ingeniería de Sistemas e Informática
Universidad Nacional Mayor de San Marcos
Av. Carlos Germán Amezaga 375, Lima, Perú
Email: alfredo.soto.as@gmail.com

1. INTRODUCTION

Climate change has become one of the greatest global challenges, significantly affecting various sectors, with agriculture being one of the most vulnerable due to its reliance on climatic conditions. This phenomenon profoundly impacts the growth cycle, yield, and quality of crops, jeopardizing not only food production but also the economic and social stability of rural communities in critical regions such as Africa and developing countries, where arable land is increasingly degraded and scarce [1]-[7].

Similarly, the agricultural sector not only suffers from the effects of climate change but also contributes to its worsening through significant greenhouse gas emissions [8]. In this scenario, microfinance has emerged as a key tool to mitigate agricultural risks, promote sustainable practices, empower vulnerable communities, and promote gender equality [9]. However, the growing complexity of risks associated with climate change demands more sophisticated solutions, such as the use of artificial intelligence (AI), which enables the assessment and management of credit risks, prediction of crop failures, and optimization of financial resource allocation [10].

This paper aims to explore the potential of AI in strengthening agricultural microfinance, evaluating its capacity to transform climate risk management and promote economic and environmental sustainability. AI is a rapidly growing technology that has the potential to transform various sectors, including microfinance. Microfinance has played a crucial role in empowering millions of people around the world to launch their businesses and achieve their aspirations. Microfinance organizations are constantly exploring innovative strategies to expand their reach and attract more clients. Many of these organizations are investing in AI tools to help provide loans to individuals with limited or no access to traditional banking services [10].

The rationale for this research stems from the pressing necessity to devise innovative frameworks that synergize technological advancements with financial mechanisms to effectively confront the multifaceted challenges posed by climate change in agriculture. This endeavor is pivotal for safeguarding global food security while fostering sustainable rural development.

2. LITERATURE REVIEW

This section details the basic theory for a better understanding of the research topic, focusing on related work regarding credit prediction in microfinance using environmental data. A literature review was conducted with the following PICO questions, see Table 1.

Table 1. PICO questions

Code	Questions
P–Q1	To what extent do machine learning algorithms predict overdue credit to improve the credit portfolio in microfinance institutions?
I–Q2	What is the performance of machine learning algorithms in predicting overdue credit compared to traditional statistical credit scoring in microfinance institutions?
C–Q3	What is the level of precision or accuracy of the results from machine learning algorithms in predicting overdue credit in microfinance institutions?
O–Q4	To what extent does the application of machine learning algorithms in predicting overdue credit reduce credit scoring errors compared to the traditional credit analysis method?
C–Q5	What are the new features introduced by these models?

The study selection process in a systematic literature review, following the preferred reporting items for systematic reviews and meta-analyses (PRISMA) methodology, is carried out through four main phases. The first phase, identification, involves gathering the total number of relevant studies through searches in scientific databases and exploring other complementary sources, ensuring a comprehensive reach of the available literature. Subsequently, in the preliminary assessment phase, an initial filtering of the articles is conducted to identify those that are likely to contribute to the specific objectives of the research, considering criteria such as thematic relevance and alignment with the research question. Next, in the eligibility stage, the selected studies undergo a detailed analysis to verify if they meet the pre-established inclusion criteria, ensuring their relevance and methodological quality. Finally, in the inclusion phase, the studies that will be part of the final synthesis are determined, integrating both qualitative and quantitative evidence to robustly address the objectives of the systematic review. This systematic and rigorous process ensures transparency and reproducibility in the selection of scientific literature.

As a result of the search, a total of 87 articles were obtained, of which 52 are from Scopus and 35 from Web of Science. Then, 27 duplicate articles were removed, leaving 32 articles for review. After a final filter based on the inclusion and exclusion criteria, 11 articles were removed, leaving 21 articles for this review.

As a result of the systematic literature review, it is concluded that the implementation of advanced machine learning models, such as neural networks, genetic algorithms, and hybrid techniques (e.g., rotation forest and logit boosting), offers significantly higher accuracy in predicting credit risks in the agricultural sector compared to traditional methods. These approaches not only address climatic, economic, and sustainability variables but also integrate considerations such as financial education, enhancing inclusion and proactive risk management. Additionally, systems based on algorithms like XGBoost allow for early warnings of up to six months, providing financial institutions with stronger tools to mitigate losses. However, their success depends on the quality of the data and the transparency of the models, critical aspects for their acceptance and effectiveness in credit evaluation in agricultural microfinance.

The study search result showcasing the systematic review process utilizing PRISMA methodology, showcasing how initial studies were considered and selected, providing transparency in literature section process, see Figure 1.

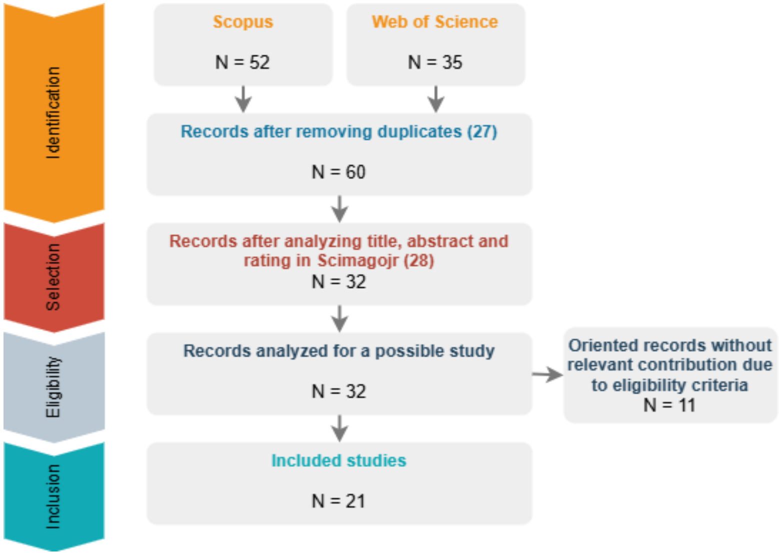

Figure 1. Study search result showcasing the systematic review process utilizing PRISMA methodology

2.1. Basic concepts

2.1.1. Temperature

Agricultural species respond differently to temperature changes throughout their life cycles. Each species has defined maximum, minimum, and optimal limits for growth. Beyond the optimal range, growth rates decrease and may stop once the maximum threshold is exceeded. Vegetative growth tolerates higher temperatures than reproductive stages and increases in temperature can accelerate phenological development up to a specific point. However, this accelerated development may reduce plant size, shorten the reproductive phase, and decrease yields by limiting light capture.

High temperatures, especially during the reproductive stage, affect pollen viability and fruit formation, significantly reducing yield potential. The relationship between temperature and yield is non-linear: yield losses due to high temperatures are more severe than gains from low temperatures. Understanding this dynamic is crucial for predicting the impacts of climate change on agricultural productivity [11].

2.1.2. Humidity

Humidity is essential for photosynthesis, allowing plants to keep their stomata open, regulate temperature through evaporative cooling, and absorb carbon dioxide. Adequate humidity reduces water loss through evaporation, promoting optimal plant function even with limited light. However, low humidity conditions can increase water loss through the leaves, exceeding the roots' ability to replenish it. Under water stress conditions, plants partially close their stomata to conserve water, which restricts photosynthesis and slows growth. This phenomenon highlights the importance of proper humidity management to optimize productivity [11].

2.1.3. Precipitation

Extreme rainfall alters the water dynamics in the soil, increasing surface runoff and modifying the antecedent moisture. High temperatures intensify evapotranspiration, which can dry out the soil before heavy rains, limiting its absorption capacity. Dry and compacted soils generate more runoff, exacerbated in areas with steep slopes and low permeability, such as mediterranean and semi-arid regions. Current models underestimate the role of runoff due to excessive infiltration, which is crucial in these environments [11].

2.1.4. Soil
Climate change, through high temperatures and altered precipitation patterns, destabilizes soil organic carbon reserves, compromising its fertility and biodiversity. Extreme events such as floods and droughts reduce soil quality by accelerating the decomposition of organic matter, leading to nutrient loss. These combined effects negatively impact agricultural growth and the sustainability of ecosystems [12].

2.1.5. Crop growth cycle
Extreme temperatures, coupled with elevated CO_2 levels, significantly impact crop growth. While northern regions may temporarily benefit from warmer and wetter conditions, tropical and subtropical regions will face severe droughts and hotter climates. In Africa, climate change could expand arid areas by 5–8%, adding between 58 and 92 million hectares of marginal lands by 2080, while up to 51 million hectares of favorable agricultural lands could be lost [12].

2.1.6. Crop yield
Climate change presents complex challenges for food production. While higher levels of CO_2 may enhance the growth of crops like wheat and rice, maize will only benefit under drought conditions. Higher temperatures could increase yields in colder regions but will reduce them in warm tropical areas. Extreme weather events, such as droughts and floods, also impact crop productivity, resulting in mixed outcomes depending on the region and crop type [13].

2.1.7. Crop quality
Climate change negatively affects crop quality, including nutritional value and flavor. High levels of CO_2 and elevated temperatures reduce the protein content of staple foods, lowering the quality and demand for agricultural products and impacting farmers' profits [14]. These concepts highlight the importance of environmental data in credit risk modeling, taking into account the implications of climate change on agricultural productivity and food security.

2.2. Related works
2.2.1. Role of microfinance in agriculture
Microfinance represents an essential tool for rural development in the face of various socioeconomic challenges. It helps mitigate agricultural risks through financial education, assisting farmers in managing loans and reducing over-indebtedness [15]. Furthermore, microfinance plays a key role in poverty reduction by extending financial services to those without access to traditional banking, promoting asset accumulation and economic activities. It also stimulates job creation by supporting the growth of agricultural businesses and fostering rural entrepreneurship, which reduces unemployment and boosts economic growth. An important aspect is its contribution to the economic empowerment of rural women by providing access to credit and promoting gender equality [16].

2.3. Artificial intelligence and credit risk assessment in microfinance agriculture
AI is an emerging technology with the potential to transform multiple sectors, including microfinance. This field has empowered millions of people worldwide, enabling them to start businesses and achieve their goals. Microfinance institutions are constantly seeking innovative strategies to expand their reach and attract new clients, highlighting investment in AI tools to provide loans to individuals with limited or no access to traditional banking [17].

Microfinance institutions that offer loans to agricultural businesses must assess the risks associated with the impact of climate change on borrowers' ability to meet their obligations. On the other hand, agriculture significantly contributes to greenhouse gas emissions, exacerbating climate change. In light of this reality, the pressure to adopt sustainable agricultural practices that reduce environmental impact and promote a circular economy is increasing. Given the level of uncertainty and volatility associated with climate change, lenders need to analyze factors such as the probability of crop failures, supply chain disruptions, and reduced market demand. This evaluation is crucial for making informed decisions regarding loan terms, collateral requirements, and risk management strategies [18].

3. METHOD
3.1. Methodology
In this study, the knowledge discovery in databases (KDD) methodology was used, which has been implemented in many financial prediction studies, including Latas' [19], where a big data analysis was performed using the KDD framework to better understand the correlations between environmental, economic, and social indicators.

3.2. Objective of the model design

The objective is to predict whether a borrower applying for an agricultural loan will default (i.e., be unable to repay the loan) based on various characteristics, including the borrower's traits, loan details, and climate-induced environmental data. See the Figure 2 to understand the steps of the KDD methodology applied in this article.

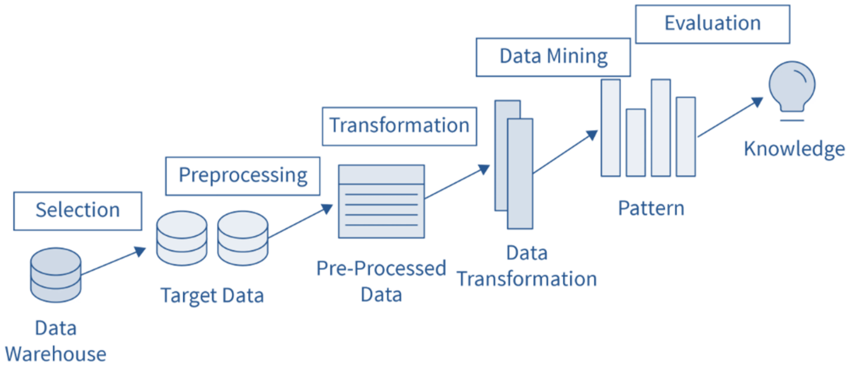

Figure 2. KDD methodology

3.3. Selection of the machine learning technique

Regression techniques for model development. The selection of regression techniques for evaluating credit risk in agricultural microfinance loans is based on the nature of the problem and the specific goals of the analysis. In this study, the models considered are: decision tree, random forest, logistic regression, and XGBoost. Regression models are highly effective in predicting a continuous credit risk score, as they estimate the probability of loan default on a numerical scale rather than a binary classification. This allows lenders to quantify risk levels, differentiating between low, moderate, and high-risk borrowers, facilitating more precise decisions, such as defining loan terms.

In microfinance, where loan conditions must be flexible and tailored to individual risk profiles, a continuous risk score allows for risk-based pricing. Regression methods, including advanced techniques such as gradient boosting and random forest regression, can incorporate multiple influencing factors, including financial, agricultural, and environmental data, to model complex relationships. These models also provide insights into the most influential risk factors, enhancing the explainability of decisions and helping refine lending strategies. Additionally, they are scalable, easy to implement in microfinance systems, and capable of handling large datasets typical of these institutions. XGBoost, a popular regression model, is particularly effective in capturing nonlinear relationships and managing complex interactions between features, making it a standout option for credit risk assessment in microfinance [20].

3.4. Algorithm

The algorithm minimizes a regularized objective function, balancing prediction error with model complexity to avoid overfitting. One of these models is XGBoost, which uses a second-order Taylor expansion approach, simplifying the objective function into a quadratic form, optimizing the tree structure calculations and leaf weights [21].

Some of its key features include support for custom loss functions, the ability to handle missing values, and parallel computing capabilities. These advantages make it more precise and efficient than traditional decision trees enhanced by gradient boosting. Widely adopted in data mining competitions and real-world industrial applications, XGBoost is recognized for delivering cutting-edge results in various machine learning tasks. Its scalability and high performance make it a go-to tool for data scientists across multiple industries [22].

3.5. Key variables to include in the model

Variables commonly considered to predict loan defaults, especially in microfinance and agricultural loans that incorporate environmental factors, should include borrower characteristics (e.g., age, gender, education level, agricultural experience, and credit history), loan characteristics (such as loan amount, interest rate, term, and payment schedule), and agricultural factors (e.g., type of crops grown, use of irrigation or fertilizers).

Environmental data, such as precipitation levels, temperature, and soil moisture, also play a crucial role. These variables help build a comprehensive model to assess credit risk, as they capture both the borrower's personal financial situation and the external environmental conditions that influence the borrower's ability to repay a loan.

3.6. Mathematical formulation

Given a dataset with n features (X_1, X_2, X_3, ..., X_n) and a target variable Y (default status), the goal is to estimate the probability of default P(Default=1|X) which is described in (1).
The XGBoost model can be defined as:

$$P(Default=1|X)=1/(1+e^{-F(X)})) \tag{1}$$

Where: $F(X)=\sum_{m=}^{M} fm(X)$ (sum of predictions from multiple decision trees); f_m (X) is a weak learner (decision tree); and M is total number of boosting rounds.

3.7. Model implementation

3.7.1. Selection-extraction, transformation, and loading

The process of extraction, transformation and loading of data consist of several steps described below, those steps consist on performing actions with user interface as well there are other steps which are handled using appropriated commands for managing the data.

a. Data extraction

Historical information was collected from the credit portfolio database of Access Bank, using SQL Server to extract data from approximately 50 tables. For this research, real data was used. After data cleaning, we have: 142,017 loan application records. Database and list of tables are shown in Figure 3, 23 columns were selected, as described in Figure 4.

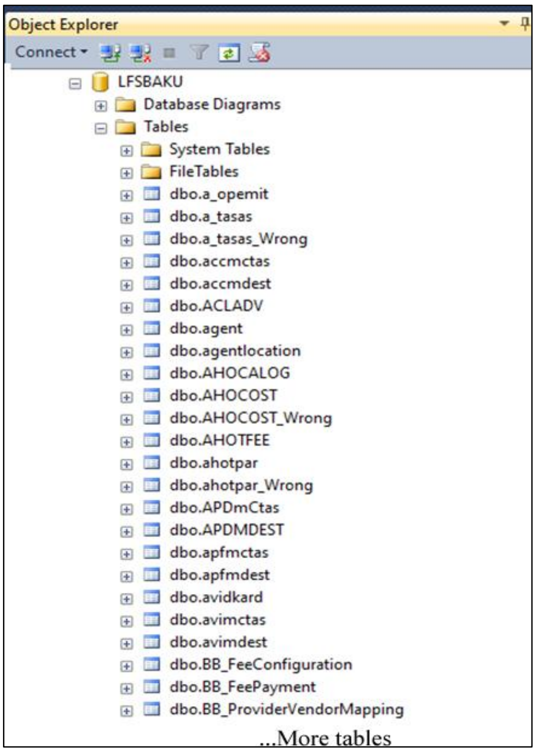

Figure 3. Database and source tables of the information used in this research

```
data.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 142017 entries, 0 to 157442
Data columns (total 23 columns):
 #   Column                Non-Null Count   Dtype
---  ------                --------------   -----
 0   pkMonthEnd            142017 non-null  int64
 1   pkLoanDeliquency      142017 non-null  int64
 2   average_Overdue_days  142017 non-null  float64
 3   clientCode            142017 non-null  object
 4   familyMembers         142017 non-null  int64
 5   clientAge             142017 non-null  object
 6   clientGender          142017 non-null  object
 7   maritalStatus         142017 non-null  object
 8   economicSector        142017 non-null  object
 9   clientEducation       142017 non-null  object
 10  clientIdType          142017 non-null  object
 11  clientAddressDistrict 142017 non-null  object
 12  LoanAmountBracket     142017 non-null  object
 13  LoanBranch            142017 non-null  object
 14  CreditLine            142017 non-null  object
 15  Currency              142017 non-null  object
 16  Deliquency            142017 non-null  object
 17  InterestRate          142017 non-null  object
 18  LoanNewRepeated       142017 non-null  object
 19  Product               142017 non-null  object
 20  Purpose               142017 non-null  object
 21  Status                142017 non-null  object
 22  amount_Credit_Request 142017 non-null  float64
dtypes: float64(2), int64(3), object(18)
```

Figure 4. Selection of relevant columns for this research

b. Data transformation

For the processing of the information from the tables, stored procedures were executed, which are described below. Cleaning, deletion, and modification of records were carried out to obtain clean and specific data in each of the tables. To understand the code used for data transforming please see the Figure 5.

```
if (SELECT count(*) FROM DW_AccessBI.dbo.D_EDUCATION
where Education like '%other%') > 1
        BEGIN
        Print 'Warning with default value other'
          update DW_AccessBI.dbo.D_EDUCACION
        set Education=loanClientEducation+' '+Education
        where Education like '%other%'
        End
else if (SELECT count(*) FROM DW_AccessBI.dbo.D_EDUCATION where
Education like '%other%') = 0
        Begin
          INSERT INTO DW_AccessBI.dbo.D_EDUCATION
             ([loanClientEducation])
             ,[Education]
             ,[dateProcess]
             ,[dateUpdated])
          VALUES
             ('999'
             ,'other'
             ,getdate()
             ,getdate())
        End
GO
```

Figure 5. Data transformation and cleaning process

c. Data loading

To structure and define the data loading in the data warehouse, the snowflake dimensional model was used, where the tables are defined as shown in Figure 6. Once the extraction, transformation, and loading (ETL) process is completed, the data warehouse is ready to proceed with the stages of the KDD methodology applied in this research.

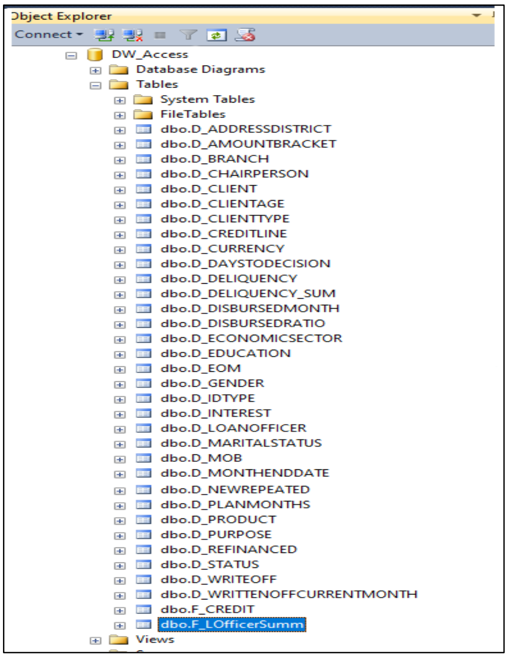

Figure 6. Data warehouse structure for data loading, which will be used as the data source for the machine learning model

3.7.2. Preprocessing

A process is executed to generate the heatmap of the variable correlation. The correlation of predominant variables are shown in the heatmap, darker colors show stronger positive or negative correlations, see Figure 7.

```
# mapa de calor
import seaborn as sns
sns.heatmap(creditos_corr,
            xticklabels=creditos_corr.columns,
            yticklabels=creditos_corr.columns,
            cmap='coolwarm')
```

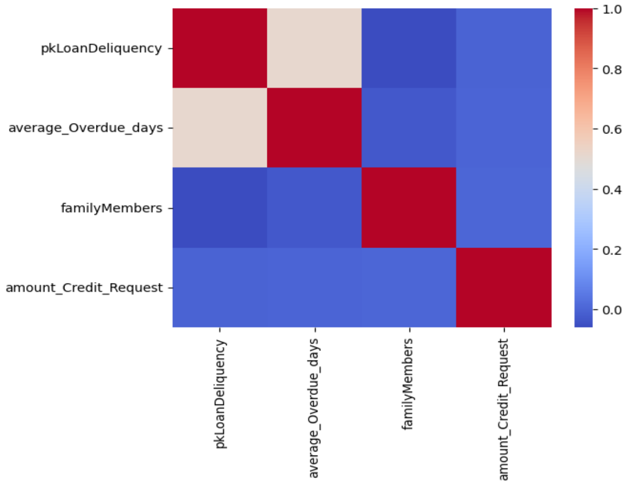

Figure 7. Heatmap showing the correlation of predominant variables

3.7.3. Transformation

In this stage, the feature selection process is executed, indicating the most significant variables and extracting these variables to be used in the generation of the model. Feature selection shows which variable adds most of the prediction of loan defaults, to describe an essential part of the model using Python see the Figure 8, feature selection. Feature selection is processed with the code written in Python, see Figure 9.

```
feature_imp = pd.DataFrame(list(zip(cols,xgb.feature_importances_)),columns=['feature','importance'])
feature_imp_sorted = feature_imp.sort_values(by='importance',ascending=False)
fig,ax = plt.subplots(figsize=(15,8))
feature_imp_sorted[:20].plot(kind='bar',x='feature',y='importance',ax=ax)
plt.tight_layout()
```

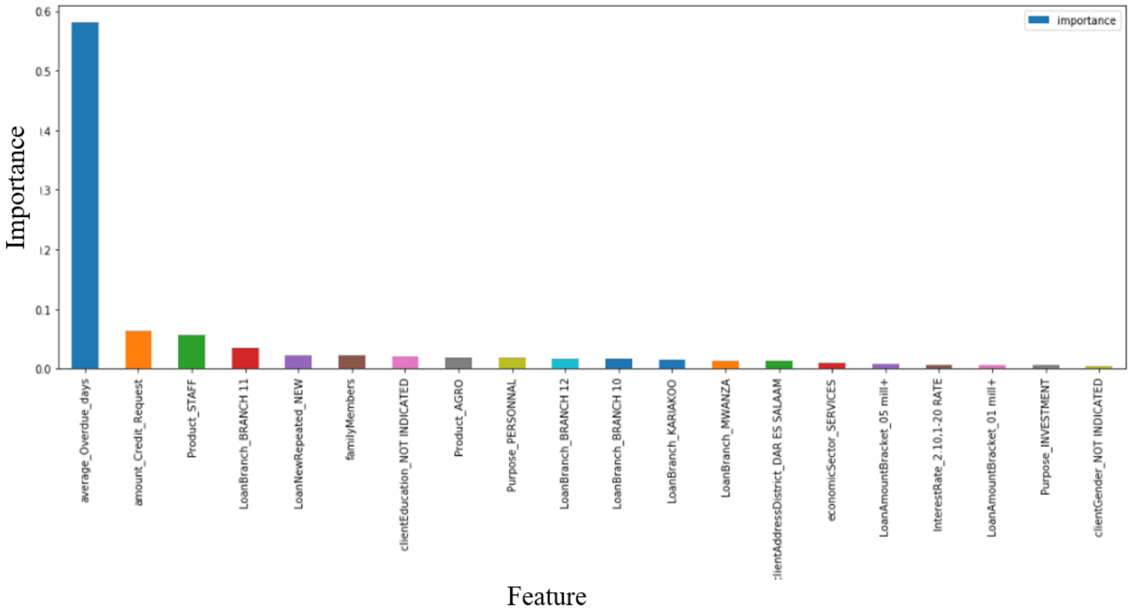

Figure 8. Feature selection

```python
# Extract Features and Target Variable
features = data.iloc[:, :-1].values  # All columns except the last one (assumed to be the target)
target = data['CreditRiskScore'].values  # Last column as target variable
```

```python
# Step 3: Feature Scaling
# Normalize the features to ensure all have the same scale
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
features = scaler.fit_transform(features)
```

Figure 9. Feature selection code

3.7.4. Datamining - pattern

In this stage, the data splitting process is carried out. For the present research, the practice of using 70% for training and 30% for validation was applied. To examine the code responsible for data splitting and training invocation, the dataset is first partitioned into training and testing subsets to enable robust model validation and mitigate the risk of overfitting, as illustrated in Figure 10. To identify key drivers of loan default we need to see the importance of features, see the Figure 11.

```python
# Step 4: Split the Data into Training and Testing Sets
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(features, target, test_size=0.3, random_state=4
```

```python
# Step 5: Train a Gradient Boosting Regression Model
from sklearn.ensemble import GradientBoostingRegressor
model = GradientBoostingRegressor(n_estimators=100, learning_rate=0.1, random_state=42)
model.fit(X_train, y_train)
```

```
GradientBoostingRegressor
GradientBoostingRegressor(random_state=42)
```

Figure 10. Main code for splitting and calling data training

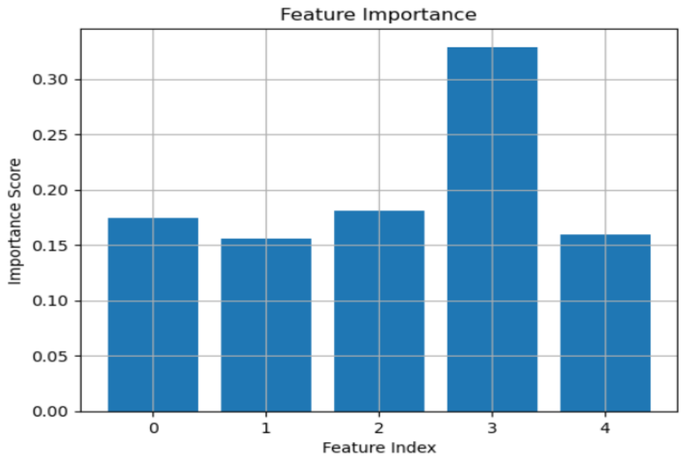

Figure 11. Feature importance

3.7.5. Evaluation-knowledge

After executing the training and testing/validation process, the Table 2 was obtained. Table 2 presents the evaluation metrics information applied to the same data for four different machine learning models i.e., random forest, logistic regression, decision tree, and XGBoost. Included metrics are accuracy, precision, recall, and F1-score for both classes.

Table 2. Comparison table of the different models applied to the same data

Model	Accuracy	Good				Bad				Total records
		Precision	Recall	F1-score	Support	Precision	Recall	F1-score	Support	
Decision tree	0.98	0.99	0.99	0.99	27,194	0.82	0.80	0.81	1,210	28,404
Random forest	0.99	0.99	1	0.99	27,194	0.93	0.81	0.85	1,210	28,404
Logistic regression	0.96	0.96	1	0.98	27,194	0.00	0.00	0.00	1,210	28,404
XGBoost	0.99	0.99	1	0.99	27,195	0.89	0.84	0.86	1,209	28,404

The XGBoost method allows for more accurate classification of information to assess credit, significantly contributing to the reduction of credit risk in clients. Similarly, these techniques can be implemented in cooperatives across the country to segment clients into differentiated groups, facilitating more efficient credit management based on advanced supervised learning models.

4. RESULTS AND DISCUSSION

In this study, a comparison of the XGBoost algorithm with other algorithms was not conducted, as various studies have established that, in predictive models of loan default with large and complex datasets, XGBoost outperforms other high-performance ensemble methods such as random forest and gradient boost.

This study proposes a predictive machine learning model based on XGBoost for agricultural microfinance loan defaults during the application stage, incorporating environmental factors induced by climate, such as temperature, precipitation, humidity, among others. XGBoost was selected due to its high efficiency, portability, and flexibility as a distributed system for gradient boosting. It leverages the gradient boosting framework, which encompasses generalized linear models and gradient-boosted decision trees.

Key features include support for custom loss functions, robust handling of missing values, and parallel processing capabilities, making it more accurate and efficient than traditional gradient-boosted decision trees. Studies have confirmed that XGBoost is the most efficient algorithm when working with nonlinear variables such as climate-induced environmental factors. Although with different variables and weights, previous studies have clearly demonstrated the effectiveness of the XGBoost algorithm in various loan default prediction models [23].

The research indicates that optimal results are achieved by allocating 30% of the data for testing, reserving 70% for training. Consequently, we divided the dataset into the two corresponding subsets [24]. The XGBoost model showcased great performance as it attained an overall accuracy of 99%, a recall of 84%, a precision of 89%, and an F1-score of 86%. As compared to other models utilized in this research study, such as decision tree, logistic regression, and random forest, this model's results outperformed than these models and identified loan defaults accurately [25]. For example, logistic regression model did not identify defaults accurately and attained an F-1 score of 0% while XGBoost model attained an F-1 score of 86% as this study prioritized minority class prediction performance [26].

The strength of the proposed model lies in its ability to tackle complex and non linerar interactions between environmental variables and borrower data. Unlike traditional models that fail to capture effectively, XGBoost model provides a robust and reliable approach for credit risk prediction that is influenced by environmental factors significantly.

This type of study faces several limitations, particularly related to data availability. Similar to the challenges highlighted in other research [27], access to specific borrower data by financial institutions or at the farm level was restricted due to confidentiality concerns. As a result, key financial indicators such as debt-to-asset ratios, insurance coverage, and other risk metrics, which are essential for evaluating a borrower's repayment capacity, were excluded from our model.

Additionally, the research has shown that factors such as crop prices and agricultural yields can significantly enhance predictive accuracy [27]. However, such granular data was not available in this study. If future research can integrate this type of detailed financial and agricultural data, the predictive accuracy of models in agricultural finance is likely to improve.

5. CONCLUSION

This XGBoost-based model leverages borrower profiles, loan characteristics, agricultural data, and climate-induced environmental data to accurately predict agricultural loan defaults. By integrating climate data, the model can help microfinance institutions assess risks more accurately, thereby improving their credit evaluation process.

XGBoost was selected as the most suitable model for predicting credit default due to its ability to capture complex, non-linear relationships between variables, which is essential when incorporating factors such as environmental data, borrower characteristics, and loan details. Its performance-optimized design allows it to handle large volumes of high-dimensional data, typical of microfinance institutions, and its parallel computing capability significantly speeds up the training time. Additionally, XGBoost minimizes overfitting by balancing predictive accuracy and model complexity through its regularized objective function. These features, combined with its flexibility to implement custom loss functions and its efficient handling of missing data, make it a benchmark tool in risk assessment problems, providing accurate and explainable predictions that support robust financial decisions.

This approach can be further expanded through geospatial analysis to assess region-specific climate impacts or through time series analysis if historical climate data is available over time.

FUNDING INFORMATION

The authors express their gratitude to the Universidad Nacional Mayor de San Marcos, Peru. This study forms part of the author's doctoral thesis work and was conducted as a requirement for the completion of the degree program; no external funding was received for its development.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Eladio Alfredo Soto Alvarez	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
Miguel Angel Cano Lengua		✓		✓	✓			✓		✓	✓		✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] Y. Yang *et al.*, "Climate change exacerbates the environmental impacts of agriculture," *Science*, vol. 385, no. 6713, 2024, doi: 10.1126/science.adn3747.

[2] X. Yuan *et al.*, "Impacts of Global Climate Change on Agricultural Production: A Comprehensive Review," *Agronomy*, vol. 14, no. 7, pp. 1–19, Jun. 2024, doi: 10.3390/agronomy14071360.

[3] A. Rawat, D. Kumar, and B. S. Khati, "A review on climate change impacts, models, and its consequences on different sectors: a systematic approach," *Journal of Water and Climate Change*, vol. 15, no. 1, pp. 104–126, Jan. 2024, doi: 10.2166/wcc.2023.536.

[4] S. Y. Chia and M. W. Lim, "A critical review on the influence of humidity for plant growth forecasting," *IOP Conference Series: Materials Science and Engineering*, vol. 1257, no. 1, pp. 1–7, Oct. 2022, doi: 10.1088/1757-899x/1257/1/012001.

[5] J. P. C. Eekhout, J. E. Hunink, W. Terink, and J. De Vente, "Why increased extreme precipitation under climate change negatively affects water security," *Hydrology and Earth System Sciences*, vol. 22, no. 11, pp. 5935–5946, 2018, doi: 10.5194/hess-22-5935-2018.

[6] F. Bibi and A. Rahman, "An Overview of Climate Change Impacts on Agriculture and Their Mitigation Strategies," *Agriculture*, vol. 13, no. 8, pp. 1–15, 2023, doi: 10.3390/agriculture13081508.

[7] G. R. Dixon, "Climate change-impact on crop growth and food production, and plant pathogens," *Canadian Journal of Plant Pathology*, vol. 34, no. 3, pp. 362–379, 2012, doi: 10.1080/07060661.2012.701233.

[8] M. S. Céu and R. M. Gaspar, "A Review on Climate Change, Credit Risk and Agriculture," *Rural Sustainability Research*, vol. 51, no. 346, pp. 38–49, 2024, doi: 10.2478/plua-2024-0004.

[9] K. Sandeep, Shilpa, A. Sharma, and S. Sharma, "Microfinance in Indian agriculture: Cultivating prosperity in rural India," *Vigyan Varta an International E-Magazine for Science Enthusiasts*, vol. 5, no. 3, pp. 34–37, 2024.

[10] S. Mondol, I. Satpathy, B. C. M. Patnaik, A. Nayak, A. Patnaik, and A. Khang, *Impact of AI and Data in Revolutionizing Microfinance in Developing Countries: Improving Outreach and Efficiency*, Auerbach Publications, 2024, doi:

10.1201/9781032618845-5.
[11] F. Noma and S. Babu, "Predicting climate smart agriculture (CSA) practices using machine learning: A prime exploratory survey," *Climate Services*, vol. 34, 2024, doi: 10.1016/j.cliser.2024.100484.
[12] A. Jemal, A. Araba, and A. Dima, "Integration of satellite data for predicting crop yields in Eastern Ethiopia using machine learning," *Scientific Reports*, vol. 15, 2025, doi: 10.1038/s41598-025-00810-z.
[13] H. Ritchie, "How will climate change affect crop yields in the future?," *Our World in Data*. 2024. [Online]. Available: https://ourworldindata.org/will-climate-change-affect-crop-yields-future. (Accessed: Dec. 21, 2024).
[14] National Aeronautics and Space Administration, "Climate change and its environmental impacts on crop growth," NASA at Your Table, [Online]. Available: https://climate.nasa.gov/news/. (Accessed: Dec. 19, 2024).
[15] N. Metawa, S. Afchal, and N. El-Kanj, "AI-based model for Enhancing Credit Risk and Delinquency Management in Banks," *Journal of Cybersecurity and Information Management*, vol. 16, no. 1, pp. 282-300, 2025, doi: 10.54216/JCIM.160120.
[16] S. Zheng, J. Wang, C. Rizos, W. Ding, and A. El-Mowafy, "Simultaneous Localization and Mapping (SLAM) for Autonomous Driving: Concept and Analysis," *Remote Sensing*, vol. 15, no. 4, p. 1156, 2023, doi: 10.3390/rs15041156.
[17] J. Wang, W. Rong, Z. Zhang, and D. Mei, "Credit Debt Default Risk Assessment Based on the XGBoost Algorithm: An Empirical Study from China," *Wireless Communications and Mobile Computing*, pp. 1–14, 2022, doi: 10.1155/2022/8005493.
[18] Y. Chang, Y. Jin, Q. Xing, Y. Zhang, S. Guo, and S. Meng, "Advanced user credit risk prediction model using LightGBM, XGBoost and Tabnet with SMOTEENN," *arXiv preprint*, 2024, doi: 10.48550/arXiv.2408.03497.
[19] W. Wu, "Machine Learning Approaches to Predict Loan Default," *Intelligent Information Management*, vol. 14, no. 05, pp. 157–164, 2022, doi: 10.4236/iim.2022.145011.
[20] N. Suhadolnik, J. Ueyama, and S. D. Silva, "Machine Learning for Enhanced Credit Risk Assessment: An Empirical Approach," *Journal of Risk and Financial Management*, vol. 16, no. 12, pp. 1–21, 2023, doi: 10.3390/jrfm16120496.
[21] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
[22] Y. Jiang, "Predicting Loan Default: A Comparative Analysis of Multiple Machine Learning Models," *Highlights in Science, Engineering and Technology*, vol. 85, pp. 169–175, 2024, doi: 10.54097/10dk2m95.
[23] Z. Li, S. Li, Z. Li, Y. Hu, and H. Gao, "Application of XGBoost in P2P Default Prediction," *Journal of Physics: Conference Series*, vol. 1871, no. 1, pp. 1–10, 2021, doi: 10.1088/1742-6596/1871/1/012115.
[24] M. Zhu, Y. Zhang, Y. Gong, K. Xing, X. Yan, and J. Song, "Ensemble Methodology: Innovations in Credit Default Prediction Using LightGBM, XGBoost, and LocalEnsemble," in *2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI)*, Changchun, China, 2024, pp. 421–426, doi: 10.1109/ICETCI61221.2024.10594630.
[25] H. Pallathadka, M. Mustafa, D. Sanchez, G. Sajja, S. Gour, and M. Naved, "Impact of machine learning on management, Healthcare and Agriculture," *Materials Today: Proceedings*, vol. 80, part 3, pp. 2803-2806, 2023, doi: 10.1016/j.matpr.2021.07.042.
[26] J.-Y. Shih, "Application of XGBoost Algorithm to Develop Mutual Fund Marketing Prediction Model for Banks Wealth Management," *Engineering Proceedings*, vol. 89, no. 1, pp. 1-9, 2025, doi: 10.3390/engproc2025089003.
[27] J. Chen, A. L. Katchova, and C. Zhou, "Agricultural Loan Delinquency Prediction Using Machine Learning Methods," *International Food and Agribusiness Management Review*, vol. 24, no. 5, pp. 797–812, 2021, doi: 10.22434/IFAMR2020.0019.

BIOGRAPHIES OF AUTHORS

Eladio Alfredo Soto Alvarez is a senior IT advisor and international consultant in finance and digital transformation. He has worked for several years where he has led technological projects for financial institutions across Europe, Eastern Europe, South Asia, and Africa. He holds a Master's in IT Governance and is currently a Ph.D. candidate in Systems and Informatics Engineering at Universidad Nacional Mayor de San Marcos (UNMSM), Peru. His research focuses on applying machine learning to credit risk prediction in microfinance, with a strong emphasis on digital inclusion and technological innovation for development. He can be contacted at email: alfredo.soto.as@gmail.com.

Miguel Angel Cano Lengua is a Ph.D. Engineering of Systems and Computer Science from the Universidad Nacional Mayor de San Marcos, a Master's in systems engineering from the Universidad Nacional del Callao (UNAC), a professor at the Universidad Tecnologica del Peru (UTP) and teaching researcher in the Universidad Nacional Mayor de San Marcos (UNMSM), has a degree in Mathematics. He works on continuous optimization, artificial intelligence algorithms, conical programming, numerical methods, methodology, and software design. He can be contacted at email: mcanol@unmsm.edu.pe.