Management Studies and Entrepreneurship Journal
Vol 6(6) 2025:617-632

Artificial Intelligence In Predictive Analytics For Advancing Credit Risk Management In The Digital Economy

Kecerdasan Buatan Dalam Analisis Prediktif Untuk Meningkatkan Pengelolaan Risiko Kredit Dalam Ekonomi Digital

Putri Sarah Olivia[1*], Adam Puspabhuana[2], Dwi Winarno[3]
Universitas Cakrawala, Jakarta, Indonesia[1,2,3]
putri.sarah@cakrawala.ac.id[1*], adambhuana@cakrawala.ac.id[2], dwiwin@cakrawala.ac.id[3]

**Coresponding Author*

ABSTRACT
The rapid advancement of digital technologies has encouraged the banking sector to adopt Artificial Intelligence (AI)-based approaches for credit risk management. Traditional credit scoring methods often lack accuracy in identifying default risks, particularly for unbanked and underbanked groups, leading to higher Non-Performing Loan (NPL) rates. This research addresses the need for a more adaptive, accurate, and inclusive credit risk assessment system in the digital economy era. This research aims to develop and evaluate an AI-driven predictive analytics model for credit risk assessment by comparing the performance of machine learning algorithms, such as Logistic Regression, Random Forest, XGBoost, and Deep Learning. The dataset comprises customer demographics (such as age and income), details of their banking relationship (including mortgage and securities account), and their response to the most recent personal loan campaign. The comparative analysis indicates that Random Forest substantially outperformed the other models, demonstrating superior accuracy (98.80%) alongside balanced precision (93.75%) and recall (93.75%), as well as the highest ROC-AUC (99.86%). These results highlight its robustness in both classification performance and discriminatory power. XGBoost and Deep Learning followed, showing competitive but lower predictive capabilities. In contrast, Logistic Regression exhibited clear limitations, yielding the lowest accuracy (90.40%) and precision (50%), despite achieving a relatively high recall (92.71%) and ROC-AUC (96.77%). This suggests that while Logistic Regression can identify positive cases, its overall reliability and precision are insufficient compared to advanced ensemble and deep learning methods.
Keywords : *Predictive Analytics, Credit Risk, Digital Banking, Machine Learning, Accuracy*

ABSTRAK
Perkembangan teknologi digital yang pesat telah mendorong sektor perbankan untuk mengadopsi pendekatan berbasis Kecerdasan Buatan (AI) dalam manajemen risiko kredit. Metode penilaian kredit tradisional seringkali kurang akurat dalam mengidentifikasi risiko gagal bayar, terutama untuk kelompok yang tidak memiliki akses perbankan dan kurang terlayani, yang mengakibatkan tingkat Kredit Macet (NPL) yang lebih tinggi. Penelitian ini bertujuan untuk memenuhi kebutuhan akan sistem penilaian risiko kredit yang lebih adaptif, akurat, dan inklusif di era ekonomi digital. Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model analitik prediktif berbasis AI untuk penilaian risiko kredit dengan membandingkan kinerja algoritma pembelajaran mesin, seperti Regresi Logistik, Hutan Acak, XGBoost, dan Pembelajaran Mendalam. Data yang digunakan mencakup demografi pelanggan (seperti usia dan pendapatan), rincian hubungan perbankan mereka (termasuk pinjaman hipotek dan rekening sekuritas), serta respons mereka terhadap kampanye pinjaman pribadi terbaru. Analisis perbandingan menunjukkan bahwa Random Forest secara signifikan outperformed model-model lain, dengan akurasi yang superior (98,80%) disertai presisi yang seimbang (93,75%) dan recall (93,75%), serta nilai ROC-AUC tertinggi (99,86%). Hasil ini menyoroti ketahanan model dalam kinerja klasifikasi dan daya diskriminatifnya. XGBoost dan Deep Learning mengikuti, menunjukkan kemampuan prediktif yang kompetitif namun lebih rendah. Di sisi lain, Regresi Logistik menunjukkan keterbatasan yang jelas, dengan akurasi terendah (90,40%) dan presisi (50%), meskipun mencapai recall yang relatif tinggi (92,71%) dan ROC-AUC (96,77%). Hal ini menunjukkan bahwa meskipun Regresi Logistik dapat mengidentifikasi kasus

positif, keandalan dan presisi secara keseluruhan tidak memadai dibandingkan dengan metode ensembel canggih dan deep learning.
Kata Kunci: Analisis Prediktif, Risiko Kredit, Perbankan Digital, Pembelajaran Mesin, Akurasi

1. Introduction

There should be a top and a bottom margin of 3.0 cm, with right and left margins of length 3 cm. The digital transformation of the financial services sector has accelerated the adoption of intelligent technologies, particularly Artificial Intelligence (AI), across various operational functions of banking, including credit risk management (de Keijzer et al., 2021; Georgewill & Gabriel, 2024; Lee et al., 2022; Sadok et al., 2022). A central challenge currently faced by the banking industry is how to maintain credit quality amid the pressures of the digital economy, financial inclusion initiatives, and the diversification of customer profiles (Boobier, 2020a; Thomas & Raphael, 2022). One of the most critical indicators of credit portfolio health is the Non-Performing Loan (NPL) ratio (Widyastuti et al., 2023), which often rises due to inaccuracies in the initial credit risk assessment process (Umamaheswari, 2024).

Conventional approaches that rely primarily on historical data-based credit scoring remain insufficient to capture the changing nature of risk, especially in the case of customers with limited or undocumented credit histories (Gatla, 2024). Because of this, using AI-driven predictive analytics to combine different data sources and digital transaction behaviors looks like a good way to make default probability estimation more accurate (C. Li et al., 2024; Wang et al., 2021).

Nonetheless, a deficiency persists in the financial literature concerning the efficacy of AI applications in credit risk assessment when directly implemented within the framework of Indonesian banking practices (Kurniawan, 2024). This study seeks to address that gap by conducting a quantitative analysis within a national banking institution. The urgency of this research stems from the need of the banking industry to develop adaptive, efficient, and context-specific credit risk management systems that align with the characteristics of Indonesia's digital financial market.

2. Literature Review

Prior research has demonstrated that, in comparison to traditional techniques like logistic regression, the use of machine learning for credit scoring can greatly increase accuracy (Addo et al., 2018a; Karami & Igbokwe, 2025; Shi et al., 2022; Wang et al., 2021). For instance, Lessmann et al. showed that algorithms like gradient boosting and random forest can offer better predictive performance when applied to credit data (Challa, 2024; Kurniawan, 2024). However, most of these studies are highly technical in nature and conducted in the context of developed countries, which may not necessarily align with the characteristics of customers and data in developing countries such as Indonesia (Dogra et al., 2022b).

In several developing countries in Asia, a study by Dastile et al. (2020) found that most credit scoring approaches still rely on traditional statistical models, which have limitations in accommodating nonconventional data such as digital behavior or customers' electronic footprints. This research introduces several novelties, including:

1. An empirical study based on actual data from a national banking institution, rather than simulations or secondary datasets;
2. A managerial and financial perspective, in contrast to the majority of prior studies that primarily focus on algorithmic or technical aspects (Addo et al., 2018b; Badawy et al., 2023);
3. The incorporation of alternative digital data to enhance predictive model performance (Bi & Liang, 2022; Dastile et al., 2020);
4. Practical model validation through expert judgment from risk practitioners (Bose et al., 2023; Challa, 2024; Karami & Igbokwe, 2025).

Furthermore, an explainable AI (XAI) approach will be employed to identify and interpret the key variables influencing credit risk within the model, thereby supporting transparency in decisionmaking (W. Bi & Liang, 2022; Bose et al., 2022; Galaz et al., 2021).

3. Research Methods

In the era of the digital economy, this research methodology is methodically designed to investigate how Artificial Intelligence (AI) technology can be applied in predictive analytics to improve credit risk management systems in the banking industry (Aldboush & Ferdous, 2023; Dogra et al., 2022a; Hassan et al., 2021). Using a quantitative methodology, the study focuses on applying and assessing machine learning algorithms on empirical data gathered from a national banking organization that is the research partner. This research aims to objectively and measurably evaluate credit risk prediction models through analysis of historical financial data available in the bank's information system.

The information used in this study is secondary data sourced from the internal banking credit information system, including loan history data, customer profiles, transaction data, and other credit risk parameters (Fares et al., 2023; Langenbucher & Corcoran, 2021; Lau & Leimer, 2019). The research population comprises all individual customers of the bank who have credit application histories during the past three years. The sampling technique employs purposive sampling, with criteria for customers who possess complete and relevant data for the variables under investigation (Biswas et al., 2020; Boobier, 2020b). From the available population, a sample of 5,000 customers was determined, which is considered adequate for statistical analysis and machine learning model training with accountable accuracy (Khan et al., 2024). For reference, research flowchart is displayed in the following figure:

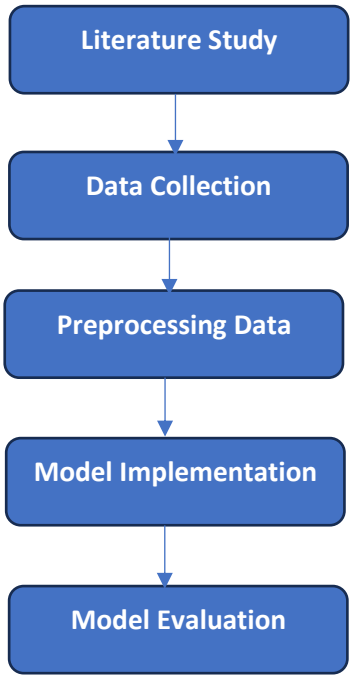

Fig. 1. Research and Methodology Flow Chart

To create a theoretical framework and methodological mapping, the research process starts with a thorough literature review. Data collection and preprocessing, which includes data cleaning, variable transformation, normalization, and data separation for model training and testing, are the next steps (S. Bi & Bao, 2024; Debbadi & Boateng, 2025; Dumitrescu et al., 2022; Uddin et al., 2022). AUC-ROC, precision, recall, and confusion matrix are among the performance metrics that will be used to evaluate the predictive models that will be tested, which include

Logistic Regression, Random Forest, and XGBoost (Y. Li et al., 2022; Qin et al., 2021; Rao et al., 2023; Wang et al., 2021).

Data Collection

The dataset comprises information on 5,000 customers, encompassing demographic characteristics (e.g., age, income), banking relationships (e.g., mortgage ownership, securities accounts), and responses to the most recent personal loan campaign. Out of these 5,000 samples, 480 individuals (9.6%) took advantage of the loan that was made available during the prior campaign.

The dataset is complete, with no missing or undefined (NaN) values. It consists of both numerical and categorical variables, with all categorical attributes encoded numerically. Additionally, several predictor variables exhibit substantial skewness (long-tailed distributions), which introduces an element of complexity to the preprocessing stage, although it does not pose significant challenges.

Table 1. Attribute Description

Attribute Name	Description	Example Values
ID	Unique identifier assigned to each customer	1, 2, …
Age	Age of the customer in completed years	25, 45, …
Experience	Number of years of professional experience	1, 19, …
Income	Customer's annual income (in thousands of USD)	49, 34, …
Zip Code	Residential ZIP code of the customer	91107, 90089,…
Family	Customer's family size (1 = single, up to 4 = four or more members)	4, 3, …
Attribute Name	**Description**	**Example Values**
CCAvg	Average monthly credit card expenditure (in thousands of USD)	1.60, 1.50, …
Education	Educational attainment (1 = Undergraduate; 2 = Graduate; 3 = Advanced/Professional)	1, 2, …
Mortgage	If the customer has a home mortgage, its value (in thousands of USD)	0, 101, …
Personal Loan	Indicator of whether the customer accepted the personal loan in the last campaign (0 = No, 1 = Yes)	0, 1
Securities Account	Indicator of whether the customer holds a securities account with the bank (0 = No, 1 = Yes)	0, 1
CD Account	Indicator of whether the customer holds a certificate of deposit (CD) account (0 = No, 1 = Yes)	0, 1
Online	Indicator of whether the client makes use of online banking services (0 = No, 1 = Yes)	0, 1
Credit Card	Indicator of whether the customer uses a credit card issued by Universal Bank (0 = No, 1 = Yes)	0, 1

The explanation of the variables presented in Table 1 can be further grouped into several categories for clarity. Personal Loan is the study's target variable, and it indicates whether a customer accepted a personal loan that was offered during the prior campaign (0 = No, 1 = Yes). This constitutes the dependent variable to be predicted.

Age (the customer's age in years) and Experience (the number of years of professional experience) are among the demographic characteristics, serving as a proxy for financial stability), Income (annual income in thousands of USD, representing the primary repayment capacity), ZIP

Code (residential area indicator reflecting socio-demographic characteristics), and Family (family size, which may influence financial needs and obligations).

CCAvg, or average monthly credit card spending in thousands of USD, is a representation of the behavioral characteristics, reflecting spending patterns and financial capacity) and Online (whether the customer uses internet banking facilities, which may indicate technological adoption and engagement with the bank).

The product ownership variables, which include securities account (ownership of a securities account), certificate of deposit (ownership of a certificate of deposit), credit card (ownership of a bank-issued credit card), and mortgage (outstanding mortgage value, which serves as an indicator of relationship depth and financial sophistication), capture the customer's financial relationship with the bank.

Lastly, the education variable represents the customer's educational attainment (1 = Undergraduate, 2 = Graduate, 3 = Advanced/Professional), which may correlate with income potential and financial literacy.

Preprocessing Data

Data Structure

Table 2. Data Structure

RangeIndex: 5000 entries, 0 to 4999			
Data columns (total 14 columns):			
#	**Kolom**	**Non-Null Count**	**Dtype**
0	ID	5000 non-null	int64
1	Age	5000 non-null	int64
2	Experience	5000 non-null	int64
3	Income	5000 non-null	int64
4	ZIP Code	5000 non-null	int64
5	Family	5000 non-null	int64
6	CCAvg	5000 non-null	object
7	Education[2]	5000 non-null[3]	int64[4]
8	Mortgage[6]	5000 non-null[7]	int64[8]
9	Personal Loan[10]	5000 non-null[11]	int64[12]
10	Securities Account[14]	5000 non-null[15]	int64[16]
11	CD Account[18]	5000 non-null[19]	int64[20]
12	Online[22]	5000 non-null[23]	int64[24]
13	CreditCard[26]	5000 non-null[27]	int64
dtypes: int64(13), object(1)			
memory usage: 547.0+ KB			

The dataset consists of 5,000 observations with a total of 14 variables. As shown in the data summary, all variables are complete with no missing values, indicating a fully populated dataset. Among these variables, 13 are stored as integer data types (int64), while one variable (CCAvg) is recorded as an object type, which suggests the need for data type conversion during preprocessing.

The variables include customer identifiers (ID), demographic attributes (Age, Experience, Income, ZIP Code, Family, Education), behavioral indicators (CCAvg, Online), product ownership information (Mortgage, Securities Account, CD Account, CreditCard), and the target variable (Personal Loan). The dataset occupies approximately 547 KB of memory, making it computationally manageable for further statistical analysis and machine learning applications.

Even after they have been defined in the abstract, define acronyms and abbreviations when they are used for the first time in the text. It is not necessary to define abbreviations like IEEE, SI, MKS, CGS, sc, dc, and rms. Unless absolutely necessary, avoid using abbreviations in the title or headings.

Data Structure

Table 3. Univariate Analysis

Statistik	Age	Experience	Income	ZIP Code	Family	CCAvg	Education
count	5000.00	5000.00	5000.00	5000.00	5000.00	5000.00	5000.00
mean	45.34	20.10	73.77	93152.50	2.40	1.94	1.88
std	11.46	11.47	46.03	2121.85	1.15	1.75	0.84
min	23.00	-3.00	8.00	9307.00	1.00	0.00	1.00
25%	35.00	10.00	39.00	91911.00	1.00	0.70	1.00
50%	45.00	20.00	64.00	93437.00	2.00	1.50	2.00
75%	55.00	30.00	98.00	94608.00	3.00	2.50	3.00
max	67.00	43.00	224.00	96651.00	4.00	10.00	3.00

Statistik	Mortgage	Personal Loan	Securities Account	CD Account	Online	CreditCard
count	5000.00	5000.00	5000.00	5000.00	5000.00	5000.00
mean	56.50	0.10	0.10	0.06	0.60	0.29
std	101.71	0.29	0.31	0.24	0.49	0.46
min	0.00	0.00	0.00	0.00	0.00	0.00
25%	0.00	0.00	0.00	0.00	0.00	0.00
50%	0.00	0.00	0.00	0.00	1.00	0.00
75%	101.00	0.00	0.00	0.00	1.00	1.00
max	635.00	1.00	1.00	1.00	1.00	1.00

Descriptive information about the dataset's properties is provided by the univariate analysis. With a mean age of 45.34 years and a range of 23 to 67 years, the sample is primarily composed of middle- aged people. Work experience averages 20.10 years, though the presence of negative values suggests inconsistencies that require data cleaning. The annual income shows substantial variation, with a mean of USD 73,770 and a maximum of USD 224,000, reflecting heterogeneity in repayment capacity. Family size averages 2.40 members, and credit card spending (CCAvg) averages USD 1,940 per month, though the distribution is highly skewed with extreme values.

Regarding education, the mean value of 1.88 suggests that most customers are between the undergraduate and graduate levels, which may correlate with income potential and financial literacy. The mortgage variable exhibits wide dispersion, with a mean of USD 56,500 and a maximum of USD 635,000, indicating significant differences in housing-related debt. For the target variable, only 9.6% of customers accepted the personal loan, highlighting a considerable class imbalance that must be addressed in predictive modeling.

The binary features show further distinctions: only 10% of customers hold a securities account, 6% maintain a CD account, and 29% own a credit card issued by the bank, reflecting relatively low penetration of certain financial products. By contrast, online banking adoption is comparatively high, with 60% of customers using digital banking services.

Outlier Detection

Table 4. Outlier Detection

Variable	Outlier Count	Outlier Percentage	Lower Bound	Upper Bound
Age	0	0.00	5.0	85.0
Experience	0	0.00	-20.0	60.0
Income	96	1.92	-49.5	186.5
CCAvg	324	6.48	-2.0	5.2
Family	0	0.00	-2.0	6.0

The outlier detection analysis was conducted to assess the distributional properties of key numerical variables. As presented in Table, no outliers were identified in the variables Age, Experience, and Family, indicating that the observed values fall within the expected range of variation.

For Income, 96 cases (1.92% of the total observations) were detected as outliers, exceeding the calculated upper bound of 186.5 (in thousands of USD). Similarly, CCAvg (average monthly credit card expenditure) exhibits the highest number of outliers, with 324 cases (6.48%) identified beyond the upper bound of 5.2 (in thousands of USD). These findings suggest the presence of extreme spending behavior and income variability among a subset of customers.

Although the proportion of outliers in Income and CCAvg remains relatively small, their potential influence on statistical modeling and predictive performance warrants consideration. Appropriate preprocessing strategies, such as transformation, capping, or robust modeling techniques, may be applied to mitigate the effects of these extreme values while preserving the integrity of the dataset.

Correlation Analysis

Fig. 2. Correlation Analysis

Model The correlation analysis with the target variable (Personal Loan acceptance) reveals several notable relationships. Income has the strongest positive correlation (0.502) of any predictor, indicating that borrowers who earn more money annually are more likely to accept personal loan offers. This is followed by average monthly credit card expenditure (CCAvg) and CD account ownership, which show moderate positive correlations of 0.367 and 0.316, respectively, indicating that spending behavior and product diversification play an important role in loan acceptance decisions.

Other variables, including mortgage value (0.142), education level (0.137), and family size (0.061), exhibit weak positive correlations, suggesting that while these factors contribute to loan acceptance, their influence is relatively limited. Similarly, securities account ownership and

online banking usage show very weak positive associations with the target variable.

Conversely, age (–0.008) and work experience (–0.007) display weak negative correlations, indicating that these demographic factors have minimal and slightly inverse effects on personal loan uptake.

Overall, the results suggest that financial capacity indicators—particularly income, spending behavior, and product ownership—are the most influential factors associated with personal loan acceptance, whereas demographic and minor product usage variables play a lesser role.

Implementation

Logistic Regression

In predictive analytics, logistic regression is a basic statistical learning technique that is widely applied to binary and multinomial classification tasks. In contrast to linear regression, logistic regression models the likelihood of class membership using the logistic function (sigmoid function), guaranteeing that output values stay within the range of 0 and 1.

The mathematical foundation of logistic regression is expressed through the logistic function:

$$P(Y=1|X) = 1 / (1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n)})$$

Where P(Y=1|X) denotes the probability of the positive class given input features X, and β coefficients represent the model parameters estimated through maximum likelihood estimation [1]. The findings generally imply that the most significant factors linked to the acceptance of personal loans are financial capacity indicators, specifically income, spending patterns, and product ownership, while demographic and minor product usage variables have a smaller impact.

Random Forest

Random Forest is an ensemble learning technique that reduces overfitting and improves predictive accuracy by bootstrap aggregating (bagging) multiple decision trees. Developed by Breiman, this algorithm introduces randomness at two levels: randomly selecting features at each node split and randomly sampling training observations.

Algorithmic Process:

1. Generate B bootstrap samples from the original training dataset
2. Construct decision trees using random feature subsets at each split
3. Aggregate predictions through majority voting (classification) or averaging (regression) The mathematical representation for classification is:

$$\hat{y} = mode\{T_1(x), T_2(x), ..., T_\beta(x)\}$$

where $T_\beta(x)$ depicts the prediction of the b-th tree and B symbolizes the total number of trees.

Extreme Gradient Boosting (XGBoost)

In order to achieve state-of-the-art performance in structured data prediction tasks, XGBoost is an optimized implementation of gradient boosting algorithms that incorporates sophisticated regularization techniques and computational optimizations. Unlike bagging methods, XGBoost employs sequential learning where each subsequent model rectifies errors made by previous models.

Mathematical Framework:

$$Obj = \Sigma_i L(y_i, \hat{y}_i) + \Sigma_k \Omega(f_k)$$

where L represents the loss function, Ω denotes the regularization term, and f_k represents individual tree models.

Deep Learning

Deep Learning encompasses a family of artificial neural network architectures characterized by multiple hidden layers that enable automatic feature extraction and representation learning from raw data [39]. These models are excellent at identifying hierarchical patterns and intricate non-linear relationships in high-dimensional datasets.

Architectural Components:

1. Input Layer: Receives raw feature vectors
2. Hidden Layers: Numerous interconnected layers of neurons with non-linear activation functions
3. Output Layer: Produces final predictions through appropriate activation functions The forward propagation process is mathematically expressed as:

$$a^l = f(W^l a^{l-1} + b^l)$$

where a^l represents layer l activations, W^l denotes weight matrices, b^l represents bias vectors, and f is the activation function [39].

Training Process: Deep networks employ backpropagation algorithm with gradient descent optimization to minimize loss functions through iterative weight updates.

Model Evaluation

Confusion Matrix

By showing the correlation between actual and predicted class labels, the confusion matrix is a basic tabular representation that offers thorough visualization of classification model performance. This matrix provides comprehensive insights into model behavior across classes and forms the basis for calculating a variety of performance metrics.

The confusion matrix for a binary classification problem is organized as a 2x2 table: The Pearson correlation coefficient between variables X and Y is mathematically expressed as:

Table 5. Confusion Matrix

	Predicted	
	Positive	**Negative**
Actual Positive	TP	FN
Negative	FP	TN

where:

1. True Positives (TP): Correctly predicted positive instances
2. True Negatives (TN): Correctly predicted negative instances
3. False Positives (FP): Incorrectly predicted positive instances (Type I error)
4. False Negatives (FN): Incorrectly predicted negative instances (Type II error)

Accuracy

The primary performance indicator that measures the percentage of accurate predictions made in relation to all of a classification model's predictions is accuracy [45]. The mathematical formulation is:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

where:

1. TP: True Positives (correctly predicted positive instances)
2. TN: True Negatives (correctly predicted negative instances)
3. FP: False Positives (incorrectly predicted positive instances)
4. FN: False Negatives (incorrectly predicted negative instances)

Precision

Precision highlights the model's capacity to reduce false positive errors by calculating the percentage of correctly identified positive instances among all instances predicted to be positive. The mathematical expression is:

$$Precision = TP / (TP + FP)$$

Recall

Recall measures the percentage of correctly identified positive instances among all actual positive instances in the dataset. It is also referred to as sensitivity or true positive rate [48]. The mathematical formulation is:

$$Recall = TP / (TP + FN)$$

Receiver Operating Characteristic Area Under the Curve (ROC-AUC)

ROC-AUC provides a comprehensive performance metric that assesses classifier performance across all possible classification thresholds [49]. The True Positive Rate (Recall) is plotted against the False Positive Rate using the ROC curve:

$$False\ Positive\ Rate = FP / (FP + TN)\ True\ Positive\ Rate = TP / (TP + FN)$$

The Area Under the Curve (AUC) quantifies the overall discriminative ability of the classifier, with values ranging from 0 to 1.

1. **AUC = 1.0:** Perfect classifier with complete separation between classes
2. **AUC = 0.9-1.0:** Excellent classification performance
3. **AUC = 0.8-0.9:** Good classification performance
4. **AUC = 0.7-0.8:** Fair classification performance
5. **AUC = 0.6-0.7:** Poor classification performance
6. **AUC = 0.5:** Random classifier (no discriminative ability)
7. **AUC < 0.5:** Worse than random (predictions can be inverted) [50]

5. Results and Discussions

Logistic Regression

Table 6. Confusion Matrix of Logistic Regression

	Predicted	
	No	**Loan**
Actual No	815	89
Loan	7	89

The confusion matrix summarizes the classification outcomes for the loan prediction model. Out of the 904 customers who did not accept the loan, the model correctly classified 815 cases as “No Loan” (true negatives), while 89 cases were misclassified as “Loan” (false positives). Conversely, among the 96 customers who accepted the loan, the model successfully identified 89 cases as “Loan” (true positives), with only 7 cases incorrectly predicted as “No Loan” (false negatives).

These findings show that the model does well in identifying clients who are willing to take out loans (high recall for the “Loan” class), while still maintaining strong performance in identifying non-loan customers. However, the relatively higher number of false positives compared to false negatives suggests that the model has a tendency to overpredict loan acceptance, which may impact precision for the “Loan” class.

Overall, the confusion matrix shows how well the model predicts both classes, with stronger reliability in minimizing false negatives—a desirable property in credit risk applications, where failing to detect potential loan customers may represent a lost business opportunity.

Table 7. Logistic Regression Result

Kategori	Precision	Recall	F1-Score	Support
No Loan	0.99	0.90	0.94	904
Loan	0.50	0.93	0.65	96
Accuracy	-	-	0.90	1000
Macro Avg	0.75	0.91	0.80	1000
Weighted Avg	0.94	0.90	0.92	1000

The classification performance results, as summarized in the confusion matrix metrics, indicate an overall accuracy of 90%. For the “No Loan” class, the model achieved very high performance with a precision of 0.99, recall of 0.90, and an F1-score of 0.94, reflecting the model’s strong ability to correctly identify customers who did not accept the loan.

In contrast, the performance for the “Loan” class is less balanced. While the model obtained a relatively high recall of 0.93, designating that most loan-accepting customers were correctly detected, the precision is only 0.50, resulting in a lower F1-score of 0.65. This suggests that the prototype tends to produce a considerable number of false positives, predicting that some customers would accept the loan when in fact they did not.

The macro average F1-score of 0.80 highlights this imbalance between classes, whereas the weighted average F1-score of 0.92 remains high due to the dominance of the majority class (No Loan). These results show that the model is biased in favor of the majority class, but retains strong detection capability for the minority class at the expense of precision.

Random Forest

Table 8. Confusion Matrix of Random Forest

Kategori	Precision	Recall	F1-Score	Support
No Loan	0.99	0.90	0.94	904
Loan	0.50	0.93	0.65	96
Accuracy	-	-	0.90	1000
Macro Avg	0.75	0.91	0.80	1000
Weighted Avg	0.94	0.90	0.92	1000

The confusion matrix shows how well the loan classification model performs. Out of 904 customers who did not accept the loan, the model correctly classified 898 cases as “No Loan” (true negatives), with only 6 cases misclassified as “Loan” (false positives). Similarly, among the 96 customers who accepted the loan, the model accurately identified 90 cases as “Loan” (true positives), while misclassifying 6 cases as “No Loan” (false negatives).

These findings point to a model that is very accurate and has a good predictive balance between the two classes. Both the "Loan" and "No Loan" categories exhibit high precision and high recall, as evidenced by the low number of false positives and false negatives. Such balanced performance is particularly important in credit risk modeling, as it reduces the likelihood of misclassifying loan applicants while maintaining reliability in identifying non-loan customers.

Overall, the confusion matrix demonstrates the model's resilience, demonstrating its suitability for practical implementation in personal loan acceptance prediction.

Table 9. Random Forest Result

Kategori	Precision	Recall	F1-Score	Support
No Loan	0.99	0.99	0.99	904
Loan	0.94	0.94	0.94	96
Accuracy	-	-	0.99	1000
Macro Avg	0.97	0.97	0.97	1000
Weighted Avg	0.99	0.99	0.99	1000

The classification report offers a thorough assessment of how well the model predicts the acceptance of personal loans. With precision, recall, and F1-score all at 0.99 for the "No Loan" category, the model demonstrated near-perfect accuracy in identifying customers who declined the loan, resulting in exceptionally high scores across all metrics. Likewise, with precision, recall, and F1-score all at 0.94 for the "Loan" category, the model showed excellent performance, indicating its ability to accurately identify loan-accepting customers with few misclassifications.

The overall model accuracy reached 0.99, signifying that 99% of predictions were correct across the dataset. Additionally, the weighted averages were 0.99, highlighting the model's resilience even in the face of class imbalance, while the macro average values for precision, recall, and F1-score were all 0.97, indicating consistent performance across both classes.

These findings demonstrate the model's high degree of dependability and generalization, with strong predictive power for both majority (No Loan) and minority (Loan) classes. Such balanced and accurate performance underscores the model's potential applicability in practical banking scenarios, particularly for customer targeting and loan acceptance prediction.

XG Boost

Table 10. Confusion Matrix of XGBoost

	Predicted	
	No	**Loan**
Actual No	895	9
Loan	4	92

The confusion matrix presented above summarizes the performance of the predictive model in classifying loan eligibility. Out of the total instances, 895 were correctly identified as "No Loan" and 92 were correctly classified as "Loan". Meanwhile, 9 cases were incorrectly predicted as "Loan" when the actual outcome was "No Loan" (false positives), and 4 cases were incorrectly predicted as "No Loan" when the actual outcome was "Loan" (false negatives).

These findings show that the model attained a high degree of precision in distinguishing between applicants eligible and not eligible for loans, with relatively few misclassifications. The model appears to be more cautious when approving loans, giving priority to minimizing credit risk, as evidenced by the comparatively lower number of false negatives as opposed to false positives.

Table 11. XGBoost Result

Kategori	**Precision**	**Recall**	**F1-Score**	**Support**
No Loan	1.00	0.99	0.99	904
Loan	0.91	0.96	0.93	96
Accuracy	-	-	0.99	1000
Macro Avg	0.95	0.97	0.96	1000
Weighted Avg	0.99	0.99	0.99	1000

Additional information about the model's predictive performance can be found in the classification report. With a precision of 1.00, recall of 0.99, and F1-score of 0.99 for the "No Loan" class, the model successfully identified almost all applicants who were not eligible for a loan with few false positives. Although there were a few more misclassifications than in the majority class, the model's precision, recall, and F1-score for the "Loan" class were all 0.91, 0.96, and 0.93, respectively, indicating a strong ability to identify loan-eligible applicants.

Despite the imbalance between the classes, the model's robustness was confirmed by its overall accuracy of 0.99, which was backed by both a weighted-average F1-score of 0.99 and a macro- average F1-score of 0.96. These findings imply that the model operates with high

reliability, successfully striking a balance between recall and precision, and is thus well-suited for real-world use in credit risk prediction.

Deep Learning

Table 12. Confusion Matrix of Deep Learning

	Predicted	
	No	**Loan**
Actual No	884	20
Loan	4	92

The model's classification performance in predicting loan eligibility is depicted in the confusion matrix. A total of 884 instances were correctly classified as “No Loan”, while 92 instances were correctly identified as “Loan.” However, 20 cases were misclassified as “Loan” when the actual class was “No Loan” (false positives), and 4 cases were misclassified as “No Loan” when the actual class was “Loan” (false negatives).

The comparatively small number of false negatives shows that the model is very successful in identifying applicants who qualify for loans. Nonetheless, the slightly higher number of false positives indicates that the model occasionally predicts applicants as eligible for loans when they are not. From a financial risk management perspective, this trade-off implies that the model prioritizes capturing potential loan-eligible customers, even at the cost of a small increase in misclassification for non-eligible applicants.

Table 13. Deep Learning Result

Kategori	Precision	Recall	F1-Score	Support
No Loan	1.00	0.98	0.99	904
Loan	0.82	0.96	0.88	96
Accuracy	-	-	0.98	1000
Macro Avg	0.91	0.97	0.94	1000
Weighted Avg	0.98	0.98	0.98	1000

The classification report offers a thorough assessment of the predictive power of the model. The model obtained a precision of 1.00, recall of 0.98, and F1-score of 0.99 for the "No Loan" class, meaning that almost all ineligible applicants were correctly classified with very few false positives. While the majority of loan-eligible applicants were correctly identified (high recall), a comparatively higher proportion of false positives occurred compared to the majority class, as indicated by the model's precision of 0.82, recall of 0.96, and F1-score of 0.88 for the "Loan" class.

Overall, the model reached an accuracy of 0.98, supported by a macro-average F1-score of 0.94 and a weighted-average F1-score of 0.98. These findings demonstrate that the model performs with strong reliability across both classes despite class imbalance, with a slight tendency to predict more false positives in the minority class. From a practical perspective, this trade-off indicates that the model emphasizes minimizing false negatives in the loan-eligibility prediction task, which is favorable for reducing credit risk.

5. Conclusion

Important trade-offs between precision, recall, and overall accuracy are highlighted by the comparison of four machine learning models for predicting personal loan acceptance: Deep Learning, XGBoost, Random Forest, and Logistic Regression.

Logistic Regression achieved reasonable predictive capability, particularly with high recall for the “Loan” class, ensuring that most loan-accepting customers were identified.

However, its relatively low precision for the minority class reflects a tendency to generate false positives, thereby reducing reliability in practical applications.

Random Forest demonstrated superior and balanced performance, with near-perfect results across both classes. The model attained an overall accuracy of 0.99, along with consistently high precision, recall, and F1-scores, underscoring its robustness and suitability for real-world deployment in credit risk prediction tasks.

XGBoost also exhibited strong predictive accuracy (0.99) and balanced performance, particularly in minimizing false negatives, making it a highly reliable option for loan eligibility prediction. Its effectiveness in maintaining both high precision and recall across classes confirms its applicability in practical financial decision-making.

Deep Learning achieved strong results (accuracy of 0.98), with excellent recall for the "Loan" class, reducing the risk of missing potential customers. However, the lower precision for this class indicates a greater incidence of false positives, which could affect operational efficiency.

Overall, the most successful models overall were Random Forest and XGBoost, which provided balanced performance across majority and minority classes along with high accuracy. Logistic Regression and Deep Learning, while effective in certain aspects (notably recall for loan-accepting customers), exhibited limitations in precision. These findings suggest that ensemble-based methods, particularly Random Forest and XGBoost, provide the most reliable predictive performance for personal loan acceptance, making them strong candidates for practical adoption in banking and credit risk management.

Aknowledgement (Ucapan Terima Kasih)

We would like to express our deepest gratitude to Universitas Cakrawala for the academic and institutional support provided during the research process. We also acknowledge the valuable assistance from colleagues and banking practitioners who shared insights and expertise that enriched this study. Special appreciation is extended to the Directorate General of Higher Education, Research, and Technology for granting the Penelitian Dosen Pemula (PDP) research funding, which made this work possible.

References

Addo, P. M., Guegan, D., & Hassani, B. (2018a). Credit risk analysis using machine and deep learning models. *Risks*, *6*(2), 38.

Aldboush, H. H. H., & Ferdous, M. (2023). Building Trust in Fintech: An Analysis of Ethical and Privacy Considerations in the Intersection of Big Data, AI, and Customer Trust. *International Journal of Financial Studies*, *11*(3). https://doi.org/10.3390/ijfs11030090

Badawy, M., Ramadan, N., & Hefny, H. A. (2023). Healthcare predictive analytics using machine learning and deep learning techniques: a survey. *Journal of Electrical Systems and Information Technology*, *10*(1), 40.

Bi, S., & Bao, W. (2024). Innovative Application of Artificial Intelligence Technology in Bank Credit Risk Management. *International Journal of Global Economics and Management*, *2*(3), 76–81. https://doi.org/10.62051/ijgem.v2n3.08

Bi, W., & Liang, Y. (2022). Risk Assessment of Operator's Big Data Internet of Things Credit Financial Management Based on Machine Learning. *Mobile Information Systems*, *2022*. https://doi.org/10.1155/2022/5346995

Biswas, S., Carson, B., Chung, V., Singh, S., & Thomas, R. (2020). AI-bank of the future: Can banks meet the AI challenge. *New York: McKinsey & Company*.

Boobier, T. (2020). *AI and the Future of Banking*. John Wiley & Sons.

Bose, S., Dey, S. K., & Bhattacharjee, S. (2022). *Big Data, Data Analytics and Artificial Intelligence in Accounting: An Overview*. Edward Elgar Publishing.

Bose, S., Dey, S. K., & Bhattacharjee, S. (2023). Big data, data analytics and artificial intelligence in accounting: An overview. *Handbook of Big Data Research Methods*, 32–51.

Challa, K. (2024). *Enhancing credit risk assessment using AI and big data in modern finance* (Vol. 1, Issue 1).

Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, *91*, 106263.

de Keijzer, I. N., Vistisen, S. T., & Scheeren, T. W. L. (2021). Artificial Intelligence and Predictive Analytics. In *Advanced Hemodynamic Monitoring: Basics and New Horizons* (pp. 287–293). Springer.

Debbadi, R. K., & Boateng, O. (2025). Optimizing end-to-end business processes by integrating machine learning models with UiPath for predictive analytics and decision automation. *Int J Sci Res Arch*, *14*(2), 778–796.

Dogra, V., Verma, S., Verma, K., Jhanjhi, N. Z., Ghosh, U., & Le, D. N. (2022). A comparative analysis of machine learning models for banking news extraction by multiclass classification with imbalanced datasets of financial news: Challenges and solutions. *International Journal of Interactive Multimedia and Artificial Intelligence*, *7*(3), 35–52.

Dumitrescu, E., Hué, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research*, *297*(3), 1178–1192. https://doi.org/10.1016/j.ejor.2021.06.053

Fares, O. H., Butt, I., & Lee, S. H. M. (2023). Utilization of artificial intelligence in the banking sector: a systematic literature review. *Journal of Financial Services Marketing*, *28*(4), 835–852. https://doi.org/10.1057/s41264-022-00176-7

Galaz, V., Centeno, M. A., Callahan, P. W., Causevic, A., Patterson, T., Brass, I., Baum, S., Farber, D., Fischer, J., Garcia, D., McPhearson, T., Jimenez, D., King, B., Larcey, P., & Levy, K. (2021). Artificial intelligence, systemic risks, and sustainability. *Technology in Society*, *67*. https://doi.org/10.1016/j.techsoc.2021.101741

Gatla, T. R. (2024). *Anovel APPROACH TO DECODING FINANCIAL MARKETS: THE EMERGENCE OF AI IN FINANCIAL MODELING*.

Georgewill, I. A., & Gabriel, P. D. I. (2024). Artificial Intelligence and Predictive Analytics: Revolutionizing Strategic Business Insights in Digital Era. *"INSIGHT TO IMACT-LEVERAGING ADMINISTRATIVE AND MANAGEMENT KNOWLEDE FOR ECONOMIC TRANSFORMATION AND SUSTAINABILITY NOVEMBER, 20–21, 2024*, 449.

Hassan, S., Dhali, M., Zaman, F., & Tanveer, M. (2021). Big data and predictive analytics in healthcare in Bangladesh: regulatory challenges. *Heliyon*, *7*(6).

Karami, A., & Igbokwe, C. (2025). The impact of big data characteristics on credit risk assessment. In *International Journal of Data Science and Analytics* (Vol. 20, Issue 5, pp. 4239–4259). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s41060-025-00753-8

Khan, A. A., Chaudhari, O., & Chandra, R. (2024). A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation. In *Expert Systems with Applications* (Vol. 244). Elsevier Ltd. https://doi.org/10.1016/j.eswa.2023.122778

Kurniawan, R. (2024). Application of Random Forest Algorithm on Credit Risk Analysis. *Procedia Computer Science*, *245*(C), 740–749. https://doi.org/10.1016/j.procs.2024.10.300

Langenbucher, K., & Corcoran, P. (2021). Responsible AI Credit Scoring - A Lesson from Upstart.com. In *Digital Finance in Europe: Law, Regulation, and Governance* (pp. 141–179). De Gruyter. https://doi.org/10.1515/9783110749472-006

Lau, T., & Leimer, B. (2019). The era of connectedness: How AI will help deliver the future of banking. *Journal of Digital Banking*, *3*(3), 215–231.

Lee, C. S., Cheang, P. Y. S., & Moslehpour, M. (2022). Predictive analytics in business analytics: decision tree. *Advances in Decision Sciences*, *26*(1), 1–29.

Li, C., Wang, H., Jiang, S., & Gu, B. (2024). THE EFFECT OF AI-ENABLED CREDIT SCORING ON FINANCIAL INCLUSION: EVIDENCE FROM AN UNDERSERVED POPULATION OF OVER ONE MILLION1. *MIS Quarterly: Management Information Systems*, *48*(4), 1803–1834. https://doi.org/10.25300/MISQ/2024/18340

Li, Y., Stasinakis, C., & Yeo, W. M. (2022). A Hybrid XGBoost-MLP Model for Credit Risk Assessment on Digital Supply Chain Finance. *Forecasting*, *4*(1), 184–207. https://doi.org/10.3390/forecast4010011

Qin, C., Zhang, Y., Bao, F., Zhang, C., Liu, P., & Liu, P. (2021). XGBoost optimized by adaptive particle swarm optimization for credit scoring. *Mathematical Problems in Engineering*, *2021*. https://doi.org/10.1155/2021/6655510

Rao, C., Liu, Y., & Goh, M. (2023). Credit risk assessment mechanism of personal auto loan based on PSO-XGBoost Model. *Complex and Intelligent Systems*, *9*(2), 1391–1414. https://doi.org/10.1007/s40747-022-00854-y

Sadok, H., Sakka, F., & El Maknouzi, M. E. H. (2022). Artificial intelligence and bank credit analysis: A review. *Cogent Economics and Finance*, *10*(1). https://doi.org/10.1080/23322039.2021.2023262

Shi, S., Tse, R., Luo, W., D'Addona, S., & Pau, G. (2022). Machine learning-driven credit risk: a systemic review. In *Neural Computing and Applications* (Vol. 34, Issue 17, pp. 14327–14339). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s00521-022-07472-2

Thomas, B., & Raphael, D. (2022). *Artificial intelligence for financial markets: the polymodel approach*. Springer.

Uddin, M. S., Chi, G., Al Janabi, M. A. M., & Habib, T. (2022). Leveraging random forest in micro-enterprises credit risk modelling for accuracy and interpretability. *International Journal of Finance and Economics*, *27*(3), 3713–3729. https://doi.org/10.1002/ijfe.2346

Umamaheswari, D. D. (2024). Role of Artificial Intelligence in Marketing Strategies and Performance. *Migration Letters*, *21*(S4), 1589–1599.

Wang, K., Li, M., Cheng, J., Zhou, X., & Li, G. (2021). Research on personal credit risk evaluation based on XGBoost. *Procedia Computer Science*, *199*, 1128–1135. https://doi.org/10.1016/j.procs.2022.01.143

Widyastuti, M., Ferdinand, D. Y. Y., & Hermanto, Y. B. (2023). Strengthening Formal Credit Access and Performance through Financial Literacy and Credit Terms in Micro, Small and Medium Businesses. *Journal of Risk and Financial Management*, *16*(1). https://doi.org/10.3390/jrfm16010052.