Global Journal of Accounting
Department of Accounting, Faculty of Management Sciences
University of Lagos

Credit Decision Making process in FinTech Services in Nigeria: An Application of Logistic Regression Credit Scoring

Lukman Abolaji Ajijola[1]
Department of Actuarial Science and Insurance, University of Lagos.
Email: lajijola@unilag.edu.ng

&

Sunday Lawrence Solanke
Department of Actuarial Science and Insurance, University of Lagos.

Abstract

A Robust Enterprise Risk Management (ERM) framework is very critical for FinTech sustainability and continuity as it helps to manage potential losses from lending activities. To this end, the objective of this study is to identify the factors influencing credit risk and to examine credit scoring process for credit risk decision making in FinTech companies in Nigeria. Logistic regression-based methodology was employed to improve and optimized the traditional approach of credit risk decision making. The study utilizes secondary data from a FinTech company over three-years, focusing on both corporate and individual clients who have closed loans with varying tenors. Python software was used to process and analyzed the retrieved data and the key predictors impacting loan repayment behaviour are identified. Loan amount, frequency of repayment and number of dependents are strong predictors while marital status, net pay after statutory deductions, disbursement turn-around-time (TAT) and years in service show moderate predictive strength. This study contributes to the literature by demonstrating the effectiveness of logistic regression in improving credit risk assessment models for FinTech companies in Nigeria. The findings emphasize the need for FinTech companies to integrate logistic regression models into credit scoring systems to enhance risk assessment accuracy and business value creation.

.

Keywords: *Credit scoring, Credit risk, Logistic Regression, Fintech Services, Loan Repayment.*

1. Introduction

The management of loan repayment behavior has become increasingly critical for financial institutions. With the proliferation of credit facilities and the diversification of financial products, understanding the factors influencing loan repayment behavior has become a paramount concern (Akins, 2019). In recent years, financial institutions have witnessed a surge in non-performing loans, which directly impacts their profitability and stability. The inability to accurately assess the creditworthiness of borrowers and predict their repayment behavior has been identified as a major contributor to this problem (Akram & Hussain, 2020). This issue underscores the importance of employing robust credit scoring models to enhance the accuracy of loan approval decisions and mitigate the risk of default. Another critical aspect of the problem is the dynamic nature of borrowers' financial profiles and economic conditions. Factors such as income fluctuations, employment stability, and macroeconomic indicators can significantly influence borrowers' ability to repay loans (Le, Nguyen & Schinckus, 2022).

The large proportion of loans in the overall operating assets of lending institutions highlights the critical role of healthy loan portfolios in maintaining their liquidity, lending capacity, earnings, and profitability (Emile, 2021). For microfinance institutions (MFIs), robust loan portfolios are essential, as they directly influence financial stability and growth prospects. Kofi, (2017) also posited that extension of credit facilities is one of the major activities of all

Microfinance institutions (MFIs) including savings and loans companies, financial non-governmental organizations and credit unions. Existence of high levels of loan delinquency problem in microfinance industry negatively affect the level of investment, increase in deposit liabilities and constrain the scope of microfinance institution credit to borrowers through reduction of MFIs' capital, following falling accumulation of losses to compensate for loan delinquency losses. The success of MFIs largely depends on the effectiveness of their credit management systems because these institutions generate most of their income from interest earned on loans extended to small and medium entrepreneurs (Lamichhane, 2022).

As the financial landscape becomes increasingly complex, institutions are required to employ sophisticated techniques to make informed credit decisions, thereby minimizing potential losses and maximizing returns (Ogunyele and Akanni, 2021). One widely adopted technique for loans problem is credit scoring. Credit scoring consists of the assessment of risk associated with lending to an organization or an individual (Sum et al, 2022). A study conducted by Einav, Jenkins and Levin (2018) described the magnitude and channels by which the adoption of credit scoring affected loan originations, repayment and defaults, and profitability at a large auto finance company. The adoption of credit scoring technology led to a large increase in profitability.

Credit scoring models have evolved significantly over time, transitioning from simple algorithms to sophisticated machine learning models capable of analyzing vast amounts of data. These advancements have enhanced the accuracy and predictive power of the models, enabling financial institutions to better assess the risk associated with lending to various individuals or entities (Thomas, 2021; Hand & Henley, 2019). Machine learning techniques, such as decision trees, neural networks, and ensemble methods, have been particularly influential in refining the models' ability to detect subtle patterns and trends that traditional methods might overlook (Lessmann, Baesens, Seow, & Thomas, 2015). This evolution has not only improved credit risk assessment but also contributed to more inclusive lending practices by incorporating a wider range of data sources, such as social media activity and transaction history (Khandani, Kim, & Lo, 2020).

Moreover, the use of credit scoring models in credit risk decision making supports regulatory compliance and reporting. Financial institutions are often required to justify their lending decisions to regulators and stakeholders, and credit scoring models provide a transparent and auditable framework for decision-making (Hand & Henley, 2019). This ensures that lending practices align with regulatory standards and industry best practices, as the models offer a consistent methodology for evaluating creditworthiness (Thomas, 2021). By systematically applying these models, institutions can demonstrate that their decisions are based on objective criteria, reducing the risk of regulatory scrutiny and potential penalties (Lessmann et al., 2015; Khandani, Kim, & Lo, 2020). This transparency and adherence to regulatory requirements help maintain the institution's reputation and trust with stakeholders. Hence, it is on this background that the study seek to examine credit scoring model for credit risk decision making. With this background, the aim of the study is to identify the factors influencing credit risk and to examine credit scoring process for credit risk decision making in FinTech companies in Nigeria.

2. Literature Review

Credit risk modeling has evolved significantly over the years, adapting to advances in technology and changes in economic conditions. Traditional credit risk models primarily utilized statistical methods that relied on linear relationships between variables. Recent developments have seen the integration of machine learning techniques into credit risk modeling.

Altman's (1968) study employed a quantitative longitudinal research design focusing on publicly traded companies. The study sampled 66 bankrupt firms and 66 non-bankrupt firms using a non-probability sampling technique. Data analysis was conducted using discriminant analysis, leading to the development of the renowned Altman Z-score model. The major findings underscored the model's efficacy in predicting corporate bankruptcy based on financial ratios, affirming that these metrics serve as reliable indicators in assessing financial distress and potential bankruptcy risks among businesses.

Adebayo and Abolaji's (2023) compared various credit scoring techniques utilized by commercial banks in Nigeria. The study focused on commercial banks as the population of interest, although specific details regarding the sample and sampling technique were not provided. Data analysis involved a comparative study of credit scoring techniques to evaluate their effectiveness in assessing creditworthiness and managing risks within Nigerian banks. The study concluded that the choice of credit scoring models significantly influences the ability of banks to make informed lending decisions and effectively manage credit risks in the Nigerian banking environment.

Beaver's (2021) utilized a quantitative cross-sectional research design focusing on publicly traded companies. The study sampled 79 failed firms and 79 non-failed firms using a non-probability sampling technique. Data analysis employed logistic regression to identify significant financial ratios for predicting corporate failure. The findings highlighted specific financial metrics that effectively predict corporate failure, emphasizing their role as crucial indicators of corporate financial health and risk.

Okafor and Agiomoh's (2022) explore the relationship between credit scoring models and loan performance within the Nigerian banking sector. The study focused on commercial banks as the population of interest, although specific details regarding the sample and sampling technique were not provided. Data analysis employed regression analysis and performance evaluation methods to assess how credit scoring models impact loan performance and credit risk management in Nigerian banks. Their conclusion is that there is a positive correlation between effective credit scoring models and improved loan performance, emphasizing their role in mitigating credit risks and enhancing overall financial stability within the Nigerian banking sector.

Ojeka's (2019) employed a mixed methods research design to investigate the adoption and impact of credit scoring models within Nigerian commercial banks. The study targeted commercial banks as the population of interest, although specific details regarding the sample and sampling technique were not provided. Data analysis utilized qualitative interviews and quantitative analysis to explore how credit scoring models influence credit risk management practices and lending decisions in Nigerian banks. The findings emphasized the significant role of credit scoring models in enhancing credit risk management by providing more accurate assessments of borrower creditworthiness.

Adigun's (2021) study titled "Development of Credit Scoring Models for SMEs in Nigeria" focused on conducting quantitative analysis to develop and validate credit scoring models specifically tailored for small and medium-sized enterprises (SMEs) in Nigeria. The study targeted SMEs as the population of interest. Data analysis involved model development and validation processes to ensure the accuracy and applicability of these models within the Nigerian SME lending context. The study concluded that adopting tailored credit scoring approaches is crucial for facilitating financial inclusion and supporting sustainable growth among SMEs in Nigeria by providing reliable tools for assessing creditworthiness and managing lending risks effectively.

3. Materials and Methods

Material

The population of this study consists of all 303 corporate clients who received loans from the FinTech Company with varying repayment tenors, over a period of three years (April 2021 – April 2024). We focused on both corporate and individual clients who have closed loans with varying tenors (ranging from 1 to 36 months). All the loans have been closed out, providing a complete dataset for analysis. To ensure a refined dataset, several filtering techniques were employed during the preprocessing stage. Initially, a data analysis filter was used to include only clients who borrowed money during the specified period. An information retrieval filter subsequently focused on relevant borrower records, followed by a machine learning-based filter to select the most pertinent client data for modeling and evaluation.

Method

Credit Scoring and Logistic Regression

Let $(x_i, y_i), i = 1,2,3,\dots,n$ be a sample of size n of independent and identically distributed observations, where $x_i \in \mathbb{R}^p$ is a p −dimensional vector of predictors and $y_i \in \{0,1\}$ is a binary variable taking the value one when the i-th borrower defaults and zero otherwise. The goal of a credit scoring model is to provide an estimate of the posterior probability $Pr\ (y_i = 1\ |x_i\ )$ that borrower i defaults given its attributes x_i. The relevant characteristics of the borrower vary according to its status: household or company. For corporate credit risk scoring, the candidate predictive variables $x_{i,j} = 1,2,3,\dots,p$, may include balance-sheet financial variables that cover various aspects of the financial strength of the firm, such as the firm's operational performance, its liquidity, and capital structure (Altman, 1968). For retail loans, financial variables such as the number and amount of personal loans, normal repayment frequency of loans, the number of credit cards, the average overdue duration of credit cards and the amount of housing loans are combined with socio-demographic factors.

Regardless of the type of borrower, the conditional probability of default is generally modelled using a logistic regression with the following specifications:

$$Pr\ (y_i = 1\ |x_i\ ) = F\ (\eta\ (x_i;\ \beta)) = \frac{1}{1 + \exp(-\eta\ (x_i;\ \beta))}, \qquad (1)$$

where $F\ (.)$ is the logistic cumulative distribution function and $\eta\ (x_i;\ \beta)$ is the so-called index function defined as

$$\eta\ (x_i;\ \beta) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{i,j},$$

where $\beta\ =\ (\beta_0, \beta_1, \ldots\ \beta_p) \in \mathbb{R}^{p+1}$ is an unknown vector of parameters. The estimator $\hat{\beta}$ is obtained by maximizing the log-likelihood function

$$\mathcal{L}(y_i;\ \beta)\ = \sum_{i=1}^{n} \left\{ y_i log\{F\ (\eta\ (x_i;\ \beta))\} + (1 - y_i) log\{1 - F\ (\eta\ (x_i;\ \beta))\} \right\}$$

The main advantage of the logistic regression model is its simple interpretation. Indeed, this model searches for a single linear decision boundary in the predictors' space. The core assumption for finding this boundary is that the index $\eta\ (x_i;\ \beta)$ is linearly related to the predictive variables. In this framework, it is easy to evaluate the relative contribution of each predictor to the probability of default. This is achieved by computing marginal effects as

$$\frac{\partial Pr(y_i = 1\ |x_i)}{\partial x_{i,j}} = \beta_j \frac{\exp(\eta\ (x_i;\ \beta))}{[1 + \exp(\eta\ (x_i;\ \beta))]^2},$$

with estimates obtained by replacing β with $\hat{\beta}$. Thus, a predictive variable with a positive (negative) significant coefficient has a positive (negative) impact on the borrower's default probability.

Weight of Evidence (WoE) and Information Value (IV)

Based on Information Theory conceived in the later 1940s and initially developed for scorecard development, WoE and IV have been gaining increasing attention in recent years for such uses as segmentation and variable reduction (Lin, 2015). This method of analysis is usually simple and comparatively consumes less time (Alsabhan et al, 2022)]. WoE works by recoding variable values into discrete categories and assigning a unique WoE value to each category with the aim of generating the largest difference between the recoded ones. An important assumption here is that the dependent variable must be binary to indicate the occurrence and non-occurrence of an event. In the example of food insecurity analysis where households are neither food insecure (good) nor food insecure (bad), the WoE for each household segment is calculated as follows.

$$WoE\ = \left[ln \left(\frac{\%bad_i}{\%good_i} \right) \right] \times 100 \qquad (2)$$

While WoE analyzes the predictive ability of a variable in relation to its targeted outcome, IV assesses the overall predictive ability of the variables that have been used. IV can be used to compare the predictive ability among competing variables. The following is the calculation of IV.

$$IV = \sum_{i=1}^{n}\left((\%bad_i - \%good_i) \times \left(\frac{\%bad_i}{\%good_i}\right)\right) \qquad (3)$$

3. Results

Presentation of Data

Table 1: Demographic Statistics of the Customer

Category		%
Gender		
	Female	44%
	Male	56%
Age Group		
	26-30 years	16%
	31-35 years	29%
	36-40 years	24%
	41-45 years	19%
	46-50 years	9%
	51-55 years	5%
Marital Status		
	Divorced	4%
	Married	51%
	Separated	1%
	Single	43%
	Widow	1%
	Widower	1%
Loan Tenure in Months		
	1-6	2%
	7-12	9%
	13-18	11%
	19-24	20%
	25-30	3%
	31-36	55%
Payment Behavior		
	Good	89%
	Bad	11%
Credit Score Range		
	600 – 660	7%
	661 – 721	59%
	722 – 782	31%
	783 – 850	3%

Source: *The FinTech Company (2024)*

Table 2: Class-Specific Data Classification

Variable	Categories	Frequency	Percent	Valid Percent	Cumulative Percent	Purpose
Sex	Female	134	44%	44%	44%	To analyze gender distribution in the data.
	Male	169	56%	56%	100%	
Age	26-30	47	16%	16%	16%	To investigate age distribution and its potential impact on credit risk.
	31-35	88	29%	29%	45%	
	36-40	72	24%	24%	68%	
	41-45	56	19%	19%	87%	
	46-50	26	9%	9%	96%	
	51-55	14	5%	5%	100%	
Marital Status	Divorced	13	4%	4%	4%	To explore how marital status might affect creditworthiness.
	Married	154	51%	51%	55%	
	Separated	3	1%	1%	56%	
	Single	129	43%	43%	99%	
	Widow	2	1%	1%	99%	
	Widower	2	1%	1%	100%	
Tenor	1--6	7	2%	2%	2%	To evaluate how the loan tenure affects credit risk.
	7--12	28	9%	9%	12%	
	13-18	33	11%	11%	22%	
	19-24	60	20%	20%	43%	
	25-30	8	3%	3%	45%	
	31-36	167	55%	55%	100%	
Year in Service	3 --4	15	5%	5%	5%	To understand how years of service affect creditworthiness.
	5 --6	43	14%	14%	19%	
	7--8	42	14%	14%	33%	
	9 --10	30	10%	10%	43%	
	11 --12	21	7%	7%	50%	
	13-14	21	7%	7%	57%	
	15-16	13	4%	4%	61%	
	17-18	6	2%	2%	63%	
	19-22	7	2%	2%	65%	
	23 - 24	2	1%	1%	66%	
Number of Dependents	0	77	25%	25%	25%	To determine how the number of dependents impacts credit risk.
	1	81	27%	27%	52%	
	2	97	32%	32%	84%	
	3	41	14%	14%	98%	
	4	7	2%	2%	100%	

Source: *The FinTech Company*

Table 2 (Continued): Class-Specific Data Classification

Variable	Categories	Frequency	Percent	Valid Percent	Cumulative Percent	Purpose
Number of Loans Already Obtained	0	126	42%	42%	42%	To explore the effect of previous loans on credit risk.
	1	107	35%	35%	77%	
	2	65	22%	22%	98%	
	3	5	2%	2%	100%	
Reimbursement of Past Loans	No	126	42%	42%	42%	To examine the impact of loan reimbursement on current credit risk.
	Yes	177	58%	58%	100%	
Years Practicing Economic Activity	0-1	5	2%	2%	2%	To analyze the impact of years of economic activity on credit risk.
	2-4	71	24%	24%	25%	
	5-6	64	21%	21%	46%	
	7-8	52	17%	17%	63%	
	9-10	26	9%	9%	72%	
	11-12	34	11%	11%	83%	
	13-15	24	8%	8%	91%	
	16-18	13	4%	4%	95%	
	19-22	8	3%	3%	98%	
	24-30	7	2%	2%	100%	
Time Between the Demand and Loan Disbursement	1 day	14	5%	5%	5%	To evaluate how the disbursement time affects loan performance.
	1 month	7	2%	2%	7%	
	1 week	114	38%	38%	45%	
	2 days	69	23%	23%	67%	
	2 weeks	63	21%	21%	88%	
	3 days	27	9%	9%	97%	
	3 weeks	9	3%	3%	100%	
Payment Behavior	Bad	32	11%	11%	11%	To assess the relationship between payment behavior and credit risk.
	Good	271	89%	89%	100%	
Credit Bureau Data	600 – 660	21	7%	7%	7%	To determine the impact of credit score on credit risk.
	661 – 721	179	59%	59%	66%	
	722 – 782	93	31%	31%	97%	
	783 – 850	10	3%	3%	100%	

Source: *The FinTech Company*

Table 2 (Continued): Class-Specific Data Classification

Variable	Categories	Frequency	Percent	Valid Percent	Cumulative Percent	Purpose
State/Zone	North Central	34	11%	11%	11%	To analyze regional differences in credit risk.
	North East	42	14%	14%	25%	
	North West	38	13%	13%	38%	
	South East	36	12%	12%	50%	
	South South	43	14%	14%	64%	
	South West	42	14%	14%	78%	
	Special	18	6%	6%	84%	
Loan Product Type	DPL	15	5%	5%	5%	To assess the impact of different loan products on credit risk.
	DPL - I'nt.	1	0%	0%	5%	
	Edu. Inst Loan	2	1%	1%	6%	
	Federal Loan	234	77%	77%	83%	
	Nig. Army & P. M. Loan	23	8%	8%	91%	
	State Business	28	9%	9%	100%	

Source: *The FinTech Company (2024)*

Descriptive Analysis

The dataset comprising 303 observations provides a detailed overview of key financial variables. The ages of the participants range from 26 to 55 years, with a mean age of 37.6 years and a standard deviation of 6.58 years, indicating a predominantly mid-career demographic. Net pay varies significantly from 34,563.52 to 2,287,500.00, averaging 274,819.24 with a standard deviation of 245,176.77, reflecting considerable income disparity among participants. Loan amounts also show substantial variability, ranging from 100,000.00 to 5,000,000.00, with a mean value of 1,317,544.55 and a standard deviation of 862,680.52, highlighting diverse loan sizes.

The tenor of loans spans from 1 to 36 months, with an average duration of 28.11 months and a standard deviation of 9.63 months, suggesting a predominance of medium-term loans. The years of service range from 2.97 to 24.15 years, with a mean of 10.07 years and a standard deviation of 4.20 years, reflecting a broad spectrum of professional experience. Borrowers have an average of 1.41 dependents, with values ranging from 0 to 4 and a standard deviation of 1.08, indicating varied family sizes.

The number of loans already obtained ranges from 0 to 3, with a mean of 0.83 and a standard deviation of 0.82, suggesting that most borrowers hold a few additional loans. The years spent practicing economic activity range from 0 to 30, with a mean of 8.17 years and a standard deviation of 5.19 years, demonstrating a wide range of experience levels. Finally, credit scores range from 600 to 850, with an average score of 712.18 and a standard deviation of 39.22, indicating relatively consistent creditworthiness among the borrowers. These descriptive statistics provide a comprehensive understanding of the financial characteristics of the dataset, forming a crucial basis for developing an effective credit scoring model for credit risk assessment.

Table 3: Descriptive Statistics

	N	Min	Max	Sum	Mean	Std. Error	Std. Deviation
Age	303	26	55	11.39	37.6	0.38	6.58
Net pay	303	34,563.52	2,287,500	83,270.23	274,819.24	14,085.04	245,176.77
Loan amount	303	100,000	5,000,000	399,216	1,317,544.55	49,559.70	862,680.52
Tenor	303	1	36	8.52	28.11	0.55	9.63
Years in service	303	2.97	24.15	3.05	10.07	0.24	4.2
Number of dependents	303	0	4	0.43	1.41	0.06	1.08
No. of prev. loans	303	0	3	0.25	0.83	0.05	0.82
Years of experience	303	0	30	2.48	8.17	0.3	5.19
Credit bureau data	303	600	850	215.79	712.18	2.25	39.22

Source: *Author's Computation (2024)*

Evaluation of Key Variables Influencing the Model

The logistic regression analysis identified several predictors with varying levels of predictive power based on Weight of Evidence (WOE) and Information Value (IV) scores. The analysis of predictive factors reveals key insights based on Information Value (IV) and Weight of Evidence (WOE). Strong predictors of credit risk include loan amount, disbursement turnaround time (TAT), frequency of repayment, and the number of dependents. Higher loan amounts and longer disbursement times are linked to an increased likelihood of default, while more frequent repayments and fewer dependents reduce the risk. These variables demonstrate a strong influence on the credit risk model, with the loan amount and TAT showing the highest predictive power.

Moderate predictors include marital status, net pay, the number of previous loans, and years of experience. While marital status and net pay moderately influence the risk of default, fewer previous loans and more years of experience correlate with lower credit risk. On the other hand, gender and age were identified as weak predictors, contributing minimally to the overall model. These findings highlight the significance of financial and repayment-related factors, while demographic variables like gender and age have less predictive value.

Table 5: Credit Risk Features Categorization by Predictive Strength

Variable	Category	Bad (B)	Good (G)	% B	% G	(%G – %B)	WOE	(Sub) IV	Total (IV)	Predictive Power
Disbursement TAT	**< 1 week**	17	93	53.10%	34.30%	-18.80%	-0.44	0.0822	0.39	Strong
	1 week	7	107	21.90%	39.50%	17.60%	0.59	0.104		
	2 weeks	7	56	21.90%	20.70%	-1.20%	-0.06	0.0007		
	3 weeks	1	8	3.10%	3.00%	-0.20%	-0.06	0.0001		
	1 month	0	7	0.00%	2.60%	2.60%	7.86	0.203		
Loan Amount (# '000)	**100 – 499**	1	14	3.10%	5.20%	2.00%	0.5	0.0103	0.33	Strong
	500 – 899	9	75	28.10%	27.70%	-0.40%	-0.02	0.0001		
	900 – 1299	12	85	37.50%	31.40%	-6.10%	-0.18	0.011		
	1300 – 1699	1	47	3.10%	17.30%	14.20%	1.71	0.2437		
	1700 – 2099	3	16	9.40%	5.90%	-3.50%	-0.46	0.016		
	2100 – 2499	3	11	9.40%	4.10%	-5.30%	-0.84	0.0445		
	> 2499	3	23	9.40%	8.50%	-0.90%	-0.1	0.0009		
No. of Dependents	**0**	5	72	15.60%	26.60%	10.90%	0.53	0.0581	0.33	Strong
	1	8	73	25.00%	26.90%	1.90%	0.07	0.0014		
	2	12	85	37.50%	31.40%	-6.10%	-0.18	0.011		
	3	7	34	21.90%	12.50%	-9.30%	-0.56	0.0519		
	4	0	7	0.00%	2.60%	2.60%	7.86	0.203		
Freq. of Repayment	**Bi – Weekly**	3	12	9.40%	4.40%	-4.90%	-0.75	0.0371	0.31	Strong
	Monthly	29	250	90.60%	92.30%	1.60%	0.02	0.0003		
	Quarterly	0	9	0.00%	3.30%	3.30%	8.11	0.2693		
	Divorced	1	12	3.10%	4.40%	1.30%	0.35	0.0045		
	Married	20	134	62.50%	49.40%	-13.10%	-0.23	0.0306		

Source: *Authors Computation, 2024*

Table 5: Credit Risk Features Categorization by Predictive Strength (Cont'd)

Variable	Category	Bad (B)	Good (G)	% B	% G	(%G – %B)	WOE	(Sub) IV	Total (IV)	Predictive Power
Marital Status	**Separated**	1	2	3.10%	0.70%	-2.40%	-1.44	0.0344	0.25	Medium
	Single	9	120	28.10%	44.30%	16.20%	0.45	0.0733		
	Widow	0	2	0.00%	0.70%	0.70%	6.61	0.0487		
	Widower	1	1	3.10%	0.40%	-2.80%	-2.14	0.0589		
Years of Experience	**0 – 5**	10	97	31.30%	35.80%	4.50%	0.14	0.0062	0.25	Medium
	6 – 11	11	115	34.40%	42.40%	8.10%	0.21	0.017		
	12 – 17	8	44	25.00%	16.20%	-8.80%	-0.43	0.0378		
	18 – 23	3	10	9.40%	3.70%	-5.70%	-0.93	0.053		
	24 – 30	0	5	0.00%	1.80%	1.80%	7.52	0.1388		
	0	7	119	21.90%	43.90%	22.00%	0.7	0.1536		
No of Loans Obtained	**1**	15	92	46.90%	33.90%	-12.90%	-0.32	0.0417	0.23	Medium
	2	9	56	28.10%	20.70%	-7.50%	-0.31	0.023		
	3	1	4	3.10%	1.50%	-1.60%	-0.75	0.0124		
Repayment of Past Loans	**Yes**	25	152	78.10%	56.10%	-22.00%	-0.33	0.073	0.23	Medium
	No	7	119	21.90%	43.90%	22.00%	0.7	0.1536		
	Digital	1	14	3.10%	5.20%	2.00%	0.5	0.0103		
	Digital 'Int	0	1	0.00%	0.40%	0.40%	5.91	0.0218		

Source: *Authors Computation, 2024*

Table 5: Credit Risk Features Categorization by Predictive Strength

Variable	Category	Bad (B)	Good (G)	% B	% G	(%G – %B)	WOE	(Sub) IV	Total (IV)	Predictive Power
Product Type	Educational Inst.	0	2	0.00%	0.70%	0.70%	6.61	0.0487	0.18	Medium
	Federal	28	206	87.50%	76.00%	-11.50%	-0.14	0.0162		
	Armed F & Para	2	21	6.30%	7.70%	1.50%	0.21	0.0032		
	State	1	27	3.10%	10.00%	6.80%	1.16	0.0793		
Net Pay (# '000)	< 35	0	1	0.00%	0.40%	0.40%	5.91	0.0218	0.15	Medium
	35 – 134	6	54	18.80%	19.90%	1.20%	0.06	0.0007		
	135 – 234	13	112	40.60%	41.30%	0.70%	0.02	0.0001		
	235 – 334	5	44	15.60%	16.20%	0.60%	0.04	0.0002		
	335 – 434	6	26	18.80%	9.60%	-9.20%	-0.67	0.0613		
	435 – 534	1	9	3.10%	3.30%	0.20%	0.06	0.0001		
	> 534	1	25	3.10%	9.20%	6.10%	1.08	0.066		
Tenor	1 – 6	1	6	3.10%	2.20%	-0.90%	-0.34	0.0031	0.15	Medium
	7 – 12	2	26	6.30%	9.60%	3.30%	0.43	0.0143		
	13 – 18	3	30	9.40%	11.10%	1.70%	0.17	0.0028		
	19 – 24	3	58	9.40%	21.40%	12.00%	0.83	0.0993		
	25 – 30	1	7	3.10%	2.60%	-0.50%	-0.19	0.001		
	31 – 36	22	144	68.80%	53.10%	-15.70%	-0.28	0.0333		
Channel	Web	2	50	6.30%	18.50%	12.20%	1.08	0.1321	0.15	Medium
	Toolkit	29	216	90.60%	79.70%	-10.90%	-0.13	0.014		
	USSD	1	5	3.10%	1.80%	-1.30%	-0.53	0.0067		

Source: *Authors Computation, 2024*

Table 5: Credit Risk Features Categorization by Predictive Strength

Variable	Category	Bad (B)	Good (G)	% B	% G	(%G – %B)	WOE	(Sub) IV	Total (IV)	Predictive Power
Years in Service	1 – 6	8	62	25.00%	22.90%	-2.10%	-0.09	0.0019	0.07	Weak
	7 – 12	12	132	37.50%	48.70%	11.20%	0.26	0.0293		
	13 –18	11	65	34.40%	24.00%	-10.40%	-0.36	0.0374		
	19 – 24	1	12	3.10%	4.40%	1.30%	0.35	0.0045		
Age	26 – 31	7	59	21.90%	21.80%	-0.10%	0	0	0.07	Weak
	32 – 37	9	88	28.10%	32.50%	4.30%	0.14	0.0062		
	38 – 43	8	73	25.00%	26.90%	1.90%	0.07	0.0014		
	44 – 49	4	37	12.50%	13.70%	1.20%	0.09	0.001		
	50 – 55	4	14	12.50%	5.20%	-7.30%	-0.88	0.0648		
Credit Bureau Score	600 – 660	1	20	3.10%	7.40%	4.30%	0.86	0.0366	0.05	Weak
	661 – 721	21	158	65.60%	58.30%	-7.30%	-0.12	0.0087		
	722 – 782	9	84	28.10%	31.00%	2.90%	0.1	0.0028		
	783 – 850	1	9	3.10%	3.30%	0.20%	0.06	0.0001		
Sex	Male	16	153	50.00%	56.50%	6.50%	0.12	0.0078	0.02	Weak
	Female	16	118	50.00%	43.50%	-6.50%	-0.14	0.0089		

Source: *Authors Computation, 2024*

Model Evaluation

This section covers the evaluation of logistic regression model after prediction using the test dataset. The dataset consisted of 303 observations comprising of 271 and 32 good and bad repayment behaviours respectively.

Logistic Regression Coefficients and Statistical Significance

The logistic regression model examines factors impacting credit risk, with the Intercept (- 4.1401) which represents the baseline risk. Negative coefficients like Sex (gender) (- 0.4824) and Net Pay (- 0.4794), suggest that male borrowers and those with higher net pay are less likely to pose credit risks. Conversely, positive coefficients for Loan Amount (0.2264) and Tenor (0.4067) imply that higher loan amounts and extended repayment tenor marginally increase default probability and potentially impact credit risk. Additionally, the significant coefficient of Reimbursement of Past Loans (1.5028) indicates that, while past repayments are indicative of good credit character, it may increase risk if borrowers become over-burdened. These findings help FinTech companies manage credit risk effectively.

Table 6: Logistic Regression Model Analysis

Logit Regression Results									
===================									
	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**	**P-value**	**Sig**	**VIF**
Const	-4.14	2.07	-2	0.05	-8.2	-0.09	0.0454	True	87.519
Sex (gender)	-0.482	0.46	-1.05	0.3	-1.39	0.422	0.2959	False	1.087
Age	-0.376	0.3	-1.26	0.21	-0.96	0.208	0.2068	False	2.6648
Marital status	0.0771	0.25	0.308	0.76	-0.41	0.568	0.7581	False	2.1841
Occupation	0.0025	0.01	0.398	0.69	-0.01	0.015	0.6909	False	1.059
Net Pay	-0.479	0.83	-0.58	0.56	-2.10	1.137	0.5610	False	2.977
Loan Amount	0.2264	0.53	0.429	0.67	-0.81	1.261	0.6678	False	2.2925
Tenor	0.4067	0.36	1.135	0.26	-0.3	1.109	0.2565	False	1.6935
Purpose of the loan	0.0263	0.06	0.458	0.65	-0.09	0.139	0.6471	False	1.3173
Frequency of repayment	-0.412	0.84	-0.49	0.62	-2.06	1.237	0.6244	False	1.7205
Years in service	-0.080	0.57	-0.14	0.89	-1.19	1.030	0.8873	False	7.654
Number of dependents	0.2758	0.51	0.541	0.59	-0.72	1.276	0.5888	False	5.8736
Number of loans already obt	-0.217	0.56	-0.39	0.7	-1.32	0.885	0.6997	False	7.3359
Reimbursement of past loans	1.5028	0.89	1.68	0.09	-0.25	3.256	0.0929	False	5.0215
Number of yrs practicing an eco	-0.176	0.51	-0.34	0.73	-1.18	0.832	0.7322	False	6.7801
Disbursement TAT	0.4301	0.18	2.382	0.02	0.076	0.784	0.0172	True	1.4607
Credit bureau data	-0.381	0.32	-1.21	0.23	-1.00	0.236	0.2260	False	2.5656
Channel	1.1652	0.71	1.646	0.1	-0.22	2.553	0.0998	False	1.1524
State	-0.102	0.03	-3.96	0	-0.15	-0.05	0.0001	True	1.4642
Product	0.0558	0.32	0.175	0.86	-0.57	0.68	0.8610	False	1.3981

Source: *Author's Computation, 2024*

The logistic regression model analyzes various factors influencing loan approval, with coefficients representing the relationship between each variable and the likelihood of approval. Key variables such as gender, age, and net pay show minimal or no significant impact, while disbursement turnaround time (TAT) and state have significant effects, suggesting their importance in credit risk decision making. Z calculates the log-odds of loan approval based on factors like sex, age, loan amount, and repayment frequency. A Z value of 0 means a 50% chance of approval, with values above 0 increasing the likelihood and below 0 decreasing it. The logistic function then transforms Z into a probability, enhancing credit risk decision making by providing a more accurate loan approval estimate. Variables like loan amount, tenure, and reimbursement of past loans exhibit some influence but lack statistical significance. Multicollinearity was observed, leading to the exclusion of some variables, resulting in a reduced logistic regression model that was estimated afterward. Overall, the model provides useful insights into credit risk but emphasizes the importance of key factors like disbursement TAT and location in making informed lending decisions.

Table 7: Refined Logistic Regression Model Analysis

Logistic Regression Model									
	Coef	std err	z	P>|z|	[0.025	0.975]	P-value	Sig	VIF
Const	-3.8337	1.394	-2.751	0.006	-6.565	-1.102	0.005948	True	46.007
Marital status	0.0911	0.228	0.399	0.69	-0.356	0.538	0.689574	False	1.8459
Net Pay	-0.4502	0.621	-0.725	0.468	-1.667	0.767	0.468491	False	2.3601
Loan Amount	0.2881	0.419	0.688	0.491	-0.532	1.109	0.491397	False	2.1257
Tenor	0.2092	0.296	0.706	0.48	-0.372	0.79	0.480293	False	1.5096
Frequency of repayment	0.1838	0.67	0.274	0.784	-1.13	1.497	0.783859	False	1.2354
Number of dependents	0.3743	0.291	1.285	0.199	-0.197	0.945	0.198859	False	2.2563
Disbursement TAT	0.4039	0.166	2.428	0.015	0.078	0.73	0.015162	True	1.1767
Credit bureau score	-0.4837	0.238	-2.031	0.042	-0.951	-0.017	0.042242	True	1.5076
Channel	1.4683	0.655	2.243	0.025	0.185	2.751	0.024889	True	1.1038
State	-0.0871	0.022	-3.971	0	-0.13	-0.044	0.000072	True	1.2120

Source: *from the study, Python, 2024*

The refined logistic regression model demonstrates a robust and reliable framework for credit risk assessment, balancing interpretability with predictive power. The analysis reveals that the model effectively classifies credit outcomes, with significant predictors offering valuable insights for strategic decision-making. Among the key findings, the Disbursement Turnaround Time (TAT) emerged as a critical factor, indicating that faster loan disbursements positively influence credit performance. Similarly, Credit Bureau Score showed a significant negative relationship, underscoring the importance of reliable credit histories in reducing default risks. The model also highlights the influence of the loan disbursements acquisition channel, which significantly impacts credit behavior, suggesting targeted strategies for optimizing distribution pathways. Additionally, state-level variations reveal the geographic disparities in credit risk patterns, calling for region-specific interventions.

Although several variables, such as loan amount, net pay, and repayment frequency, were retained in the model, their statistical insignificance indicates a lesser direct impact on credit outcomes. By addressing multicollinearity concerns using Variance Inflation Factor (VIF) measures, the model ensures reliable and unbiased results, further enhancing its credibility. With an overall accuracy of 77%, the model achieves a balanced performance across classes, making it a dependable tool for operational risk assessment. While alternative models, such as decision trees, offer higher accuracy, the refined logistic regression model stands out for its transparency and practical interpretability, making it particularly valuable for organizations seeking actionable insights and strategic improvements in credit risk management.

4. Conclusion and Recommendations

This study reveals that key variables such as loan amount, repayment frequency, and the number of dependents have substantial predictive power for credit risk, which strengthens the model's ability to assess applicant profiles accurately. The integration of Weight of Evidence (WOE) and Information Value (IV) not only improved model performance but also highlighted feature importance, allowing for a more targeted and efficient risk assessment process. The refined model effectively addresses multicollinearity issues, further enhancing its stability and reliability for credit risk prediction.

Credit risk management practices can be improved by prioritizing predictive variables identified in this analysis, such as applicant state, loan amount, channel, and credit bureau score, to streamline the credit assessment process. Additionally, implementing a regular review of these variables in response to market changes can ensure the model remains aligned with real-world risk factors. FinTech companies are encouraged to adopt WOE and IV methodologies to continuously evaluate and adjust feature importance, maintaining model effectiveness in various economic conditions.

Future research could explore the inclusion of dynamic variables, such as macroeconomic indicators or industry-specific risk factors like; interest rate, technology disruption, to capture broader influences on credit risk. Additionally, assessing the model's performance across different applicant demographics and loan types may uncover unique insights for segment-specific risk management. Further studies could also investigate alternative decision algorithms, like decision tree, neural network, to enhance predictive accuracy and resilience in credit risk assessment frameworks.

References

Adebayo, O., & Abolaji, O. (2023). A comparative analysis of credit scoring techniques in commercial banks in Nigeria. *Journal of Banking & Finance*, 35(2), 301-325.

Adigun, O. (2021). Development of credit scoring models for SMEs in Nigeria. *Small Business Economics*, 19(2), 201-220.

Akins, T. (2019). The management of loan repayment behavior. *Journal of Financial Studies*, 34(2), 115-129.

Akram, M., & Hussain, A. (2020). The challenge of non-performing loans: Predicting and managing loan repayment behavior. *Financial Management Review*, 45(3), 201-216.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609.

Beaver, W. H. (2021). Financial ratios as predictors of failure. *Journal of Business Research*, 15(3), 401-419.

Einav, L., Jenkins, M., & Levin, J. (2018). The adoption of credit scoring and its effects on loan originations, repayment, and defaults. *Finance Journal*, 56(1), 78-102.

Emile, A. (2021). The role of healthy loan portfolios in maintaining liquidity and profitability. *Banking Operations Review*, 52(1), 133-149.

Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 160*(3), 523-541.

Khandani, A. E., Kim, A. J., & Lo, A. W. (2020). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11), 2767-2787.

Kofi, T. (2017). Extension of credit facilities by Microfinance institutions. *Microfinance Journal*, 19(3), 110-123.

Lamichhane, B.D. (2022). Loan Delinquency in Microfinance Institutions (MFIs): Ways to Overcome the Problem. *Nepalese Journal of Management Research*, 2, 37-43

Le, H.N.Q., Nguyen, T.V.H., Schinckus, C., 2022. Bank lending behaviour and macroeconomic factors: A study from strategic interaction perspective. *Heliyon ScienceDirect*, 8(11) https://doi.org/10.1016/j.heliyon.2022.e11906

Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 124-136.

Ogunyele, T., & Akanni, O. (2021). Techniques for informed credit decisions in a complex financial landscape. *Journal of Credit Risk Analysis*, 34(2), 189-205.

Ojeka, S. (2019). The impact of credit scoring models on credit risk management in Nigerian commercial banks. *Journal of Risk Management*, 12(2), 201-215.

Okafor, C., & Agiomoh, U. (2022). Credit scoring and loan performance in Nigerian banking sector. *Journal of Financial Stability*, 18(1), 102-120.

Sum, R.M.D., Ismail, W., Abdullah, Z.H., Noor Shah, N.F. and Hendradi, R. (2022). A New Efficient Credit Scoring Model for Personal Loan Using Data Mining Technique Toward for Sustainability Management. *Journal of Sustainability Science and Management,* 17(5), 60-76.

Thomas, J. (2021). Credit scoring and its applications. *International Journal of Finance & Economics*, 28(4), 512-530.