Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 9, No. 2, April, 2025

Early Prediction Detection of Retail and Corporate Credit Risks Using Machine Learning Algorithms

Mohamed A. Hamada [1*], Karim Farag [2], Adejor E. Abiche [3]

[1] *Abu Dhabi University, Abu Dhabi, United Arab Emirates.*

[2] *Berlin School of Business and Innovation, Berlin, Germany.*

[3] *Astana IT University, Astana, Kazakhstan.*

Abstract

Nowadays, banks operate in a highly dynamic environment where substantial vulnerability to credit risk exposures threatens their performance by affecting the quality of bank portfolios and increasing their vulnerability to insolvency. In this context, the paper reviews the existing literature and finds no studies investigating the determinants of retail and corporate credit risk using machine learning techniques to enhance the predictive performance of bank credit risk exposures. Consequently, the paper aims to utilize machine learning algorithms, regression analysis, and classification models to identify the most effective predictive model that can improve banks' credit risk prediction capabilities. It will cover the period from 2011 to 2023 and analyze a sample of 26 banks operating in Egypt. Additionally, it classifies credit risk into retail and corporate categories to develop more robust predictive models tailored for the retail and corporate sectors of bank credit risk management, thereby underscoring the paper's novelty. The findings showed that the Random Forest and Kernel SVM can be used to improve the prediction of corporate and retail credit risk by utilizing bank-specific factors like profitability, liquidity, income diversification, capital, asset size, and operating efficiency, as well as macroeconomic factors like external debt, inflation, exchange rate, GDP, interest rate, and foreign direct investment.

Keywords:

Predictive Models;
Corporate and Retail Credit Risk;
Random Forest; Decision Tree;
Naïve Bayes; Kernel SVM;
Support Vector Machine;
K-Nearest Neighbors.

Article History:

Received:	26	January	2025
Revised:	15	March	2025
Accepted:	23	March	2025
Published:	01	April	2025

1- Introduction

Credit risk is a significant concern in the banking system, threatening banks' profitability and solvency. It compels banks to invest excessively in enhancing loan screening performance, adopting modern software, and hiring more qualified professionals and experts to manage risks that could jeopardize their growth and survival within the economy [1]. Credit risk refers to the possibility that borrowers will default on their loans, which diminishes the value of assets as losses are translated into lousy debt expenses. This situation increases expenses relative to income and diminishes profitability [2]. Furthermore, it reduces capital by absorbing default losses, rendering banks more susceptible to insolvency risk, threatening their ability to operate in credit markets, and diminishing their effectiveness in fostering economic growth [3]. Additionally, compared to the other types of risks, Koju et al. [4] demonstrated that credit risk has the highest probability of occurrence and impact size, which signifies that credit risk is the most important type of risk that needs to be carefully managed to prevent bankruptcy risk. Credit risk is the possibility that a borrower won't repay his loans to the banks as agreed, which would cause the banks to turn these losses from assets into lousy debt expenses that should be borne by bank capital, endangering the bank's solvency and degrading the quality of bank portfolios [5].

* **CONTACT**: mohamed.hamada@adu.ac.ae; mohamed_abohamada@yahoo.com

DOI: http://dx.doi.org/10.28991/ESJ-2025-09-02-025

Furthermore, Kryzanowski et al. [2] argued that banks classify their loans based on borrowers' repayment performance. In other words, loans whose installments are paid on time are classified as performing loans.

In contrast, loans with installment delays exceeding three months are classified as non-performing loans (NPLs). Banks use the NPL ratio—defined as the proportion of NPLs to total gross loans—to measure their exposure to credit risk. Karaaslan [6] noted that a higher NPL ratio indicates increased credit risk and a lower quality of the bank's loan portfolio. Banks with elevated NPL ratios are expected to increase their capital reserves to absorb potential loan losses; otherwise, they risk insolvency. To prevent such outcomes, central banks enforce regulations requiring financial institutions to maintain minimum capital levels relative to their assets. Accordingly, banks must closely monitor the factors influencing corporate and retail credit risk to accurately predict their trends and improve credit risk management.

Over the past decade, Egypt has experienced significant economic challenges, including sharp currency devaluation, rising inflation, high interest and unemployment rates, and a notable slowdown in economic growth. These factors have made it increasingly difficult for individuals and businesses to repay bank loans, contributing to a rise in the NPL ratio and threatening the solvency and stability of banks [7]. Furthermore, Egypt's growing foreign debt burden has aggravated currency devaluation. In response to economic hardship, the country relaxed credit regulations to stimulate lending, which inadvertently increased corporate credit risk exposure [8]. This study compiled data on corporate and retail NPLs (CNPL and RNPL) in Egyptian banks from 2011 to 2023 and plotted two graphs (Figures 1 and 2) to illustrate the trends in credit risk over time. The data show a consistent decline in NPL ratios from 2011 to 2018, followed by a reversal during the COVID-19 pandemic between 2018 and 2021. Recently, the ratios have improved, indicating better control of credit risk. Moreover, both figures highlight a strong correlation in the historical patterns of CNPL and RNPL, although with different magnitudes, with CNPL generally exceeding RNPL.

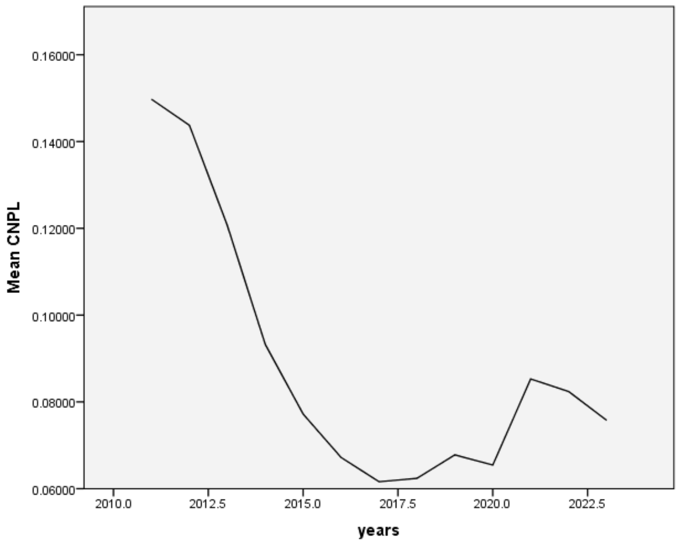

Figure 1. **CNPL (Annual reports of the commercial banks of Egypt)**

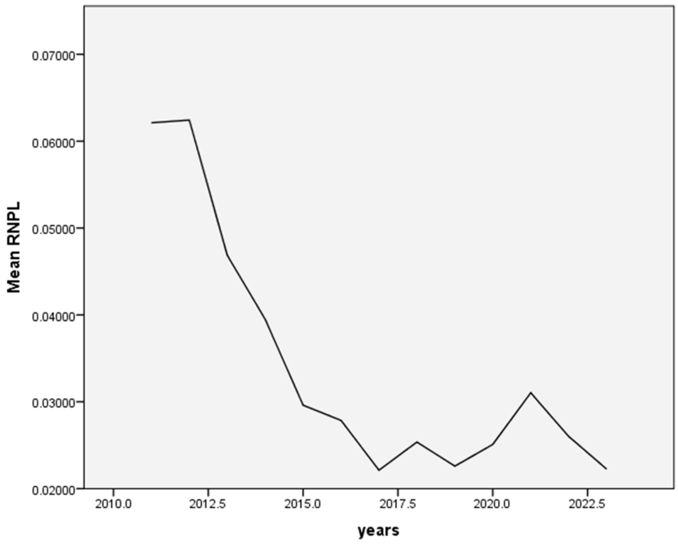

Figure 2. **RNPL (Annual reports of the commercial banks of Egypt)**

Bao et al. [9], Li [10], Bhatore et al. [11], Bussmann et al. [12], Shi et al. [13], Montevechi et al. [14], Sugozu et al. [15], and Gasmi et al. [16] utilized the machine learning algorithms to enhance credit risk prediction, and it turns out that the models generated high accuracy scores, proving the machine learning algorithms' efficiency in improving credit risk prediction performance. However, Isaev & Masih [17], Kjosevski et al. [18], Farag et al. [8], and Farag [7] employed the Generalized Method of Moments (GMM) with the enablement to analyze credit risk determinants by classifying credit risk into retail and corporate categories. The results revealed notable differences, underscoring the importance of categorizing credit risk before analysis. However, recent research has not identified any studies employing machine learning to enhance the predictive performance of retail and corporate credit risk in a comparative context. Therefore, the research aims to apply machine learning algorithms and regression models to improve the predictive models of retail and corporate credit risk, which are prerequisites for improving the financial performance of banks, guaranteeing the continuity of boosting the growth of the economies.

2- Literature Review

Banks play a crucial intermediate role in supporting economic growth by efficiently allocating their financial resources to the most productive investments. However, unexpected credit risk threatens this pivotal role and raises the risk of insolvency, which can force banks to declare bankruptcy. Therefore, banks must carefully and efficiently manage their loan portfolios to survive and thrive in the economy [8]. ElGaliy [19] supports the diversification theory, which encourages banks to spread their assets into alternative investments to lower credit risk. On the other hand, the theory of “too big to fail” suggests that large banks may engage in risky loan investments, believing that governments are always prepared to bail them out in cases of financial distress [20]. In this respect, large banks should closely monitor their risk appetite while managing their portfolios to avert the threats of insolvency.

Furthermore, Farooq et al. [21] advocate the theory of “bad management," asserting that banks with high expense-to-income ratios experience poor operational efficiency, making them more vulnerable to credit risk. Therefore, banks must do their utmost to control and manage their resources efficiently to achieve greater profits while minimizing overall credit risk. Moreover, an opposing theory, "Cost Skimping," is presented by Naili & Lahrichi [1], who argue that when banks reduce costs to enhance profitability, it often undermines the quality of credit risk management, thereby increasing credit risk exposure. However, when banks invest in technology to save money in the future by replacing or reducing staff and branches, it can sometimes enhance portfolio quality. This, in turn, improves operational efficiency and reduces exposure to credit risk.

The literature finds that the bank-specific variables also play a sensitive role in reshaping the volume of the NPL relative to the gross loans, which indicates the importance of studying these internal factors to defend against any insolvency risk that could affect bank profitability and solvency and, thus, lower the growth of the economies. In this regard, Gulati et al. [22] proved that bank profitability hurts NPLs, arguing that high-profit banks have less incentive to engage in risky investments, while those banks with low-profit margins are looking for those risky borrowers to charge them higher interest rates to receive higher profit as compensation for bearing higher risk compared to the other types of loans. Additionally, based on the “diversification theory,” Karadima & Louri [23], Barra & Ruggiero [3], and Naili & Lahrichi [1] argued that the size of assets also affects the credit risk size because large-sized banks always have more potential to diversify their risk and hire more professionals to enhance their management performance and thus achieve lower credit risk exposure than small-sized banks. However, some other studies adopt the theory of “too big to fail,” which explains that large-sized banks always have the incentive to engage in risky investments to improve their profitability, believing that the central banks always protect and bail out those large-sized banks from failure to protect public funds and the economy as well.

In addition, Wood & Skinner [24], Ozili [25], Barra & Ruggiero [3], and Naili & Lahrichi [1] find that bank capital also significantly affects credit risk, arguing that large-capitalized banks have high protection against insolvency and always prudently manage their portfolios to protect their capital from dilution and consequently achieve lower credit risk exposure. However, some articles stated that large-capitalized banks have more incentive to engage in risky investments because they have a high buffer against insolvency, resulting in more bank NPLs and deteriorating bank portfolio quality. Additionally, Mehmood & De Luca [26] unveiled that lending interest rates play a vital role in affecting credit risk, but they have a vague relationship with the NPLs; some studies find a positive effect on credit risk, arguing that higher interest rates place more financial pressure on borrowers, weakening their repayment capacity and thus raising the NPL ratio, and others find a negative effect on credit risk, stating that higher levels of interest rates discourage borrowers from taking loans, reducing the growth of loans and thereby reducing the NPLs.

On the other hand, macroeconomic factors have a sensitive role in affecting the degree of credit risk in banks. Mpofu & Nikolaidou [27], Naili & Lahrichi [1], and Christodoulou-Volos [28] illustrated that the growth of the Gross Domestic Product (GDP) hurt NPLs because increases in GDP growth rate imply that most industries will have better financial positions due to the higher sales and thus enhance firm abilities in repaying their debts, which consequently lowers the credit risk exposure in banks. In addition, public and foreign debts also play a crucial role in affecting the size of the credit risk in banks. Those countries with substantial public debt burdens will contribute less to the growth of the economies and place more pressure on banks, affecting the amount of their reserves and thus reducing the quality of bank portfolios. Further, inflation also plays a pivotal role in the growth of loans and its impact on credit risk exposure. Radivojević et al. [29] contend that increases in inflation weaken borrowers' financial positions, leading to higher loan default rates. Moreover, Gulati et al. [22], Niepmann & Schmidt-Eisenlohr [30], and Ofria & Mucciardi [31] clarified that the exchange rate is a double-edged sword; if the country devalues its currency without having a stable source of foreign income and production increase over the years, it could exacerbate the economic problem and raise the credit risk exposure in banks, threatening their solvency and growth. Further, Shehata [32] conducted a comparative analysis examining macroeconomic factors' effect on credit risk in Egypt's listed and non-listed banks. The results indicated that the listed banks' credit risk was more sensitive to macroeconomic variables than the non-listed ones. As a result, the listed banks in Egypt were forced to pay closer attention to the external economic environment and take precautions against any threats arising from credit risk that could jeopardize their survival and expansion. In Egypt, ElGaliy [19] examined the effect of macroeconomic factors on the NPLs covering the period of 2011-2020. The results show that the Egyptian economy's growth severely affects the credit risk exposure of banks. Moreover, Farag et al. [8] employ the fixed-effect and GMM models to conduct a comparative study between the determinants of the retail and corporate credit risks in the banks of Egypt, selecting twenty-eight commercial banks out of thirty-seven covering the period of 2011-2020. The findings reveal that both models are closely the same and conclude that the corporate credit risk is more sensitive to the macro- and microeconomic factors compared to the retail credit risk, which showed differences in the results of the corporate and retail NPL ratios, which illustrates the importance of classifying the credit risk to enhance the prediction of the overall credit risk.

Tajik et al. [33] investigated the impact of U.S. housing prices on bank non-performing loans (NPLs) using the system Generalized Method of Moments (GMM) with a sample of U.S. banks covering the period from 1999 to 2012. They categorized loans into distinct types to determine which categories were more sensitive to housing price fluctuations. Their findings highlight the importance of classifying loans to produce clearer and more accurate results. Specifically, changes in housing prices were found to influence NPL dynamics, with the degree of impact varying across loan types. In the Republic of Macedonia, Kjosevski et al. [18] examined the influence of both bank-specific and macroeconomic factors on credit risk using the Autoregressive Distributed Lag (ARDL) model for the period 2009–2014. They classified NPLs into retail and corporate segments and found that different factors influenced each category. Elnaggar et al. (2020) applied machine learning algorithms—including Support Vector Machines (SVM), Decision Trees (DT), and Meta-classifiers—to predict agricultural NPLs using data from Egyptian agricultural banks. The objective was to improve credit risk management and support investment in the agricultural sector for better economic outcomes.

In Turkey, the rising NPL ratio—from 2.74% in 2014 to 5.02% in 2019—underscored the need for advanced predictive tools. Serengil et al. [34] utilized machine learning techniques such as SVM, Random Forest, Bagging Classifier, XGBoost, and LSTM to forecast NPL trends in Turkish banks. Their approach enhanced credit risk prediction, helping maintain higher portfolio quality, mitigate insolvency risk, and promote credit market growth. Similarly, in Spain, Alonso and Carbó [35] employed Logistic Regression (Logit), Classification and Regression Trees (CART), Random Forest, XGBoost, and Deep Neural Networks to predict NPLs in a large Spanish bank. Their findings showed that XGBoost could reduce regulatory capital requirements under the IRB approach by 12.4% to 17%, reinforcing the effectiveness of machine learning in managing bank insolvency risk and fostering market growth. Additionally, Abdullah et al. [36] applied machine learning to predict NPLs in 322 banks across 15 emerging markets. Their results demonstrated that machine learning models outperformed traditional linear models in forecasting NPLs, thereby improving credit risk management performance. Beltman et al. [37] also explored the use of various machine learning algorithms—Logistic Regression, Random Forest, and XGBoost—to predict retail NPLs in the largest Dutch commercial banks using customer-level data such as credit card, mortgage, and checking account activity. The study found that the Random Forest model achieved the highest accuracy (based on AUC), making it a highly effective early warning system for retail NPL prediction.

Farag et al. [38] conducted a comparative study between listed and non-listed banks of the stock exchange of Egypt (EGX) by employing the GMM to investigate the effect of retail and corporate credit risk on capital adequacy covering the period from 2011 to 2023. The results demonstrate that the retail credit risk is negatively related to the capital risk of the non-listed banks. In contrast, the listed ones are found to be insignificant, which confirms the importance of classifying the credit risk into retail and corporate risk before being examined to provide more unbiased models and accurate prediction performance. Further, Waleed et al. [39] studied the determinants of credit risk in the banks of Egypt by employing the GMM covering the period from 2012 to 2022. According to the results, credit risk is negatively impacted by profitability, capital adequacy, and income diversification. It was claimed that banks with high profit margins are less likely to make risky loan investments and that large capital-to-asset ratios force banks to make rational credit decisions. Additionally, higher diversification ratios result in less reliance on loans when combined with investments in non-traditional assets, which lowers exposure to credit risk and improves the quality of bank portfolios. Moreover, Getinet et al. [40] found that income diversification plays a vital role in reducing credit risk exposure in banks by relying less on traditional investments and reallocating bank reserves more to non-interest-bearing investments to strengthen banks' survival and growth in times of recessions and crises.

Furthermore, Permana & Rahyuda [41] employed regression to study the credit risk determinants using a sample from listed banks in the Indonesian stock exchange covering the period from 2019 to 2023. The findings demonstrated that profitability positively relates to credit risk, arguing that banks with high-profit margins are discouraged from engaging in risky investments and thus lower credit risk. At the same time, the operating efficiency was found to be positively significant, claiming that a higher level of operating expenses reduces operating efficiency and deteriorates bank profitability. Additionally, Christodoulou-Volos [28] employed the GMM on a sample of Cyprus' banks from 2013 to 2019 and discovered that increased economic growth and profitability lower credit risk exposure. He argued that banks that can make large profits during economic boom periods are less likely to engage in riskier investments, which keeps them with low credit risk vulnerability.

2-1- Literature Gap Analysis

Based on the literature reviewed and to the best of the authors' knowledge, this paper concludes that while several studies—including those by Bao et al. [9], Li [10], Bhatore et al. [11], Bussmann et al. [12], Shi et al. [13], Montevechi et al. [14], Sugozu et al. [15], and Gasmi et al. [16]—have applied machine learning algorithms to predict credit risk, they did so without differentiating between categories of credit risk. In contrast, studies by Isaev & Masih [17], Kjosevski et al. [18], Farag et al. [8], and Farag [7] utilized regression models to examine the determinants of corporate and retail credit risk in the banking sector, identifying significant differences between the two. Notably, the literature lacks studies that apply machine learning algorithms specifically to predict both retail and corporate credit risks separately. This highlights a critical gap that this research aims to address by leveraging machine learning techniques to improve the predictive accuracy of credit risk models and reduce insolvency risk, which poses a threat to banks' stability and growth. To this end, the study applies both regression and machine learning models to a dataset from Egyptian banks, distinguishing between retail and corporate credit risk. The objective is to develop robust predictive models that can serve as early warning tools for credit managers, enabling improved credit risk management, stronger portfolio quality, and reduced insolvency rates—thereby enhancing banks' intermediation roles and supporting long-term economic sustainability.

3- Research Methods and Dataset

The research aims to employ regression models and other machine learning algorithms to improve the prediction performance of the retail and corporate credit risk in the banks of Egypt by using a sample of 26 banks out of 36 registered with the Central Bank of Egypt (CBE). The research has macroeconomic data such as EDEBT, IDEBT, INF, EXR, GDP, INT, and FDI collected from the published annual reports of the CBE. In contrast, the bank-specific data, such as CNPL, RNPL, ROA, CAR, SIZE, OPEFF, DIV, and LTD, were gathered from the published annual reports of the commercial banks of Egypt, using panel data from 2011 to 2023. Moreover, the research employed various techniques and algorithms to analyze the research problem and extract accurate results comprehensively. Regression classification models analyze the credit risk dataset for CNPL/RNPL. Each model was preprocessed, trained, and evaluated using the R^2 score to assess accuracy and predictive power. The missing values in the dataset were replaced using mean imputation. The dataset was then split into training and testing sets. Also, polynomial regression is used in model training by incorporating polynomial features for non-linear relationships. It is trained with polynomial transformations, and predictions are compared with actual values. In performance evaluation, the R^2 score was employed for regression and the confusion matrix with accuracy score for classification. The models used in this research are:

Support Vector Regression (SVR) uses a kernel trick for regression in higher-dimensional feature space.

Decision Tree Regression captured complex patterns by splitting data based on decision rules.

Random Forest Regression is an ensemble method that combines multiple decision trees for improved accuracy. It was trained with optimized parameters and showed robust predictive performance.

Multiple Linear Regression modeled relationships using various independent variables and preprocessed data with feature scaling and encoding.

Polynomial regression extends linear regression by incorporating polynomial features to capture non-linear relationships. It extends linear regression by fitting a polynomial equation to the data, allowing it to capture non-linear relationships. In this respect, the model has trained the model with polynomial feature transformations, predicted outcomes for test data, and compared with actual values, and the R^2 score was calculated to evaluate fit quality. Further, The Support Vector Regression (SVR) is utilized in the paper to employ a kernel trick to perform regression in higher-dimensional feature space, ideal for non-linear relationships. It predicts continuous values by finding a hyperplane in the feature space that adjusts the shape of the target variable for compatibility with the SVR model. The scaled features use standardization for improved SVR performance, train the SVR model, and make predictions using the R^2 score to quantify model accuracy on test data.

The Decision Tree Regression splits data based on decision rules, capturing complex patterns. A tree-based algorithm that divides data into branches based on feature values effectively modeling complex relationships trains the decision tree on the entire training dataset and predicts test outcomes using the trained model, using the R^2 score to evaluate the accuracy and overfitting concerns. Moreover, the Random Forest Regression was adopted in the paper as an ensemble learning method combining multiple decision trees to improve accuracy and reduce variance. An ensemble method combines numerous decision trees to improve accuracy, reduce variance, and handle overfitting. It trains the model on the dataset with optimized parameters, predicting new and test data outcomes and providing the R^2 score calculated for evaluation to show robust predictive performance. In addition, Multiple Linear Regression is used to indicate a dependent variable based on multiple independent variables to study the relationship between the dependent variable and the independent variables using a linear equation, preparing data for regression, importing libraries, and extracting features and target variables to predict a dependent variable based on multiple independent variables.

3-1- Preprocessing Techniques and Procedures

The imported libraries NumPy, Pandas, and Matplotlib are loaded and prepared for the dataset. The extracting features of the matrix (X) and target or dependent variable vector(y) are designated appropriately at the extraction stage. Pre-processed missing data and split the dataset into training and test subsets and the dependent vector (y) encoding. Lastly, the feature scaling of the matrix (X) features is done after splitting the dataset to avoid data leakage before building the respective models for training and testing the dataset.

Accordingly, the paper developed the following econometric models for the regression analysis:

$$\begin{aligned} \Delta CNPLit = \alpha i + \sum \beta 1\, ROA + \sum \beta 2\, CAR + \sum \beta 3\, SIZE + \sum \beta 4\, OPEFF + \sum \beta 5\, DIV + \sum \beta 6\, LTD \sum \beta 7\, EDEBT + \sum \beta 8\, INF \\ + \sum \beta 9\, EXR + \sum \beta 10\, GDP + \sum \beta 11\, IDEBT + \sum \beta 12\, INT + \sum \beta 13\, FDI + eit \\ \Delta RNPLit = \alpha i + \sum \beta 1\, ROA + \sum \beta 2\, CAR + \sum \beta 3\, SIZE + \sum \beta 4\, OPEFF + \sum \beta 5\, DIV + \sum \beta 6\, LTD \sum \beta 7\, EDEBT + \sum \beta 8\, INF \\ + \sum \beta 9\, EXR + \sum \beta 10\, GDP + \sum \beta 11\, IDEBT + \sum \beta 12\, INT + \sum \beta 13\, FDI + e_{it} \end{aligned} \tag{1}$$

$\Delta CNPL_{it}$ denotes the Corporate Non-performing loan ratio for bank ***i*** at time ***t***, **αi** denotes the constant term specific to each bank ***i***. It accounts for bank features that remain constant throughout time, $\sum$ **β** denotes the beta coefficient of the impact of each explanatory variable on the CNPL and RNPL models, **$_{it}$** denotes the error term, which captures the impact of other factors not included in the model.

The above-formulated econometric models consist of two models investigating the effect of the macroeconomic and bank-specific variables on retail and corporate credit risk. Equation 1 is the CNPL model that measures the impact of the explanatory variables on the corporate credit risk, which is measured by the corporate NPL ratio (CNPL), while Equation 2, for the retail credit risk, is a proxy for the retail NPL ratio. Each econometric model contains 13 independent variables divided into six bank-specific variables and seven macroeconomic variables, as shown in Table 1.

Table 1. Variables and Measurements

Variables	Symbols	Measurements
Corporate Non-performing Loans	CNPL	Corporate Non-Performing Loan ratio = corporate nonperforming loans / total corporate loans
Retail Non-Performing Loan	RNPL	Retail Non-Performing Loan ratio = retail nonperforming loans / total retail loans
Return on Assets	ROA	ROA = Net income / total assets
Capital Adequacy Ratio	CAR	CAR = capital / total assets
Bank size	SIZE	Bank size = log of total assets
Operating efficiency	OPEFF	Operating efficiency = total expenses / total income
Income Diversification	DIV	Income diversification ratio = total fees / (total interest income + total fees)
Bank Liquidity	LTD	Loan-to-deposit ratio = total loans / total deposits
External Debt	EDEBT	Amount of external debt in dollars
Inflation	INF	Inflation rate = percentage change in consumer price index
Exchange rate	EXR	Egyptian pound / US dollars
Economic Growth	GDP	Economic growth = percentage change in real gross domestic product (GDP)
Internal debt	IDEBT	Log of internal debt in dollars
Lending Interest rate	INT	Average annual market interest rate
Foreign Direct Investment	FDI	Foreign direct investment in percentage of GDP

The following flowchart as shown in Figure 3 depicts the scenario of the research methodology till extracting the research findings.

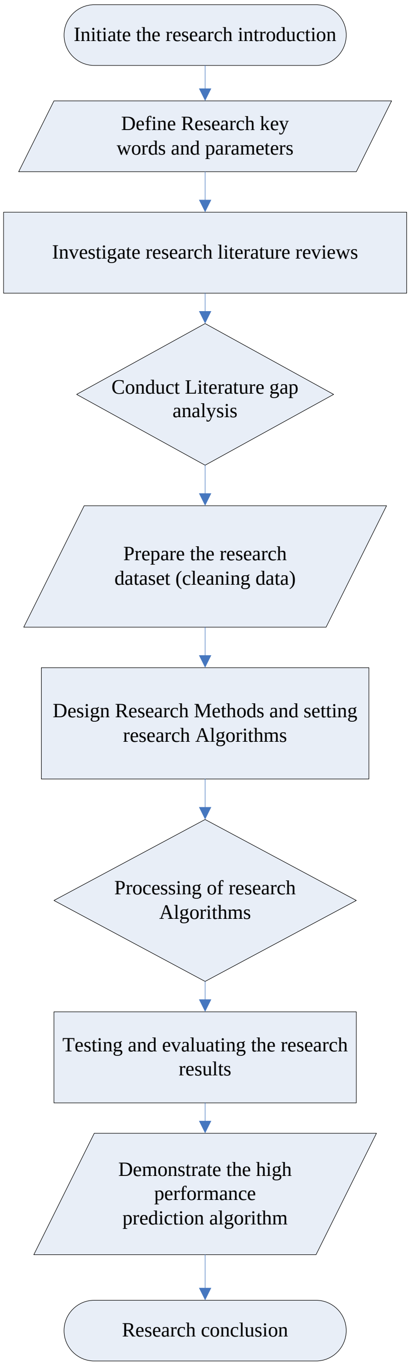

Figure 3. Research Methodology Flowchart

4- Results and Discussion

4-1- Descriptive Analysis

In this section, the paper conducted a descriptive analysis by describing the collected data regarding observations, minimum, maximum, mean, and standard deviation to illustrate the dataset on a more comprehensive scale, as shown in Table 2. The dataset contains 338 observations with a sample of panel data from 2011 to 2023, covering the last 13 years, considered a new economic transition stage for Egypt, selecting 26 banks out of 36. The mean of the CNPL is 8.87%, while the RNPL is 3.4%, indicating that the CNPL accounts for the majority of the total credit risk in the banks. Additionally, the standard deviation of CNPL is 9.3% compared to 4.2% in RNPL, indicating that the corporate credit risk has higher volatility than the RNPL. Furthermore, the mean ROA is 1.55%, along with a 1.3% standard deviation, which shows the high stability of the bank's profitability on average of 1.5%. Moreover, the mean capital adequacy ratio is 17.6%, meeting and far exceeding the required ratio of 12.5% of regulated capital, which shows intense solvency levels in the commercial banks of Egypt.

Table 2. Descriptive analysis

	N	Minimum	Maximum	Mean	Std. Deviation
ROA	338	-0.05870	0.08989	0.0155141	0.01291733
CAR	338	0.00000	0.44800	0.1763905	0.05617606
SIZE	338	6.34664	12.35702	10.1334061	1.18255980
OPEFF	338	0.19071	1.04526	.5800032	0.15222979
DIV	338	0.01322	0.31948	.0858628	0.04317281
LTD	337	0	1	0.48	0.152
EDEBT	338	1.52763	2.22531	1.8896604	0.24403063
INF	338	4.66000	33.70000	13.5329645	8.34045862
EXR	338	5.90000	30.94000	14.0196154	7.40357586
GDP	338	1.80000	6.60000	3.8281065	1.36397372
IDEBT	338	0.76200	1.03000	0.8872604	0.06570528
INT	338	9.50000	19.80000	13.7130178	3.35947339
FDI	338	-0.20454	3.26026	1.9374899	0.96314257
CNPL	338	0.00008	0.73505	0.0886647	0.09326356
RNPL	338	0.00004	0.30098	0.0340644	0.04229522
Valid N (listwise)	337				

However, such a high level indicates that the banks injected more capital relative to their risky assets as a precaution against expecting high fluctuations in their overall risk exposure in the near future due to the current economic instability that the country faces. In addition, the OPEFF has a mean of 58%, indicating that the commercial banks have expenses that account for 58% of their total income, which is considered very high and indicates poor efficiency in bank management. The mean of the DIV is 8.6%, illustrating that the non-interest income is 8.6% of the total income, signifying a low level of diversification. In other words, it shows that the banks rely more on traditional investments than fee-based income and off-balance sheet activities, which could make the banks more vulnerable to the overall risk, threatening their survival and growth in times of recessions and crises. Additionally, the LTD has a mean of 48%, indicating that the loans account for 48% of the total deposits, which is a normal percentage compared to other banks in Europe, the USA, and Gulf markets. Moreover, in terms of comparing all the variables with the standard deviation, the paper concluded that the inflation rate, exchange rate, and interest rate have the highest STDEV compared to the other variables, indicating that the macroeconomic variables are highly fluctuated compared to the bank-specific variables, while the ROA and CAR have the lowest STDEV, showing more stability in bank profitability and solvency during 2011-2023. Based on the results of the descriptive analysis, the paper found no anomalies in the dataset, proving the accuracy of data collection.

4-2- The Results of Machine Learning Algorithms

Figures 4 and 5 present the feature importance scores derived from the Random Forest model for predicting CNPL and RNPL, respectively. These visualizations highlight the relative contribution of each feature to the model's predictive accuracy. Features with higher importance scores have a greater influence on the model's classification outcomes. As illustrated in both figures, the variables ROA, LTD, and DIV exhibit the highest importance scores, indicating their significant roles in predicting CNPL. These are followed by SIZE, OPEFF, and CAR, suggesting that bank-specific variables are more critical than macroeconomic variables in forecasting both CNPL and RNPL. This finding underscores the pivotal role of internal bank management in mitigating corporate and retail credit risk. The results align with those of Farag et al. [8], who used the GMM approach and emphasized the importance of effective profit management in controlling CNPL and RNPL to sustain high-quality portfolios and minimize insolvency risks—risks that could undermine bank stability and their contributions to economic growth. Banks must also monitor their loans-to-deposits ratios closely to ensure adequate liquidity, preventing potential losses that could devalue assets. Furthermore, a greater reliance on income diversification—particularly in generating fee income relative to interest income—can reduce the adverse impact of credit risk on financial performance. Additionally, cost control relative to earnings should be pursued without compromising the effectiveness of credit screening processes, thereby enhancing operational efficiency and reducing the likelihood of unexpected increases in CNPL and RNPL.

Larger banks can leverage their scale to diversify loans and investments, thereby reducing overall risk. Regular monitoring of CNPL and RNPL is essential to ensure adequate capital relative to assets, maintaining financial health and enabling banks to support sustainable economic development. Conversely, the feature importance results reveal that macroeconomic indicators such as EDEBT, INF, EXR, GDP, INT, and FDI have the lowest predictive power for CNPL and RNPL. This suggests that while these external variables are less influential compared to bank-specific factors, they should not be overlooked. Given that macroeconomic factors are largely beyond banks' control, it is imperative that banks proactively adjust to changing economic conditions. This includes regularly aligning credit policies and internal risk management strategies with current macroeconomic environments to mitigate unexpected credit risk surges and ensure the continued quality of their investment portfolios.

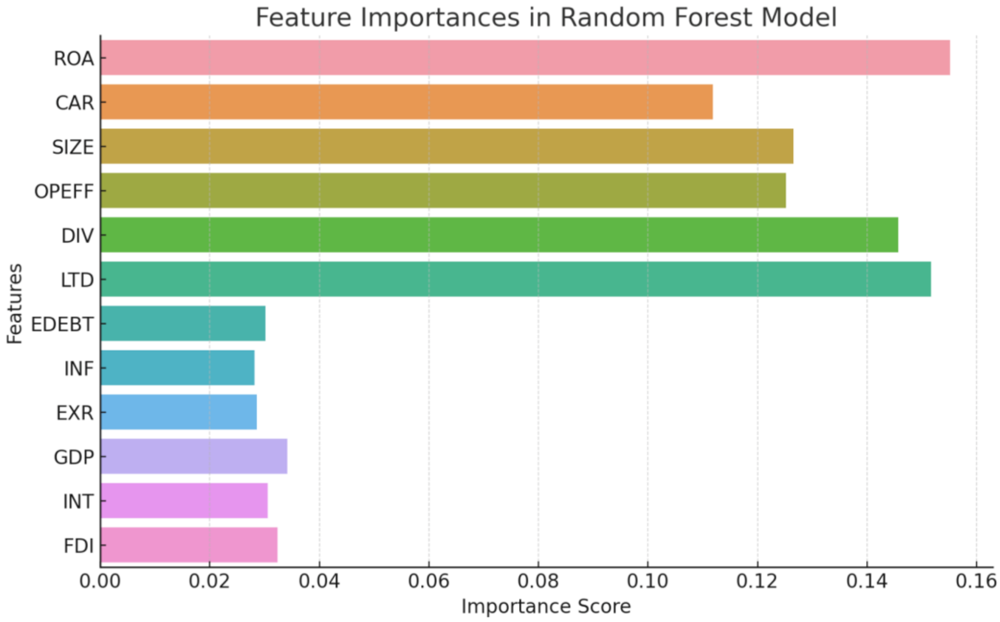

Figure 4. Features Importance in Random Forest Model in the CNPL

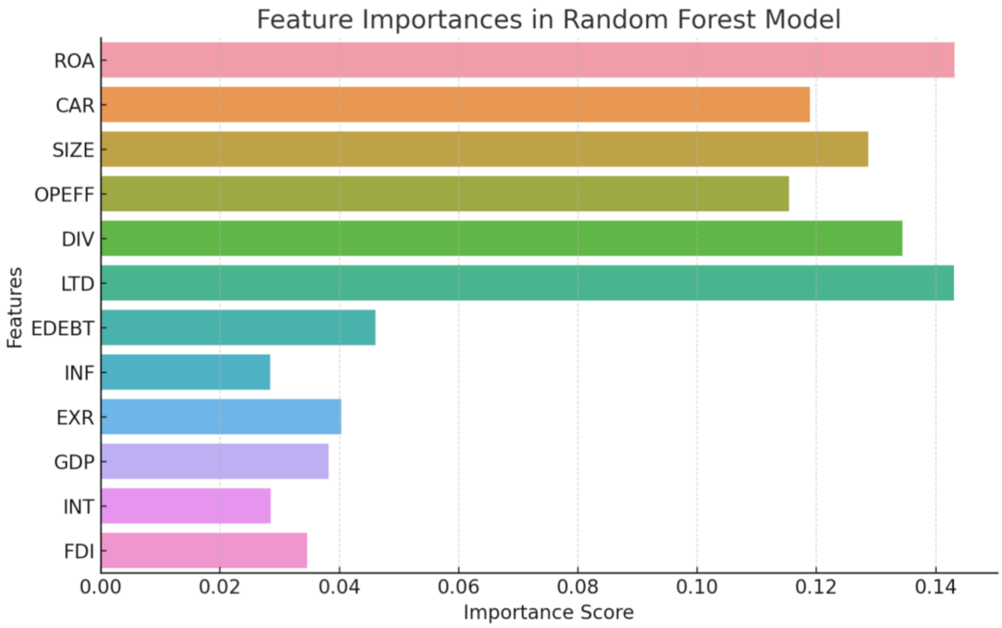

Figure 5. Features Importance in Random Forest Model in the RNPL

The paper runs the collected data into different machine learning algorithms to find the best predictive model that can predict the CNPL and RNPL based on the variations of the bank-specific and macroeconomic variables. This study underscores the strengths and limitations of various regression, classification, and Artificial Neural Network (ANN) techniques on a credit risk dataset. Balancing trade-offs between accuracy, speed, interpretability, and scalability, aligning with both technical and business needs.

Advanced models like Random Forest and simple vector regression (SVR) showed strong performance, while simple models provided valuable interpretability. These insights help guide optimal model selection in classification analyses. The R-square (R^2) Scores were used for regression model performances, while accuracy scores were used for classification models. Random Forest Regression excelled due to its ensemble nature. Simpler models like Polynomial Regression offered insights into non-linear relationships.

In this context, the study employed several machine learning algorithms—Random Forest, Decision Tree, Naïve Bayes, Kernel SVM, Support Vector Machine (SVM), and K-Nearest Neighbors—to identify the models with the lowest prediction error and highest accuracy for CNPL and RNPL forecasting, as presented in Tables 1 and 2. As shown in Table 3, the data collected can effectively predict CNPL using the Random Forest and Kernel SVM models, both achieving an accuracy of 64%, outperforming the other models. This underscores the importance of carefully managing bank-specific variables to better control corporate credit risk and enhance banks' role in supporting the broader economy. Accordingly, banks may adopt the Random Forest and Kernel SVM models as reliable tools to forecast future CNPL levels, thereby strengthening their corporate credit risk management practices. Doing so would help improve corporate credit screening processes, sustain high-quality portfolios, and reduce exposure to insolvency risks.

Conversely, Table 3 reveals that all models performed poorly in predicting RNPL, with average accuracy around 50%. This indicates that the selected bank-specific and macroeconomic variables are not sufficiently significant for predicting RNPL. These findings are consistent with those of Farag et al. [8], emphasizing the importance of classifying credit risk into corporate and retail segments before analysis to develop more accurate and robust prediction models. The results also highlight that CNPL is more sensitive to bank-specific variables than RNPL, likely due to the inherently stronger linkage between corporate credit behavior and macroeconomic and institutional factors, in contrast to the more individual-driven nature of retail credit.

Table 3. **Model Accuracy scores for CNPL & RNPL**

Model Type	Accuracy scores for CNPL	Accuracy scores for RNPL
Random Forest	64%	54%
Decision tree	55%	51%
Naïve Bayes	52%	47%
Kernel SVM	64%	55%
Support Vector Machine	62%	53%
K-Nearest Neighbors	55%	48%

4-3- The Results of Regression Model Performance and Evaluation

In this section, the research provides an overview of various regression models implemented to analyze a dataset (Credit Risk CNPL/RNPL). Each model underwent preprocessing steps, training, and evaluation using standard metrics like R^2 score to assess accuracy and predictive power. Moreover, the replacement of missing values is accomplished by using mean imputation, and the data is split into training and testing sets. The research uses various regression models, along with performance evaluation steps. Notable sections for evaluation metrics include evaluating the model performance, predicting test set results, and training and testing various regression models.

Table 4 shows the models that are used in the analysis of credit risk management with their respective predictions and accuracy. The CNPL models have different accuracy evaluated using R^2 score: This metric assesses how well the model explains the variance of the dependent variable. No explicit accuracy or predictions were found in this section, such as the Multiple Linear Regression model = 45%, Polynomial Regression model = 81%, Support Vector Regression (SVR) = 72%, Decision Tree Regression model = 86%, and Random Forest Regression model = 80%. As can be seen, the highest accuracy score is 86%, and therefore, the most recommended model is the Decision Tree Regression model.

Table 4. **R_2 of Regression models for CNPL and RNPL**

Model Type	R_2 of CNPL	R_2 of RNPL
Polynomial Regression	81%	77%
Support Vector Regression (SVR)	72%	65%
Decision Tree Regression	86%	89%
Random Forest Regression	80%	80%
Multiple Linear Regression	45%	56%

The following results were obtained: Multiple Linear Regression model = 0.08357, Polynomial Regression model = 0.01213, Support Vector Regression (SVR) = 0.30005901, Decision Tree Regression model = 0.01213, Random Forest Regression model = 0.02931. with the polynomial regression and decision tree regression having a perfect accuracy of the actual target value of 0.01213 on the CNPL dataset, as shown in Figures 6 and 7:

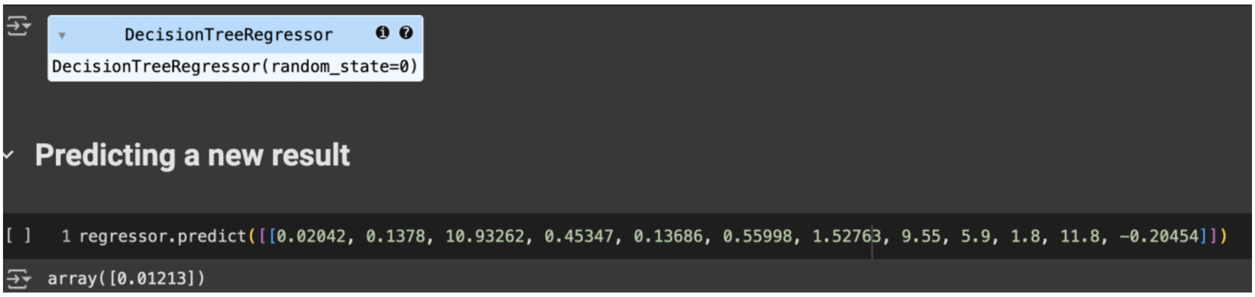

Figure 6. CNPL Decision Tree Regression model predicting Result

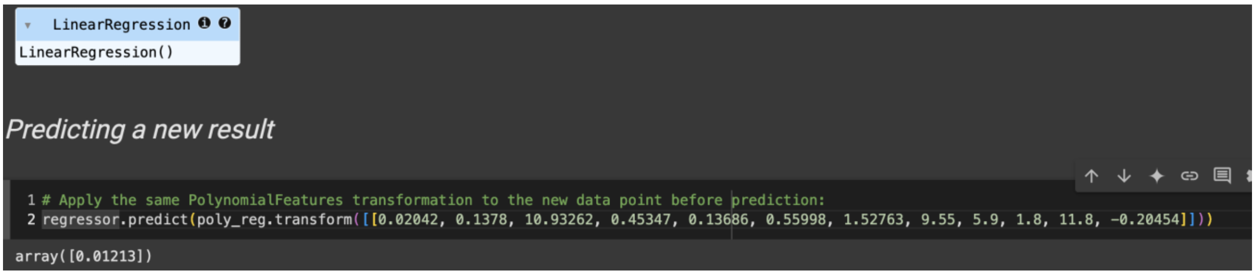

Figure 7. CNPL Linear Regression model Predicting Result

With the RNPL dataset analysis, polynomial regression and decision tree regression predicted with 100% accuracy when tested for RNPL target values of 0.00942, respectively, as shown in Figure 8:

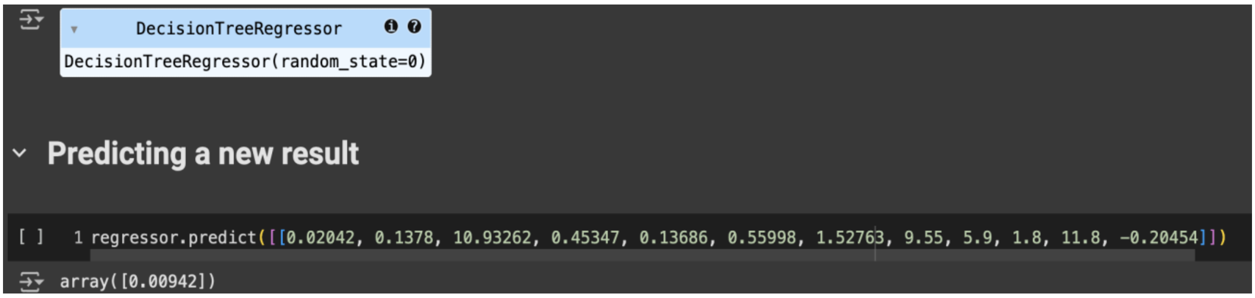

Figure 8. RNPL Decision Tree Regression Model Predicting Result

On the retail non-performing loan (RNPL) credit risk side, the accuracy results of the respective models are presented in Table (5) as follows: Multiple Linear Regression = 56%, Polynomial Regression = 77%, Support Vector Regression (SVR) = 65%, Decision Tree Regression = 89%, and Random Forest Regression = 80%. Among these, the Decision Tree Regression model achieved the highest accuracy at 89%.

4-4- The Results of Classification Model Performance and Evaluation

When contrasting the regression models with the classification models that are the primary models used for the research, the true/test values of the independent vector (y) are classified under the following feature engineering nomenclature that facilitates the classification analysis, as indicated by Figure 9:

IF y_test< 0.02 = "low" encoded with the value of **0**

y_test <0.05 = "moderate", encoded with the value of **1**

y_test >0.05 = "high" and encoded with the value of **2**

ROA	CAR	SIZE	OPEFF	DIV	LTD	EDEBT	INF	EXR	GDP	INT	FDI	CNPL
0.02042	0.1378	10.93262	0.45347	0.13686	0.55998	1.52763	9.55	5.9	1.8	11.8	-0.20454	low
0.02333	0.1571	10.975	0.46111	0.10728	0.52314	1.58883	4.66	6.1	2.2	12.2	1.00234	low
0.02299	0.1355	11.05596	0.42377	0.12163	0.43896	1.66087	11.66	6.9	2.2	11.9	1.45343	low
0.02539	0.1677	11.1573	0.41271	0.12627	0.4102	1.61595	10.13	7.1	2.9	11.4	1.50925	low
0.0259	0.1272	11.25332	0.41621	0.11571	0.37452	1.67943	11.06	7.7	4.4	11.8	2.10258	low
0.02255	0.1397	11.42136	0.45212	0.09311	0.38047	1.82802	23.27	10	4.3	16.3	2.43856	moderate
0.02561	0.193	11.46949	0.53564	0.08539	0.36457	1.91855	21.9	17.8	4.2	19.8	3.14283	moderate
0.02791	0.1909	11.53456	0.4963	0.08338	0.37291	1.98498	11.97	17.8	5.3	17.8	3.26026	low
0.03052	0.2607	11.58737	0.48192	0.07495	0.43109	2.05308	7.1	16.8	5.6	13.8	2.97284	moderate
0.02393	0.3141	11.63128	0.3979	0.0676	0.39747	2.11059	5.42	15.8	3.6	9.7	1.60212	high

Figure 9. CNPL-Dependent Vector Column Table

As shown in the screenshot above, classification is performed on the CNPL-dependent vector column based on the specified classification conditions. The following classification models were trained using the dataset: Random Forest, Decision Tree, Naïve Bayes, Kernel SVM, Support Vector Machine, and K-Nearest Neighbor. When predicting the encoded target column using the topmost input row with a corresponding value of 0, the Decision Tree, Naïve Bayes, Kernel SVM, and Support Vector Machine accurately predicted the class as 0. However, the Random Forest and K-Nearest Neighbor models misclassified the input, predicting values of 1 and 2, respectively, as illustrated in Figures 10 and 11.

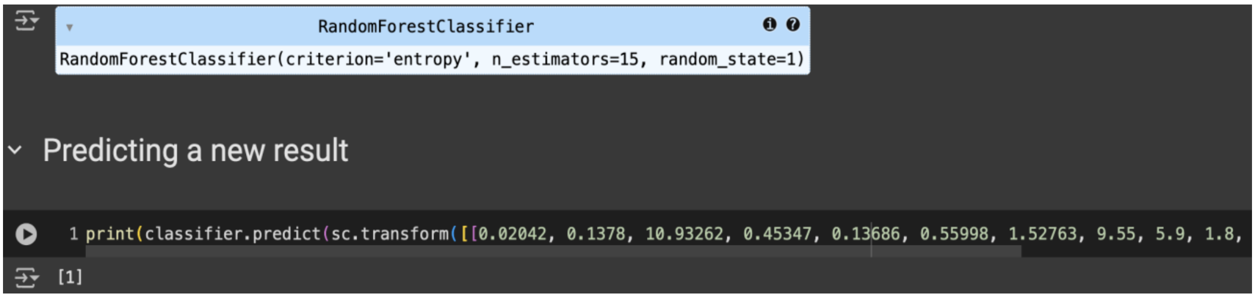

Figure 10. **CNPL Random Forest Classification Predicting Result**

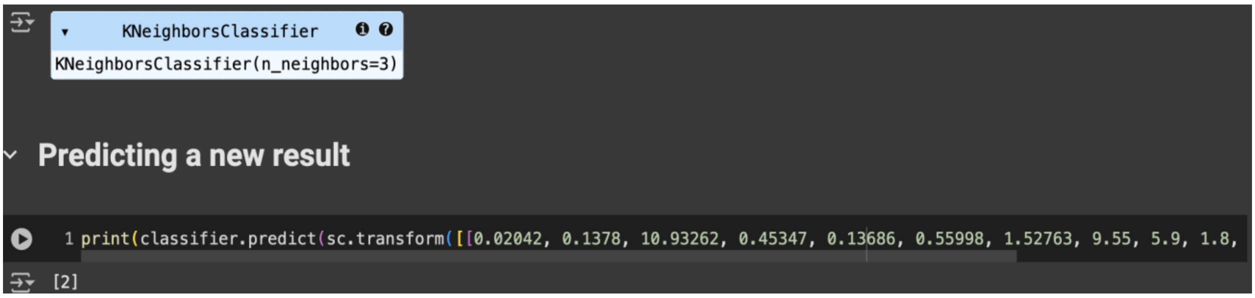

Figure 11. **CNPL K-Neighbor Classification Predicting Result**

In terms of accuracy scores, the CNPL results for each model are as follows, as presented in Table 5: Random Forest = 84%, Decision Tree = 75%, Naïve Bayes = 76%, Kernel SVM = 84%, Support Vector Machine = 82%, and K-Nearest Neighbor = 75%. On the other hand, for RNPL, the corresponding accuracy scores are: Random Forest = 84%, Decision Tree = 81%, Naïve Bayes = 77%, Kernel SVM = 85%, Support Vector Machine = 83%, and K-Nearest Neighbor = 78%.

Table 5. **Accuracy scores of the Classification models for CNPL and RNPL**

Model Type	R_2 of CNPL	R_2 of RNPL
Random Forest	84%	84%
Decision Tree	75%	81%
Naïve Bayes	76%	77%
Kernel SVM	84%	85%
Support Vector Machine	82%	83%
K-Nearest Neighbor	75%	78%

In predicting the new results with each of the model's column-wise, the following predicted values results are obtained with the corresponding y_test/true value of **0.** Random Forest, Decision Tree, and Kernel SVM predicted the value of **1,** which depicts the corresponding classification of **moderate** instead of **low,** which is incorrect (see Figures 12 to 14).

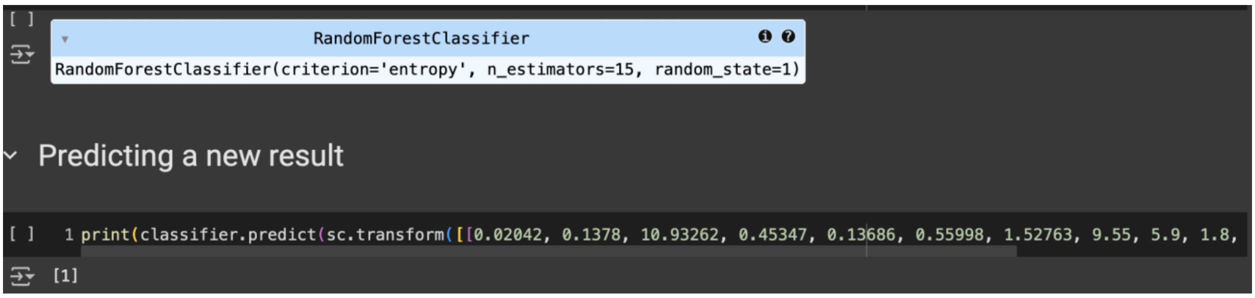

Figure 12. **CNPL Random Forest Classifier Predicting Result**

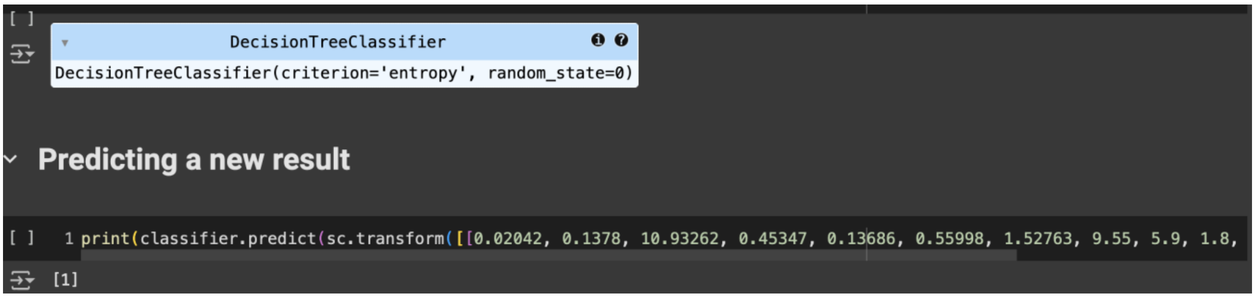

Figure 13. CNPL Decision tree Classifier Predicting Result

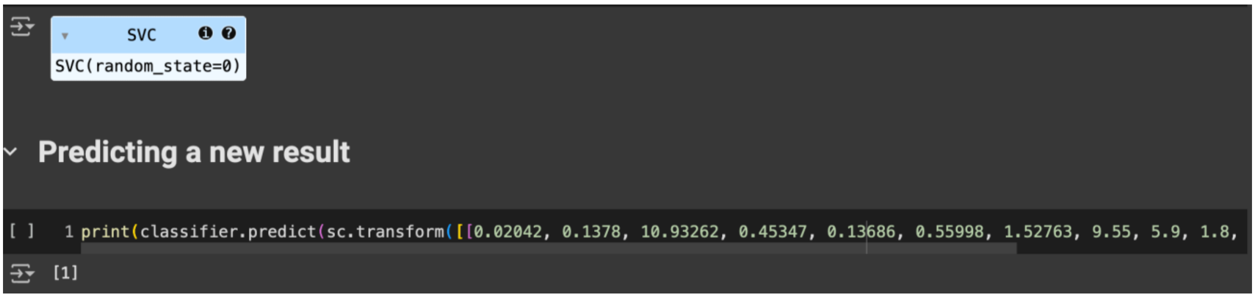

Figure 14. CNPL Kernel SVM Classifier Predicting Result

While Naïve Bayes, Support Vector Machine, and K-Nearest Neighbor have the value of 0, which correspondingly means **low**, which the actual value of the target column RNPL (y_test) (see Figures 15 to 17).

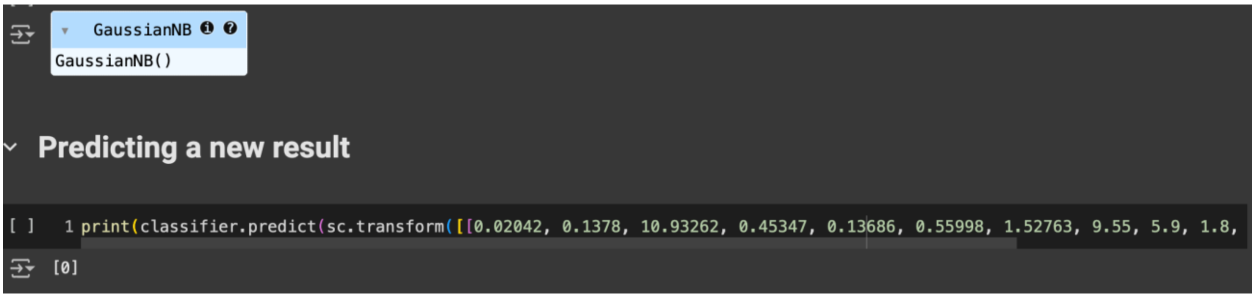

Figure 15. RNPL Naïve Bayes Classifier Predicting Result

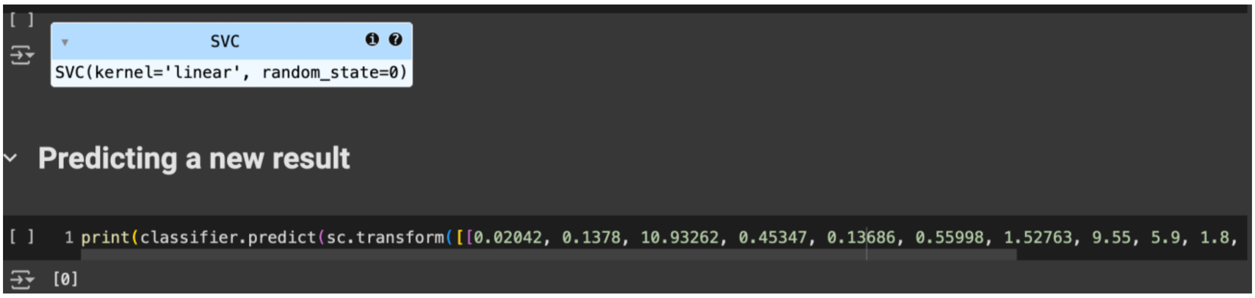

Figure 16. RNPL Kernel SVM Classifier Predicting Result

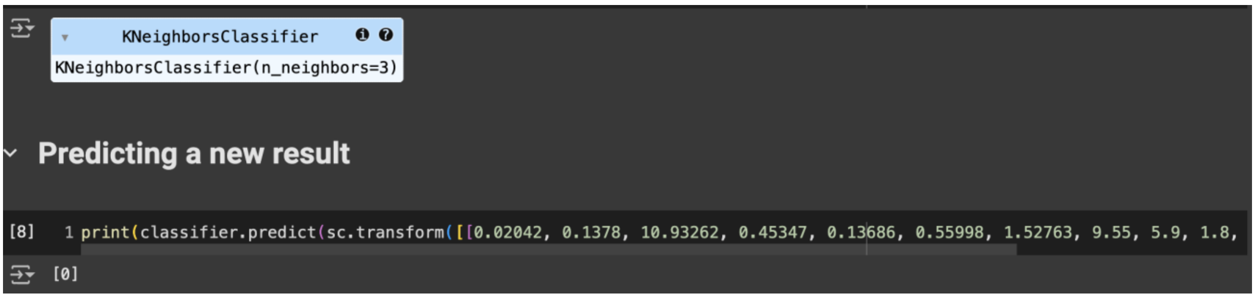

Figure 17. RNPL K-Neighbor Classifier Predicting Result

The R^2 score was used to evaluate the performance of the regression models, while accuracy scores were applied to assess the classification models. Random Forest Regression outperformed others due to its ensemble nature. Simpler models, such as Polynomial Regression, offered valuable insights into non-linear relationships. For each regression model, the R^2 score served as the primary metric for measuring performance, whereas in classification models, accuracy was used as the standard evaluation metric. These metrics effectively captured the explanatory power of each model on unseen test data. Models like Random Forest Regression achieved high accuracy thanks to their ensemble approach, while simpler models such as Multiple Linear Regression offered greater interpretability.

According to the results from the regression and machine learning algorithms, the study found high accuracy scores, indicating that the selected bank-specific and macroeconomic variables are significant predictors of both retail and corporate credit risks. In other words, bank profitability is a key indicator for predicting CNPL and RNPL. This finding aligns with previous research by Gosh [42], Wood & Skinner [24], Kjosevski et al. [18], Gulati et al. [22], Karadima & Louri [23], Naili & Lahrichi [1], Waleed et al. [39], Permana & Rahyuda [41], and Christodoulou-Volos [28], who argue that banks with higher profit margins are less likely to engage in risky investments to boost returns, thus reducing their credit risk exposure. Additionally, the Capital Adequacy Ratio (CAR) was found to be significant, supporting findings by Wood & Skinner [24] and Farag et al. [8], who suggest that higher CAR levels lead banks to adopt more conservative lending practices to mitigate credit risk and avoid insolvency. Bank asset size also emerged as a relevant predictor for CNPL and RNPL, consistent with the studies of Al-Khazali & Mirzaei [43], Karadima & Louri [23], Naili & Lahrichi [1], and Farag et al. [8]. Larger banks are typically better able to diversify loan portfolios, invest in advanced technology, and attract qualified personnel, all of which contribute to lowering overall credit risk. Moreover, the Loan-to-Deposit (LTD) ratio was found to influence CNPL and RNPL, in line with Al-Khazali & Mirzaei [43], Mpofu & Nikolaidou [27], Ozili [25], and Farag et al. [8]. These studies highlight that a high LTD ratio can strain liquidity, thereby increasing credit risk. Income diversification also plays a critical role in predicting credit risk, as supported by Al-Khazali et al. [43], Ozili [25], Gulati et al. [22], and Mehmood & Luca [26]. Banks with diversified income sources from both fee-based and interest-based activities are generally less vulnerable to economic downturns, unlike banks that rely heavily on traditional lending. Finally, operating efficiency was shown to be a strong predictor of CNPL and RNPL. This finding is consistent with studies by Wang et al. [44], Nguyen et al. [45], Barra & Ruggiero [3], and Naili & Lahrichi [1], which indicate that banks with lower cost-to-income ratios tend to have more competent management, enabling better control over expenses and more effective credit risk management.

On the other hand, macroeconomic variables were also found to be significant, as evidenced by the high accuracy scores achieved through both regression and machine learning algorithms. This highlights the need for banks to closely monitor macroeconomic indicators to avoid unexpected surges in credit risk and to enhance the predictive accuracy of CNPL and RNPL. Specifically, GDP emerged as a crucial factor in forecasting credit risk at both corporate and retail levels. This finding aligns with the work of Mpofu & Nikolaidou [46] and Naili & Lahrichi [1], who suggest that GDP growth reflects increased national income, which strengthens borrowers' repayment capacity and thereby reduces credit risk exposure in banks. Additionally, interest rates were shown to be predictive of CNPL and RNPL, consistent with findings by Isaev & Masih [17], Wood & Skinner [24], Shehata [32], and ElGaliy [19]. These studies argue that rising market interest rates elevate financing costs, thereby weakening borrowers' repayment ability and increasing banks' credit risk. Foreign Direct Investment (FDI) was also found to impact credit risk, supporting the findings of Konstantakis et al. [47] and Giammanco et al. [48], who note that FDI brings additional capital, expertise, and technology into an economy. This, in turn, strengthens banks' loan portfolios through increased activity with foreign-qualified corporate borrowers, ultimately lowering credit risk.

Furthermore, the analysis reaffirmed that interest rates are reliable predictors of CNPL and RNPL trends, as also supported by Mahrous et al. [49]. Exchange rate fluctuations were found to influence credit risk as well, corroborating results from Haniifah [50], Beck et al. [51], Gulati et al. [22], Niepmann & Schmidt-Eisenlohr [30], and Ofria & Mucciardi [31]. These studies emphasize that in emerging markets, currency devaluation exerts inflationary pressures that undermine borrowers' repayment performance, thereby raising credit risk levels. The study also identified public debt as a relevant predictor of CNPL and RNPL. This finding aligns with the research of Ghosh [42] and Naili & Lahrichi [1], who argue that increased public debt, driven by government issuance of treasury securities to finance budget deficits, places added pressure on banks to absorb these securities. This, in turn, reduces banks' liquidity available for private lending, increases interest rates, and raises credit risk. Similarly, external debt was found to influence credit risk, consistent with the results of Kauko [52], Maltritz & Molchanov [53], and Nikolaidou & Vogiazas [54]. Finally, inflation was also found to be a significant factor in predicting credit risk levels. This is supported by Radivojević et al. [29], who argue that rising inflation in emerging economies erodes purchasing power, thereby impairing borrowers' ability to repay loans and increasing credit risk exposure. In summary, the study contributes to the existing literature by employing machine learning algorithms to train and analyze the dataset, thus enhancing credit risk prediction performance. Moreover, it offers tailored predictive models for both retail and corporate credit risk management departments, enabling more accurate forecasting and improved risk mitigation strategies.

5- Conclusion

The paper concludes that bank-specific variables—such as profitability, liquidity, income diversification, capital adequacy, asset size, and operating efficiency—and macroeconomic indicators—such as external debt, inflation, exchange rate, GDP, interest rate, and foreign direct investment (FDI)—can effectively enhance the prediction of corporate and retail credit risk when using models like Random Forest and Kernel SVM. Accordingly, it is recommended that credit risk analysts in both corporate and retail departments adopt these models to improve prediction accuracy, thereby strengthening credit screening processes. This, in turn, supports the maintenance of high-quality loan portfolios with reduced exposure to insolvency and credit risks, ultimately enabling banks to continue playing a vital role in fostering economic growth.

The findings emphasize the importance of prudent earnings management by banks to effectively control CNPL and RNPL levels, maintain sound portfolios, and prevent the risk of bankruptcy, which can jeopardize their market presence and reduce their contribution to economic development. Banks should also closely monitor their loan-to-deposit ratios to sustain adequate liquidity and mitigate potential asset value losses. Additionally, they are encouraged to enhance income diversification by generating fee-based revenue alongside interest income, thereby reducing the adverse effects of credit risk on overall bank performance and financial outcomes.

Moreover, banks must prioritize operational efficiency by controlling costs relative to income without compromising credit evaluation standards. This will help prevent unexpected increases in CNPL and RNPL. Leveraging their size, banks should diversify their loan and investment portfolios to lower overall risk and consistently assess CNPL and RNPL levels to ensure sufficient capital relative to assets. This strategy will support financial stability and enable banks to contribute to sustainable economic development.

Lastly, these findings provide valuable insights for policymakers, offering a clearer understanding of the key determinants influencing corporate and retail credit risks. This can guide the formulation or refinement of credit regulations aimed at enhancing risk management practices and securing long-term economic sustainability.

6- Declarations

6-1- Author Contributions

Conceptualization, M.H.; methodology, K.F. and A.E.; software, A.E.; validation, M.H. and A.E.; formal analysis, A.E. and K.F.; investigation, M.H. and K.F.; resources, K.F.; data curation, A.E.; writing—original draft preparation, M.H., K.F., and A.E.; writing—review and editing, M.H., K.F., and A.E.; visualization, A.E. and K.F; supervision, M.H.; project administration, M.H. and K.F; funding acquisition, M.H., K.F., and A.E. All authors have read and agreed to the published version of the manuscript.

6-2- Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6-3- Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6-4- Institutional Review Board Statement

Not applicable.

6-5- Informed Consent Statement

Not applicable.

6-6- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

[1] Naili, M., & Lahrichi, Y. (2022). Banks' credit risk, systematic determinants and specific factors: recent evidence from emerging markets. Heliyon, 8(2), 8960. doi:10.1016/j.heliyon.2022.e08960.

[2] Kryzanowski, L., Liu, J., & Zhang, J. (2023). Effect of COVID-19 on non-performing loans in China. Finance Research Letters, 52, 103372. doi:10.1016/j.frl.2022.103372.

[3] Barra, C., & Ruggiero, N. (2021). Do microeconomic and macroeconomic factors influence Italian bank credit risk in different local markets? Evidence from cooperative and non-cooperative banks. Journal of Economics and Business, 114, 105976. doi:10.1016/j.jeconbus.2020.105976.

[4] Koju, L., Abbas, G., & Wang, S. (2019). Do Macroeconomic Determinants of Non-Performing Loans Vary with the Income Levels of Countries? Journal of Systems Science and Information, 6(6), 512–531. doi:10.21078/jssi-2018-512-20.

[5] Kishimba, K. J., Akande, J. O., & Muzindutsi, P. F. (2024). Macro Credit Risk Stress Testing in Tanzanian Banking Sector: A Global Vector Autoregressive (GVAR) Approach. Journal of African Business, 25(3), 531–554. doi:10.1080/15228916.2023.2215579.

[6] Karaaslan, İ. (2023). Determinants of Non-Performing Loan Rate in the Turkish Banking Sector. Journal of Current Researches on Social Sciences, 13(2), 235-254. doi:10.26579/jocress.13.2.1.

[7] Farag, K. M. (2024). The impact of Macroeconomic and Microeconomic variables on Credit Risk: A Comparative Study between Retail and Corporate Credit Risks in the banks of Egypt. MSA-Management Sciences Journal, 3(3), 116–128. doi:10.21608/msamsj.2024.283469.1062.

[8] Farag, K., Kassem, T., & Ramzy, Y. (2023). The Crucial Macroeconomic and Microeconomic Determinants of Retail and Corporate Credit Risks. Journal of Accounting Finance and Auditing Studies, 9(3), 148-161. doi:10.32602/jafas.2023.028.

[9] Bao, W., Lianju, N., & Yue, K. (2019). Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. Expert Systems with Applications, 128, 301–315. doi:10.1016/j.eswa.2019.02.033.

[10] Li, Y. (2019). Credit Risk Prediction Based on Machine Learning Methods. 2019 14th International Conference on Computer Science & Education (ICCSE), 1011-1013. doi:10.1109/iccse.2019.8845444.

[11] Bhatore, S., Mohan, L., & Reddy, Y. R. (2020). Machine learning techniques for credit risk evaluation: a systematic literature review. Journal of Banking and Financial Technology, 4(1), 111–138. doi:10.1007/s42786-020-00020-3.

[12] Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable Machine Learning in Credit Risk Management. Computational Economics, 57(1), 203–216. doi:10.1007/s10614-020-10042-0.

[13] Shi, S., Tse, R., Luo, W., D'Addona, S., & Pau, G. (2022). Machine learning-driven credit risk: a systemic review. Neural Computing and Applications, 34(17), 14327–14339. doi:10.1007/s00521-022-07472-2.

[14] Montevechi, A. A., Miranda, R. de C., Medeiros, A. L., & Montevechi, J. A. B. (2024). Advancing credit risk modelling with Machine Learning: A comprehensive review of the state-of-the-art. Engineering Applications of Artificial Intelligence, 137, 109082. doi:10.1016/j.engappai.2024.109082.

[15] Sugozu, I. H., Verberi, C., & Yasar, S. (2025). Machine learning approaches to credit risk: Evaluating Turkish participation and conventional banks. Borsa Istanbul Review, 497–512. doi:10.1016/j.bir.2025.02.001.

[16] Gasmi, I., Neji, S., Mansouri, N., & Soui, M. (2025). Bank credit risk prediction using machine learning model. Neural Computing and Applications, 1–18. doi:10.1007/s00521-025-11044-5.

[17] Isaev, M., & Masih, M. (2017). Macroeconomic and bank-specific determinants of different categories of non-performing financing in Islamic banks: Evidence from Malaysia. MPRA Paper No. 79719, The International Center for Education in Islamic Finance University, Kuala Lumpur, Malaysia.

[18] Kjosevski, J., Petkovski, M., & Naumovska, E. (2019). Bank-specific and macroeconomic determinants of non-performing loans in the Republic of Macedonia: Comparative analysis of enterprise and household NPLs. Economic Research-Ekonomska Istrazivanja , 32(1), 1185–1203. doi:10.1080/1331677X.2019.1627894.

[19] ElGaliy, N. (2022). Macroeconomic Shocks and Credit Risk Stress Testing: Evidence from the Egyptian Banking Sector. Master Thesis, American University, Cairo, Egypt.

[20] Chaibi, H., & Ftiti, Z. (2015). Credit risk determinants: Evidence from a cross-country study. Research in International Business and Finance, 33, 1–16. doi:10.1016/j.ribaf.2014.06.001.

[21] Farooq, M. O., Elseoud, M. S. A., Turen, S., & Abdulla, M. (2019). Causes of non-performing loans: The experience of gulf cooperation council countries. Entrepreneurship and Sustainability Issues, 6(4), 1955–1974. doi:10.9770/jesi.2019.6.4(29).

[22] Gulati, R., Goswami, A., & Kumar, S. (2019). What drives credit risk in the Indian banking industry? An empirical investigation. Economic Systems, 43(1), 42–62. doi:10.1016/j.ecosys.2018.08.004.

[23] Karadima, M., & Louri, H. (2020). Non-performing loans in the euro area: Does bank market power matter? International Review of Financial Analysis, 72, 101593. doi:10.1016/j.irfa.2020.101593.

[24] Wood, A., & Skinner, N. (2018). Determinants of non-performing loans: evidence from commercial banks in Barbados. The Business & Management Review, 9(3), 44-64.

[25] Ozili, P. K. (2018). Bank loan loss provisions, investor protection and the macroeconomy. International Journal of Emerging Markets, 13(1), 45–65. doi:10.1108/IJoEM-12-2016-0327.

[26] Mehmood, A., & De Luca, F. (2023). How does non-interest income affect bank credit risk? Evidence before and during the COVID-19 pandemic. Finance Research Letters, 53, 103657. doi:10.1016/j.frl.2023.103657.

[27] Mpofu, T. R., & Nikolaidou, E. (2018). Determinants of credit risk in the banking system in Sub-Saharan Africa. Review of Development Finance, 8(2), 141–153. doi:10.1016/j.rdf.2018.08.001.

[28] Christodoulou-Volos, C. (2025). Determinants of Non-Performing Loans in Cyprus: An Empirical Analysis of Macroeconomic and Borrower-Specific Factors. International Journal of Economics and Financial Issues, 15(1), 190–201. doi:10.32479/ijefi.17398.

[29] Radivojević, N., Cvijanović, D., Sekulic, D., Pavlovic, D., Jovic, S., & Maksimović, G. (2019). Econometric model of non-performing loans determinants. Physica A: Statistical Mechanics and Its Applications, 520, 481–488. doi:10.1016/j.physa.2019.01.015.

[30] Niepmann, F., & Schmidt-Eisenlohr, T. (2022). Foreign currency loans and credit risk: Evidence from U.S. banks. Journal of International Economics, 135, 103558. doi:10.1016/j.jinteco.2021.103558.

[31] Ofria, F., & Mucciardi, M. (2022). Government failures and non-performing loans in European countries: a spatial approach. Journal of Economic Studies, 49(5), 876–887. doi:10.1108/JES-01-2021-0010.

[32] Shehata, K. M. F. (2019). The macroeconomic determinants of the non-performing loans: comparative research between non-listed & listed commercial banks of the stock exchange of Egypt. Fifth International Conference of Business Innovation and Economic Sustainability Challenges, 1-49.

[33] Tajik, M., Aliakbari, S., Ghalia, T., & Kaffash, S. (2015). House prices and credit risk: Evidence from the United States. Economic Modelling, 51, 123–135. doi:10.1016/j.econmod.2015.08.006.

[34] Serengil, S. I., Imece, S., Tosun, U. G., Buyukbas, E. B., & Koroglu, B. (2021). A Comparative Study of Machine Learning Approaches for Non-Performing Loan Prediction. 6th International Conference on Computer Science and Engineering (UBMK), 326–331. doi:10.1109/ubmk52708.2021.9558894.

[35] Alonso, A., & Carbo, J. M. (2021). Understanding the Performance of Machine Learning Models to Predict Credit Default: A Novel Approach for Supervisory Evaluation. SSRN Electronic Journal. doi:10.2139/ssrn.3774075.

[36] Abdullah, M., Chowdhury, M. A. F., Uddin, A., & Moudud-Ul-Huq, S. (2023). Forecasting nonperforming loans using machine learning. Journal of Forecasting, 42(7), 1664–1689. doi:10.1002/for.2977.

[37] Beltman, J., Osterrieder, J., & Machado, M. (2024). Predicting Retail Customers' Distress: Early Warning Systems and Machine Learning Applications. SSRN Electronic Journal, 1-26. doi:10.2139/ssrn.4730470.

[38] Farag, K., Mamdouh, B., Gubbi, N., Luqman, R., Ray, S. (2024). Measuring the Impact of the Macroeconomic and Bank-specific Variables on Capital Risk: A Comparative Study between the Listed and Non-listed Banks. Computer Fraud and Security, (12), 237-248.

[39] Waleed, S., Soliman, M. M., & Elfaham, K. (2025). The Impact of Internal and External Factors on Credit Risk during Periods of Economic Instability: A Study of the Egyptian Banking Sector. Alexandria Journal of Accounting Research, 9(1), 1-46. doi:10.21608/aljalexu.2025.397680.

[40] Getinet, B., Dereja, B., Akinida, T., Yohannes, D., & Mengistu, S. (2024). Income Diversification, Credit Risk and Profitability of Ethiopian Commercial Banks. Jefore Ethiopian Journal of Applied Sciences, 1(1), 23–37. doi:10.70984/8yrkh266.

[41] Permana, I. K. G., & Rahyuda, H. (2025). The Effect of Liquidity, Bank Capital, Profitability, and Operational Efficiency On Credit Risk (A Case Study on Banking Sub-Sector Companies Listed on the Indonesia Stock Exchange for the Period 2019-2023). International Journal of Management Research and Economics, 3(1), 323-347. doi:10.54066/ijmre-itb.v3i1.2817.

[42] Ghosh, A. (2015). Banking-industry specific and regional economic determinants of non-performing loans: Evidence from US states. Journal of Financial Stability, 20, 93–104. doi:10.1016/j.jfs.2015.08.004.

[43] Al-Khazali, O. M., & Mirzaei, A. (2017). The impact of oil price movements on bank non-performing loans: Global evidence from oil-exporting countries. Emerging Markets Review, 31, 193–208. doi:10.1016/j.ememar.2017.05.006.

[44] Wang, R., Liu, J., & Luo, H. (2021). Fintech development and bank risk taking in China. European Journal of Finance, 27(4–5), 397–418. doi:10.1080/1351847X.2020.1805782.

[45] Nguyen, T. B. N., Lin, L. F., Su, X. Q., & Yu, J. H. (2021). Does Managerial Education Matter for Credit Risk? Evidence from Taiwan. Finance Research Letters, 41, 101812. doi:10.1016/j.frl.2020.101812.

[46] Mpofu, T. R., & Nikolaidou, E. (2019). Macroeconomic and bank-specific determinants of non-performing loans in sub-Saharan Africa (No. 2019-02). School of Economics, University of Cape Town, Cape Town, South Africa.

[47] Konstantakis, K. N., Michaelides, P. G., & Vouldis, A. T. (2016). Non-performing loans (NPLs) in a crisis economy: Long-run equilibrium analysis with a real time VEC model for Greece (2001-2015). Physica A: Statistical Mechanics and Its Applications, 451, 149–161. doi:10.1016/j.physa.2015.12.163.

[48] Giammanco, M. D., Gitto, L., & Ofria, F. (2023). Government failures and non-performing loans in Asian countries. Journal of Economic Studies, 50(6), 1158–1170. doi:10.1108/JES-06-2022-0348.

[49] Mahrous, S. N., Samak, N., & Abdelsalam, M. A. M. (2020). The effect of monetary policy on credit risk: evidence from the MENA region countries. Review of Economics and Political Science, 5(4), 289–304. doi:10.1108/REPS-07-2019-0099.

[50] Haniifah, N. (2015). Economic Determinants of Nonperforming Loans in Ugandan Commercial Banks. Ph.D. Thesis, Asia Pacific University, Forest Grove, United States.

[51] Beck, R., Jakubik, P., & Piloiu, A. (2015). Key Determinants of Non-performing Loans: New Evidence from a Global Sample. Open Economies Review, 26(3), 525–550. doi:10.1007/s11079-015-9358-8.

[52] Kauko, K. (2012). External deficits and non-performing loans in the recent financial crisis. Economics Letters, 115(2), 196–199. doi:10.1016/j.econlet.2011.12.018.

[53] Maltritz, D., & Molchanov, A. (2014). Country credit risk determinants with model uncertainty. International Review of Economics & Finance, 29, 224–234. doi:10.1016/j.iref.2013.05.018.

[54] Nikolaidou, E., & Vogiazas, S. (2017). Credit risk determinants in Sub-Saharan banking systems: Evidence from five countries and lessons learnt from Central East and South East European countries. Review of Development Finance, 7(1), 52–63. doi:10.1016/j.rdf.2017.01.003.