Journal of Statistical Research *https://doi.org/10.3329/jsr.v58i2.80616*
2024, Vol. 58, No. 2, pp. 335-351 *ISSN 0256 - 422 X*

PREDICTIVE MODELING FOR NON-PERFORMING ASSETS IN THE INDIAN BANKING SECTOR

SURESH B. PATHARE*
School of Mathematics, Applied Statistics & Analytics, School of Commerce
SVKM's NMIMS (Deemed to be University), Navi Mumbai, Maharashtra, INDIA
Email: suresh.pathare@nmims.edu

MOHNEET SANDHU
School of Mathematics, Applied Statistics & Analytics
SVKM's NMIMS (Deemed to be University), Navi Mumbai, Maharashtra, INDIA
Email: mohneetkaursandhu@gmail.com

ABSTRACT

This paper explores the financial and operational factors that contribute to India's Nonperforming Assets (NPA) problem and discusses practical solutions for mitigating the risk of future NPAs. Descriptive statistics, regression analysis, and time series analysis are used to identify the main drivers of NPAs of Indian banks, revealing that high levels of NPAs have resulted in lower profitability, increased provisioning requirements, and higher borrowing costs. The findings and recommendations of this study provide valuable insights for policymakers, regulators, and banking practitioners seeking to reduce the risk of NPAs in India.

Keywords and phrases: Nonperforming assets, Indian banking sector, Credit underwriting, Monitoring systems, Risk management, financial impact, Profitability, Provisioning requirements.

AMS Classification: 62G10, 62M10,

1 Introduction

In the last decade, Indian banks have often made headlines with mounting Nonperforming Assets (NPAs) and subsequent mergers. "An asset, including a leased asset, becomes non-performing when it ceases to generate income for the bank. According to the Reserve Bank of India (2007), a 'non-performing asset' (NPA) was defined as a credit facility in respect of which the interest and/or instalment of principal has remained 'past due' for a specified period of time." The lender bears a financial burden as a result of NPAs, and a significant number of NPAs over time may indicate to regulators that the bank's financial health is jeopardized. The banking industry is regarded as the backbone of any economy. With the growing need for banks to participate in the economic growth

* Corresponding author

process, the problem of non-performing assets has reached an epic proportion. As emerging-market governments push for greater financial inclusion, the risk associated with bank assets increases.

India's NPAs have increased due to multiple reasons, including excessive lending during economic booms, inadequate credit appraisal mechanisms, and sector-specific downturns. Public sector banks have reported an increase in NPAs linked to large corporate defaults and high exposure to priority sector lending (Kumar et al., 2018). Other contributing factors include weak corporate governance, policy changes, and economic slowdowns, which have compounded the issue (Bhaskaran et al., 2016; Mishra & Pawaskar, 2017). These factors, when not managed effectively, lead to decreased trust in financial institutions and broader economic repercussions (Singh, 2016). Das (2023) conducted a comprehensive analysis of the trends and determinants of the Non-Performing Assets (NPA) crisis in India's banking sector.

The adverse effects of NPAs on banks' profitability and stability have been well-documented globally and within India. For instance, Masood and Ashraf (2012) studied Islamic banks across 12 countries and found that nonperforming loans adversely affect bank performance and profitability. Ongore and Kusa (2013) analyzed commercial banks in Kenya, identifying a negative relationship between NPAs and profitability. Similar findings have been reported globally, including studies by Bace (2016) and Etale et al. (2016), which highlight the universal challenges posed by NPAs. Indian researchers have also emphasized these challenges, with Singh (2016) noting the severe NPA crisis in public sector banks compared to private sector banks, and Kadanda and Raj (2018) identifying major determinants of NPA growth in Indian banks.

The problem of NPAs is not only a concern for the banking sector but also has broader economic implications. Rising NPAs undermine depositor confidence, which may result in potential withdrawals, causing liquidity crunches. This hampers the banks' ability to lend and affects other economically productive activities. Consequently, the economy may experience a slowdown, marked by reduced investments, increased unemployment, inflation, and a bear market. Serrano (2021) and Park and Shin (2021) have emphasized the destabilizing effect of NPAs on banking systems, impairing financial health and lending activities.

Despite the extensive research on NPAs, certain gaps remain. While global studies provide a broad understanding of the challenges posed by NPAs, the Indian context presents unique factors, including policy changes, industrial sickness, and procedural flaws in lending (Kumar et al., 2018). Bhaskaran et al. (2016) compared public and private sector banks and highlighted the need for targeted solutions. Mishra and Pawaskar (2017) suggested the importance of a robust legal framework and effective credit appraisal systems, while Mukhopadhyay (2018) called for innovative approaches from the Reserve Bank of India to address specific cases. However, there remains a lack of predictive models tailored to the Indian banking sector, particularly in assessing potential borrowers' creditworthiness and estimating loan default probabilities.

To address these gaps, the objectives of this study are threefold. Firstly, it aims to develop a comprehensive financial projection of NPAs across public, private, state, and foreign banks operating in India, providing an overarching view of their financial health. Secondly, the study seeks to specifically assess the gross NPA ratio for IDBI Bank, which recorded the highest NPA ratio among public sector banks in 2021, and forecast NPAs for the bank by considering its loan portfolio and asset quality. Lastly, the research aims to create a predictive financial model that can evaluate potential borrowers' creditworthiness and estimate the likelihood of loan defaults, contributing to more informed credit risk management practices.

2 Research Design and Methodology

This study utilizes a dataset spanning from 2004 to 2021, sourced from the Reserve Bank of India (RBI): Economic Data Warehouse for banking performance metrics and Statista for macroeconomic indicators such as GDP, inflation, and interest rates. The study considers three categories of banks: Public Sector Banks (PSBs), Private Sector Banks (PvSBs), and Foreign Banks (FBs).

The dataset includes both bank-level and macroeconomic variables relevant to Non-Performing Asset (NPA) analysis, such as: Gross NPA Ratios (by sector and bank),Financial indicators (loan advances, profits, capital adequacy ratios),Economic variables (GDP growth, inflation, interest rates).

To achieve the study's objectives, predictive analysis was conducted. Historical data from 2004 to 2018 was used to train statistical models and machine learning algorithms, identifying patterns and trends. These models were validated using recent data (2019–2021) to assess prediction accuracy and robustness, followed by forecasts for NPA ratios and sector-specific trends for 2022–2023.

Statistical and machine learning methods

A combination of statistical and machine learning techniques was used. Stationarity test was conducted using the Augmented Dickey-Fuller test to ensure data suitability for time series modeling. The best-fit Autoregressive Integrated Moving Average (ARIMA) model was chosen based on Akaike Information Criterion (AIC) while Exponential Smoothing was applied to assign higher weight to recent data for trend forecasting.

Machine Learning models such as Logistic Regression was used as a baseline model for classifying loan default likelihood. Random Forest, Gradient Boosting Trees, and XGBoost were used to enhance prediction accuracy. These models were evaluated using metrics like Root Mean Square Error (RMSE), accuracy, Receiver operating characteristic (ROC), and Area Under the Curve (AUC). Here Root Mean Square Error (RMSE) is calculated as

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2},$$

where y_i: Actual observed values, $\hat{y}_i$: Predicted values from the model, n: Number of data points.

The NPA ratio represents the percentage of a bank's loans that are classified as non-performing (i.e., loans for which the borrower is not making interest or principal payments on time) and it is calculated as

$$\text{NPA Ratio} = \frac{\text{Gross NPAs}}{\text{Total Advances}} \times 100,$$

where Gross NPAs refer to the total value of loans classified as non-performing, and Total Advances represent the total loans disbursed by the bank. Multiple software tools (R, Python, Tableau) were used for different parts of the analysis. However, future iterations may consolidate the workflow into R for streamlined execution.

3 Trend and Sector-wise Comparison of NPA in India

3.1 The NPA's Timeline in India

The trajectory of NPAs in the Indian banking sector from 1992 to 2021 exhibited a three-phase pattern, as shown in Figure-1, with each phase marked by specific economic and regulatory factors.

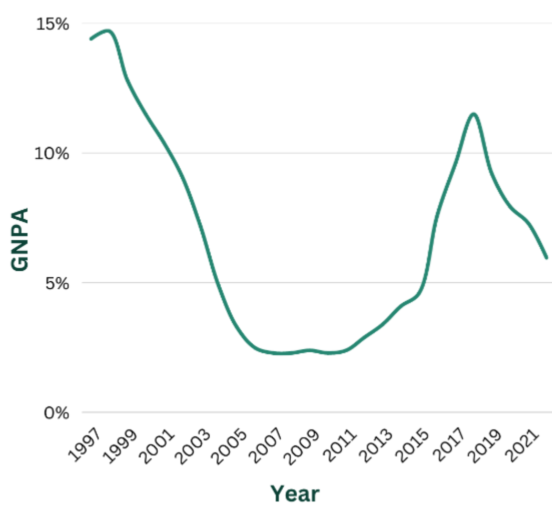

Figure 1: Gross NPA in India from 1997-2021.

In the first phase (1992-2008), NPAs showed a consistent downwards trend until the North Atlantic Financial Crisis (NAFC). Expansionary economic policies and aggressive lending by banks during the period of rapid economic growth from 2004 to 2008 resulted in a surge of credit growth, which set the stage for the subsequent phase.

In the second phase (2008-2018), NPAs began to rise slowly at first, and then rapidly from 2014 to 2018. The Reserve Bank of India (RBI) issued a circular in 2018 stating that any loan that is 90 days past due will be classified as a default, which resulted in the highest NPA levels ever recorded.

In the third phase (2018-2021), NPAs exhibited a downwards trend due to the improved recovery of stressed assets, primarily through the Insolvency and Bankruptcy Code (IBC). This code has enabled faster resolution of stressed assets, resulting in improved asset quality for banks and a reduction in NPAs.

Overall, the trajectory of NPAs in the Indian banking sector can be attributed to a complex interplay of macroeconomic factors, regulatory policies, and institutional reforms. The persistent issue of NPAs highlights the need for continued regulatory scrutiny and prudent lending practices by banks to maintain financial stability in the banking system.

3.2 A Sectorwise Comparison of NPA in the Indian Banking Sector

The number of NPAs in India's banking sector has fluctuated significantly across public sector banks (PSBs), private sector banks (PVSBs), and foreign banks (FBs).

Public sector banks

In 2015, an asset quality review (AQR) of banks was conducted by the RBI, and it was observed that gross NPAs of PSBs had increased from 5.43% in March 2015 to 9.32% in December 2015. This increase in NPAs was attributed to loan mismanagement, asset-liability mismatches, and ineffective credit appraisal processes. The five public sector banks with the highest NPAs were State Bank of India(SBI), Panjab National Bank (PNB), Bank of Baroda (BOB), Canara Bank (CB), and Union Bank of India (UBI) with the infrastructure and metals sectors contributing the most. Some real-life incidents that contributed to the NPA problem in public sector banks in India include the Vijay Mallya case, the IL & FS crisis, the Bhushan Steel case, and the telecom sector crisis. In recent years, the RBI has implemented stricter rules for loan classification, introduced the PCA framework, and set up a stressed asset fund to address NPAs in public sector banks, improving their asset quality and reducing exposure to stressed assets.

Private sector banks

Since 2011, there has been an increase in NPAs in private sector banks, which has coincided with a decline in global commodity prices. The financial sheets of Indian banks were also stretched by past price decrease episodes that occurred in the late 1990s and 2009, albeit these two episodes were not as severe and prolonged as the ones that lasted from 2011 to 2016. Banks' profitability and solvency have decreased due to the increase in the number of institutions dealing with risk or NPAs. Even the 2012 restructuring of corporate debt, which allowed borrowers to restructure their loans and defer interest payments, contributed to a significant increase in the number of NPAs in private sector banks. Although the number of NPAs in private sector banks slightly decreased after the COVID-19 pandemic, this modest decrease can be attributed to the availability of more efficient borrowing alternatives, such as the corporate bond market, as well as the industry's turn to the market to generate funds. The Yes Bank crisis is one of several incidents that have affected the NPAs of private sector banks in India. In March 2020, Yes Bank, a prominent private sector bank in India, faced a severe liquidity crisis due to a significant amount of bad loans on its books. In addition to public sector banks, private sector banks such as Axis Bank and ICICI Bank also extended loans to Vijay Mallya's Kingfisher Airlines. In the Videocon case several private sector banks, including ICICI Bank and Axis Bank, extended loans to the Videocon Group, and when the company defaulted on its debt payments, the banks left a significant number of NPAs.

Foreign banks

Due to their stronger credit management strategies and use of better technology, the number of foreign banks operating in our nation with NPAs has never been particularly high. Even if foreign banks report high profits, which may be the result of manipulating balance sheets, it has also been claimed that they are under pressure from global issues in their own markets, particularly in Europe.

4 Data Analysis and Interpretations

4.1 Seasonality test

Seasonality refers to periodic patterns or fluctuations that occur at regular intervals within a dataset. Identifying seasonality is critical for improving the accuracy of forecasting models and understanding underlying trends. The seasonality test in this study checks whether the dataset exhibits such periodic patterns. To identify whether there is a pattern of seasonal variation in a time series dataset, we use the seasonality test using R- software, and it is found that there is no seasonality in the dataset.

4.2 Exponential Smoothing Method

Based on the analysis of the time series data, it was observed that there was a persistent trend but no noticeable seasonality. Therefore, double exponential smoothing is deemed to be the most suitable method for forecasting future values of a series. Upon applying double exponential smoothing, as shown in Figure-2, to the time series data, the model produced favourable outcomes with a root mean squared error (RMSE) of 1.8041. This indicates that the model's predictions are, on average, 1.8041 units away from the actual values of the series.

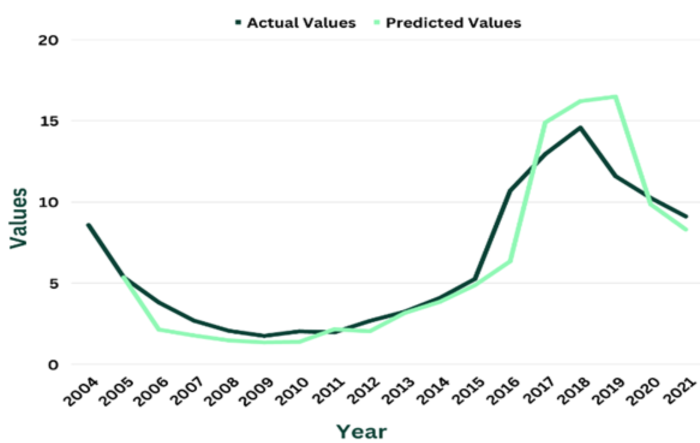

Figure 2: Fitting of the double exponential smoothing curve

4.3 The Durbin-Watson(DW) test

This test is used to determine the presence of autocorrelation in the regression residuals. The DW test statistic D is

$$D = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2},$$

where e_t: Residual (error) at time t, e_{t-1}: Residual at time $t-1$, n: Number of observations.

Here we obtained D = 0.38227 and p-value is 3.595e-07. P-value less than 5% indicates that there is no autocorrelation in the data.

4.4 Augmented Dickey-Fuller (ADF) test

This test is used to determine whether a time series dataset is stationary or not. By using ADF test, we obtained Dickey-Fuller test statistic value -2.3842 with Lag order 2 and p-value is 0.426. P-value is greater than 5% indicates that the data are not stationary. As the data are not stationary, we use differencing to make ensure that the data are stationary. After differencing again, we apply augmented Dickey-Fullet test to check the stationarity of the data, and we obtained Dickey-Fuller test statistic value -3.7157 with Lag order 2 and p-value 0.04174. P- value is smaller than 5%, indicates that the data are stationary.

4.5 Fitting of ARIMA models

ARIMA is a time series forecasting technique used to model and forecast data that has a trend and/or seasonality.

ARIMA(p, d, q):

p: Order of the autoregressive (AR) term.

d: Degree of differencing (to make the series stationary).

q: Order of the moving average (MA) term.

The ARIMA model with drift and when $d > 0$ is:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots$$

where:

Y_t: The value of the time series at time t.

c: The drift term, representing a constant component or trend in the model.

$\phi_1, \phi_2, \ldots$: Autoregressive (AR) coefficients, capturing the influence of previous values of the time series.

ϵ_t : White noise (random error term) at time t, assumed to have zero mean and constant variance.

$\theta_1, \theta_2, \ldots$: Moving Average (MA) coefficients, capturing the influence of past errors on the current value.

The drift term allows for a linear trend in the differenced series (before integration). Without drift, the series is assumed to revert to a mean of zero after differencing.

Since the data are stationary, we fit several ARIMA models (using R commond auto.arima) to the data and choose the best fit ARIMA models among them based on Akaike's information criterion (AIC) values. The results obtained are as follows:

Table 1: Various ARIMA models fitted and their AIC score

ARIMA model fitted	**AIC value**
ARIMA(2,1,2) with drift	73.92266
ARIMA(0,1,0) with drift	75.50182
ARIMA(1,1,0) with drift	72.13002
ARIMA(0,1,1) with drift	74.24539
ARIMA(0,1,0)	73.50602
ARIMA(2,1,0) with drift	74.09773
ARIMA(1,1,1) with drift	74.11644
ARIMA(2,1,1) with drift	Inf
ARIMA(1,1,0)	**70.21555**
ARIMA(2,1,0)	72.19856
ARIMA(1,1,1)	72.20863
ARIMA(0,1,1)	72.25196
ARIMA(2,1,1)	Inf

The value listed next to each ARIMA model is the AIC score. A lower AIC indicates a better balance between goodness of fit and model complexity. Models with Inf AIC were likely discarded due to non-convergence or other issues. Models "with drift" indicate that a constant term is included, which helps capture trends in the differenced data. For instance, ARIMA(2,1,2) with drift, Includes both AR(2) and MA(2) terms with a drift term. Models "without drift" assume the differenced series has no linear trend and reverts to zero. Also drift is useful for non-stationary time series with a consistent trend after differencing. For example, a stock price series might show a slight upward trend even after differencing, which the drift term can capture.

Here ARIMA(1,1,0) is chosen because it has the lowest AIC (70.21555). This means it provides the best fit with the fewest parameters. $ARIMA(1, 1, 0)$ with AIC value 70.21555 implies that the original series is non-stationary (since $d = 1$). A first difference was sufficient to remove the non-stationarity.The series relies on its lagged value (AR(1)) for predictions, but no trend or MA terms are significant. The lack of "with drift" in the best model suggests that after differencing, the series does not exhibit a significant linear trend.

Our research findings suggest that the ARIMA method produced the most favourable results for all four bank sectors analysed, as it demonstrated the lowest RMSE. Using this prediction model, we have forecasted NPA ratios for various banks for the year 2023, as shown in Table-2.

Table 2: Forecasted NPA ratios for the banks for the year 2023.

Sr. No	Bank Sector	Exponential Smoothing Method RSME	ARIMA Model RSME	NPA Ratio
1	Public	1.81	1.63	8.48
2	Private	0.52	0.43	4.07
3	Foreign	1.10	0.79	2.96
4	SBI & associates	1.14	1.09	6.34

5 IDBI Bank Case Study

The public sector banks exhibited the highest NPA ratio, and within this group, the IDBI Bank recorded the highest NPA ratio (see Figure-3). Subsequently, following its inclusion in the Prompt Corrective Action (PCA) framework by the RBI in 2017, IDBI Bank's shares have experienced a 50% decline (see Figure-4). The negative impact of the rising NPA ratio on the profitability of IDBI Bank was apparent in the same fiscal years in which the NPAs exhibited a significant increase.

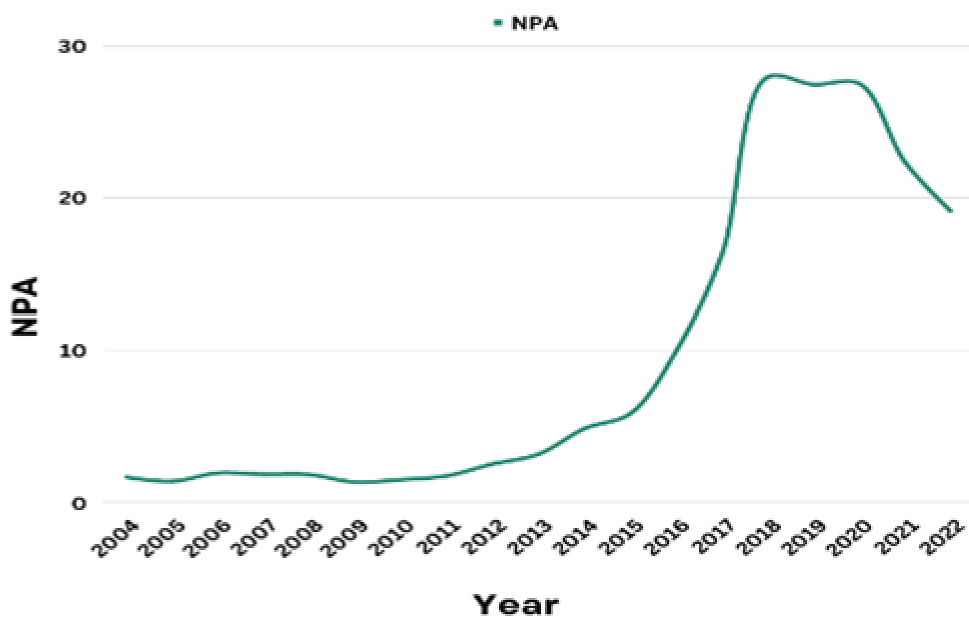

Figure 3: The IDBI Bank's yearwise NPA ratio

In 2021, a significant proportion of NPAs in the IDBI Bank were derived from the micro, small, and medium-sized enterprises (MSMEs) sector, accounting for 79.52% of the total NPAs (Figure-5).

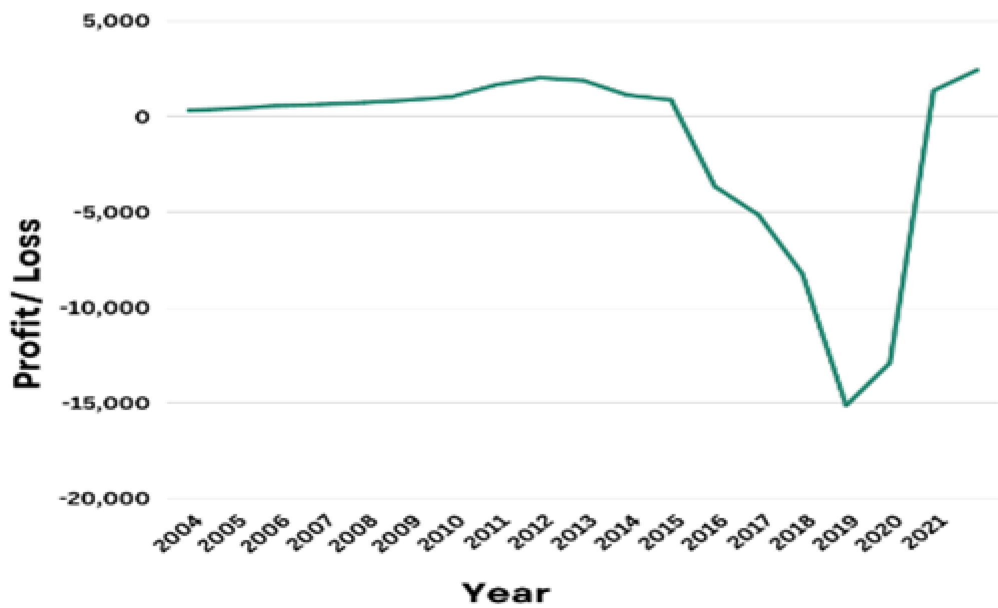

Figure 4: IDBI Bank's yearwise profit/loss

The service industry contributed 12.8% of the NPA, while agricultural allied activities contributed 3.78% of the total NPA.

6 Fitting of the Multiple Linear Regression Model

To predict NPA(Y), multiple linear regression is used. The Response(Indepedent) variables used for multiple linear regression are profit and loss, loan advances, GDP and inflation.

The fitted MLR equation is as follows with an R-square of 0.7784:

$$Y = 0.0695\text{–}0.0012 \times \text{(Profit and Loss)} + 0.5321 \times \text{(Loan Advances)} \\ - 0.1137 \times \text{(GDP)–}0.8045 \times \text{(Inflation)}$$

This model provides a quantitative relationship between the dependent variable (Y, the NPA) and the independent variables. Each coefficient represents the expected change in Y (NPA) for a unit increase in the corresponding predictor variable, holding all other variables constant.

Coefficients and Their Interpretations:

(i) Intercept (0.0695): When all independent variables (Profit and Loss, Loan Advances, GDP, and Inflation) are zero, the predicted NPA value is 0.0695.

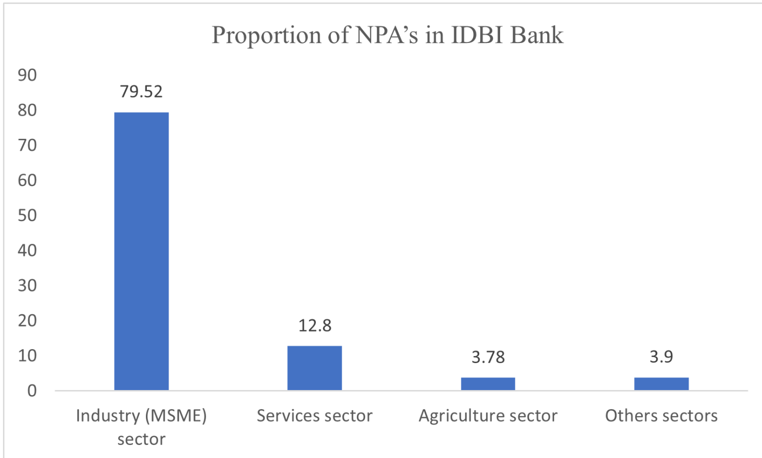

Figure 5: Proportion of NPA's in IDBI Bank

(ii) Profit and Loss (- 0.0012): For each additional unit of profit and loss (increase in profit or reduction in loss), the NPA decreases by 0.0012 units, assuming other factors remain constant. This negative relationship suggests that higher profitability reduces NPAs, which aligns with financial expectations.

(iii) Loan Advances (0.5321): For every unit increase in loan advances, the NPA increases by 0.5321 units, assuming other factors remain constant.This positive coefficient indicates that higher loan disbursements are associated with higher NPAs, likely due to increased risk exposure.

(iv) GDP (- 0.1137): For every unit increase in GDP, the NPA decreases by 0.1137 units, assuming other factors remain constant. This negative relationship implies that economic growth reduces NPAs, as borrowers are better positioned to repay loans during periods of economic expansion.

(v) Inflation (- 0.8045): For every unit increase in inflation, the NPA decreases by 0.8045 units, assuming other factors remain constant. This finding is counterintuitive, as higher inflation often increases financial stress on borrowers. The negative coefficient could indicate peculiarities in the dataset or context that merit further investigation.

The R^2 value of 0.7784 indicates that the model explains 77.84% of the variance in the NPA. This suggests the model fits the data well, but some variation in NPA remains unexplained by these predictors.

6.1 Loan Default Prediction

A loan default prediction predicts the likelihood of borrowers defaulting on their loans based on various factors. Machine learning models are trained on historical loan default data to identify patterns and make predictions. Factors influencing the prediction include, Borrowers' income and financial ratios (e.g. Annual Income and Debt-to-Income Ratio), Loan-specific characteristics (e.g., Loan Term, Purpose), Borrowers' credit behavior (e.g., Revolving Utilization Rate, Last Delinquency). The bar chart (Figure 6) visualizes the relative importance of each variable in the loan default prediction model. Importance is a measure of how much each variable contributes to the model's predictive accuracy.

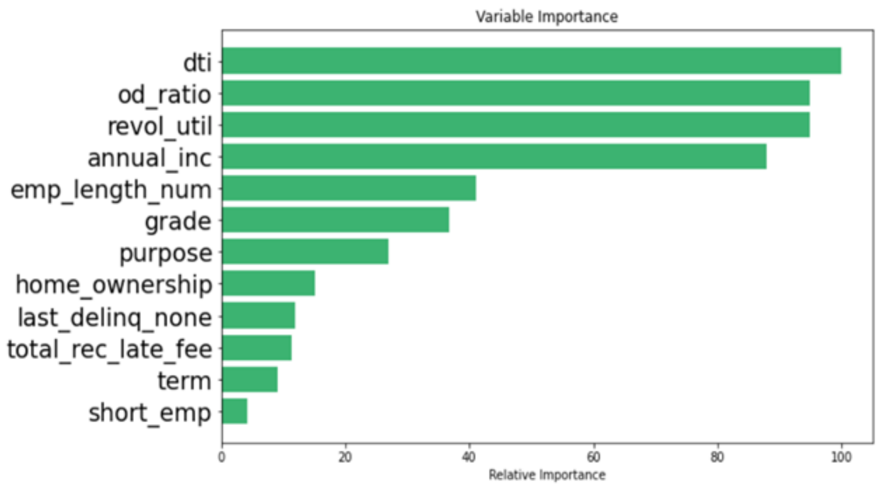

Figure 6: The graph shows the importance of the variables.

Most Important Variables:

(i) Debt-to-Income Ratio (DTI): The most significant variable. It reflects the borrower's debt obligations relative to their income and directly impacts their ability to repay loans.

(ii) Overdraft Ratio (Od_ratio): Indicates the borrower's financial behavior in terms of overdrafts, which can signal financial instability.

(iii) Revolving Utilization Rate (Revol_util): Measures the credit usage compared to the available limit. High utilization often correlates with financial stress.

(iv) Annual Income (Annual_inc) of the borrower : Highlights the borrower's earning capacity, a key factor in assessing repayment ability.

Moderately Important Variables:

(i) Employment Length (emp_length_num): Employment stability impacts repayment ability.

(ii) Grade: Indicates the lender-assigned credit grade.

Less Important Variables:

(i) Purpose: A category provided by the borrower for the loan request. Credit card, house, education, etc.

(ii) Home Ownership(Home_ownership): Denotes the borrower's housing status (e.g., rent, own), which provides some insight into financial stability.

(iii) Last Delinquency (Last_delinq_none) : Indicates whether the borrower had prior delinquency, which is important but not as influential as other variables.

(iv) Variables like term, Short employement (short_emp) and Late fees received to date (total_rec_late_fee) contribute less to prediction accuracy but still add value to the model.

6.2 Explanation of the ROC Curve and Findings

After evaluating various classification algorithms, including support vector machines (SVM), k-nearest neighbours (KNN), random forests, logistic regression, gradient boosted trees (GBTree) and extreme gradient boosting (XGBoost), it was observed that logistic regression yielded the highest performance. Specifically, after training and testing the models on a given dataset, logistic regression yielded the lowest error rate and the highest accuracy among all the models. This finding suggested that logistic regression is the most suitable algorithm for the given dataset and classification task. This finding was also supported by the Receiver Operating Characteristic (ROC) curve shown in Figure- 7.

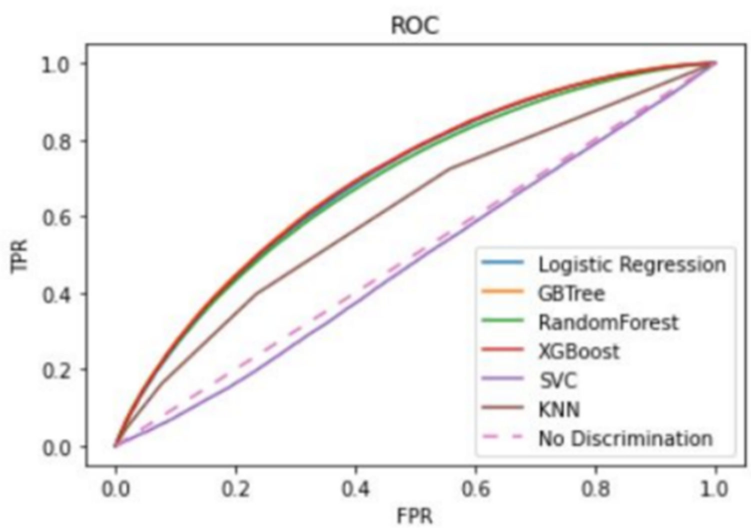

Figure 7: ROC curve

The ROC curve is a graphical representation of the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR) for different classification thresholds across multiple machine learning models. Logistic regression yields the highest True Positive Rate while maintaining a low False Positive Rate, resulting in the best performance. This is evident from the ROC curve where the logistic regression line is closer to the top-left corner, indicating better classification performance. Models like GB Tree, Random Forest, and XGBoost also perform well, as their curves are close to the logistic regression curve. However, Support Vector Classifier, k-Nearest Neighbors, and others are farther from the top-left corner, indicating comparatively lower performance. The dashed purple line represents the baseline model, which makes random predictions.

All the models outperform this baseline, but logistic regression demonstrates the largest improvement. The ROC curve confirms that logistic regression has the highest area under the curve (AUC), further validating its reliability for the given dataset and task.

7 Discussion and Conclusion

The findings of this study align with and diverge from existing literature on the determinants and prediction of Non-Performing Assets (NPAs) in the Indian banking sector. The study identifies Debt-to-Income Ratio (DTI), Overdraft Ratio (OD_Ratio), and Revolving Utilization Rate (Revol_Util) as the most influential predictors of loan defaults. These findings are consistent with prior studies, such as Bhaskaran et al. (2016), which highlighted borrower-specific financial ratios as significant determinants of NPAs. The observed importance of Annual Income (Annual_Inc) aligns with Kumar et al. (2018), who emphasized income levels as critical for assessing repayment capacity. The strong predictive performance of Logistic Regression is comparable to findings in Das and Uppal (2021), who demonstrated that simpler models often outperform complex algorithms in datasets with linear relationships.

Unlike Serrano (2021) and Park and Shin (2021), who highlighted the destabilizing effects of macroeconomic variables like GDP growth and inflation on NPAs, this study found a more limited role for these variables in the predictive model. This divergence may stem from dataset-specific factors or differences in the granularity of borrower-level data. While Mukhopadhyay (2018) suggested institutional reforms as a primary driver for reducing NPAs, this study focuses more on borrower-centric financial and credit behavior metrics.

The findings underscore the necessity of enhancing borrower-level credit evaluations by prioritizing highly significant predictors. The study's emphasis on financial ratios and income highlights a shift towards borrower accountability, diverging from broader economic factors. Non-Performing Assets (NPAs) remain one of the most critical challenges affecting credit growth and the overall health of the Indian banking sector. They are a significant concern for both the banking system and the Indian economy. The rising volume of NPAs can be attributed to various factors, including poor credit appraisal processes, inadequate vigilance, and cyclical business downturns. This persistent issue has prompted the government to implement stringent measures aimed at addressing India's NPA crisis.

The Insolvency and Bankruptcy Code (IBC) has shown promise in improving asset recovery and reducing NPAs. If banks continue to perform under the discipline of the IBC framework, a further decline in NPAs can be expected. Furthermore, positive developments involving LIC and government initiatives in the past year suggest a potential improvement in the NPA ratio, particularly for IDBI Bank. Predictors such as the Debt-to-Income (DTI) Ratio, Overdraft Ratio, and Annual Income should be prioritized during the loan sanctioning process. Assigning greater weight to these variables can help reduce the risk of loan defaults and improve overall credit quality.

Despite various initiatives by the government, RBI, and financial institutions, the NPA issue remains unresolved. Addressing this challenge requires a combination of well-researched, practical legal frameworks and strict adherence to regulatory policies. By implementing these measures, the Indian banking system can work towards achieving sustainable credit growth and fostering inclusive economic development.

Suggestions

This study primarily explores the future trajectory of NPAs while emphasizing the need for immediate measures to curb their growth, particularly in the short term. The following suggestions are proposed:

Public sector banks (PSBs) should incorporate proven measures used by foreign banks to manage NPAs effectively. These practices could enhance profitability and operational efficiency, enabling PSBs to achieve results comparable to their foreign counterparts. Introducing greater accountability through oversight mechanisms, such as the Office of the Central Vigilance Commissioner, can motivate public sector loan officers to take proactive measures in managing loan approvals and recoveries.

The government should implement flexible compensation packages and performance-based incentives for middle and lower-level management in PSBs. Extending such incentives to senior management could further align their efforts toward reducing NPAs and improving overall efficiency. Private banks should consider expanding their operations internationally to offset losses caused by domestic NPAs. Both public and private banks should also focus on establishing a presence in rural areas, diversifying their customer base and increasing profitability by tapping into untapped markets.

A robust credit appraisal system is essential for minimizing NPAs. It is recommended to adopt a committee-based screening system for loan applications. This approach encourages shared responsibility and reduces the likelihood of poor credit decisions. Banks like IDBI should strategically minimize lending to sectors such as MSMEs, which contribute disproportionately to their NPAs. Instead, banks should prioritize sectors with lower default risks.

Future research directions

This study serves as an initial exploration of the long-term effects of NPAs. However, given the limited variables included, its findings should be interpreted cautiously. Future research should include a broader range of loan categories, such as home loans, gold loans, and education loans, to gain deeper insights into NPAs across various sectors. Develop more comprehensive predictive models that consider macroeconomic and borrower-specific variables under changing economic conditions.

By implementing these measures, the banking sector can work toward controlling NPAs, fostering profitability, and ensuring long-term financial stability.

References

Bace E. (2016), Bank profitability: liquidity, capital and asset quality. J Risk Manag Financ Inst 9(4), pp. 327–331.

Bhaskaran R., Bhalla L., Sarin V., Kaur S., Rahman A., Singh G., Bhattacharya A.M., Jha A.K. and Verma P. (2016), Nonperforming assets of public and private sector banks in India – a comparative study, International Journal of Services and Operations Management, Vol. 25 No. 2, pp. 155-172.

Das, S.K. (2023), NPAs in India's banks: trends and determinants, Journal of Money and Business, Vol. 3 No. 2, pp. 147-158.

Das S.K. and Uppal K. (2021), NPAs and profitability in Indian banks: an empirical analysis, Future Business Journal, Vol. 7 No. 53, pp. 1-9.

Etale L.M., Ayunku P.E., Etale E. (2016), The impact of nonperforming loans and bank performance in Nigeria. Int J Humanit Soc Sc Invent 5(4), pp.1–5

Kadanda D., Raj K. (2018), Non-performing assets (NPAs) and its determinants: a study of Indian public sector banks. J. Soc. Econ. Dev. 20, pp. 193–212.

Kumar R.B., Subba Rao B.V.S.S. and Kusuma G.D.V. (2018), Genesis for increase of NPAs in Indian banks – an empirical analysis, Journal of Banking and Finance Management, Vol. 1 No. 1, pp. 1-8.

Masood O., Ashraf M. (2012), Bank-specific and macroeconomic profitability determinants of Islamic banks: the case of different countries. Qual Res Financ Mark 4(2/3), pp.255–268.

Mishra, M.K. and Pawaskar, J.R. (2017), A study of nonperforming assets and its impact on banking sector, Journal for Research, Vol. 3 No. 1, pp. 1-5.

Mukhopadhyay, A. (2018), Finding innovative solutions to India's NPA woes, available at: `www.orfonline.org/research/finding-innovative-solutions-to-indias-npa-woes/`.

Ongore V.O., Kusa G.B. (2013), Determinants of financial performance of commercial banks in Kenya. Int J Econ Financ Issues 3(1), pp. 237–252.

Park, C.Y. and Shin, K. (2021), COVID-19, nonperforming loans, and cross-border bank lending, Journal of Banking and Finance, Vol. 133, 106233.

Serrano, A.S. (2021), The impact of non-performing loans on bank lending in Europe: an empirical analysis, The North American Journal of Economics and Finance, Vol. 55, 101312.

Singh, V.R. (2016), A study of nonperforming assets of commercial banks and its recovery in India, Annual Research Journal of SCMS, Vol. 4, pp. 110-125.

RBI Circular (2007), https://www.rbi.org.in/Scripts/BS_ViewMasCirculardetails.aspx?id=3702#non

Received: April 10, 2024

Accepted: February 10, 2025