Design and Development of Robust Credit Risk Management and Risk Appetite Framework for the Banking Sector

THESIS

Submitted in partial fulfillment
of the requirements for the degree of
DOCTOR OF PHILOSOPHY

By
ANKUR JOSHI
2015PHXF0503P

Under the Supervision of
Prof. N. V. Muralidhar Rao

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE
PILANI CAMPUS, PILANI
(RAJASTHAN) INDIA
2024

Dedicated to

*My dearest late father, **Dr. SN Joshi, Babba,** as I called him who I lost recently in December 2023. He has been my inspiration and would be super proud and smiling from heaven today.*

*Gratitude to my mother, **Mrs. Beena Joshi** who has been the pillar of our family and my younger brother, **Abhinav Joshi** who has been a constant source of support.*

*Sincere thanks to my wife, **Pratibha** without whose support this journey would not have been possible.*

*And to my kids, **Ananya** and **Praayan** who will make me prouder of their achievements one day…!*

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE
PILANI CAMPUS, PILANI, RAJASTHAN

CERTIFICATE

This is to certify that the thesis entitled '**Design and Development of Robust Credit Risk Management and Risk Appetite Framework for the Banking Sector',** submitted by Mr. Ankur Joshi, ID No. 2015PHXF0503Pfor the award of Ph.D. degree of the Institute embodies original work done by him under my supervision.

N. V. Muralidhar Rao

Signature in Full of the Supervisor : N V MURALIDHAR RAO
Name of the Supervisor : **Prof. N. V. MURALIDHAR RAO**
Designation : **Professor**
Department of Economics and Finance
BITS Pilani, Pilani Campus, Rajasthan.

Date: **25 APR 2024** :

ACKNOWLEDGEMENT

I would like to extend my sincere and heartfelt thanks to my supervisor, Prof. N. V. Muralidhar Rao for his invaluable and esteemed guidance, encouragement, and un-tiring moral support and motivation, without which sailing through the formidable task of thesis completion would not have been possible. It has been a great pleasure to be associated with him on this work.

I would like to express my deep sense of gratitude to Prof. Ramagopal Rao, Vice Chancellor, BITS Pilani for providing me an opportunity to work in an area of my choice. I also want to acknowledge my sincere gratitude to Prof. S K Barai, Director of BITS, Pilani, Pilani Campus, Prof. M B Srinivas, Dean, Academic Graduate Student Research Division, and Prof. S Chakraborty, Associate Dean, Academic Graduate Student Research Division of Pilani Campus.

I would also like to extend my thanks to my Doctoral Advisory Committee members, Dr. Krishna M., Head of the Department and Dr. Rajan Pandey for taking out time from their busy schedules and giving their valuable comments and feedback on the thesis work that helped me to improve it further. I would also like to express my gratitude to Dr. Arun Kumar Giri, Former Head of Department, and Prof Arya Kumar, Senior Professor of Economics and Finance, for their continuous support in all ways. My special thanks to all the faculty members of the Department of Economics and Finance and my fellow research scholars for their help and interest in my work. I also thank the supporting staff of the Department of Economics and Finance, BITS-Pilani for their cooperation and help. I would like to extend my gratitude to the BITS Pilani Library for their excellent service. I again thank all the people who were directly and indirectly involved and helped me in writing this thesis.

Ankur Joshi

ABSTRACT

Risk management in banks has changed significantly over the years. The regulations that emerged from the global financial crisis and the fines that were levied in its wake triggered a wave of change in risk functions. These included more comprehensive and stringent capital, leverage, liquidity, and funding requirements, as well as higher standards for risk reporting, such as BCBS 239. Stress testing emerged as a major supervisory tool, in addition to the rise of expectations for bank risk-appetite statements. Banks were also forced to invest in strengthening risk culture and involved their boards in key risk decisions. The rapid adoption of a new breed of models is offering much deeper insights into data. Machine learning identifies complex, nonlinear patterns in large data sets and makes more accurate risk models possible.

Importance of having heightened standards of risk management were first tested during the 2008 Global Financial Crisis which started with the US housing market and turned into a global recession. Availability of cheap credit and relaxed lending rules that resulted in creating real estate bubble in 2008 were the root cause of the crisis. The financial institutions were left with trillions of US dollars of substandard mortgage assets when the housing market bubble broke (Singh, 2023).

The most recent financial crisis events of 2023 that resulted in failure of Credit Suisse and eventual takeover by UBS Bank and failure of Silicon Valley Bank (SVB), First Republic Bank and Signature Bank in United States has again to the fore-front the need for heightened standards of risk management.

India’s banking sector is at present sufficiently capitalized and well-regulated. The financial and economic conditions in the country are at present superior to the rest of the world. The central bank, Reserve Bank of India (RBI) has played a pivotal oversight role in managing the financial performance of banks in India. Credit risk, market risk and liquidity risk studies suggest that Indian banks are generally resilient and have withstood the global downturn and have performed better compared to its peers in developed and other developing markets. The Indian banking industry has witnessed the successful rollout of innovative banking models like payments and small finance banks.

Public sector banks in India, often referred to as the government banks of India, play a pivotal role in India's financial landscape, exerting a profound influence on its economic development. These banks have driven financial inclusion and socio-economic progress by supporting government initiatives like the Jan Dhan Yojana and Mudra Yojana. Their extensive branch networks extend banking services to even the remotest corners, ensuring accessibility for all. With such a network and mandate of these banks to reach out to the remotest and the unbanked areas of this vast country brings unique challenges of maintaining credit quality and sustainability to continue to carry out its mandate to be the universal bank for all.

RBI has played an important role in de-risking the Indian Banking System. It has successfully navigated the Banking Crisis in India by bringing down Non-Performing Loans which peaked in 2018. Mergers of public sector banks in India have resulted in reduction of public sector banks from 27 to 12 during 2017-2022 period. Loan advances share of public sector banks has reduced from 76 percent in 2013 to 58 percent in 2022.

The Indian banking system at present consists of 12 public sector banks and 21 Private Sector Banks, 46 foreign banks and 43 regional rural banks, in addition to cooperative credit institutions. Public-sector banks comprise of nearly 58 percent of total credit advances. Standard and Poor’s estimates that credit growth in India’s banking sector would continue at greater than 10 percent. The government and the regulator have undertaken several measures to strengthen the Indian banking sector.

The Indian banking sector accounts for a major portion of financial intermediation and is considered to be the main channel of monetary policy transmission, credit delivery, and payment systems. The stability and sound health of the banking system hence is a key pre-requisite for overall economic development and financial stability. The Non-Performing Assets (NPA) is an important indicator to assess the financial health of the banking sector. NPAs reflect the credit risk management soundness. There is a near unanimity in the literature that asset quality is a critical determinant of the sound functioning of the banking system. NPAs affect operational efficiency, which in turn affects the profitability, liquidity, and solvency position of banks (Michael, et al, 2006). The consequences of NPAs would be a reduction in interest income, high level of provisioning, stress on profitability, increased pressure on net interest margin (NIM) thereby reducing competitiveness, steady erosion of capital resources and increased difficulty in augmenting capital resources. The contagious nature of loan losses emanates from the fact that their downside impact can quickly transmit to earnings, capital, and liquidity. They are insidious in the sense that it is often difficult to know that there is a problem until it's too late. Moreover, these problems prey on weak banks, which are vulnerable and have relatively small amounts of capital to absorb unanticipated losses.

NPAs generate a vicious cycle of effects on the sustainability and growth of the banking system, and if not managed properly could lead to bank failures. Empirical evidence indicates a relationship between bank failures and higher NPAs worldwide. Such crisis tend to arise primarily from deteriorating macro-economic indicators, thus causing significant impact in asset quality of banks that have higher concentration in riskier assets. The issue is of particular importance after the global financial crisis and the failure of large institutions and bailouts that followed. In the Indian context, empirical research suggests that asset quality is one of the main determining factors of credit, besides time deposits and lending interest rate (RBI, RCF, 2006-08). A recent example of poor NPA management and asset quality issues leading to the crisis in an Indian private sector bank was YES Bank. Asset quality surveys conducted by Yes Bank in the years 2017 and 2018 revealed significant administrative shortcomings and a sharp increase in lending, which ultimately contributed to bank failure and government bailout. From April and September 2019, the bank witnessed a remarkable increase in its non-performing assets (NPAs). Yes Bank lacked governance and risk management oversight that led to its collapse. The bank witnessed a steady withdrawal of deposits, which further worsened its asset liability ratios and ultimately led to its default (Sarkar 2020).

Despite the deterioration in asset quality of banks worldwide during GFC 2008, Indian banks largely remained insulated and continued to declare good quality of assets and earnings. However, in little over 5 years from the time global financial crisis of 2008 ended its impact on large economies like the US, Indian banks came under intense pressure on account of high NPAs with gross NPAs at 10%+ levels, in many public sector banks. During the time when indian public sector banks started witnessing an increase in non-performing loans, nearly 70% of gross advances were held with public sector banks. Any stress in these banks was detrimental to the Indian banking sector and the economy at large.

Credit risk management is the backbone for prudent management of banks. Risk management is based on lending principles and an efficient risk framework is required to manage the risk in a financial organization. In banking, risk management is regulated by the supervision of the

Risk management committees where different policies, industry-specific standards, and guidelines along with risk concentration limits are prioritized. These policies, standards, and procedures are governed for credit risk measurement, monitoring and reporting.

According to commonly used definition of risk appetite, it is measured as "the amount and type of risk that an organization is able and willing to accept in pursuit of its business objectives". For the management of risk, the risk appetite is known as the "key performance indicator (KPI)" system.

The global financial crisis has highlighted that number of banks lacked a proper understanding of their true risk profile and realized too late that it was not in line with their desired risk appetite. This forced senior management to explain losses that were a multiple of what shareholders had expected to face. The key lessons learned from this crisis is that financial institutions need to have a comprehensive risk appetite framework in place that helps them better understand and manage their risks by translating risk metrics and methods into strategic decisions, reporting, and day-to-day business decisions.

Risk appetite is considerably more than a sophisticated key performance indicator system for risk management. It's the core instrument for better aligning overall corporate strategy, capital allocation, and risk. Regulators, rating agencies, and professional investors are aggressively pushing banks to advance their risk management practices. A comprehensive risk appetite framework is the cornerstone of risk management architecture. It is no doubt that more studies on the design and development of robust credit risk management and risk appetite framework for the Indian banking sectors are required as a stepping stone and this research will be performed as a benchmark for researchers, academicians, regulatory authorities, bankers, etc. concerning the prevention of financial loss in the banking sectors.

Based on the review of the literature, research gaps were identified. There are limited studies that have tried to conduct a deep-dive cause and effect of weak risk management practices in public sector banks. There is no definitive research available that has covered the impact of banks' mandate for inclusive growth which requires banks to increase their fund/non-fund based credit limits into the priority sector to 40% by 2019. Therefore, the major research issues relevant to us for this study were:

To identify and evaluate Indian banking system's risk management practices and design a robust risk framework for sustainable asset quality of banks

To assess the potency of risk identification, risk assessment and analysis, risk monitoring, and control within the banks to understand the causality between risk management framework and its impact on performance of banks.

Finally, the design and development of a credit risk management framework by studying the causality of various macro-economic variables and bank's internal key metrics to probability of defaulting loans and thus leverage these relationships to set up risk appetite framework.

The research constitutes studying key public sector banks (PSUs) and private sector banks in India. The selected 12 banks from public and private sector comprised of approximately 70% of total gross advances and is considered a reasonable sample size for research. The work describes an empirical framework to assess the cause and effect of various macro-economic and internal variables on non-performing loans. Various machine learning models and traditional regression models were used to assess the correlation between dependent variable (non-performing loans) and independent variables (macro-economic variables and key internal metrics).

The trend analysis (2008-2019) was conducted based on secondary data for bank-related parameters such as PSL, TL, STA, TE, TP, TA, USTA (Key Internal Metrics) and GDP, RR and CPI (macro-economic variables). These variables were analysed to understand the impact of bank-related factors and macroeconomic variables on credit risk measured by non-performing loans (GNPA/NNPA). Period of 2020 and 2023 was also reviewed as out of sample research but the limitation of this period is the Covid impact where economic indicators of markets globally, including India, were significantly impacted thus not providing meaningful conclusions of the impact of this period on the research work. This period also witnessed a substantial improvement in Indian banks' performance, as a result of merger of weak public sector banks into strong banks and sale of non-performing loans to asset reconstruction companies.

Machine learning models and traditional multi-variate statistical techniques have been used to quantify the impact of bank-related factors and macroeconomic variables on the credit risk. The selected variables have varying degree of impact on credit risk. The research confirms the reasonableness of selected variables to assess, measure and monitor credit risk and usage of these variables in designing the agile risk assessment framework.

Regression coefficient between GNPA/NNPA (dependent variable) and macro-economic and key internal variables (independent variables) revealed varying degrees of significance between the independent variables such as unsecured/total advances, gross domestic product (GDP), consumer price index (CPI), total profit/total advances, term loans (TL), GDP-1 (lag), total advances, repo (interest) rate (RR), secured to total advances (STA), total earnings/total advances, priority sector lending (PSL) and CPI-1 (lag). In regression analysis, it can be inferred that GNPA/NNPA has varying degrees of correlation to the selected independent variables.

Accuracy and prediction of dependent variable (GNPA/NNPA) through advanced machine learning algorithms were assessed by computing linkages between macroeconomic variables and key internal variables. Various machine learning regression models were used such as Random Forest, Decision Tree, XG Boost, Artificial Neural Network (ANN) and Support Vector Machine (SVM). Machine learning regression models were used to predict the non-performing asset percentage (GNPA/NNPA) and based on significance of the independent variables, these variables have been recommended to be used in risk appetite framework to set risk concentration limits in the newly designed risk appetite framework.

Machine learning classification models such as Random Forest, Decision Tree, XG Boost, Naïve Bayes, Artificial Neural Network (ANN), Support Vector Machine (SVM) and K-Nearest Neighbour (KNN) were used to compute the accuracy of relationship between independent variables and dependent variable. The research has concluded varying degree of significance of these selected variables and classification model results show varying degrees of accuracy of causal relationship between selected dependent and independent variables.

The results of the Regression and Classification Models showed different degree of Accuracy measured as 1-MAPE (Mean Absolute Percentage Error) when the testing the relationship of independent variables with Gross Non-Performing Assets (GNPA). Random Forest model showed the most reasonable MAPE of 11.5% while Decision Tree showed MAPE of 0%, which could be attributed to over-fitting of data set. Other models such as XG Boost, SVM and ANN have shown high MAPE values, which have not been considered due to high error values. However, given that the data set used in research is limited, and machine learning models results are most optimal with large data sets, it's recommended that these models should be considered by Banks for setting up risk appetite framework.

Machine learning regression model results to predict the net non-performing assets (NNPA) were similar with Random Forest Model giving MAPE of 22.7% (Accuracy = 77.3%), which was the most reasonable fit amongst all the Models that were tested. Decision Tree results showed the lowest MAPE of 0% (Accuracy = 100%), as it was observed in GNPA predicted values. This can be attributed to over-fitting of data set. Other models such as XG Boost, SVM and ANN have shown high MAPE values, which have not been considered due to high error values. However, given that the data set used in research is limited, and Machine learning models results are most optimal with large data sets, it's recommended that these models should be considered by banks for designing new generation agile risk appetite framework.

Machine learning regression Random Forest model was trained to further assess whether the predicted value of GNPA further improves by segregating data set based on risk-tiering (cohorts). For this purpose, data set comprising of 8 public Sector Banks was segregated into 4 cohorts. Cohorts are designed in terms of size of banks' credit advances to ensure that similar sized banks are clubbed together. This cohort approach has been used to test whether different risk limits be designed by type of assets in risk appetite framework. Cohort 1 comprises of State Bank of India, Bank of Baroda and Canara Bank. Cohort 2 comprises of Bank of India, Andhra Bank. Cohort 3 is comprised of Central Bank of India and Allahabad Bank. Cohort 4 comprised of Bank of Maharashtra. GNPA predicted values have varying degree of MAPE, with Cohort 1 MAPE of 8.45% (Accuracy = 93.55%) to Cohort 4 with MAPE of 21.33% (Accuracy = 79.67%).

Machine learning regression Random Forest model was trained to further assess whether the predicted value of NNPA further improves by segregating data set based on risk-tiering (cohorts). For this purpose, data set comprising of 8 public sector banks was segregated into 4 cohorts. NNPA predicted values have varying degree of MAPE, with cohort 1 MAPE of 13.07% (Accuracy = 76.93%) to Cohort 2 with MAPE of 39.5% (Accuracy = 60.5%).

Machine learning classification models were tested to compute the accuracy of whether the dependent variable (GNPA) was correctly classified as GNPA, based on relationship of the variable with independent variables. The accuracy rates, which is defined by Precision value were found to be highest in case of Random Forest (100%) followed by Decision Tree and XG Boost (97.2%), SVM (95%), KNN (93%), Naïve Bayes (86%) and ANN (65%). It can be concluded that XG Boost and Decision Tree have yielded optimum accuracy rates and can be considered as good fit models. Random Forest with precision of 100% indicates model over-fitment. SVM, KNN and Naïve Bayes models should be considered in risk appetite framework development, as these models have high precision values and given the limited data sets available in this research, should be considered.

Machine learning classification models were tested to compute the accuracy of whether the dependent variable (NNPA) was correctly classified as NNPA, based on relationship of the variable with independent variables. The accuracy rates, which is defined by Precision value were found to be highest in case of Random Forest (100%) followed by Decision Tree and XG Boost (96.5%), SVM (95%), KNN (92%), Naïve Bayes (88%) and ANN (78%). It can be concluded that XG Boost and Decision Tree have yielded optimum accuracy rates and can be considered as good fit models. Random Forest with precision of 100% indicates model over-fitment. SVM, KNN and Naïve Bayes models should be considered in risk appetite framework development, as these models have high precision values and given the limited data sets available in this research, should be considered.

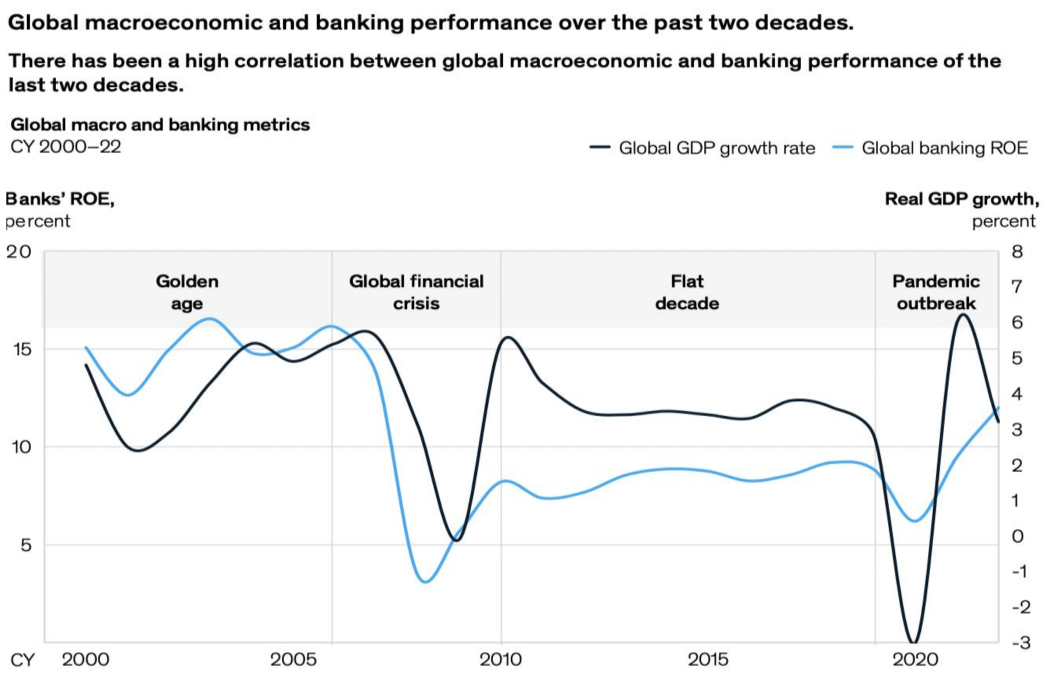
Machine learning classification Decision Tree model was trained to further assess whether the classification of GNPA further improves by segregating data set based on risk-tiering (cohorts). For this purpose, data set comprising of 8 public sector banks was segregated into 4 cohorts. GNPA classification further improved by using the cohort approach with each cohort estimating 100% precision value.

Machine learning classification Decision Tree model was trained to further assess whether the classification of NNPA further improves by segregating data set based on risk-tiering (cohorts). For this purpose, data set comprising of 8 public sector banks was segregated into 4 cohorts. GNPA classification further improved by using the cohort approach with each cohort estimating 100% precision value, except Cohort 1 which estimated precision value of 97.2%.

Besides the empirical analysis, Machine learning algorithm models have provided better insights in predicting the dependent variable values and its accuracy. ML algorithm models such as Linear Regression, Logistic Regression, Multivariate Regression, Naïve Bayes, Decision tree, Random Forest, SVM, ANN, KNN and hybrid models (simple hybrid models and class-wise classifier) have been successfully used in credit scoring, operational efficiency, but the predictive analysis of understanding the relationship of macro-economic variables and key internal variables with non-performing assets (GNPA/NNPA) is still not widely used in Indian Banks in its the risk appetite framework.

Global macro-economic environment and banks' performance show a high degree of correlation. Data from period 2000-2022 shows strong correlation between global real GDP growth percent and return on equity of banks globally. Thus, it's imperative that any risk appetite framework should fundamentally incorporate GDP as one of the key variables while computing set risk limits. Impact of changes in GDP growth percent should be analysed and used in risk appetite framework while setting risk concentration limits.

Table: Source (Mc Kinsey)



During the period of 2013-2018, Indian banks performed in sharp contrast to banks globally with sharp decline in asset quality and negative to low return on equity (ROE). This was a period marked by slow GDP growth in India and public sector banks exposure increase in riskier, lower quality assets, higher exposure in long gestation projects, higher exposures to large corporates. There are reports of lack of sound risk appetite framework based governance and weak risk management practices in some public and private sector banks that resulted in significant increase in credit defaults in banks. Since 2018-2019, Reserve Bank of India has taken number of actions which has resulted in marked improvement in asset quality in Indian banks. Exposure to wholesale lending credit has reduced from 54 percent in 2017 to 47 percent in 2023. Secured lending exposure share of banks in housing loans has increased from 58 percent in 2017 to 62 percent in 2023. Unsecured lending exposure share of banks has reduced from 83 percent to 79 percent during 2017-2023 period. All these changes directed by RBI validates the research findings where monitoring of exposures to priority sector, unsecured lending, exposure to large corporates have been found to be significant independent variables to be used in design of new generation agile risk appetite framework.

The research work carried out and elaborated in the thesis document conclusively indicate the significance of select independent variables and its correlation to dependent variable and its use in setting up a robust credit risk appetite framework by using latest Machine learning algorithm models to set risk concentration limits. The risk appetite framework will ensure that Indian banking system remains robust and can proactively identify emerging risks and thus maintain an optimal level of credit defaults & non-performing assets (GNPA/NNPA) for sustainable profitability of returns on equity and healthy banking system in India.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF SELECT ABBREVIATIONS

ABS	Asset backed security
AFI	Annual financial inspection
AIFI	All-India financial institution
ALM	Asset-liability management
AS	Accounting standard
BCBS	Basel committee on banking supervision
BIFR	Board for industrial and financial reconstruction
BIS	Bank for international settlements
BSE	Mumbai stock exchange
CALCS	Capital adequacy, asset quality, liquidity, compliance and system
CAMELS	Capital adequacy, asset quality, management, earnings, liquidity, systems and control
CBLO	Collateralized borrowing and lending obligation
CBS	Consolidated banking
CDR	Corporate debt restructuring
CFS	Consolidated financial statement
CIBIL	Credit Information Bureau of (India) Limited
CPI	Consumer price index
CRA	Credit reporting agency
CRR	Cash reserve ratio
DRI	Differential rate of interest
DRT	Debt Recovery Tribunal
ECB	External commercial borrowing
ECS	Electronic clearing service
FCNR	Foreign currency non-resident
FCRB	Financial companies regulation bill
FEMA	Foreign exchange management act
FII	Foreign institutional investor
GDP	Gross domestic product
IBA	Indian Banks' Association
IDBI	Industrial Development Bank of India
IFCI	Industrial Finance Corporation of India
IIBI	Industrial Investment Bank of India Limited
IMF	International monetary fund
IPO	Initial public offering
IRF	Interest rate future
IRS	Interest rate swap
IT	Information technology
JLG	Joint liability group
JPC	Joint parliamentary committee
KYC	Know your customer
LAF	Liquidity Adjustment Facility
MBS	Mortgage Backed Security
ML	Machine learning algorithms
MEI	Macro economic indicator
MFI	Micro finance institution

NRO	Non-resident ordinary accounts
NRSR	Non-resident special rupee
NSDL	National Securities Depository Limited
OBC	Oriental Bank of Commerce
OBU	Off-shore banking unit
OCB	Overseas corporate body
OMO	Open market operation
OTCEI	Over the Counter Exchange of India
OTS	One-time settlement
PAC	Public accounts committee
PAT	Profit after tax
PBT	Profit before tax
PCA	Prompt corrective action
PIS	Portfolio investment scheme
PLR	Prime lending rate
PSB	Public sector bank
PSU	Public sector undertaking
RBS	Risk based supervision
RFC	Resident foreign currency
RIDF	Rural Infrastructural Development Fund
RoA	Return on asset
RoE	Return on equity
RRB	Regional rural bank
RTGS	Real time gross settlement
RWA	Risk weighted asset
SCB	Scheduled commercial bank
SEBI	Securities and Exchange Board of India
SEFT	Special electronic funds transfer
SFC	State financial corporation
SHG	Self help group
SIDBI	Small Industries Development Bank of India
SIDC	Small Industries Development Corporation
SLBC	State level bankers' committee
SLR	Statutory liquidity ratio
SME	Small and medium enterprise
SPV	Special purpose vehicle
SSI	Small scale industry
VaR	Value-at-risk
WADR	Weighted average discount rate
WPI	Wholesale price index
YTM	Yield to maturity

CHAPTER 1

INTRODUCTION

1.1Background of the Indian banking system

The Central Banking Enquiry Committee (1931) mentioned that money lending activities in India have been observed from the Vedic period, *i.e.*, from 1400 BC. Since 500 BC, the existence of professional banking in India has been traced to the concept of "*Kautilya's Arthashastra*", which was earlier from 400 BC and referenced employing creditors, lenders, and lending rates. In general, the banking system was fairly based on the need for credit for trading, commercial aspects, agricultural activities, and the development of a person for a better economy (RBI, 2008).

As per RBI (2008), before the independence period, the existence of private banks was noticed as an organized form of "joint-stock companies". Majority of banking sectors were small-sized and had private shareholding with holding varieties that were localized in nature and most of them failed. In 1935, these sectors came under the purview of the Reserve Bank, which was established as a central bank for the country. It was found with the limited process of regulation and supervision as per the provisions of the Reserve Bank of India Act, 1934, and the Companies Act, 1913. From the organizational part of the system, the native bankers and moneylenders were found isolated, and the cooperative credit system was the only hope for credit. Fourteen significant private banks were nationalized on July 19, 1969. The justification for bank nationalization was straightforward: as credit symbolizes control over money, the distribution of this control among different social groups, economic sectors, and geographic areas influences the course of social growth. Given the significant social ramifications of financial resource distribution, financial institutions must be held by the State in order to ensure social authority over the distribution of credit (Patnaik, 2019). Later, the most significant evolutionary phase in banking was observed as per financial sector reforms, which were initiated in 1991-92 and further divided into two sub-phases such as 1991-1992 to 1997-1998 followed by 1998-99 and beyond.

Rajeswari (2014) reported on different banking sectors of India. Among Indian banking sectors, 28 state-owned banks are dominating in which the operations are found through a network of 50,000 branches and 13,000 ATMs. Among public sector banks, the “State Bank of India” (SBI) is the largest bank in India and seven associate banks are linked, which has an asset base of above Rs, 7,000 billion (approximately US$ 150 billion). The other largest public sector banks are “Punjab National Bank”, “Canara Bank”, “Bank of Baroda” and “IDBI Bank”. It is known that the public sector banks are running overseas operations with “Bank of Baroda” connected with 52 branches, subsidiaries, joint ventures, and representative offices outside India, followed by SBI (45 overseas branches/offices) and Bank of India (26 overseas branches/offices). The “Indian banks”, which also include private sector banks, have 171 branches/offices outside India. Some of the well-known private banking sectors are recorded as “ICICI Bank”, “HDFC Bank” and “IndusInd Bank”. Besides these, “Yes Bank” is the latest entrance to the private sector banking organization. The private banking sector are found with an asset value of over Rs. 5,700 billion (above US$ 124 billion) operates through a network of 6,500 branches and over 7,500 ATMs. Some of the leading international banking sectors, which are operating a business in India include “Standard Chartered Bank”, “HSBC Bank” and “Citibank N.A.”.

As of 2022, the Indian banking system is made up of 12 government-owned banks, 22 banks in the private sector, 46 foreign banks, 56 regional rural banks, 1485 urban co-operative banks, and 96,000 rural cooperative banks in addition to cooperative credit institutions. There were 213,145 ATMs across the nation as of September 2021; 47.5% of these were located in semi-urban and rural locations. Bank assets increased in all industries in 2020–2022. In 2022, the combined assets of the banking industry—which includes both public and private banks—rose to US$ 2.67 trillion. Public and private banking sectors' combined assets in 2022 were US$ 1,594.51 billion and US$ 925.05 billion, respectively (Sreenivasamurthy, 2022).

Barman and Samanta (2007) reported in their review article that the banking sectors are classified as banks and other related financial organizations. In India, several sub-sectors have been discussed by many researchers (Barman, 1995; CSO, 1989; 1999; 2006; Barman and Samanta, 2007), which are as follows:

1. The banking department of the Reserve Bank of India (RBI).
2. Commercial banks.
3. Public non-banking financial corporations.
4. Organised non-banking financial companies engaged in shares trading, investment holdings, loan finance, etc.
5. Unorganised non-banking financial authorities viz. professional moneylenders and pawnbrokers.
6. Post office savings banks including operations concerning cumulative time deposits and national saving certificates.
7. Cooperative credit societies.

In order to promote financial inclusion in India, particularly in rural regions, the RBI established the concept of Payment Banks, primarily targeting low-income individuals and small enterprises. The primary goal of these banks' setup is to concentrate on low-value, high-volume transactions. In a study released on January 7, 2014, the Committee on Comprehensive Financial Services for Small Businesses and Low-Income Households made recommendations for the establishment of payment banks, a brand-new category of banks. On August 19, 2015, the Reserve Bank of India granted "in-principle" licenses to 11 businesses to establish payment banks, out of the 41 applicants for the same (Shrey et al., 2018).

The Reserve Bank of India (RBI) is known as the monetary authority of the country and is also called as Central Bank of India. This organization is regulating the banking sectors and laid down guidelines for the regular functioning of banking sectors within the overall framework of the “Banking Regulation Act, 1949”, “Foreign Exchange Management Act, 1999” and “Foreign Direct Investment (FDI)” policy of the government (Rajeswari, 2014). By implementing monetary policy, the RBI manages all price fluctuations and prevents both inflation and deflation. Traditionally the Reserve Bank's operations have evolved frequently in tandem with changes in the Indian economy and the banking industry. The Reserve Bank of India carries out significant monetary tasks, such as issuing currency notes and preserving the nation's monetary stability. Unified oversight of India's financial industry is the Reserve Bank of India's primary goal. The Central Board of Directors, which is chosen by the Government of India, oversees the RBI. The Reserve Bank of India has been crucial to the nation's monetary stability and economic growth (Mahato, 2022).

In general, the function of the banking sector is mainly to receive deposits from customers and also borrow money from other sources such as loans, advances, etc. for raising funds called “working capital funds” (Limbore and Mane, 2014). It is a well-known process that banks have paid costs through paying interest on these raising funds. On the other hand, for the recovery of these costs especially used in administrative and other expenses of banks have earned a profit, as well as banks, have utilized the working capital funds through providing advances or investments. In this regard, working capital funds are observed as bank liabilities, which are converted into assets. Therefore, banking sectors must take care of asset values.

1.2 Risk factors in the banking system

The risk factors concerning monetary lending are found in any financial organization (Srinivas, and George, 2020). The banking sectors of India somehow are holding risks, but these sectors are trying to identify risks followed by the estimation and management of risks. The risk is generated mainly from credit. Globally, the banking sectors are facing credit risk, which is the primordial form of risk (Treacy and Carey, 2000).

It was mentioned in an earlier study that risk cannot be avoided in any financial organization (Fennelly and Perry, 2017). A researcher stated that without risk the world may not pose an innovative idea, rewarding aspects, and proper responsibility (Swanson, 2017). It is well known that risk is a positive force for developing growth and success, leading to uncertainty, and discovering the advantages of organizations (Swanson, 2017).

As per the 1991 economic policy, financial organizations especially banking sectors have received specific attention regarding the improvement of financial strength and functional efficiency to reach the level of international standards (Sankareswari, 2012). There had been a little substantial rivalry in the Indian banking industry prior to the beginning of the reforms in 1991 for two reasons. Initially, banks were provided with limited flexibility to distinguish themselves from the market by the RBI's stringent regulations on, for instance, interest rate setting. Second, entry requirements for emerging financial institutions were stringent in India. Since the early 1990s, there has been a notable increase in competitiveness due to the reduction of entrance restrictions. From 1994 and 2000, the banking sector saw the entry of seven new privately owned banks. Furthermore, since 1994, more than 20 foreign banks have opened offices in India. The overall proportion of foreign banks and the new privately owned banks by March 2004 was close to 20% of total assets (Roland, 2006).

In an earlier report, Englund (1999) mentioned that bankruptcies occurred due to non-performing loans, credit losses, and acute banking crises in the Swedish banking sectors. It was observed that in the late 1990s Asian crisis struck while the Russian crisis followed soon after, which are reported by Wahlstrom (2009) and Roberts (2015). Another study revealed that significant “unanticipated” and “unintended” financial risks are accumulated due to large risks at an individual, organizational, or government level. These causative factors are a lack of knowledge or understanding of the risks by stakeholders and managers within those entities (Draghi et al., 2003).

The fundamental cause of the Asian crisis was a significant, abrupt shift in capital flows. A great deal of foreign capital was being drawn to this economic growth, but of suddenly short-term loans were being withdrawn, portfolio capital was leaving, and local investors fled overseas. Most of these fresh inflows were in the form of financing from private creditors, which quadrupled in size from $25.8 billion to $78.4 billion in a matter of two years. Private debtors include commercial financial institutions as well as non-bank creditors like bondholders. However, in the second portion of 1997, the inflows abruptly switched, with a $11.9 billion net outflow of private capital. It's also critical to remember that in nations with fixed exchange rates, local investor panics can wipe out foreign exchange reservations, which can then spark a financial catastrophe (Radelet and Sachs, 1999).

Prior to a number of years of financial reform, privatization, and macroeconomic stabilization policies in Russia, the Soviet Union fell apart in 1991. The implementation of the currency peg, a form of exchange rate system whereby the value of one currency remains constant compared to the value of another, was a key component of this. Despite the changes made since 1991, Russia's basic institutional flaws persisted. The Asian financial crisis emphasized and intensified these shortcomings. Weak tax collection in Russia and an expensive war in Chechnya were made worse by a global economic downturn and a decline in prices for commodities. Due to this, there were budgetary imbalances and concerns about the government's capacity to maintain a stable currency rate and pay off its sovereign debts. The depreciation of the rouble and investment moving out of Russia became more probable as a result of the rise in default and exchange rate risk. This also resulted in severe pressure on the currency and bond market in Russia. Subsequently, on August 17, the government declared a default on its domestic obligations, a decline of the rouble's pegged rate of exchange, and a 90-day halt on repayments made by commercial lenders to foreign debtors. Two weeks afterward, on the second of September, the Bank of Russia gave up trying to keep the rouble at a fixed rate and let it fluctuate freely. The rouble had lost nearly two-thirds of its worth in just three weeks (Coyle, 2022).

The risks lead to uncertainties, which may cause the banks to lose and go bankrupt. According to the Basel Accord (Basel II, 2006), risks are categorized as “credit risk”, “market risk” and “operational risk”. It was mentioned that credit risk is caused by the risk of loss due to an obligator’s non-payment of a commitment in terms of loans or other lines of credit (Basel II, 2006; Sankareswari, 2012; Sophia, 2013).

Credit risk is the likelihood that a borrower would default on a loan, causing a loss of money. Credit risk is basically the possibility that a lender won't get paid the full amount of interest and principal due, which would cause cash flow problems and higher collection expenses. By examining details regarding the creditworthiness of a borrower, such as their earnings and present debt burden, lenders can reduce credit risk (Schmitt, 2023).

Market risk is defined as the possibility of losing money due to trading positions' unpredictability in the four main economic markets: interest-sensitive debt instruments, stocks, currencies, and commodities; it also considers changes in the risk of credit spreads. Banks are subject to variations in the worth or price of tradable securities due to the market turbulence in each of these marketplaces. Market risk assessment applies to all securities that are defined as being available for selling or fair value through profit and loss; however, holdings held to maturity do not require the recognition of changes in valuations in the market (Greuning and Bratanovic, 2020).

A wide range of factors can lead to operational risk events, including as fraud, dishonest corporate procedures, defective products, technological malfunctions, discrimination in employment, terrorism, natural disasters, and transaction and execution mistakes. Therefore, a wide range of materials should be included in operational risk evaluation and management. These avenues should include information about internal corporate shortcomings, clearly defined losses and how these sums are classified, specifics about recovery processes, and more precise definitions of the occurrence's starting and end dates (De Jongh et al. 2013).

1.3 Concept of credit risk

In the past, the risk in banking sectors was considered as credit risk and globally it is seen among the bankers. Generally, it occurs due to the risk of defaulting on loans. As per the nature of the business, the credit risk is faced by the banking sectors, which may be inherited in nature. Credit and its related risk are defined in the article by Sophia (2013), which is “credit is nothing but the expectation of a sum of money within some limited time” and “credit risk is the possibility in which the expectation will not be fulfilled”. The majority of credit risk has been created by lending (Sophia, 2013).

In the recent era, credit risk has become more prevalent. It was reported that when an organization borrows to make a purchase and develops itself may use credit to fulfil the purposes then credit risk has occurred unable to return the values (Sophia, 2013). When the credit risk exposure is continued then banking sectors have faced the leading source of problems. Punjab National Bank in India faced financial difficulties due to poor risk management and fraud. The scam is based on focuses on fraudulent activity at one of its branches. The scam ran from 2010 to 2017 and it was declared in February 2018 (Dogra and Dogra, 2019).

Swanson (2017) gives different forms of credit risk as follows:

1. As per lending – It is related to non-paying of principal and/or interest amount.

2. As per guarantees or letter of credit –It is related to funds, which are not prospective from the constituent upon crystallization of the liability.

3. As per treasury operation –It is related to the payment or series of payments, which are showing dues from counterparties within the respective contracts and are not prospective or ceases.

4. As per the securities trading business, funds/securities settlement, which is not affected.

5. As per cross-border exposure –It is related to funds availability and free transfer of foreign currency, which are ceasing or restricting and imposed by the sovereign.

1.4 Factors related to credit risk in banks

According to Lokare (2014), the Indian banking sector is accounted for a major part of financial intermediation, which is considered the main passage of “monetary policy transmission”, “credit delivery” and “payment systems”. It was reported that when the banking system maintains stability and proper health a vital role is to the necessity for overall economic development and financial stability. Among several factors, the non-performing assets (NPA) may be the vital prudential indicator in which it is known easily about the financial health of the banking sector. Besides asset quality, NPAs epitomize credit risk management and efficacy in the allocation of resources. There is an agreement in the research articles that asset quality can be a suitable determinant for the proper functioning of the banking system. Moreover, Michael et al.(2006) stated that NPAs are affecting operational efficiency, which may show an impact on the profitability, liquidity, and solvency position of banking sectors. In another study, Batra (2003) indicated the consequences of NPAs which were observed reduction in “interest income”, “high level of provisioning”, “stress on profitability”, and gradual decrease-unable to serve the steady increase in cost, along with rising pressure on net interest margin (NIM), ultimately, reduction of competitiveness, steady loss of capital resources and expanded difficulties in raising capital resources. On the other hand, the asset quality problems can be explained as contagious and lead to weakness in the consumers. In this respect, the contagious nature of loans may create an issue, furthermore, this impact can spread to earnings, capital, and liquidity. The impact on these mentioned features may pose vulnerability and have relatively small amounts of capital to absorb unexpected losses (Lokare, 2014).

It is well established that NPAs form a vicious cycle, which is affected by the sustainability and growth of the banking system, and it should be managed properly to mitigate bank failures. Global research based on empirical studies is indicated a close relationship between bank failures and increased NPAs (Chijoriga 2000; Gopalakrishnan, 2005; Dash and Kabra, 2010; Sankareswari, 2012; Lokare, 2014; Tuo, 2016; Wang, 2019). It was observed that there are close links to financial crises due to bank funding, which may be increased during banking crises. These crises have led to a rise in deteriorating economic fundamentals, ultimately declining asset quality (Borio and Lowe, 2002; Lokare, 2014). According to RBI (2006-2008), it was reported on empirical research, which suggested asset quality as the main determining factor of credit, besides time deposits and lending interest rate. Another significant factor that affects the NPA levels of the banks is whether they prefer retail investors or corporates. One of the primary causes of the rising non-performing asset (NPA) rate is the weak lending standards, particularly for corporate executives whose creditworthiness and financial standing are improperly evaluated. There is a misconception that lending to the priority sector was the primary cause of the increase in non-performing assets (NPAs) in public sector banks. However, the results of the Standing Committee on Finance indicate that NPAs in the corporate sector are significantly larger than those in the priority or farm sectors. Additionally, according to certain measures, a sizable portion of corporate debt has not gone toward high-productivity industries or profitable ventures. In addition to negatively affecting medium-term growth, this tendency has sparked worries about the viability of debt in growing nations like India (Bagadi, 2020).

1.5 Credit risk management in banks

A fact that credit risk occurs when a financial organization's borrower or counterparty fails to fulfil its agreements with agreed terms and conditions. A credit risk leads to a huge loss in the banking sector. To prevent loss in the banking sector, credit risk management is needed on an urgent basis (Sophia, 2013).

Credit risk management (CRM) is an important area for the banking sector with a wider prospect for financial growth (Sophia, 2013). CRM is started with the establishment of good lending principles and an efficient framework to manage the risk in a financial organization. In banking sectors, CRM is regulated by the supervision of the risk management committee where different policies, industry-specific standards, and guidelines along with risk concentration limits are prioritized. These policies, standards, and procedures are also governed regarding credit risk measurement, monitoring, reporting, and overall controlling. It is suggested that a quarterly review is mandatory as per market conditions. Bank executives, policymakers, as well as governments should understand the effect of risk-balancing methods within the banking sectors (Hulinsky, 2015). The credit risk management process consists of measurement of risk by using credit rating, risk pricing, controlling risk through effective loan review mechanism and management of the portfolio and calculating projected loan losses, or the entire quantity of losses on loans a bank would incur over a selected period of time, in order to quantify the risk. The risk can also be mitigated by a comprehensive underwriting process and diversification where concentration risk, a significant amount of unsystematic credit risk, is faced by lenders to a limited number of borrowers (or types of borrowers). Lenders lower this risk by spreading out the group of potential borrowers. Another mitigation measure can be deposit insurance where in order to protect bank savings held by failing banks, several governments set up deposit insurance. In order to prevent a bank run, this protection deters customers from taking money out of the bank and pushes them to keep their savings in the banking system rather than cash (Singh, 2013).

Sankareswari (2012) reported that the banking sectors determined to strengthen the CRM process, it was found with a proposal of Basel I and Basel II accord as per the RBI guidelines to manage different types of risks when banking sectors are trying to avail the CRM in the Indian banking system. Moreover, as per the issued guidelines by RBI, the banking sectors are also concerned with "Risk-Based Supervision (RBS)" and "Risk-Based Internal Audit (RBIA)".

1.6 Risk appetite framework

According to Analyst Prep (2020), the definition of risk appetite is mentioned as "the amount and type of risk that a company is able and willing to accept in pursuit of its business objectives".

In 2009, Hyde and colleagues reported that risk appetite is considered more than an advanced "key performance indicator (KPI)" system for the management of risk. It isa core mechanism to develop overall corporate strategy, capital allocation, and risk. The risk management practices are performed by regulators, rating agencies, and professional investors after forcefully pushing banking sectors.

Fig 1.1 represents the flow diagram of the risk appetite framework as per Hyde et al. (2009) and was used in the present research work. In this framework, five elements were considered.

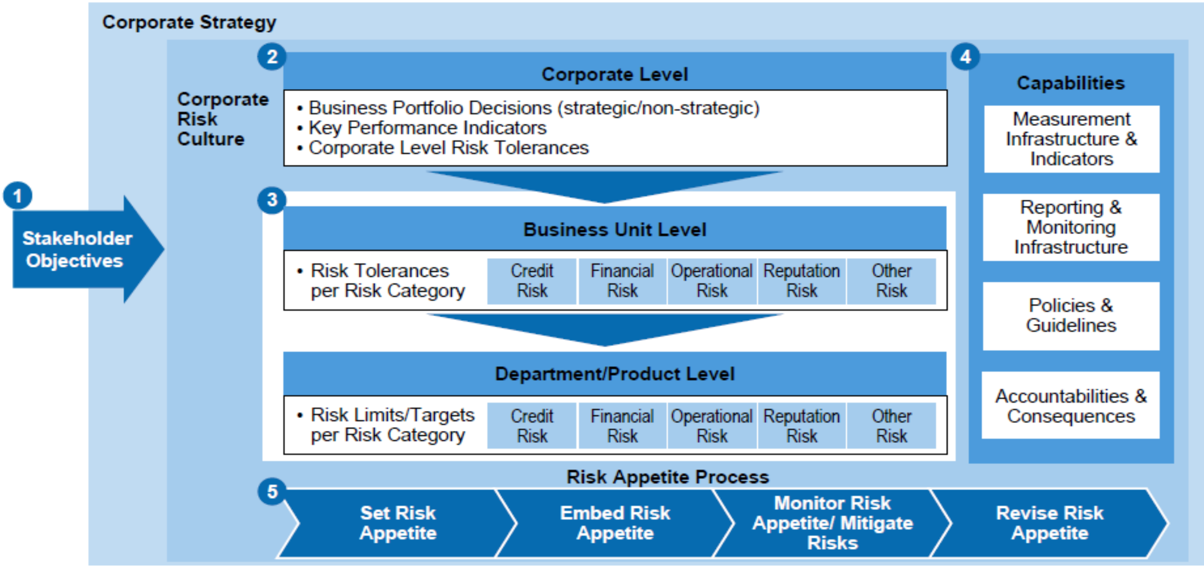


Figure 1.1: Flow diagram representing risk appetite framework (Source: Hyde et al., 2009)

According to Baldan et al. (2014) and (2016), the performance of risk appetite framework (RAF) in banking sectors is based on the quantitative methods and regulations depend largely on a financial organization with some requirements. This framework should be formed on the following criteria:

1. An adequate risk culture should be taken care of.

2. There should be a correlation between a business's risk management and other processes such as planning on strategies, "internal capital adequacy assessment process (ICAAP)", organizational, and control systems, etc.

3. It should be followed Effective methods for internal communication and cooperation.

4. Usage of sufficiently advanced IT systems.

Baldan et al. (2016) studied a quantitative model to articulate the RAF in banking sectors. Their research was based on two main objectives such as risk appetite statement (RAS) to derive a static 'picture' of the banks' risk profile and developed a quantitative method that implemented the RAF in the banking sectors.

Interestingly, RAFs have confirmed to know better risk awareness among all organizational sectors and increased understanding of internal risk profiles. According to the International Association of Credit Portfolio Managers ("IACPM") and PricewaterhouseCoopers LLP ("PwC") (2014), the study of RAF is an encouraging development for any organization in which recent supervisory expectations may be increasingly stimulated by policy related to making strong the financial organization.

The RAF is observed advantageous in the banking sectors and this helps indirect benefits to the organizations. However, after implementing RAF, it may include a lower cost of capital or decreased regulatory screening as the risk profile for the organization, which found a decreasing trend (Milliman, 2011).

The risk appetite framework consists of the following components:

- **Risk appetite statement:** The risk appetite statement is a written document that includes a description of the key facts about the risks. This declaration acts as a management tool to direct executives and staff members regarding the level of risk that the organization is willing to assume in order to pursue important goals and objectives.
- **Risk capacity:** Risk capacity, as opposed to risk tolerance, is a required statistic that characterizes two circumstances. It first describes the level of risk that an organization must assume to meet its goals. Secondly, it illustrates the highest degree of risk that the company may take on prior to a strain on its existing assets and cash flow.
- **Structure of corporate governance and responsibilities:** The staff personnel in charge of putting the risk appetite framework into practice and keeping an eye on it are outlined in the outline. To do this, some businesses employ risk analysts, risk managers,

or a group of experts in risk management. If an organization does not have these positions, a leadership figure—such as a security manager—usually assumes the duty (Indeed, 2022).

1.7 Thesis Overview and Objectives

In any financial organization especially the banking sector, credit risk is a matter of great concern. In India, several past studies regarding credit risk have been found due to the risk of defaulting on loans. As per the nature of the business, the credit risk is faced by the banking sectors, which may be inherited in nature.

In the recent era, credit risk has been found to be more prevalent. It was reported that when an organization borrows from making a purchase and developing themselves and may use credit to fulfil the purposes then credit risk has occurred unable to return the values within the stipulated time. When the credit risk exposure is continued then banking sectors have faced major losses.

In this regard, credit risk management (CRM) is a necessary recommendation for the banking sector concerning financial growth (Sophia, 2013). CRM is maintained based on lending principles and an efficient framework to manage the risk in a financial organization. In banking sectors, CRM is regulated by the supervision of the risk management committee where different policies, industry-specific standards, and guidelines along with risk concentration limits are prioritized. These policies, standards, and procedures are also governed regarding credit risk measurement, monitoring, reporting, and overall controlling. It is suggested that a quarterly review is mandatory as per market conditions. Bank executives, policymakers, as well as governments should understand the effect of risk-balancing methods within the banking sectors (Hulinsky, 2015).

According to Analyst Prep (2020), the definition of risk appetite is mentioned as “the amount and type of risk that an organization is able and willing to accept in pursuit of its business objectives”. For the management of risk, the risk appetite is known as the “key performance indicator (KPI)” system (Hyde et al., 2009).

There is no doubt that the successful studies on the design and development of robust credit risk management and risk appetite framework for the Indian banking sectors can be suitable as a steppingstone and this research was performed as a benchmark for researchers, academicians, regulatory authorities, bankers, etc. concerning the prevention of financial loss in the banking sectors.

Few studies have tried to conduct a “deep-dive” into the impact of weak risk management practices in public sector banks. There is no definitive research available that has covered the impact of banks’ mandate for inclusive growth which requires Banks to increase their fund/non-fund-based credit limits into the priority sector to 40 by 2019.

1.8 Objectives of the Research

Aim:

To identify the relationship between macroeconomic and Key Internal Variables and how these impact the Non-Performing Assets (GNPA/NNPA) of the studied banks and their usage to set risk appetite framework through machine learning models.

1. To analyse the overall bank-wise and year-wise trend of credit risk factors in select banks of India.

2. To analyse the impact of bank-related factors and macroeconomic variables on the assessment of credit risk measured by GNPA and NNPA.

3. To use Machine Learning techniques and explore the possibility of refining the accuracy and prediction through advanced Machine Learning Algorithms showing the impact of Key Internal Variables of Banks and Macro-Economic Variables.

4. To assess the accuracy of predicting GNPA/NNPA of banks through standard Multi-variate Statistical Techniques and quantify the impact of Key Internal Variables of Banks and Macro-Economic Variables.

1.9 Research Questions

After an extensive review of the literature, particularly recent research articles on credit risk assessment and management practice in the banking sectors following questions developed in the inquisitive mind.

1. What are the factors that play an important in determining credit risk in public sector and private banks?
2. What are the key variables that should be considered covering macro-economic conditions?

1.10 Thesis Chapters

Chapter – 1 describes a brief introduction in which the concept of the Indian banking system, risk factors, credit risk, management of credit risk, risk appetite framework and machine learning approach for credit risk assessment and management practice was studied in the banking sectors.

Chapter – 2 describes a thorough review of different literature published in national and international articles, RBI reports, related websites, etc. based on possible risk factors, credit risk management, risk appetite framework in the banking sectors, and machine learning approach for credit risk assessment and management practice. Moreover, research gaps were identified from the literature to help further to formulate the objectives and study design as the methodology of the present study.

Chapter – 3 elaborates on the methodologies used in the present research. This section describes in detail samples, data collection, research framework, research design, and statistical methods along with machine learning algorithms used in the study. The hypotheses formulation is done as per each objective and mentioned in this section.

Chapter – 4 describes the empirical framework and is designed for analysing the risk management practices to analyse and design risk appetite framework. In this section, various statistical methods are used to finally assess the correlation between dependent variables (asset quality) to independent variables (macro-economic variables and other banking parameters). Moreover, the predictive results of machine learning algorithm modelling especially Decision Tree, Random Forest, Secure Multi-Party Computation (SMC), and Naive Bayes were done to know the banking operational performance accuracy on overall credit risk assessment and management.

Chapter – 5 summarizes the present findings as concluding remarks and along with limitations of the present study. It also proposes a future research direction for this present study.

CHAPTER 2

INDIAN BANKING SECTOR: AN OVERVIEW

2.1 Introduction

The Indian banking sectors account for the major portion of financial intermediation and are the vital passage of monetary policy transformation, credit delivery, and payment systems. Academic research is based on credit risk and credit risk management in the banking sectors of India more important and relevant. In other words, banks are business organizations, that distribute money, and financial instruments, and also provide different financial services with some returns in terms of interest, discount, commission, and fees (Boateng, 2020).

The risk factors based on monetary lending are found in any financial organization (Srinivas, and George, 2020). Concerning the risk after credit as a credit risk in the banking sectors, several factors such as credit deposit ratio, capital adequacy ratio, non-performing loans/assets, return on assets, branch managers' perception, etc. should be maintained properly otherwise there is the potential loss on economic condition (Das and Ghosh, 2007). Moreover, in Indian banks, it is also important to know the credit efficiency in terms of banking lending behaviour, efficiency, and risk-bearing capacity (Sankareswari, 2012; Bittu and Dwivedi, 2012). It was also noted that macroeconomic as well as microeconomic development can reduce the credit risk for the improvement of financial health in a nation’s economy (Sah and Dwivedi, 2012; Gulati et al., 2019).

In banking sectors especially for public, private, and foreign banks whether located in India or abroad need to study credit risk management (CRM) for the prevention of financial loss and growth of the economy. However, the ineffectiveness of the CRM may face several difficulties in the banking sector, which is a predominant part of the main reasons behind financial loss (Boateng, 2020).

Levine (1996), has revealed that the efficacy of financial intermediation can also disturb economic growth. The health of the financial sector is a matter of policy concern, most especially in developing countries where failure in financial intermediation can critically disrupt the development process. Crucially, financial intermediation affects the net return on savings and the gross return for investment. The relationship between the financial sector and growth has been the subject of a large literature, which is analysed by Khan et al. (2001), concluded that although there is strong empirical evidence that strong financial markets support economic growth, there is little operationally relevant effort to improve the functioning of the financial sector.

Past research was revealed by Levine (1996) regarding the efficacy of financial intermediation, which affected economic growth. As per the survey, it was established that the health of the financial organization depends on policy concerns among developing countries where failure in financial intermediation is found to disrupt the development process. However, financial intermediation may affect the net return on savings, and the gross return for investment. For these reasons, it is necessary to identify the link between the activities of the banking sectors and the growth of the economy.

The "Bank for International Settlement (BIS)" is an organization that encourages "International Monetary and Financial Cooperation" and provides services like a bank for central banks. It was founded in 1930 and it is one of the world's oldest international financial organizations located in Basel-Switzerland. On the other hand, the Basel Committee on Banking Supervision (BCBS), supported by the BIS and formed as a committee through participation by the central bank governors among a group of Ten (G-10) countries such as "Belgium", "Canada", "France", "Germany", "Italy", "Japan", "Luxemburg", "Netherlands", "Spain", "Sweden", "Switzerland", "United Kingdom", and "The United States", established at the end of 1947 due to serious hazards in international currency and banking markets as per noticeable failure of

Bank Herstatt, West Germany (Sankareswari, 2012). Later, the Basel Accords referred to the banking supervision accords viz. Basel I and Basel II, which were issued by the BCBS.

In the year of 1988, the recommendations and documentation were made for the Basel Capital Accord or Basel I Accord and provided to the central banks of all the countries to implement in their national system. It was provided a level in which stipulated the monetary needs for capital, which must be maintained by active banks in an international context (Sankareswari, 2012). Sankareswari, (2012) mentioned about an indication to the Basel report some developments found in India are as follows:

a) In 1991, the “Narasimham Committee” was established to study the Basel I recommendations, which implemented in India as per recommendation by the Government of India (India. Committee on the Financial System; M. Narasimham, 1992).

b) Narasimham Committee provided its recommendations in Part I and Part II. Part I recommended implementing an urgent basis and Part II was developed after legislative measurement to implement.

c) The initiative was towards Basel I preparedness, part I recommendation implemented for “Capital Adequacy norms” according to the “Basel Norms” and stipulated “Capital Adequacy Ratio (CAR) at 9”.

d) Further recommendation was for the implementation of “Income Recognition and Assets Classification/provisions norms (IRAC)”, which should be described the transparency in bank balance sheet along with profit and loss accounts.

e) It was also reported that banks should follow strictly the prudential norms related to identified NPAs and make provisions thereof.

f) It was emphasized on weak banking sectors as per the classification of weak banks. The good banking system in which the majority of them, systems are strong. However, it was mentioned that every system may pose of weak links and the banking systems are also falling within the same. So, Part I criteria are recommended for these weak banks.

g) Moreover, it was observed that part I criteria also recommended for the reduction of "Statutory Liquidity Ratio (SLR)" and "Cash Reserve Ratio (CRR)", "Interest rate on CRR balances". In this regard, "Creation of Asset Reconstruction Funds for bad loans", "Creation of Loan Recovery Tribunals", "Reconstruction of banks", more licenses proposed to promote private sector banks, etc.

The process of banking has changed after the introduction of the Basel I document. To develop the risk management strategies along with the increasing trend of the complexity of financial activities/ or instruments handling such as options, hybrid securities, etc. proposed international supervisors for reviewing the relevance to know "regulatory capital standards under Basel I". Finally, the Basel II Accord was established to fulfil this requirement after amending and refining the Basel I Accord (Sankareswari, 2012).

The salient features of the Basel II documentation are as follows:

1. It promoted safety and soundness in the financial system through better risk assessment and capital allocation.
2. It aligned the regulatory capital to know the principal risk and induced banks to strengthen their risk management capacities.
3. It proposed incentives for increasing risk management capacities. Good risk management should be rewarded through lesser capital needs.
4. It ensured the level of the playing field for all banking sectors worldwide. Basel II provides a framework for the banks' position, which may be a benchmark to make best practices worldwide.

In the Indian context, according to RBI guidelines, Basel II norms are mainly applicable to commercial banks along with Foreign Banks operating in India but not applicable to local area banking sectors and regional rural banking sector (Basel II, 2006).

Jayadev (2013) described Basel III norms in which it was mentioned that strengthen global capital and liquidity regulations to promote a more resilient banking sector, and to improve the banking sector's ability need to absorb shocks arising from financial and economic stress, which are the main objectives to improve the capital regulation framework in Indian banks.

Several studies on credit risk and CRM have been done in national and international banking sectors. Moreover, the risk appetite framework (RAF) is an important mechanism under CRM,

2.2 Credit risk in banks

In India, the banking sector is considered robust and resilient at present but at the same time, banks are likely to get impacted with changes in macro-economic environment globally and locally, geo-political crisis, emergency situations such as pandemic etc. According to Carey (2001), financial organizations must take a risk, but they must perform consciously. However, it should be noted in mind that banks' rely on customers' ability to repay. When something goes wrong, banks may collapse, and the failure of each bank is sufficient to convey shock waves directly to the nation's economy (Rajadhyaksha, 2004). Therefore, banking sectors must identify the type and the degree of risk exposure followed by the care of those effectively.
In earlier studies, most studies have primarily investigated the US banking organizations (Berger and DeYoung, 1997; Kwan and Eisenbis, 1997), and few of these works in the Mexican banking sectors (Gonzalez-Hermosillo et al., 1997).

The studies during the periods of 1985-94 by using data on US banking sectors, Berger and DeYoung (1997) reported a decreasing trend in cost efficiency due to the increasing trend of in nonperforming loans while Kwan and Eisenbis (1997) demonstrated that inefficient banks are more prone to express the risk.

Interesting research, Gonzalez-Hermosillo et al. (1997) demonstrated that the increased level of non-performing loans to total loans led to a greater chance of failure in the banking system under the Mexican banking sector whereas Salas and Saurina (2002) indicated that, credit risk was significantly influenced by each banking-level variables after regulating macroeconomic conditions in the Spanish banking sector.

Besides these studies, in 1988, the World Bank in the banking sector at Ghana reported that the Ghanaian banking sectors were facing an increased level of operating costs due to inefficiency, and a higher level of non-performing loan documents, indicating insolvency in the banking sectors, which formed inflated profits and capital inadequacy (World Bank report, 1988).

Rajaraman et al. (1999) investigated the Indian commercial banking sectors for net NPA variations and mentioned the non-performing loan problem due to macro and microeconomic variables.

In the present section, the compilation of different opinions from several studies concerning Indian banking sectors are as follows:
According to Taori (2000), it was observed that the study of non-performing assets is important to monitor credit in an efficient way and risk management system in the Indian banking sectors.

Bhattacharya (2001) reported that due to increasing rate regulation, quality borrowers were found to switch over to other avenues like capital markets, and internal assemblage for requiring funds. For these reasons, banking sectors have found no option in which the organization diluted the quality of borrowers and there could be a high chance for the generation of NPAs.

Kaveri (2001) recommended strategies to decrease the extent of NPAs after studying the NPAs of several banks in India. The author suggests that credit ought to be improving in light of the sharp increase in new non-performing advances. To stop NPAs, the RBI should implement certain new regulations and techniques.

Swamy (2001) attempted to identify the comparative performance related to the share of rural branches, average branch size, trends in the bank's profitability, the share of public sector assets, the share of wages in expenditure, provision, and contingencies, net non-performing assets in net advances, spread in different bank groups in India and concluded that nationalized public sectors banks performed better than private sector banks, even they are much better than foreign banks.

Reddy (2002) mentioned in an article that the problem of NPA is based on a strong legal and legislative framework in which it was understood the macroeconomic variables and systematic issues related to banking sectors can solve the economic conditions.

Another study by Muniappan (2002), it was documented that NPAs have two elements such as the "overhang component" and the "incremental component". The former component is observed due to infirmities in the structural and organizational environment while the latter component is formed by factors related internally in the case of banking sector management and culture of the credit.

In 2003, Raghavan reported in detail that there are three main categories of risk such as "credit risk", "market risk", and "operational risk" in the Indian banking sectors. Moreover, he reported that credit risk is based on two elements; one is risk quantitatively in which observed outstanding loan balance as on the date of defaulting and the other is risk qualitatively which observed severity of the loss. The instruments and tools viz. exposure ceilings, review/renewal, risk rating model, risk-based scientific pricing, portfolio management, and loan review mechanism, which may help in the management of credit risk.

Another work by Misra (2003) identified that the higher rising gross and net NPAs of the Indian banking sectors may lead to recession with a heavy toll on the corporate credit segment and is further mitigated by climate recovery, legal process, initiation of lending towards the lenders and other related factors.

Malydri and Sirisha (2003) documented that the loan assets constituted a real economic cost since they reflected the application of scarce capital and credit funds for unproductive use.

In an empirical study, Das and Ghosh (2003) examined non-performing loans related to different indicators viz. asset size, credit growth and macroeconomic condition, and operating efficiency indicators for public sector banks of India. They concluded that banks should propose a device related to lending terms considering the cost of credit, funds, the maturity of loans, and credit orientation, which may increase the lower value of defaults on borrowers.

In 2004, Reddy and Bhargavi stated that NPA accumulation decreased the income level along with incurring higher expenditure for maintaining of poor-quality assets in the books in banking sectors. Besides the internal and external entities, increasing NPAs have directly affected the profitability of banks, which may hamper their existence.

Valasmma (2004) suggested that proper presentations should be done for the exchanging reports of credit information related to the defaulting customers within the banking sectors, which is an informative measure for further NPAs about other banking sectors regarding the prevention of loss.

In another work by Reddy (2004), it was revealed that the policy on lending crucially considered the influencing of non-performing loans for Indian banking sectors. He examined critically several issues related to the credit of Indian banks and argued that "the element of power has no bearing on the illegal activity. Default is not entirely an irrational decision rather a defaulter takes into the account probabilistic assessment of various costs and benefits of his decision".

Since 1990, the RBI, the Government of India (GoI), and financial institutions have been keeping an eye on the problem of "wilful default" by numerous borrowers. The regulator has been working nonstop to improve the reporting process and release the identities of SF_WDs. This is crucial to warn and protect other financial institutions from giving those who default further funds (Karthik et al., 2018). In India, wilful default is still not regarded as a crime. Big and small-time fraudsters alike abuse these evasion techniques to their detriment and carry out nefarious acts without facing legal repercussions (Singh et al., 2016).

Singh and Khan (2005) observed that "Debts Recovery Tribunals (DRTs)" are effective for recovering bank payments or dues to a certain extent. This study concluded that the establishment of the DRT Act suitable for changes in recovery suits in the banking sectors.

Naik (2006) reported that approaching Basel II norms along with different international best practices in the Indian banking sector is expected to efficiently manage NPAs. It was also known that some reforms in the legal system especially the SARFAESI Act, of 2002 could help to recover faster the NPAs.

Pal and Malik (2007) studied an empirical analysis of the different factors viz. profitability, liquidity, risk, and efficiency related to financial features of public, private, and foreign banking sectors in India. Multinomial regression analysis indicated that foreign banks were observed better performers related to business generations per the level of resources, and also wellequipped with managerial practices related to skills and technology as well as for net interest margin. On the other hand, this study reported that the public banking sector emerged as the next good performer compared to foreign banks due to providing a higher return on equity compared to foreign and private banking sectors.

Karunakar et al. (2008) pointed out that the NPAs have a detrimental effect on the return on assets in different ways such as decreasing the interest income of banks accounted only based on receipt basis, adversely affecting banks' profits due to giving of doubtful debts, which consequently determined as bad debts, reducing the return on investment (ROI), NPAs due to the capital adequacy ratio when entering into its calculation, increasing the cost of capital, mismatching the assets-liability, the "economic value added (EVA)" by banks developing equal to the net operating profit minus cost of capital, NPAs need to provisioning requirements for profits and formation to capital funds, etc.

Uppal (2009) indicated that the load of NPAs increased in public sector banking groups whereas the lower value was obtained in foreign banking sectors because advances were created majorly by public sector banking groups and this work reported an increase due to the operation of the agricultural sector within both public and private sector banking groups.

Mishra and Dhal (2010) analyzed of pro-cyclicality of bank indicators mainly focused on the non-performing loans of public sector banks in India. Their empirical analysis demonstrated that banks "non-performing assets are influenced by three major sets of factors such as terms of credit, bank-specific indicators relating to asset size, credit orientation, financial innovations or non-interest income, and regulatory capital requirement and the business cycle shocks.

In 2010, Muniswaran examined the profitability and viability of Indian banks, especially state bank groups, and found an impact directly due to the quality and performance of advances.

This study suggested that a sound NPA management system can be used for quick identification of non-performing advances, their containment at a low level, and ensured that their impingement on the finance is the least valuable.

Malyadri and Sirisha (2011) investigated the aspect of the NPAs in the public and private banking sectors in India among weaker groups for the period of 2004-2010. The results indicated that the declining ratio of NPAs, which improved the asset quality of Indian public sector banks and private sector banks, and the management of NPAs was better in public banking sectors. They concluded that the weaker groups of public banking sectors achieved a higher quality compared to the private banking sectors. However, as of 2022, at both private and public sector banks, the percentage of non-performing assets (NPAs) to total loans has been trending upward. Private banks are still superior to their public sector counterparts in managing their non-performing assets (NPAs), but they are unable to lower this ratio (Agarwal and Preeti, 2022).

Rajput et al. (2011) mentioned risk in Indian banking sectors due to higher NPAs, which can severely affect the economy in several ways. They suggested if NPAs are not properly managed then a high chance of occurrence of financial and economic degradation which in turn indicates an adverse investment environment.

Sankareswari (2012) studied the magnitude of NPAs, capital adequacy ratio, loan asset quality, etc. of four banking groups viz. public sector, private Sector and foreign banks in India. This study also examined the credit risk management system regarding the recovery of NPAs by SCBs through various measures recommended by RBI and studied the implication of NPAs on selected macroeconomic and micro-banking variables. It was concluded that CRM is done through improved loan quality and reduced the level of NPAs.

In a study by Bahtia et al. (2012), it was observed that the factors for the measure of bank profitability in the private sector banks of India. This study used return on asset (ROA) as a dependent variable while using independent variables such as credit/deposit ratio, provisions, and contingencies, non-interest income, spread ratio, operating expense ratio, profit per employee, business per employee, investment/deposit ratio, capital adequacy ratio, non-performing assets and category of bank outcome respectively. The findings indicated that positive relationship between ROA and spread ratio, credit-deposit ratio, and profit per employee, business per employee, capital adequacy ratio, and non-interest income as important profit measurements but ROA and investment to deposit ratio, non-performing assets, operating expenses, and provisions and contingencies negative influence was observed in the private sector banks of India.

Arora (2013) identified factors such as creditworthiness analysis and collateral requirements, which contributed to the credit risk measurement, and this study was compared to public and private sector banking sectors in Indore, India. In this study, primary data were collected from employees of studied banking sectors. This study concluded that the banking sectors of India effectively manage credit risks, but old and new private banking sectors were significantly better than public banking sectors.

Rawlin et al. (2014) studied and compared the determinants of profitability of India’s largest public and private banking sectors. They studied the determinant variables viz. bank metrics, risk factors, and productivity measures, and the net profit was the dependent variable. The results indicated a strong positive correlation between the critical bank metrics deposit, advances, total assets and bank size, and net profit for both banking sectors. Another positive correlation has also resulted between the risk factors capital adequacy ratio, gross and net NPAs, and net profit. It was also reported in their study with a strong correlation between productivity measures and business, profit per employee, and the net profit only for the public sector bank. It also found a significant positive correlation between asset usage, efficiency measures, return on assets, interest income, non-interest income, and operating profit related to the average working funds with net profit for the private sector bank but negatively correlated to the public sector bank.

Nag (2015) designed a study in which a comparison was made between the nationalized, private, and foreign banking sectors of India and the performance of NPAs in these banking sectors. This study also mentioned that the good health of the banking sector is mainly obtained through a good return on assets. This study concluded that nationalized and foreign banking sectors faced major problems in the case of NPAs compared to the private banking sector.

Makkar (2016) studied the least square model for analysing data from 47 commercial banks. This was mentioned that NPAs to total advance ratio was the proxy for credit risk. In this study, the researcher used the credit-risk determinant factors such as bank size, return on assets, advances to deposit ratio, credit to deposit ratio, operating profit to asset ratio, and priority sector lending to total assets ratio. The findings revealed that the bank size, return on assets, credit creation, and operating activities observed a significant adverse impact on the credit risk of commercial banks in India. It was also reported that priority sector lending had no impact on credit risk in the same banking sectors.

Kedia (2016) determined the factors of profitability such as operating expenses, NPA, net interest income, and credit-to-deposit ratio as independent variables and net profit as the dependent variable of public banking sectors of India. The results on multiple regression indicated that interest income and credit-to-deposit ratio correlated positively to net profit with a significant level while NPA also correlated positively to net profit without a significant level.

On the other hand, operating expenses, show a negative correlation and without a significant level of net profit. This study concluded that the credit deposit ratio and net interest income were affected by the net profit of the public banking sectors of India.

Bhullar and Gupta (2017) performed an empirical study of the determinants of profitability of public banking sectors in India. The profitability measure was carried out by using net profit to total funds. The independent variables such as profit before provision over total funds, investment, credit, cash to deposit ratios, other income to total income, interest income to total fund, and interest expended over interest earned were selected. The findings were observed with a significant impact on other income to total income, profit before provision to total fund, credit to deposit ratio, interest income to total fund, and cash to deposit ratio related to the profitability in the public banking sectors of India. Investments to deposit ratio and interest expended to interest earned were found without significant impact on the profitability of the same banking sectors.

Dastidar and Sarkar (2017) conducted a study on the determinants of profitability ratio such as net interest income to total income, return on assets, and return on equity of private banking sectors in India. They categorized the private banking sectors into new and old for the study of the impact of bank-specific profits. The bank-specific features were used viz. liquidity, asset quality, financial soundness, and management efficiency. The external factors viz. inflation, interest rate, and political instability were also selected. The findings revealed that all four bank-specific variables were correlated positively to profitability. Also, both GDP and Inflation were positive without a significant level of profitability within new private sector banks. Interestingly, inflation negatively correlated with profitability at a significant level in old private banking sectors. Post-crisis is positively correlated without significance to profitability in new private sector banks. About bank-specific factors except for soundness were positive and significantly correlated to profitability in old private banking sectors. GDP had a significant positive impact on profitability while inflation had without significant effect. The financial crisis negatively correlated with the profitability in old private banking sectors.

Rajput and Goyal(2019) mentioned in an article that the RBI developed the "Banking Stability Map", which was published in the "Financial Stability Report in 2010". As per the stability map, their study measured the five dimensions as "soundness(s)", "asset quality (Q)", "profitability (P)", "liquidity (L)" and "efficiency I" in different banking sectors like public, private and foreign banks. They concluded that soundness related to capital adequacy ratio and leverage ratio improved during the years 2015-2018 due to the implementation of a capital conservation buffer and decreased credit growth. The asset quality and profitability deteriorated due to the fall of NPAs and returns on assets during the study period. The liquidity was found satisfactory except for 2016 due to demonetization. Overall efficiency such as staff expenses, business per employee, and cost to income was obtained as a satisfactory result. In another work indicated that the gross NPA ratio has found an increasing trend from 8.29 in 2009 to 25.01 in 2013 with relative fluctuations during 2014-18 in IDBI Bank of India. Other parameters like doubtful debts were also found to increase as against sub-standard assets, which justified not favourable conditions for this bank (Hoorunnisa and Reddy, 2019). Zheng et al. (2018), discovered that while net interest margin and inefficiency have a favourable impact on bank credit risk, profitability, capital, and bank size have an adverse relationship. Additionally, the accurate model is constructed by adding each variable in turn, taking into account the variance and goodness of fit values in the corresponding model. Turan (2016) outlined that as long as a consumer has credit, their credit risk will increase. For this reason, a credit rating system is an effective way to assess the credit that clients are granted. One way to think of risk scoring is as a tool for risk management. The risk associated with a customer's credit lowers as it is scored. For data mining to be effective, client data needs to be trustworthy. In the event if not, the model produces inaccurate scoring results. Program tools can be used to conduct risks. Designing credit risks should take into account external risks. Managers should take into account all probability.

2.3 Credit risk management in Indian banks

Muninarayanappa and Nirmala (2004) reported on the concept of CRM in Indian banks. They determined different factors related to the direction of bank policies on CRM. They highlighted the challenges concerning internal and external factors responsible for CRM. Their study concluded that the success of CRM needs the maintenance of a proper credit risk environment, credit strategy, and policies. Finally, it should be protected, and improved loan quality in the Indian banking sector.

Arunkumar and Kotreshwar (2005) investigated and compared public banking sectors and private banking sectors on the trends of NPA level, CRM practices, the response to reforms under the Basel II accord, and risk-based supervision, respectively. Their findings reported a strong correlation between NPA level and credit portfolio diversification.

As per the Global Non-Performing Loans Report (2006), it was mentioned that the “non-performing loan ratio (NPLR) act” is a strong economic indicator. The efficient form of CRM is supported that lower NPLR is related to decreased risk and deposit rate. However, it is also implied that a relatively high deposit rate increased the deposit base in terms of funds posed by high-risk loans and consequently increased the possibility of NPLR in the long run.

Goyal and Agrawal (2010) enlightened the importance of the risk management process along with challenges and opportunities regarding the implementation of Basel II in Indian banking sectors. This study mentioned that the banking industry is continuously exposed to several risks viz. volatility on forex, interest rate variation, market scenario, an operation-based, risk-on credit, etc. may adversely affect the profitability and financial health. They proposed that risk management can easily be done as per four functional attributes such as risk identification, risk measurement or quantification, risk control, as well as monitoring, and reviewing, respectively.

Rajput et al. (2011) highlighted that regarding the management of NPAs in the Indian public banking sector through stringent asset classification norms, the use of the latest technological platform based on Core Banking Solution (CBS), recovery methods, and other bank-specific indicators with the help of stringent regulatory framework of the RBI.

Bittu and Diwedi (2012) mentioned in their empirical analyses that fundamental factors affecting periodic addition to non-performing assets (NPA's or fresh slippage), are taken as a proxy for measuring credit risk in 70 banking sectors of India. Panel regression was performed for 70 banking sectors and investigated variations in ownership dimension, aggressiveness, risk-taking behaviour, and performance of banks. The study indicated that variations in fresh slippage are inversely related to the efficiency of bank performance and directly related to the capital-adequacy ratio, credit risk for foreign banking sectors was found to have higher values compared to new-generation banking sectors. It was reported that the Standard Granger causality test based on quarterly NPA data from bigger public banking sectors indicated macroeconomic factor(s) as the gross domestic product (GDP) obtained a significant impact on CRM among banking sectors, but the direction of causality found from GDP to NPA and no "reverse causation" was obtained.

Chitra and Vani (2014) studied in detail the credit risk management for banking sectors. They mentioned in their report regarding the overall lifecycle of a typical credit risk management process. Besides these, it is emphasized based on the constitution as per a high-level credit policy committee in which the dealing with issues related to credit policies and procedures along with analysis, management, and control of credit risk banking wise. They suggested CRM is mandatory in each banking sector along with great challenges in financial institutions.

Rajeswari (2014) performed a study on CRM in scheduled banks in India. The objectives identified the specific scope of CRM improvement as well as suggested proper training to the bank managers concerning the identification of bad debts, reducing the credit limit on high-risk customers, etc. She reported that several important items such as financial instruments besides loans, including acceptances, interbank transactions, trade finances, foreign exchange transactions, financial futures, swaps, bonds, equities, options, and further extension of commitments and guarantees along with the settlement of transactions should be taken care in case of CRM in banking sectors.

Bhaskar (2014) described the tools and techniques to manage credit risks in his study. The instruments of CRM are mainly examined by the author, which are exposure ceilings, review/renewal, risk rating model, risk-based logical estimating, portfolio the executives, and advance survey system, respectively. The study underlined that the target of risk for the executives isn’t to deny or forestall hazard-taking exercises, but to guarantee those dangers are taken in the light of full information, clear reason, and seeing so it tends to be estimated and unsatisfactory misfortunes are forestalled.

Rathod and Vidyashree (2015) suggested that the banking sector needs to prescribe protocols for risk identification, measurement, and assessment. They found in their study that the ratio of gross non-performing advances (GNPAs)of scheduled commercial banks (SCBs) marginally increased. In this study, Credit risk was found in all banking sectors, and suitable measures were undertaken to recover banks’ NPAs. The study concluded that comparatively, public banking sectors found more non-performing assets, which were unable to maintain the CRM.

Singh (2015) developed solutions for risk management, which are preferred by banking sectors. It was pointed out that successful CRM can only be done through effective data, adequate control of credit provided to borrowers, and supervision of the loan transactions mostly occurred in the banking sectors along with suitable monitoring of other chances arise risks. It was suggested that new technologies for data analysis on risk and, the development of separate modules for risk management, may lead to preventing the risks.

Prakash (2016) examined the process of risk management, emphasized stress to know its importance, and further analysed the occurrence, which may deal with risk. The author reported that the risk management process can include risk identification, assessment, measurement, control, monitoring, and return trade-off. He mentioned that the risk can be understood with an opportunity and threat. He concluded the importance of banking sectors in which adequate capital may support all the expected risks in their business.

Ahsan (2018) studied on the problems and obstacles in credit risk management in Indian public sector banks, which evaluated that there were various grey areas in credit risk systems and proposed immediate attention and action if a reduction of the bank's NPAs was to be achieved. This study noted that these grey areas lead to a weak credit risk system that resulted in high NPLs and consequently low profitability or bank performance. The research also found out that the level of NPLs has intervening characteristics on the relationship between credit risk management and banking performance. Therefore, an increase in NPLs led to the deterioration of firms' balance sheets, which precipitated poor banking performance.

Das and Palit (2019) explained the various aspects such as policy and strategy, organized structure, and operations, concerning credit risk management in the banking sector of India as a simple and lucid technique. They emphasized an objective for the management of credit risk and proposed a greater value of the bank's risk-adjusted rate of return after considering credit risk exposure within acceptable parameters. They suggested following the RBI rules and the formation of an efficient credit risk management system for obtaining long-term success and prevention of banking loss.

Brahmaiah (2022) conducted an empirical study to examine the risk management techniques and practices of credit risk management in Indian commercial banks for the period from 2021-2017 to 2020-2021. Another objective was to compare risk management practices between the public sector banks (PSBs) and private sector of banks (PVBs). This study used a sample of 12 banking sectors consisting of 6 largest PSBs and 6 largest PVBs. The sample accounts for 78% of the banking business of the country. The study finds that the scheduled commercial banks (SCBs) are facing credit risk, market risk, and operational risk. This study also finds that the credit risk management process and practices include risk identification, risk assessment, risk analysis, risk evaluation, risk monitoring, and risk control. It was recorded that PVBs have better credit risk management practices in comparison with PSBs. The PSBs have more NPAs compared to PVBs whereas PVBs have better asset quality and better profitability ratios than PSBs during the study period.

The Indian government passed the Insolvency and Bankruptcy Code (IBC) 2016 in May 2016 with the goal of offering a quick and effective remedy. A firm has 270 days to "resolve or liquidate" under the terms of the code, which include an extension of 90 days and a 180-day time limit. The National Company Law Tribunal (NCLT) was established on June 1st, 2016, following the passage of the IBC resolution in Parliament in May of that same year. At the end of 2016, the NCLT began accepting cases pertaining to insolvency and bankruptcy. At the moment, it functions through 11 benches. The Insolvency and Bankruptcy Board of India (IBBI), which was founded on October 1st, 2016, is the organization under which IBC and NCLT are operated (Chandani et al., 2019).

2.4 Credit risk management in banks globally

Fatemi and Fooladi (2006), look into how the biggest financial firms in the US currently manage credit risk. It was discovered that the most crucial function performed by the risk assessment models in use is to identify partner risk of default. Models that can handle counterparty transfer risk are employed by nearly half of the participating banks. Remarkably, the majority of banks do not currently monitor their credit risk using either a vendor-marketed or proprietary approach. It's interesting to note that companies that use their own internal model also employ one that is promoted by a vendor.

Aburime (2006) examined profitability in the banking sectors of Nigeria by using a panel data set comprised of 33 banks. The results of this study indicated that capital size, size of the credit portfolio, and extension of ownership concentration were found significant for company-level determinants in the case of bank profitability. It was mentioned that the size of deposit liabilities, labour productivity, ownership, control-ownership disparity, and structural affiliation were not significant but in conclusion, no relationship was obtained between bank risk and profitability.

Guisse (2012) conducted a study on the performance of the local banking sectors of Malaysia as well as other foreign banks and compared their financial profitability. This study mentioned about the profitability of commercial banks by describing several factors viz. liquidity, asset quality, capital, operating expenses, and the size of the banks. On the other hand, profitability has measured the relation to return on asset (ROA) and return on equity (ROE). The comparison study between the two categories of banking sectors indicated foreign banking sectors are more profitable than domestic banking sectors.

Haneef (2012) studied financial performance in the banking sectors of Pakistan. This study was done with performance indicators such as Return on Assets (ROA), Return on Equity (ROE), and Dividend Pay-out that reflected the liquidity and profitability of the studied banks. The findings by using the Hausman Test for domestic and foreign banking sectors indicated domestic banks operating in Pakistan performed better than foreign banks.

Petkovski and Kjosevski (2014) performed an empirical analysis by using a generalized method of moments (GMM) dynamic panel method in the banking sectors of 16 Central and Southeast European countries such as Albania, Belarus, Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Macedonia, Moldova, Poland, Romania, Slovak Republic, Slovenia, and Ukraine, respectively. They measured the financial development in the banking sector through the study of the bank credit to the private sector, interest rates, and ratio of quasi money (RQM). Their research showed that credit to the private sector and interest margin were negatively correlated to economic growth but RQM was positively correlated to economic growth.

Zou and Li (2014), examined whether there is a connection between European commercial banks' profitability and their handling of credit risk. The results show that credit risk management does increase commercial banks' profits. When comparing the two credit risk management proxies, ROE and ROA are significantly impacted by NPLR and insignificantly impacted by CAR. However, the correlations between all the proxies are not stable; rather, they fluctuate between 2007 and 2012. Supervisors may be required to increase non-performing loan rates in order to support banks' more efficient operations, given the negative correlation between certain loan types and profitability measures.

Hulinsky (2015) evaluated the effects of risk balancing in the banking sectors of the Northern Great Plains region of the USA. A panel model was used to identify the effects of both business risk and financial risk of over 870 banking sectors in the studied area. The global financial crisis and bank policies were considered in this study. The results indicated that the risk balancing hypothesis was true in the banking sectors. This study is suggested for both bank managers and policymakers in connection to efficient policy design and these policies may be reduced.

Roberts (2015) studied risk management in the UK banking sector. The research resulted in three prominent approaches to managing risks such as a perception of risk as a measurable and quantitative construct, managers and other executives risk management groups from Glass Bank were found strong believers in risk philosophy and hence, the social subject especially agent was consciously separated from the system and the rule and protocols were viewed as being separated from the agents, which can be implemented and reproduced system.

Konovalova et al. (2016) proposed a model for credit risk assessment based on factor analysis of retail clients/borrowers concerning confirming the predictive control of the level of risk posed by potential consumers in commercial banking sectors during the post-lending situation. The study also determined the level of risk obtained by different groups of retail consumers or borrowers concerning reducing and preventing credit risk for the future along with the improvement of banking risks through management practices. The findings indicated that the creation of a model of borrowers' internal credit ratings and the development of improved methods found suitable credit risk management in commercial banks.

Cucinelli et al., (2019) examine the precision with which internal rating-based (IRB) models quantify credit risk in European banks. Utilizing a unique data panel set of 177 Western European banks that were tracked in the wake of the financial and economic crisis between 2008 and 2015, the empirical research is conducted. We discover that IRB banks performed superior to banks using the standardized strategy in reducing the rise in credit risk caused by the economic recession. This shows that, as intended by the authorities, the implementation of Basel II's internal ratings-based methodology has encouraged banks to adopt more stringent risk management procedures.

Rehman et al. (2019) identified the risk management strategies used by the commercial banking sectors of Balochistan, Pakistan for preventing or eliminating credit risk. The findings revealed the effectiveness of various risk management strategies applied in commercial banking sectors for reducing credit risk. This explanatory study indicated the opinions of the bank employees on which strategies were suitable for mitigating credit risk. The results were found on four aspects of the impact on credit risk management (CRM), which were reported as corporate governance, diversification, hedging, and capital adequacy ratio. In this case, corporate governance had a greater impact followed by diversification in the studied banking sectors.

2.5 Risk appetite framework in the banking sectors

It is very important to note that before knowing the risk appetite framework (RAF), we have to understand the risk appetite in banking sectors because without knowing of risk appetite no one can develop the risk appetite framework by approaching models. Very few studies are found on RAF and the majority of studies are carried out in international banking sectors (Hyde et al., 2009; Lam, 2014; 2015; Baldan et al., 2016).

The study by Hyde et al. (2009) indicated the RAF is developed from the risk appetite statement (RAS) that does not alone help to prevent the risk without the study of global risk governance of the organization. This study also suggested one interesting approach involved in determining the positioning of certain key performance indicators (KPIs) along with the specific objectives explained by the stakeholders in connection to the risk appetite and tolerance limits.

According to three reports by the Financial Stability Board (FSB), one mentions the topic of risk governance in a general nature (FSB, 2013a), and the other two are based on the founding principles of the RAF (FSB, 2011; 2013b). These reports developed a type of guideline, in cooperation with other standard setters, concerning the key elements in that the RAF should develop effectively.

A study was conducted by Corbellini (2013) for Italian banking sectors and implemented a quantitative model for RAF. This work suggests about regulations-based establishment, in which the RAF should be indicated different types of risk for the assumption in banking sectors, which can be identified from Pillar I and Pillar II and used these for preparing the internal capital adequacy and assessment process(ICAAP) protocol.

Baldan et al. (2014) studied the performance of a risk appetite framework (RAF) in Italian banking sectors based on quantitative methods and regulations depend largely on a financial organization with some requirements. They suggested RAF as per the following criteria:

1. An adequate risk culture should be taken care of.

2. There should be a correlation between a business's risk management and other processes such as planning on strategies, "internal capital adequacy assessment process (ICAAP)", organizational, and control systems, etc.

3. It should be followed effective methods for internal communication and cooperation.

4. Usage of sufficiently advanced IT systems.

Lam (2015) reported the key components of the RAF, which included basic concepts and definitions followed by the implementation steps and the roles and responsibilities for the board, senior management, and business and operating institutions. He studied the RAS elements such as risk capacity, risk profiles, risk-adjusted returns, risk appetite, and risk tolerance at several levels of the institution to support enterprise-wide risk. It was also reported that major risk categories can only be identified after the illustration of the risk appetite statements and risk tolerances. Finally, a RAS maturity model was developed to facilitate self-assessment and continuous improvement of the organization.

Baldan et al. (2016) reported a study in Italian banking sectors for further development through a quantitative approach to structure the RAF and suggested a possible practical theory in connection with the detailed recommendations of the supervisors. The authors proposed an approach that might be implemented to adapt a requirement of these regulations, which might be suitable for management purposes. Their research goal was to build up the risk appetite statement (RAS) as a static 'picture' of the banks' risk profile and to develop a quantitative approach with which to implement the RAF in 29 commercial banking sectors of Italy. The resulting RAS were indicated to develop the RAF, which might be especially fruitful for regulatory and banking supervisors, as well as for external analysts to know a credit institution's position in terms of its propensity/aversion to the principal types of risk that financial institutions have exposed.

An established risk-oriented culture should facilitate rapid and effective implementation of the RAF, as the RAF is mostly about harmonizing risk-related vocabulary and key figures, within and across intermediaries. This is communicated among the most relevant intermediaries but is probably less acknowledged when they are not very big. As a consequence, numerous researchers and organizations have been emphasized the importance of encouraging, developing, and circulating a risk culture to all stages for some time now (Cortez, 2011; CEBS, 2016).

According to Analyst Prep (2020), the definition of risk appetite is mentioned as "*the amount and type of risk that a company is able and willing to accept in pursuit of its business objectives*".

In the interest of providing actionable and practical guidelines, and to prevent confusion, the most important definitions used are listed below. These are based on the FSB's Principles for an effective risk appetite framework (2013) and the CRD IV framework (Directive 2013/36/EU and Regulation (EU) No 575/2013), along with minimal additions to cover all elements of the project as per RAF guideline (2018):

Bank: "A credit institution as defined in point (1) of Article 4(1) of Regulation (EU) No 575/2013: an undertaking the business of which is to take deposits or other repayable funds from the public and to grant credits for its own account."

Risk culture: "A bank's norms, attitudes and behaviours related to risk awareness, risk-taking, and risk management, that shape decisions on risks."

Risk appetite (RAP): "The aggregate level and types of risk a bank is willing to assume within its risk capacity to achieve its strategic objectives and business plan."

Risk capacity: "The maximum level of risk the bank can assume given its current level of resources before breaching constraints determined by regulatory capital and liquidity needs, the operational environment (e.g., technical infrastructure, risk management capabilities, expertise) and its obligations to depositors, policyholders, shareholders, fixed income investors, as well as other customers and other stakeholders."

Risk appetite framework (RAF): "The overall approach (including policies, processes, controls, and systems) through which risk appetite is established, communicated, and monitored. It includes a risk appetite statement, risk thresholds and limits, and an outline of the roles and responsibilities of those overseeing the implementation and monitoring of the RAF.

The RAF should consider material risks to the bank, as well as to the bank's reputation vis-à-vis policyholders, depositors, investors, and customers. It is reviewed regularly."

Risk appetite statement (RAS): "The written form of a bank's risk appetite. It includes qualitative statements as well as quantitative measures expressed relative to earnings, capital, risk measures, liquidity, and other relevant measures as appropriate. It should also address risks that are more difficult to quantify such as reputational risks, cyber and conduct risks, as well as money laundering and unethical practices. It is set on an annual basis."

Risk profile: "A point-in-time assessment of the bank's gross and as appropriate, net risk exposures (after taking into account any mitigating factors) aggregated within and across each relevant risk category based on forward-looking assumptions."

Different definitions are listed below derived from a document of RBI (2020).

Risk governance framework: "It is a significant part of the overall governance framework, the framework through which the board establishes the bank's strategy as well as risk approach and management takes decisions in adherence to the same; articulate and monitor adherence to overall risk appetite as well as specific risk limits vis-à-vis bank's strategy; and identify, measure, manage or control risks."

Risk limits: "These are specific quantitative measures or limits based on, for example, forward-looking assumptions that allocate the bank's aggregate risk appetite to business lines, legal entities as relevant, specific risk categories, concentrations, and as appropriate, other measures."

Risk management: "These are processes established to ensure that all risks, associated risk concentrations are identified, measured, limited/controlled/mitigated and reported on a timely as well as comprehensive basis."

2.6 Machine learning approach for credit risk

Besides empirical and exploratory study, machine learning algorithm (ML) models may help to know about the accuracy of the performance easily for banking sectors. Many studies have been reported about the accuracy of credit risk scoring and credit risk management in the banking sectors (Dashet al., 2006; Zhang et al., 2007; Li and Zhong, 2012; Sudhakar and Krishna Reddy, 2014; Hamid and Ahmed, 2016; Bhatia et al., 2017; Baruah, 2020; Breeden, 2020; Singh and Prasad, 2020; Kaur and Singh, 2020; Madaan et al., 2021). As well as a recent study on operational efficiency prediction of Ghanaian banks (Appiahene et al., 2020). The ML algorithm models such as linear discriminate analysis, logistic regression, multivariate adaptive regression splines, Bayesian model, Decision tree, Markov model, Artificial Intelligence methods (artificial neural networks, support vector machine, genetic algorithm, and genetic programming, K-nearest neighbour classifiers and case-based reasoning) and hybrid models (simple hybrid models and Class-wise classifier) used in credit scoring (Li and Zhong, 2012), operational efficiency (Appiahene et al., 2020), but the predictive study of operational efficiency prediction through ML modelling is lacking in Indian banking sectors.

Dash et al. (2006) performed a study after integrating data envelopment analysis (DEA) and neural networks (NNs) to analyse the relative branch efficiency of a big Canadian bank. The results were compared with the normal DEA results and obtained comparable as a whole while Neural networks predicted the short-term efficiency.

Zhang et al. (2007)carried out a study on the problem of credit scoring and they compared three powerful ML models such as genetic programming (GP), backpropagation neural networks (BPNN), and support vector machines (SVM) as well as in a combined approach (CM) on credit scoring data. They used two types of credit scoring data created by Germany and Australia. Their results showed for BP, GP, and SVM, credit data of Germany reached 79.77, 79.53 and 78.94, respectively. About 89.29, 89.47 and 88.5 achieved for the Australian credit data, respectively. Although BP and GP found better performance on average than SVM, the classification accuracy of SVM is more stable for the same data set. Moreover, SVM is relatively simple and has faster prediction ability and the performance of the three models was not good for the German credit data, the reason may be that the German credit data set has too many good customers, reaching 70. They concluded that the combined model can obtain better results compared to each model like BP, GP, and SVM.

Li and Zhong (2012) reviewed an article to learn about personal credit scoring, which helps in the application of financial risk forecasting. They studied the techniques used for credit scoring through the classified method and the new method as an ensemble learning model. According to them, classifying these methods had become a complex and difficult job. In this context, they emphasized these methods and classified them into statistical models, AI models, hybrid methods, and ensemble methods. Moreover, they discussed the benefit of different models such as linear discriminate analysis, Logistic regression, (multivariate adaptive regression splines, Bayesian classifiers, Decision tree, Markov Model, Artificial Neural Networks, Genetic Algorithm, K-Nearest Neighbour, support vector machine, Case-Based Reasoning, Simple Hybrid Models, Class-Wise Classifier, and Ensemble Classifier in credit scoring predictive study. They concluded that Ensemble learning has been widely applied to personal credit evolution and has better classification ability and prediction accuracy.

Sudhakar and Krishna Reddy (2014) mentioned that the failure and success of the banking sector mainly depend on the industry's ability to appropriately evaluate credit risk. They studied the applicability of one of the new integrated models on sample data taken from Indian banking sectors. This integrated model is a combination of models based on the methods of Logistic Regression, Multilayer Perceptron Model, Radial Basis Neural Network, Support

Vector Machine, and Decision tree (C4.5), which compared the effectiveness of the credit approval process.

Hamid and Ahmed (2016) studied a new model for classifying loan risk in different banking sectors by using a data mining approach. The model has been built using data from banking sectors to predict the loan status. Three algorithms such as J48, Bayes Net, and Naïve Bayes have been used to build the proposed model in their study by using the Weka tool for loan classification. The results have been found for the classifier's accuracy for J48 (78.3784), Bayes Net (77.4775), and Naïve Bayes (73.8739), respectively. J48 was obtained as the best algorithm based on accuracy compared to the other two algorithms.

Bhatia et al. (2017) reviewed a wide range of statistical methods in ML models, which have been applied though the datasets available to the public is limited due to confidentiality concerns. Problems particular to the context of credit scoring were examined. In their study, they used several models such as Linear discriminant analysis, Decision Trees, Logistic regression, and XG Boost. They concluded that the implementation of these ML models is helpful in creating credit risk scorecards.

Hassan and Tabasum studied ML techniques by using the WEKA tool for classification algorithms viz. Support Vector Machine, Classification and Regression Trees, and J48 algorithm for predicting risk and then categorization of loan customers into any of three risk categories i.e., “low risk”, “medium risk” and “high risk”. They obtained that these predictive models were suitable for deciding whether the loan should be recommended for sanction or not as well as if sanctioned in which risk category loan may be considered.

Zhu et al. (2019) conducted a study on the dataset of Lending Club for the first quarter of 2019 and they used the Random Forest algorithm to predict its performance accuracy and compared it with other machine learning methods such as Decision Tree, Logistic Regression and Support Vector Machine. They concluded that Random Forest has much better accuracy (98%) than other algorithms like Logistic Regression (73%), Decision Tree (95%), and Support Vector Machine (75%).

Singh and Prasad (2020) reviewed the priority usage of artificial intelligence (AI) and machine learning (ML) methods for digital credit lenders and many digital credit firms in particular claim to use AI algorithms for risk-modelling and credit underwriting decisions in the Indian context. They also reported credit risk assessment can easily be done through AI and ML techniques during the COVID-19 pandemic situation.

Appiahene et al. (2020) presented a combined Data Envelopment Analysis (DEA) with three ML approaches to evaluate the efficiency and performance of 444 branches of Ghanaian banks using Decision-Making Units (DMUs). In their results, they were compared with the corresponding efficiency ratings obtained from the DEA. Finally, the prediction accuracies of the three ML algorithm models were compared in R studio using R codes. Their results suggested that the decision tree (DT) and its C5.0 algorithm provided the best predictive model. It had 100 accuracies in predicting the 134 holdout sample dataset (30 banks) and a value of 0.00. The DT was followed closely by a random forest algorithm with a predictive accuracy of 98.5 and a value of 0.00 and finally the neural network (86.6 accuracy) with a value of 0.66. The study concluded that banks in Ghana can use the result of this study to predict their respective efficiencies.

Kaur and Singh (2020) studied client credit risk prediction in the banking sector through a data mining approach by using the WEKA tool. The prediction accuracy of defaulter instances obtained by using Cost-Sensitive Learning is considerably good as compared to the results of not using it. The overall classification results were also found relatively balanced. The parameters viz. 'Duration', 'employment', and 'age' are the most important factors for predicting the class of the loan applicant (whether the applicant would 'default' or 'not') in the case of the credit dataset. 'Job Personal_ status' and own telephone' were found the least significant attributes for prediction. Credit history and interest rate' are the most important factors for predicting the class of the loan applicant (whether the applicant would 'default' or 'not'). Finally, the prediction accuracy of three ML models was observed as Naïve Bayes algorithm (75.4) Decision tree algorithm (70.5), and the Zero R algorithm (70.0), respectively.

Madaan et al. (2021) proposed two ML models to predict the loan to an individual by assessing certain attributes and helping the banking authorities for easier processing and selecting suitable persons who apply for a loan. They conducted a comparative analysis between two algorithms (i) Random Forest, and (ii) Decision Tree. The results indicated that the Random Forest algorithm (80) with much higher accuracy compared to the Decision Tree algorithm (73.0) when both the algorithms were used on the same dataset. They concluded that exploring, analysing, and building an ML algorithm correctly identified a person, followed certain attributes, and had a high probability of defaulting on a loan.

2.7 Research Gap:

Prompted by regulator and investors, most financial institutions have completed formal "first and second generation" exercises to establish risk appetite statements (RAS). For many banks, the task of embedding and institutionalizing the essence of risk appetite frameworks, such as establishing a seamless alignment between risk appetite management and firm-wide stress testing initiatives, keeping downstream activities such as credit underwriting, limits management, and lending decision-making in the frontline—remains a work in progress. While risk management practices have become fairly advanced in developed markets, especially large banks in United States, usage of machine learning models in setting up agile enterprise wide risk appetite framework remains a gap. While there are many research studies undertaken in using machine learning models in predicting credit defaults, there are limited research analyses on using artificial intelligence and machine learning models to designing new generation agile risk management frameworks. This research work is an attempt to design a framework by understanding the causal relationship between variable key risk indicators at play that can impact financial performance of banks through the periods of economic recession, geo-political crisis, emergency crisis viz. pandemic, natural disasters etc. by leveraging basket of machine learning models to develop agile risk framework, that provides insights to bank board of directors, risk management committees and shareholders for effective management.

CHAPTER 3

MATERIALS AND METHODS

3.1 Introduction

Generally, the Indian banking sector is comprised as a major part of financial intermediation, which is considered as the main passage of “monetary policy transmission”, “credit delivery” and “payment systems”. It is well known that the banking system is maintaining permanence and proper health for the overall economic development and financial stability of the country. An important factor is the non-performing assets (NPAs), which may play a vital practical indicator where the financial health of the banking sector can easily be known. Besides asset quality, NPAs characterized the credit risk management and efficacy in the allocation of funds. In this context, the successful studies on the design and development of robust credit risk management and risk appetite framework through machine learning algorithm modelling for the Indian banking sectors can be suitable to know the overall performance accuracy in which the appropriate factor(s) of financial loss in the banking sectors can be known.

The present empirical study used quantitative statistical methods such as frequency distribution, graphical analysis viz. bars, pie charts, etc., cross-tabulation, normality test, correlation coefficient, regression, and analysis of variance (ANOVA), and logistic regression for credit risk appetite framework. Moreover, the prediction of performance accuracy of studied banking sectors was analysed through machine learning algorithms comprising of Regression and Classification Models.

This chapter considers in detail the Research Design, Research or Theoretical Framework, Data Sources, Tools and Techniques, and data analysis methods.

3.2 Research Paradigm:

A research paradigm refers to the conceptual structure upon which the investigation is predicated. It provides a framework of assumptions and knowledge that the research project's theories and procedures are based on. Due to the quantitative nature of the data, the present study is based on the Positivist Research Paradigm.

3.3 Research design

The present investigation was based on a multi-dimensional study and a primary survey was performed from the secondary data, which was collected from 8 public banking sectors and 4 private banking sectors of India for the period of 2008-2019. All data were collected for macroeconomic variable and key internal parameters of banks.

In-time validation period was 2008-2019 and out-of-time validation period used is 2020-2022. Out-of-time validation period has been used to test the soundness of the ML models analysed.

Wide array of machine learning algorithms such as Decision Tree, Random Forest, Gradient Boost, Support Vector Machine, Naïve Bayes and Artificial Neural Network were used by using Python via Jupiter Notebook (version, 3.6) as well as WEKA tool (version, 3.8.5).

The big dataset was used to predict the dependent variables (GNPA and NNPA). In addition, accuracy of the relationship between dependent and independent variables were also tested. Under this approach, both types of ML models viz. regression and classification models have been used to predict the dependent variable and to test the accuracy of dependent variable, respectively.

3.4 Research or theoretical framework

The research or theoretical framework was designed to develop a multi-variable Regression model, Statistical or Judgmental depending on the availability of statistical significance, the suitability of a model was decided. **Firstly**, the selection of banking sectors was done. **Secondly**, the Bank's key asset quality or credit risk parameters like Revenues, Earnings, NPA, YoY Revenue/PAT growth, Sectoral deployment, etc. (independent variables) were analysed by using the Multi-variable regression Model. **Thirdly**, the economic indicators or parameters viz. GDP, Unemployment, Interest Rates, Foreign Exchange reserve, etc. were considered key attributes (dependent variables). **Finally**, this model was validated by machine learning algorithms, especially regression and classification models. Fig 3.1 illustrates the flow diagram of the theoretical framework.

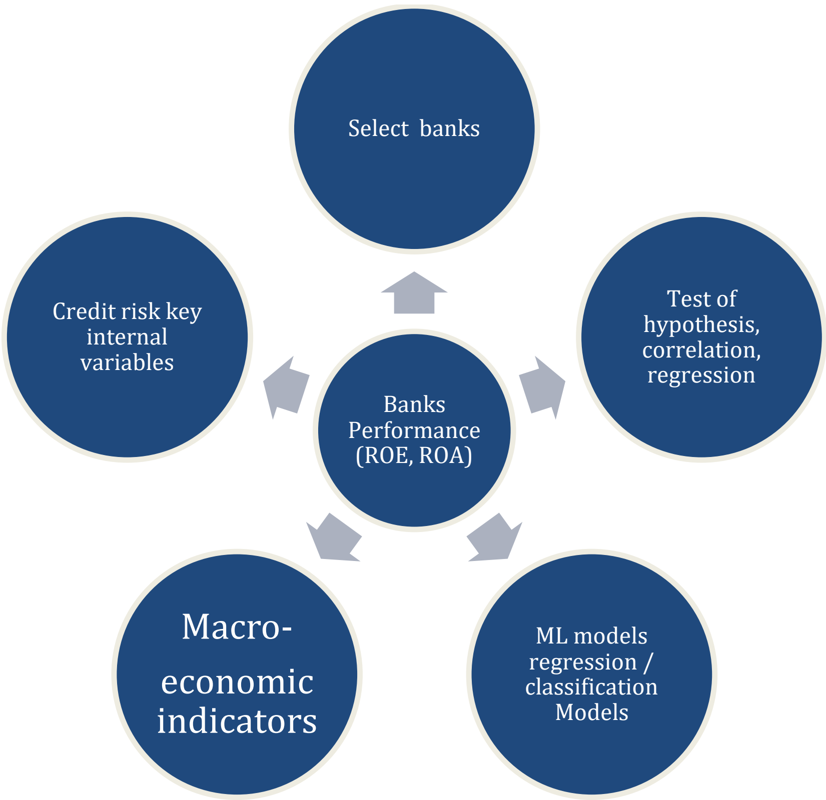


Figure 3.1 : Theoretical framework for overall banking performance

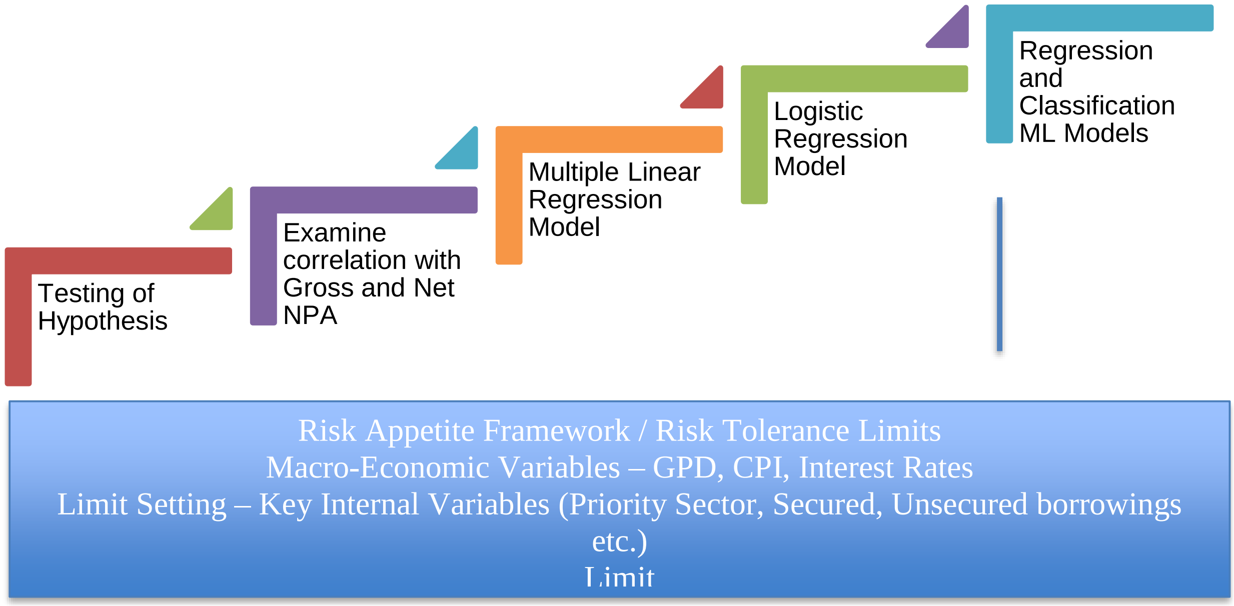


Figure 3.2: Theoretical framework for overall banking performance

3.5 Data sources

The study area was selected as 8 public banking sectors such as State Bank of India (SBI)," "Bank of India (BoI)", "Bank of Baroda (BoB)", "Bank of Maharashtra(BoM)", "Central Bank of India (CBoI)", "Andhra Bank (AB)", "Canara Bank (CB)" and "Allahabad Bank (AlB)" and 4 private banking sectors viz. "Axis Bank (AxB)", "ICICI Bank (ICB)", "HDFC Bank (HDFCB)" and " Kotak Bank (KB)" of India.

The research constitutes studying key public sector and private sector banks in India which have a significant coverage of loan exposures. Selected sample of 12 banks comprised of nearly 70% of total gross advances at the time of initiation of research. 8 public sector banks and 4 private sector banks selected were selected on stratified sampling approach depending on gross advances. These banks continue to be of significance in current context, providing meaningful research data to assess risk management practices in India.

Data collection was done primarily through secondary research based on financial performance data available in the public domain viz. Annual Reports, Analysts Presentations, and/or any other publicly available research reports on leading banks.

3.6 Research Approach:

Studies can be qualitative, quantitative, or mixed in nature depending on the nature of data collected for the research. These different research approaches can be defined as follows:

- Qualitative Research: In qualitative research, thoughts, views, or experiences are investigated through the collection and analysis of non-numerical data (written, video, or audio, for example). It can be applied to provide fresh research ideas or obtain an in-depth understanding of an issue.
- Quantitative Research: Gathering and evaluating numeric data is the process of conducting quantitative research. Finding trends and averages, generating hypotheses, examining causality, and extrapolating findings to larger populations are all possible with it.
- Mixed Method: It utilizes both qualitative and quantitative data to fulfil the research objective.

The present study utilizes a ***quantitative research*** method as the financial data is numerical in nature.

3.7 Tools and Techniques and Data Analysis Methods

3.7.1 Machine learning algorithms

3.7.1.1 Regression and classification models

Machine learning algorithms were performed by using Python via Jupiter Notebook (version 3.6). It predicted the overall banking performance accuracy through two types of models such as Regression and Classification models. In the Regression model, an Artificial Neural

Network, Decision Tree, Random Forest, and Support Vector were studied and in the Classification model, Naïve Bayes, Random Forest, and Support Vector were studied. Generally, classification predicts a label while regression predicts a quantity (Brownlee, 2017).

Regression models:

Artificial Neural Network (ANN) regression: In which the regression ANN is predicted as an output variable as a function of the inputs. The input features (independent variables) could be categorical or numeric types, however, for regression ANNs, it required a numeric dependent variable.

Decision Tree regression: Gurucharan (2020) described that the Decision Tree is a commonly used practical approach for supervised learning. It can be used to analyse the Regression.

Random Forest regression: Random Forest Regression is a well-known supervised learning algorithm, that uses an ensemble learning method for regression. The ensemble learning method is a technique, which is combines predictions from multiple machine learning algorithms where a more accurate prediction could be done compared to a single model (Bakshi, 2020).

Support Vector regression: Support Vector Regression is an established supervised learning algorithm, which is used to predict discrete values. Moreover, this is also used the same principle as the support vector machine (SVM). The basic theory behind Support Vector Regression is to find the best-fit line and the best-fit line is the hyperplane, which has the maximum number of points (Raj, 2020).

Classification models:

Naïve Bayes classifier: This classifier is a machine learning model, which is used to differentiate different objects based on certain features. Moreover, the principle defines “A

Naive Bayes classifier as a probabilistic machine learning model, which is used for classification tasks as per the Bayes theorem" (Gandhi, 2018).

Random Forest classifier: This is a well-known tree-structured classifier with three types of nodes Root Nodes, Interior Nodes, and Leaf Nodes. The Root Node is the initial node, which represents the entire sample and might get split further into additional nodes. The Interior Nodes represent the features of a dataset and the branches are represented the decision rules. The Leaf Nodes are represented for the final outcomes (Gurucharan, 2020).

Support Vector classifier: Support Vector Classification is an established supervised learning algorithm, which is used to predict a hyperplane in an N-dimensional plane. Moreover, the hyperplanes are decision boundaries, which help to classify the data points. It is well-known that the dimension of the hyperplane depends upon the number of features (Gandhi, 2018).

Accuracy in Regression model is defined as 1-MAPE and MAPE is Mean Absolute Percentage Error.

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

M = mean absolute percentage error

n = number of times the summation iteration happens

A_t = actual value

F_t = forecast value

3.7.1.2 Prediction of performance accuracy of banks through ML modeling

In the present study, data mining through the ML modelling algorithm was performed by using WEKA (Waikato Environment for Knowledge Analysis) tool (version, 3.8.5) developed by Frank et al. (2016). The WEKA explorer was developed with data pre-processing, classification, regression, and association rules. The predictive accuracy of data of different parameters of public sector banks (State Bank of India as SBI, Bank of India as BoI, Bank of

Baroda as BoB, Bank of Maharashtra as BoM, Central Bank of India as CBoI, Andhra Bank as AB, Canara Bank as CB and Allahabad Bank as AlB) and private sector banks (ICICI Bank as ICB, HDFC Bank as HDFCB, Axis Bank as AxB and Kotak Mahindra Bank as KB) were analysed through ML modeling algorithms especially, classifiers attributes viz. GNPA and NNPA (Gross / Net non-performing assets); PSL (Priority sector lending); TL (Term loan); STA (Secured to total asset); RR (repo rate); CPI (Consumer price index); TE (Total earnings); TP (Total profits); TA (Total advance); USTA (Unsecured/Tot Advances); GDP (Gross domestic product) and class (good and bad) to assess overall performance accuracy.

The performance of model accuracy of different ML algorithms such as Bayes Network (BN), Naïve Bayes (NB), Logistic Regression (LgR), Sequential minimal optimization of Support Vector Machine regression (SMOreg), Linear Logistic Regression (SL), Classification via Regression (CR) Logit Boost (LB), Pruned and unpruned decision tree C4 (J48), Logistic Model Tree (LMT), Random Forest (RF), Random tree (RT), and Class implementing minimal cost-complexity pruning (Cart) was tested.

In a pre-processing step, datasets were studied through unsupervised instance followed by segregation of data as training set (60%) and the rest data (40%) was used as cross-validation (CV) data. 50% of CV data was used as test set (Bhuvaneswari and Sarma Dhulipala, 2013). The classification results were obtained as per correctly and incorrectly classified instances, Kappa statistics (KS), mean absolute error (MAE) and root mean squared error (RMSE).

As per Bouckaert et al. (2020), the modelling summary of results was considered and retrieved from the WEKA tool. The prediction accuracy of studied ML models as per training and test set was retrieved from summary results and the statistical parameters are true positive (TP), false positive (FP), Matthew's correlation coefficient (MCC), receiver operating characteristic (ROC) and Precision-recall curve (PRC), respectively.

3.7.2 Traditional statistical analysis

For statistical analysis, the SPSS (version, 21.0) tool was used to analyse the data collected from banking sectors. Frequency distribution was studied to gather first-hand information on the various variables. All the methods of statistics were studied as per Patri (2003), Gupta (2004), Mohanty and Patel (2016) as well as SPSS tutorials.

Prior to statistical analysis and machine learning modelling, all the data were collected and incorporated in Microsoft Excel 2016 for the period of 2008-2019. All the data were converted to percentage values.

a) **Cross-tabulation analysis**

The Cross-Tabulation or Crosstab analysis or contingency table was analysed for "the basic technique to examine the relationship between two categorical variables as per the variables of additional layering" (Andersen, 1980). The protocol offers tests of independence and measures of relationship and concurrence for data. It has been mentioned that "The joint frequency distribution is examined with the Chi-square statistic to identify whether the variables were statistically independent or associated". If a dependency between variables were found, then other relative indicators such as Cramer's V, gamma, etc. can be used to determine the degree to which the estimations of one variable fluctuate with those of the other variable (Andersen, 1980).

To identify independent variables, the comparison was performed with a p-value at a significant level. Generally, "a significance level (denoted as α or alpha) of 0.05 works well but the same indicates a 5 risk of concluding that an association between the variables exists when there is no actual association". It is a well-known fact that the P-value is $\leq \alpha$, the variables have obtained an association with statistically significant value and considered H_0 rejection. It is also observed that the p-value ≤ of the significant level, the rejection of the null hypothesis, and it has been established that a statistically significantly correlated value is observed between the variables. It is also known that P-value > α then non-correlated data are obtained in the variables i.e. failure and rejection of H_0. If the p-value is > the significant value, there has been a failure and rejection of the null hypothesis and there is no confirmed evidence of the relationship of variables.

b) Mean Scoring

The mean scoring or more precisely the arithmetic mean score, is simply the arithmetic average of a group of numbers (banking data). So, the mean score of the variable x is $\bar{x}$ was calculated by adding all the values within a data set, and division was done by the total number of values in that data set.

c) Standard Deviation

As per probability theory in statistics, standard deviation or σ is identified for the variation or "dispersion" within the average (mean or expected) value. The estimation of lower standard deviation is shown when the data are directed to incline toward being near the average value, while an increased assumption deviation is demonstrated that the data obtained deviate over a huge scope of ranges.

d) Chi-Square

It was tested based on the hypothesis that the independent variables are located in rows and columns without identifying the relationship of strength or direction. In the Chi-square method, the "Pearson Chi-square" test, "Likelihood-ratio Chi-square" test, and "Linear-by-Linear" test are studied. The "Chi-square" test value is extracted through 'Fisher's exact test and Yates', which is corrected to compute 2x2 tables for identifying the differences, if any.

e) Correlation Analysis

This was measured as “the association between two continuous variables. Correlation study measures the relationships between two variables as both the size and direction” (Mohanty and Patel, 2016). The squared correlation was measured as “the strength of the relationship” (Mohanty and Patel, 2016). To understand the association between two variables the following tools are also used in this study.

f) Regression

According to Mohanty and Patel (2016), the “Regression analysis” is a “mathematical calculation” of the mean value where correlation is studied with two variables or more than two variables related to the original data. “*Regression analysis clearly indicates the cause-effect impact within the variables*”. In regression, “*the variable corresponding to the cause can be considered as an independent variable and the variable is corresponding to effect considered as a dependent variable*”. The “Regression analysis” is allowed among those who are working on market research to identify the correlation between independent versus dependent variables. Moreover, in the statistical calculations in market research, the dependent variable is usually the outputs per people care such as sales, etc. whereas the independent variables are the mechanisms, which is achieved from those outcomes or outputs (e.g., price or advertisement), (Sarstedt and Mooi, 2014). The “Regression analysis” may indicate insights of a few other techniques. According to Sarstedt and Mooi, (2014), important benefits regarding regression analysis are as follows:

1. It is indicated a close correlation between independent and dependent variables.
2. It indicates the perspective strength of various independent variables in relation to the dependent variable.

g) F statistic

The F statistic is determined by "regression mean square (MSR) divided by the residual mean square (MSE)". If the significant level of the F statistic finds a lower value of <0.05 then the independent variables perform suitable in expressing the variability in the dependent variable. If the significance value finds F >0.05 then the independent variables explained without the variation in the dependent variable and the null hypothesis indicates that all the sample values for the regression coefficients find 0, which accepts by author (Pallart, 2005).

h) Beta Coefficient

The beta coefficient finds a significant difference when the independent variable strongly associates with the dependent variable. It finds a correlation coefficient between two variables.

i) Coefficients

The measurement of independent variables considers different units. The coefficients of standards or β values studies to make the regression coefficients more comparable. If information changes to z scores before regression analysis, one can get the β coefficients as unstandardized coefficients. The "t statistics" help to determine the relative importance of individual variables in this model. The guidance regarding suitable predictors could be observed for t values suitable as below -2 or above +2 (Pallart, 2005).

j) Logistic regression

"Logistic regression" analysis for forecasting values on a dependent variable from the evidence of other (independent) variables. It is also acknowledged as a logit regression analysis, and it is carried out on a "dichotomous dependent variable" and "dichotomous independent variables". The "dependent variable" typically evaluates the occurrence of something or the probability that an upcoming event will happen. Additionally, they should be coded as "1" describing the existence of an attribute, and "0" to signify none of that attribute (Chao-Ying and Tack-Shing, 2002; Chao-Ying et al., 2002). Logistic regression is one of the most used statistical methods, which are used to model the likelihood of binary or multinomial consequences. Logistic regression not only aids in analysing the variables or factors that impact a result of interest but also the dissimilarity in the likelihood of the outcome due to unit modification in the factors. Logistic regression contingent on the nature of its dependent variable or to be more specific on the number of results of the dependent variable can be of two types as per Akinci et al. (2007).

3.9 Hypotheses development

H_0: There is no significant relationship between the measurement of credit risk using NPAs and selected bank-related factors and macroeconomic variables individually for Indian banking sectors.

H_1: There is a significant relationship between the measurement of credit risk using NPAs and selected bank-related variables and macroeconomic variables individually for Indian banking sectors.

H_2: There is no significant relationship between the measurement of credit risk using NPAs and selected bank-related factors and macroeconomic variables collectively in a multivariate setup for Indian banking sectors.

H_3: There is a significant relationship between the measurement of credit risk using NPAs and selected bank-related factors and macroeconomic variables collectively in a multivariate setup for Indian banking sectors.

H_4: Machine learning algorithm models usage does not result in increasing the accuracy of multivariate predictive models.

H_5: Machine learning algorithm models usage result in increasing the accuracy of multivariate predictive models.

To conclude, the chapter lays down various statistical methods used for the research which is a combination of most prevailing machine learning models (regression and classification) and other statistical regression and correlation analyses, to assess results from different statistical tools. Chapter also highlights the data-sets used and the rationale for choosing select banks and concludes with the hypothesis used for the research.

CHAPTER 4

RESULTS ANALYSIS AND INTERPRETATION

4.1 Data analysis

Chapter – 4 describes the results and interpretation section in which the data was analysed through machine learning modelling and validated by traditional statistical methods.

Prior to statistical analysis and machine learning modelling, all data were collected and incorporated in Microsoft Excel 2016 for the period of 2008-2019 (In-time validation) and 2020-2022 (Out-of-time validation).

Out-of-time validation period has been used to test the soundness of the models used for accuracy and consistency.

Selected independent variables consisting of key macro-economic indicators like GDP, CPI, interest rates and banks' internal indicators like unsecured loans concentration, priority sector lending exposure, secured loans concentration, net profit to advances ratio have varying degree of correlation to dependent variable, Non-Performing Loans (Gross NPA/Net NPA).

Machine learning models, using Python show different degrees of significance to CPI, Priority Sector Advances, Term Loans, Total Profits to Advances, Unsecured loans concentration, GDP, Interest rates. While GDP and Interest rates don't have a significant correlation, these are critical econometric variables which play an important in setting risk appetite framework and risk limits.

For ML algorithms, datasets were performed by using Python via Jupiter Notebook (version 3.6). It predicted the overall banking performance accuracy through two types of models such as regression and classification models. In the regression model, an Artificial Neural Network, Decision Tree, Random Forest, and Support Vector were studied and in the classification model, Naïve Bayes, Random Forest, and Support Vector were studied.

WEKA tool (version, 3.8.6) was used for validating the prediction and accuracy of various parameters used in the risk appetite framework.

For traditional statistical analysis, the SPSS (version, 21.0) tool was used, to analyse of various parameters used in the risk appetite framework.

Different bank-related and macroeconomic parameters such as GNPA and NNPA = Gross and Net non-performing assets; PSL = Priority sector lending; TL = Term loan; STA = Secured to the total asset; RR = Repo rate; CPI = Consumer price index; TE = Total earnings; TP = Total profits; TA = Total advance; USTA = Unsecured/Tot Advances*100; GDP = Gross domestic products of the public (State Bank of India, as SBI Bank of India as BoI, Bank of Baroda as BoB, Bank of Maharashtra as BoM, Central Bank of India as CBoI, Andhra Bank as AB, Canara Bank as CB and Allahabad Bank as AlB) and private (ICICI Bank as ICB, HDFC Bank as HDFCB, Axis Bank as AxB and Kotak Mahindra Bank as KB) banking sectors were studied during the period of 2008-2019.

Year-wise and bank-wise trend analyses were performed. The statistical analysis was carried out on the Kruskal Wallis test based on year and bank-wise parameters, test of normality, overall mean and standard deviation, analysis of variance, correlation, regression, and logistic regression.

4.2 Year wise trend analysis

Table 4.1: Year-wise trend analysis of bank-related parameters

Year		GNPA (%)	NNPA (%)	GDP (%)	CPI (%)	PSL (%)	TL (%)	STA (%)
2008	Mean	2.609	1.283	3.891	9.102	27.725	60.803	74.585
	Median	1.669	0.821	3.891	9.102	28.496	60.404	74.559
	Std. Deviation	3.231	1.534	0.000	0.000	8.437	15.418	3.2915
2009	Mean	2.608	1.297	8.480	11.000	26.549	60.658	76.764
	Median	1.653	0.754	8.480	11.000	26.433	60.359	77.439
	Std. Deviation	3.023	1.496	0.000	0.000	10.049	15.614	3.9113
2010	Mean	2.412	1.143	10.260	9.500	25.531	61.346	76.927
	Median	1.618	0.876	10.260	9.500	26.671	62.081	78.273
	Std. Deviation	2.229	0.985	0.000	0.000	10.620	15.234	5.919
2011	Mean	2.030	0.835	6.638	9.500	26.218	59.012	77.843
	Median	1.584	0.753	6.638	9.500	27.955	56.430	78.975
	Std. Deviation	1.638	0.5123	0.000	0.000	11.207	15.020	5.1526
2012	Mean	2.227	1.106	5.456	10.000	23.729	58.225	80.420
	Median	1.793	.9452	5.456	10.000	24.3507	54.896	82.228
	Std. Deviation	1.274	0.794	0.000	0.000	9.867	13.980	4.511
2013	Mean	2.665	1.584	6.386	9.400	23.721	57.622	83.905
	Median	2.805	1.670	6.386	9.400	24.746	56.496	85.303
	Std. Deviation	1.280	1.0289	0.000	0.000	9.86748	13.38966	4.13384
2014	Mean	3.237	2.028	7.410	5.800	24.071	57.574	84.106

Year		GNPA (%)	NNPA (%)	GDP (%)	CPI (%)	PSL (%)	TL (%)	STA (%)
	Median	3.000	1.991	7.410	5.800	22.207	56.106	83.967
	Std. Deviation	1.559	1.206	0.000	0.000	11.000	12.835	4.301
2015	Mean	3.751	2.383	8.155	4.900	24.634	58.705	84.027
	Median	4.049	2.437	8.155	4.900	22.269	58.665	84.387
	Std. Deviation	1.608	1.327	0.000	0.000	11.478	12.508	5.7438
2016	Mean	7.010	4.506	7.113	4.500	25.191	57.683	83.349
	Median	8.010	4.837	7.113	4.500	23.233	56.921	82.908
	Std. Deviation	3.814	2.631	0.000	0.000	11.144	10.704	6.3719
2017	Mean	9.194	5.897	6.741	3.602	26.586	57.100	82.434
	Median	10.650	6.545	6.741	3.602	23.885	54.696	81.641
	Std. Deviation	4.406	3.544	0.000	0.000	12.747	10.2432	6.5941
2018	Mean	8.698	6.443	7.355	4.956	27.967	56.335	80.846
	Median	9.690	6.903	7.355	4.956	24.803	53.561	80.282
	Std. Deviation	4.233	3.4484	0.000	0.000	12.7673	9.7189	7.9247
2019	Mean	9.205	6.526	5.020	7.660	29.143	57.488	79.394
	Median	9.760	5.5455	5.020	7.660	25.773	53.792	76.529
	Std. Deviation	4.203	4.622	0.000	0.000	12.390	9.1968	8.5919
Total	Mean	4.637	2.919	6.909	7.493	25.922	58.546	80.383
	Median	2.607	1.705	6.927	8.381	25.797	55.482	80.260
	Std. Deviation	4.019	3.0923	1.6104	2.4753	10.7363	12.602	6.367

GNPA and NNPA = Gross and Net non-performing assets; PSL = Priority sector lending; TL = Term loan; STA = Secured to total asset

Table 4.2: Year-wise trend analysis of bank-related parameters

Year		GDP-1 (%)	RR (%)	CPI-1 (%)	Total Earnings/Total Advances (%)	Total Profit/ Total Advances (%)	Unsecured/Tot Advances (%)
2008	Mean	9.801	7.750	6.200	13.878	0.376	54.299
	Median	9.801	7.750	6.200	12.874	0.017	28.170
	Std. Deviation	0.000	0.000	0.000	15.062	0.861	39.851
2009	Mean	3.891	5.500	9.102	10.164	0.836	54.039
	Median	3.891	5.500	9.102	13.528	0.015	26.393
	Std. Deviation	0.000	0.000	0.000	7.782	1.799	40.024
2010	Mean	8.480	4.750	11.000	9.294	0.524	54.011
	Median	8.480	4.750	11.000	12.620	0.023	30.133
	Std. Deviation	0.000	0.000	0.000	7.032	0.980	40.236
2011	Mean	10.260	6.250	9.500	8.854	2.049	52.852
	Median	10.260	6.250	9.500	12.078	0.707	27.479
	Std. Deviation	0.000	0.000	0.000	6.641	2.806	41.180
2012	Mean	6.638	8.500	9.500	9.742	0.677	51.113
	Median	6.638	8.500	9.500	13.354	0.140	24.495
	Std. Deviation	0.000	0.000	0.000	7.314	1.030	42.650
2013	Mean	5.456	7.750	10.000	9.667	0.860	50.070
	Median	5.456	7.750	10.000	12.653	0.103	20.585
	Std. Deviation	0.000	0.000	0.000	7.314	1.392	43.495
2014	Mean	6.386	8.000	9.400	9.571	0.823	49.939
	Median	6.386	8.000	9.400	12.094	0.070	21.587

Year		GDP-1 (%)	RR (%)	CPI-1 (%)	Total Earnings/Total Advances (%)	Total Profit/ Total Advances (%)	Unsecured/Tot Advances (%)
	Std. Deviation	0.000	0.000	0.000	7.287	1.311	43.683
2015	Mean	7.410	7.750	5.800	9.455	1.204	50.363
	Median	7.410	7.750	5.800	12.657	0.017	22.661
	Std. Deviation	0.000	0.000	0.000	7.091	3.628	43.390
2016	Mean	8.155	6.750	4.900	9.464	0.098	50.916
	Median	8.155	6.750	4.900	12.933	0.000	23.382
	Std. Deviation	0.000	0.000	0.000	6.979	1.193	42.922
2017	Mean	7.113	6.250	4.500	9.865	-0.628	51.557
	Median	7.113	6.250	4.500	13.102	0.002	23.309
	Std. Deviation	0.000	0.000	0.000	7.418	2.582	42.298
2018	Mean	6.741	6.250	3.602	9.387	-0.109	53.280
	Median	6.741	6.250	3.602	13.241	0.002	27.140
	Std. Deviation	0.000	0.000	0.000	6.958	3.266	40.843
2019	Mean	7.355	5.620	4.956	9.327	0.306	52.657
	Median	7.355	5.620	4.956	13.001	0.003	27.698
	Std. Deviation	0.000	0.000	0.000	6.922	1.074	41.704
Total	Mean	7.307	6.760	7.372	9.889	0.585	52.091
	Median	7.234	6.500	7.651	12.818	0.018	25.293
	Std. Deviation	1.684	1.130	2.500	7.902	2.078	40.266

RR = Repo rate; CPI = Consumer price index; TE = Total earnings; TP = Total profits; TA = Total advance; USTA = Unsecured/Tot Advances*100; GDP = Gross domestic products

Table 4.1 and Table 4.2 describes year wise (2008-2019) trend analysis of bank-related parameters such as GNPA and NNPA = Gross and Net non-performing assets; PSL = Priority sector lending; TL = Term loan; STA = Secured to total asset.

Table 4.2 describes year-wise (2008-2019) trend analysis RR = repo rate; CPI = Consumer price index; TE = Total earnings; TP = Total profits; TA = Total advance; USTA = Unsecured/Tot Advances*100; GDP = Gross domestic products.

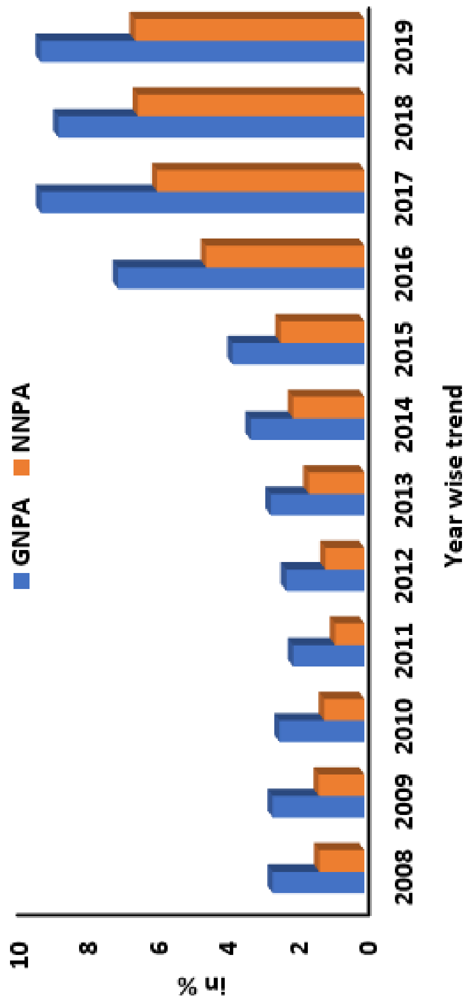


Figure 4.1: Year-wise trend analysis of GNPA and NNPA

In the case of GNPA (%), an increasing trend was observed from 2013 and a maximum increase during 2017 and 2019 while for NNPA (%), was observed an increasing trend from 2016-2019 (Fig 4.1).

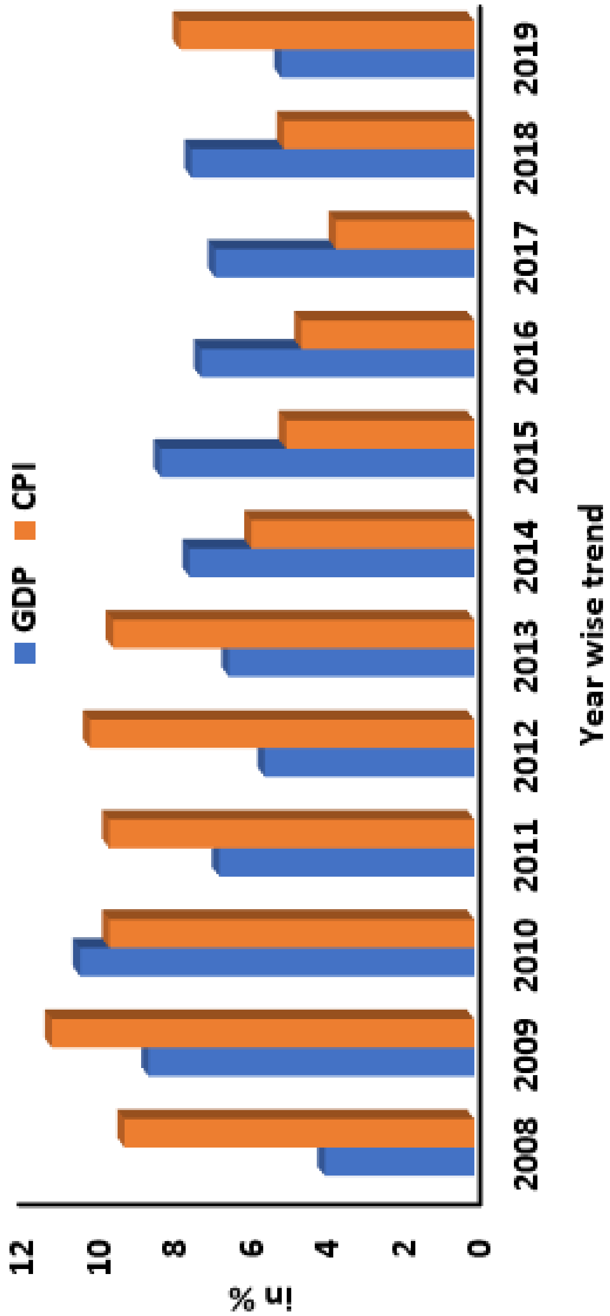


Figure 4.2: Year-wise trend analysis of GDP and CPI

In the case of GDP (%), variation was observed from 2008 and maximum decreased during 2019while for CPI (%), it was observed a decreasing trend from 2014-2017 but it was observed an increasing trend from 2018-2019 (Fig 4.2).

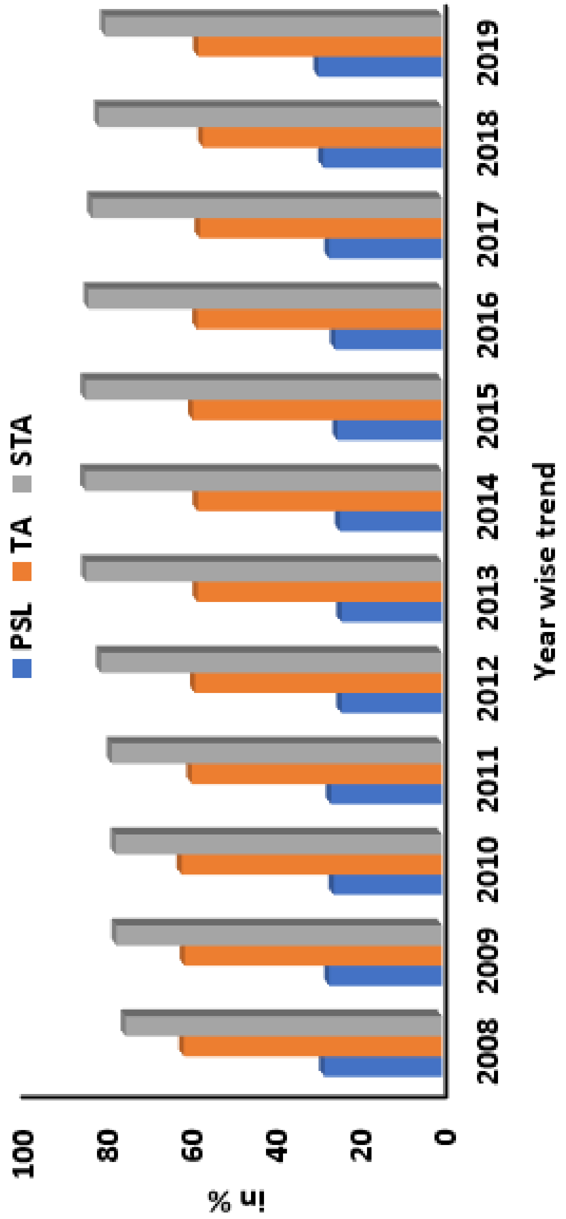


Figure 4.3: Year-wise trend analysis of PSL, TA, and STA

In the case of PSL (%), it's observed that the exposure towards priority sector (PSL) started to increase from 2014. Secured to Total Assets (STA) % exposure increased during 2013-2016 but observed a moderate decreasing trend from 2017-2019 (Fig 4.3).

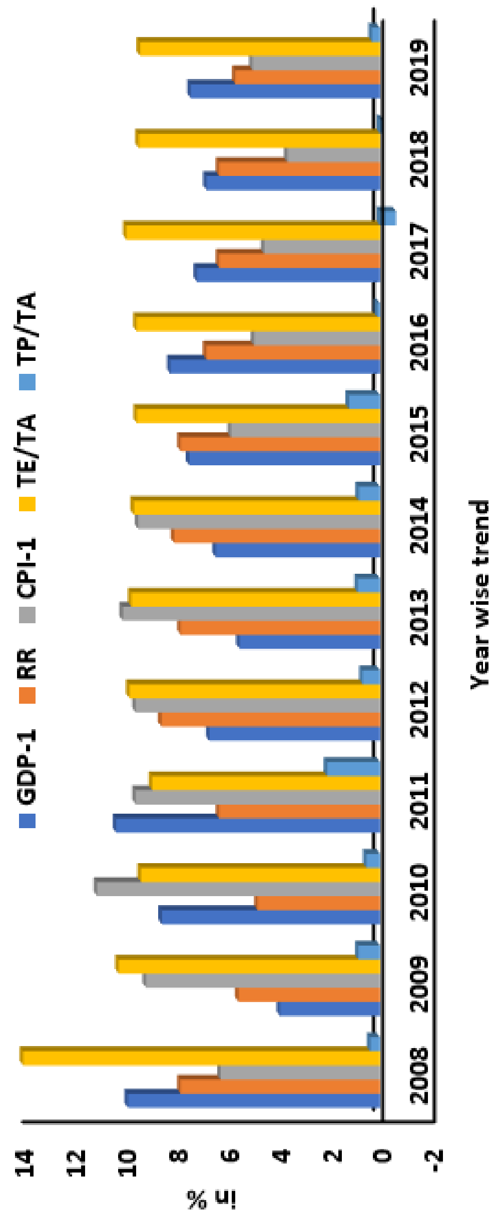


Figure 4.4: Year-wise trend analysis of GDP-1, RR, CPI-1, TE/TA, and TP/TA

In the case of GDP-1, RR, and CPI-1 (%), maximum variations were observed from 2008 to 2019. For TE/TA (%) found a decreasing trend from 2008-2019 while TP/TA (%) found negative values during 2017-2018 but it was observed slightly increase during 2019 (Fig 4.4).

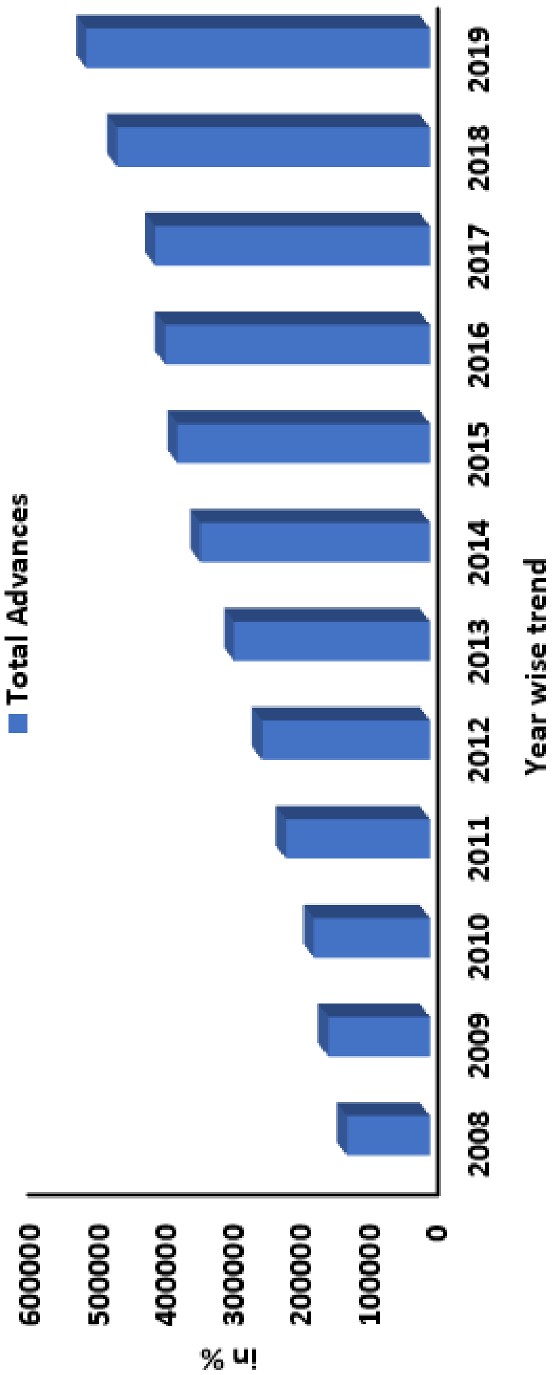


Figure 4.5: Year-wise trend analysis of Total Advances

In the case of Total Advances (%), an increasing trend was observed through the period of 2008 to 2019. Maximum value was found in the period of 2019 and minimum value was obtained in the period of 2008 (Fig 4.5).

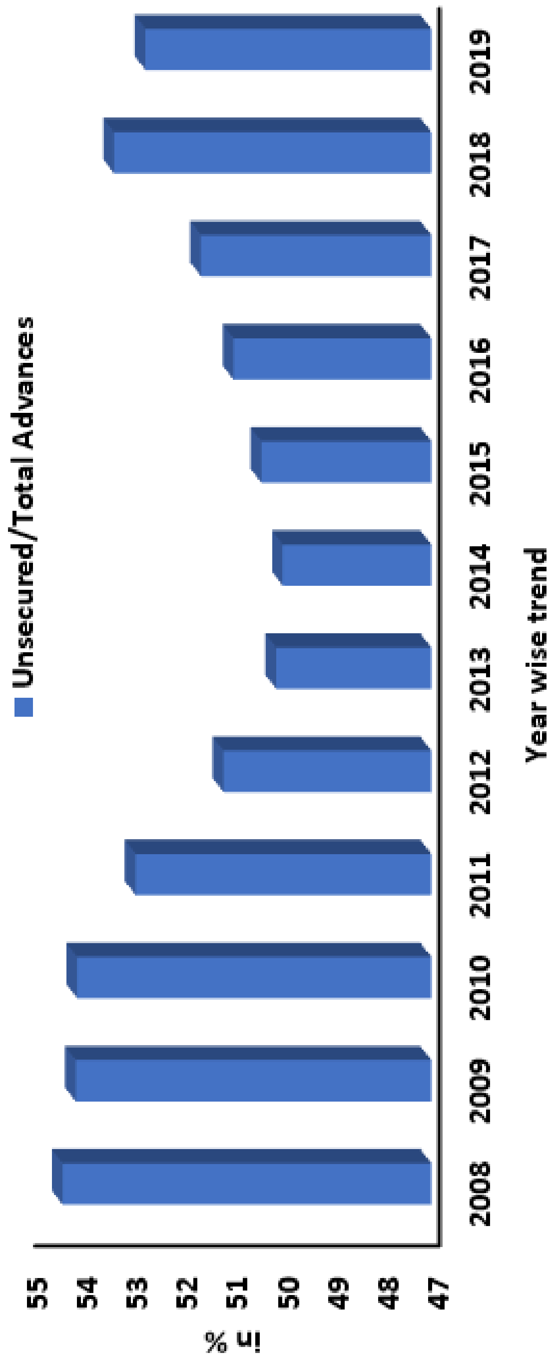


Figure 4.6: Year-wise trend analysis of Unsecured/Total Advances

In the case of Unsecured/Total Advances (%), maximum values were observed from 2008 to 2010 and in the period of 2018 but the exposures moderately declined in the period of 2019 (Fig 4.6).

4.3 Bank wise trend analysis

Table 4.3: Bank wise trend analysis of bank-related parameters

Bank		GNPA (%)	NNPA (%)	GDP (%)	CPI (%)	PSL (%)	TL (%)	STA (%)
SBI	Mean	3.932	3.057	6.909	7.493	25.283	52.341	79.320
	Median	3.488	2.110	6.927	8.381	24.550	50.787	79.924
	Std. Deviation	1.761	1.959	1.676	2.576	2.974	4.375	3.472
BoI	Mean	5.730	4.014	6.909	7.493	25.603	42.981	80.046
	Median	3.145	2.030	6.927	8.381	25.665	42.693	79.686
	Std. Deviation	4.888	4.051	1.676	2.576	3.568	4.093	3.728
BoB	Mean	4.218	2.688	6.909	7.493	25.028	44.441	81.975
	Median	2.528	1.400	6.927	8.381	25.282	43.990	82.722
	Std. Deviation	3.689	3.097	1.676	2.576	2.251	2.717	4.953
BoM	Mean	4.668	4.399	6.909	7.493	36.607	61.782	84.376
	Median	2.089	1.835	6.927	8.381	36.625	62.511	86.591
	Std. Deviation	4.237	4.548	1.676	2.576	2.676	7.015	6.382
CBoI	Mean	5.577	4.946	6.909	7.493	37.422	61.514	85.595
	Median	3.630	3.350	6.927	8.381	33.758	61.653	87.531
	Std. Deviation	4.643	4.417	1.676	2.576	9.031	6.841	7.739
AlB	Mean	6.320	3.683	6.909	7.493	12.047	52.667	83.805
	Median	4.772	3.588	6.927	8.381	11.065	52.469	83.967
	Std. Deviation	5.095	3.024	1.676	2.576	2.618	1.670	4.691
AB	Mean	5.709	3.055	6.909	7.493	14.457	44.234	85.175

Bank		GNPA (%)	NNPA (%)	GDP (%)	CPI (%)	PSL (%)	TL (%)	STA (%)
	Median	4.609	2.689	6.927	8.381	14.172	44.236	88.373
	Std. Deviation	4.941	2.961	1.676	2.576	2.438	1.890	6.900
CB	Mean	4.825	3.163	6.909	7.493	11.649	53.224	75.835
	Median	2.550	2.080	6.927	8.381	11.225	53.403	79.466
	Std. Deviation	4.203	2.486	1.676	2.576	2.706	3.486	6.614
AxB	Mean	2.561	1.014	6.909	7.493	33.996	69.567	78.946
	Median	1.315	0.431	6.927	8.381	35.660	69.816	80.145
	Std. Deviation	2.437	1.136	1.676	2.576	4.432	2.021	4.793
ICB	Mean	8.454	3.527	6.909	7.493	42.139	78.681	78.190
	Median	8.413	3.200	6.927	8.381	41.457	80.256	77.482
	Std. Deviation	3.017	2.127	1.676	2.576	3.080	5.455	5.277
HDFCB	Mean	1.213	0.325	6.909	7.493	18.701	66.811	72.574
	Median	1.060	0.298	6.927	8.381	17.043	69.997	73.068
	Std. Deviation	0.317	0.133	1.676	2.576	3.537	7.479	2.480
KB	Mean	2.438	1.161	6.909	7.493	28.135	74.309	78.764
	Median	2.213	1.022	6.927	8.381	27.345	75.510	77.725
	Std. Deviation	0.800	0.545	1.676	2.576	7.429	7.763	4.142
Total	Mean	4.637	2.919	6.909	7.493	25.922	58.546	80.384
	Median	2.607	1.705	6.927	8.381	25.797	55.482	80.260
	Std. Deviation	4.019	3.092	1.610	2.475	10.736	12.602	6.367

GNPA and NNPA = Gross and Net non-performing assets; PSL = Priority sector lending; TL = Term loan; STA = Secured to total asset; SBI = State Bank of India; BoI = Bank of India; BoB= Bank of Baroda; BoM = Bank of Maharashtra; CBoI = Central Bank of India; AB = Andhra Bank; CB = Canara Bank; AlB = Allahabad Bank; AxB =Axis bank; ICB = ICICI Bank; HDFCB = HDFC Bank; KB = Kotak Bank

Table 4.4: Bank wise trend analysis of bank related parameters

Bank		GDP-1 (%)	RR (%)	CPI-1 (%)	Total Earnings/Total Advances (%)	Total Profit/ Total Advances (%)
SBI	Mean	7.307	6.760	7.372	13.382	-0.193
	Median	7.234	6.500	7.651	13.444	0.003
	Std. Deviation	1.753	1.176	2.602	0.465	0.639
BoI	Mean	7.307	6.760	7.372	12.472	0.862
	Median	7.234	6.500	7.651	12.611	0.097
	Std. Deviation	1.753	1.176	2.602	0.675	1.733
BoB	Mean	7.307	6.760	7.372	11.836	0.677
	Median	7.234	6.500	7.651	11.800	0.019
	Std. Deviation	1.753	1.176	2.602	0.774	1.693
BoM	Mean	7.307	6.760	7.372	13.874	0.581
	Median	7.234	6.500	7.651	13.960	0.385
	Std. Deviation	1.753	1.176	2.602	0.669	5.201
CBoI	Mean	7.307	6.760	7.372	14.846	-0.001
	Median	7.234	6.500	7.651	14.394	0.001
	Std. Deviation	1.753	1.176	2.602	2.214	3.387
AlB	Mean	7.307	6.760	7.372	0.073	0.001
	Median	7.234	6.500	7.651	0.063	0.009
	Std. Deviation	1.753	1.176	2.602	0.038	0.023
AB	Mean	7.307	6.760	7.372	13.976	0.006
	Median	7.234	6.500	7.651	14.049	0.009
	Std. Deviation	1.753	1.176	2.602	0.666	0.014

Bank		GDP-1 (%)	RR (%)	CPI-1 (%)	Total Earnings/Total Advances (%)	Total Profit/ Total Advances (%)
CB	Mean	7.307	6.760	7.372	0.165	0.008
	Median	7.234	6.500	7.651	0.144	0.010
	Std. Deviation	1.753	1.176	2.602	0.091	0.010
AxB	Mean	7.307	6.760	7.372	0.168	0.020
	Median	7.234	6.500	7.651	0.153	0.024
	Std. Deviation	1.753	1.176	2.602	0.055	0.009
ICB	Mean	7.307	6.760	7.372	0.183	0.021
	Median	7.234	6.500	7.651	0.160	0.022
	Std. Deviation	1.753	1.176	2.602	0.083	0.007
HDFCB	Mean	7.307	6.760	7.372	18.627	2.593
	Median	7.234	6.500	7.651	15.871	2.634
	Std. Deviation	1.753	1.176	2.602	9.118	0.174
KB	Mean	7.307	6.760	7.372	19.068	2.443
	Median	7.234	6.500	7.651	18.036	2.604
	Std. Deviation	1.753	1.176	2.602	6.709	0.439
Total	Mean	7.307	6.760	7.372	9.889	0.585
	Median	7.234	6.500	7.651	12.818	0.018
	Std. Deviation	1.684	1.130	2.500	7.902	2.078

RR = Repo rate; CPI = Consumer price index; TE = Total earnings; TP = Total profits; TA = Total advance; USTA = Unsecured/Tot Advances*100; GDP = Gross domestic products; SBI = State Bank of India; BoI = Bank of India; BoB= Bank of Baroda; BoM = Bank of Maharashtra; CBoI = Central Bank of India; AB = Andhra Bank; CB = Canara Bank; AlB = Allahabad Bank; AxB =Axis bank; ICB = ICICI Bank; HDFCB = HDFC Bank; KB = Kotak Bank

Table 4.3 describes bank wise trend analysis of bank-related parameters such as GNPA and NNPA = Gross and Net non-performing assets; PSL = Priority sector lending; TL = Term loan; STA = Secured to the total asset.

Table 4.4 describes bank-wise trend analysis RR = repo rate; CPI = Consumer price index; TE = Total earnings; TP = Total profits; TA = Total advance; USTA = Unsecured/Tot Advances*100; GDP = Gross domestic products.

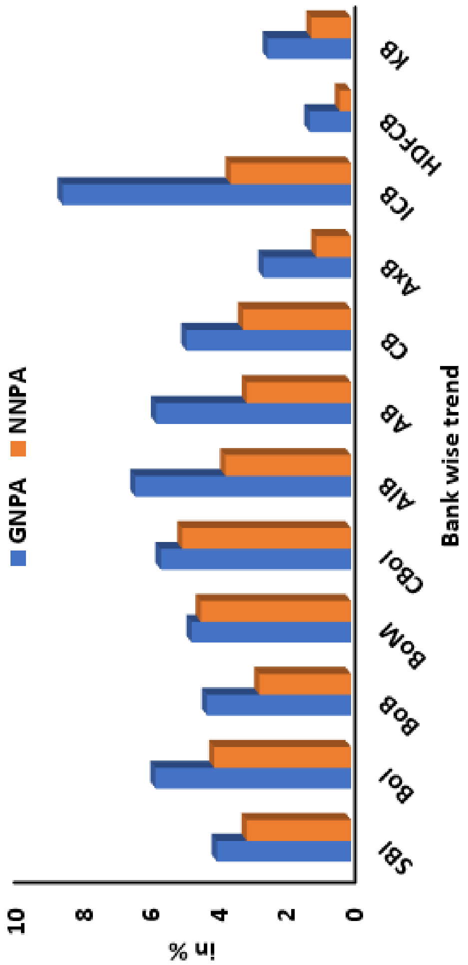


Figure 4.7: Bank-wise trend analysis of GNPA and NNPA

In the case of GNPA (%), maximum value was observed for ICB and minimum for HDFCB while much variation was observed for other banks. For NNPA (%), it was observed a maximum value for CBoI and a minimum value of HDFCB (Fig 4.7).

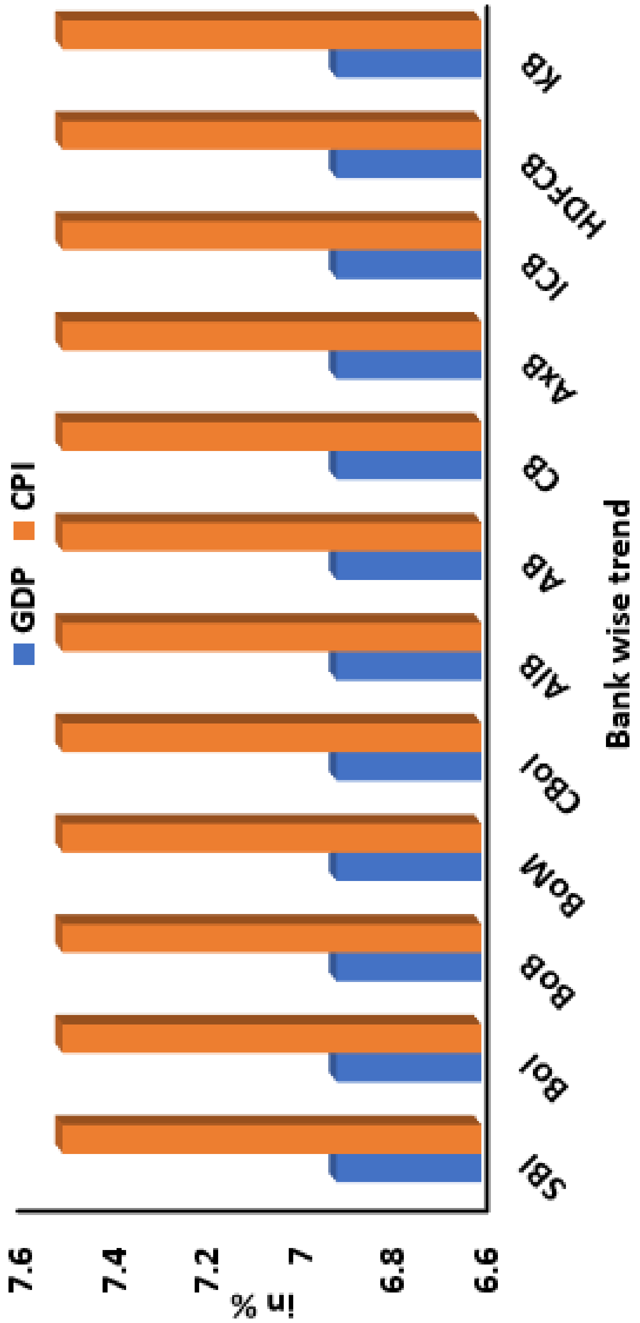


Figure 4.8: Bank-wise trend analysis of GDP and CPI

For GDP and CPI (%), similar value was observed in each case for all banks without any variation, but GDP was observed much lower compared to CPI (Fig 4.8).

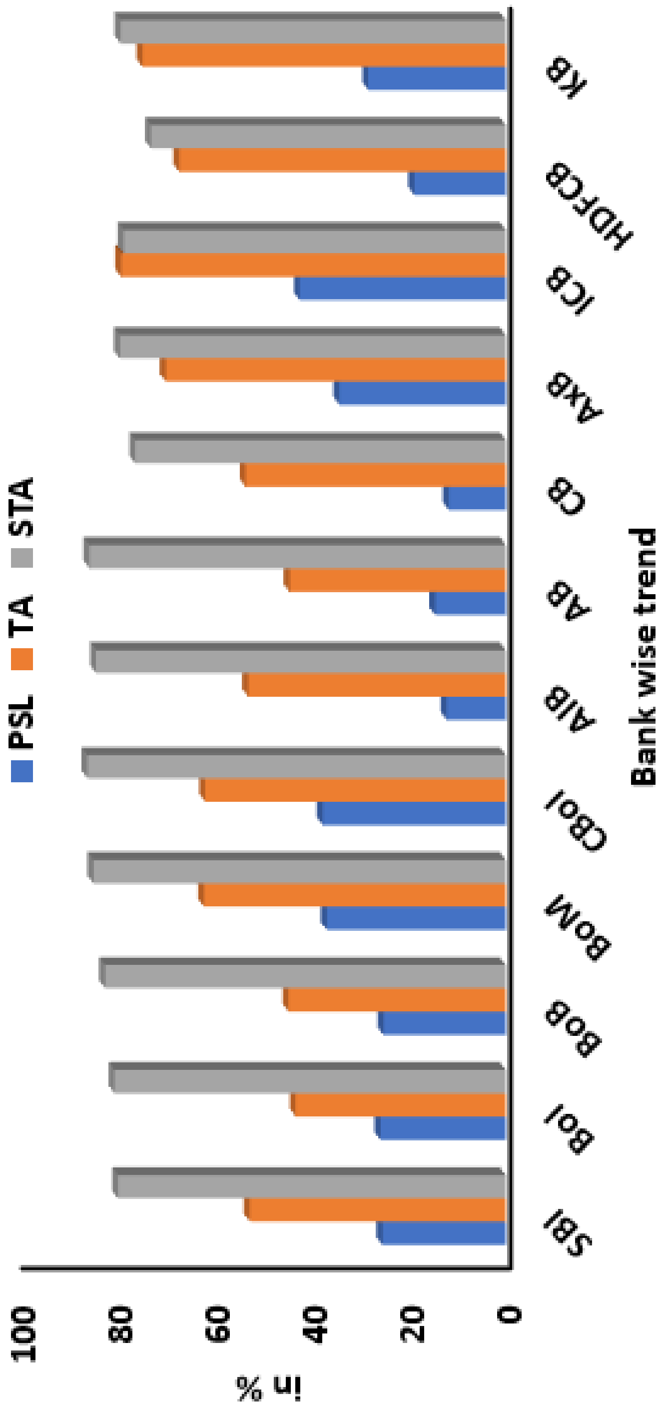


Figure 4.9: Bank-wise trend analysis of PSL, TA and STA

In the case of PSL (%), much variation was observed in the case of all banks and TA (%) was observed maximum value for ICB followed by KB and minimum value for BoI and BoB while STA (%) was increased maximum for ICB and minimum for AlB and CB (Fig 4.9).

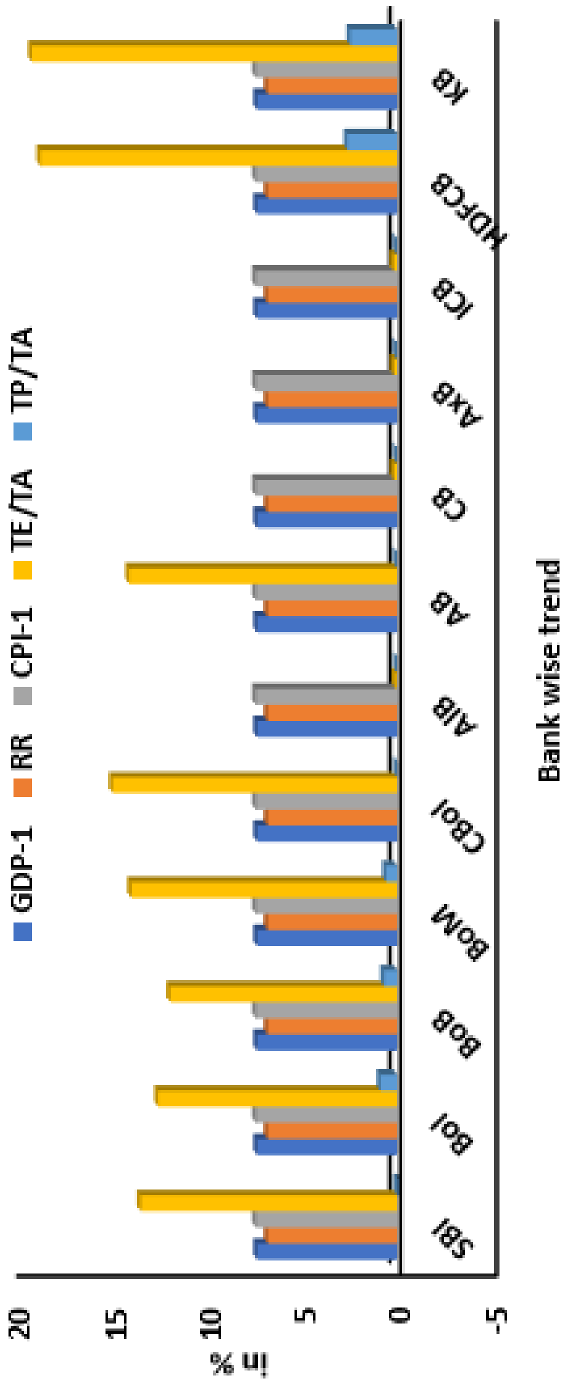


Figure 4.10: Bank-wise trend analysis of GDP-1, RR, CPI-1, TE/TA, and TP/TA

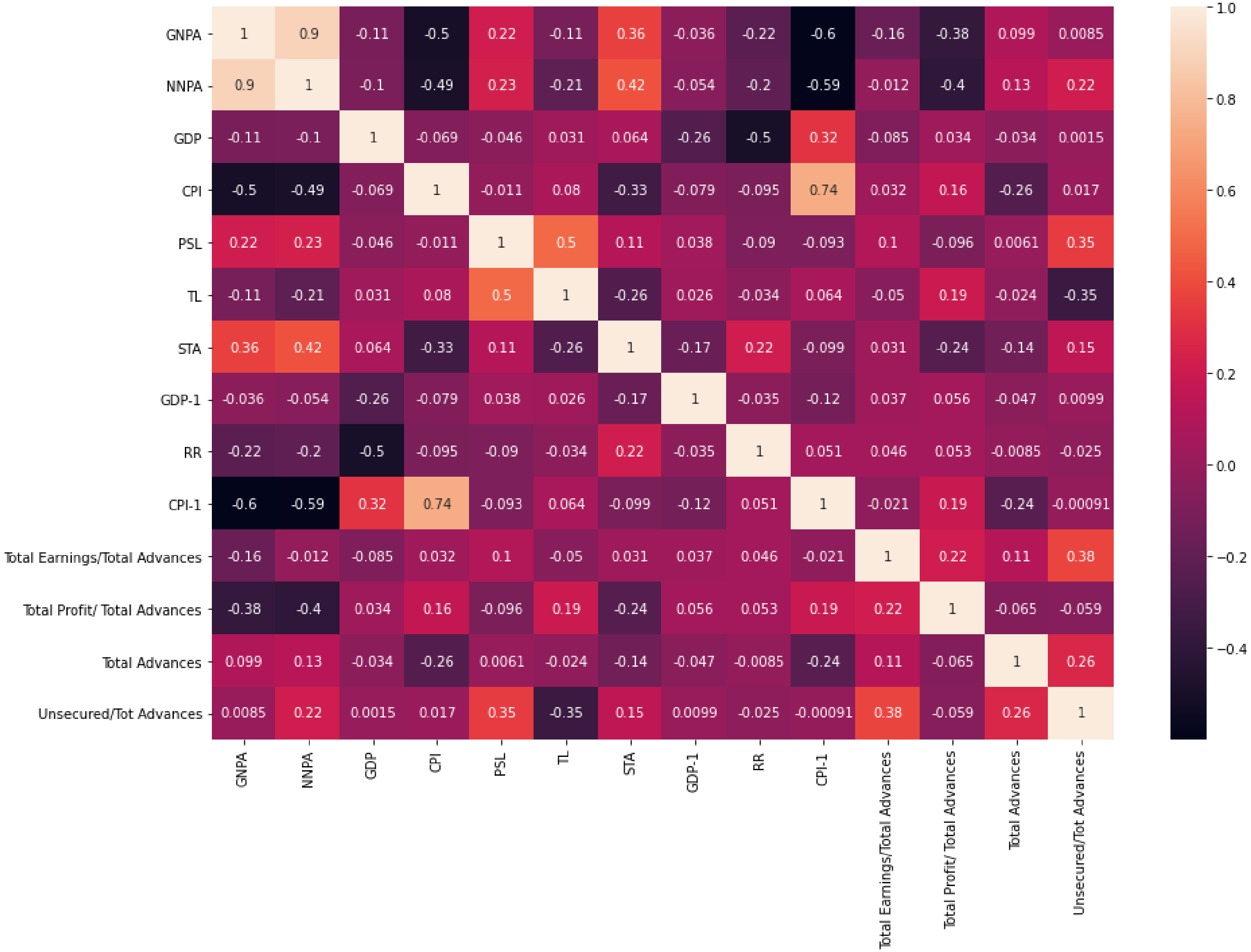
In the case of GDP-1, RR, and CPI-1 (%), a constant value was observed for these parameters for all banks. The variation was observed for TE/TA (%) in which maximum value was observed for HDFCB and minimum value was observed for BoB while TP/TA (%) was found much variation and negative values were observed for SBI and CBoI (Fig 4.10).

4.13 Prediction and accuracy of risk appetite variables using machine learning algorithms

In the present study, Artificial Neural Network (ANN) Regression was employed to predict the relation and quantification of the dependent variable as GNPA with other independent variables. The Decision Tree Regression was employed to predict the dependent variable as GNPA would be 1 (High) or 0 (Low) based on other independent variables when GNPA is 5% or more than 1 (High) else 0 (Low). The Random Forest Regression was employed to predict the relation and quantification of the dependent variable as GNPA with other independent variables. The Support Vector Regression was employed to predict discrete values and dependent variables as GNPA would be 1 (High) or 0 (Low) based on other independent variables when GNPA is 5% or more than 1 (High) else 0 (Low). The Naïve Bayes Classification employed to predict the dependent variable as GNPA would be 1 (High) or 0 (Low) based on other independent variables when GNPA is 5% or more than 1 (High) else 0 (Low). The Random Forest Classification was employed to predict the dependent variable as GNPA would be 1 (High) or 0 (Low) based on other independent variables when GNPA is 5% or more than 1 (High) else 0 (Low). The Support Vector Classification was employed to predict the dependent variable as GNPA would be 1 (High) or 0 (Low) based on other independent variables when GNPA is 5% or more than 1 (High) else 0 (Low).

The overall prediction was performed to detect how the Net GNPA is getting affected by different variables.

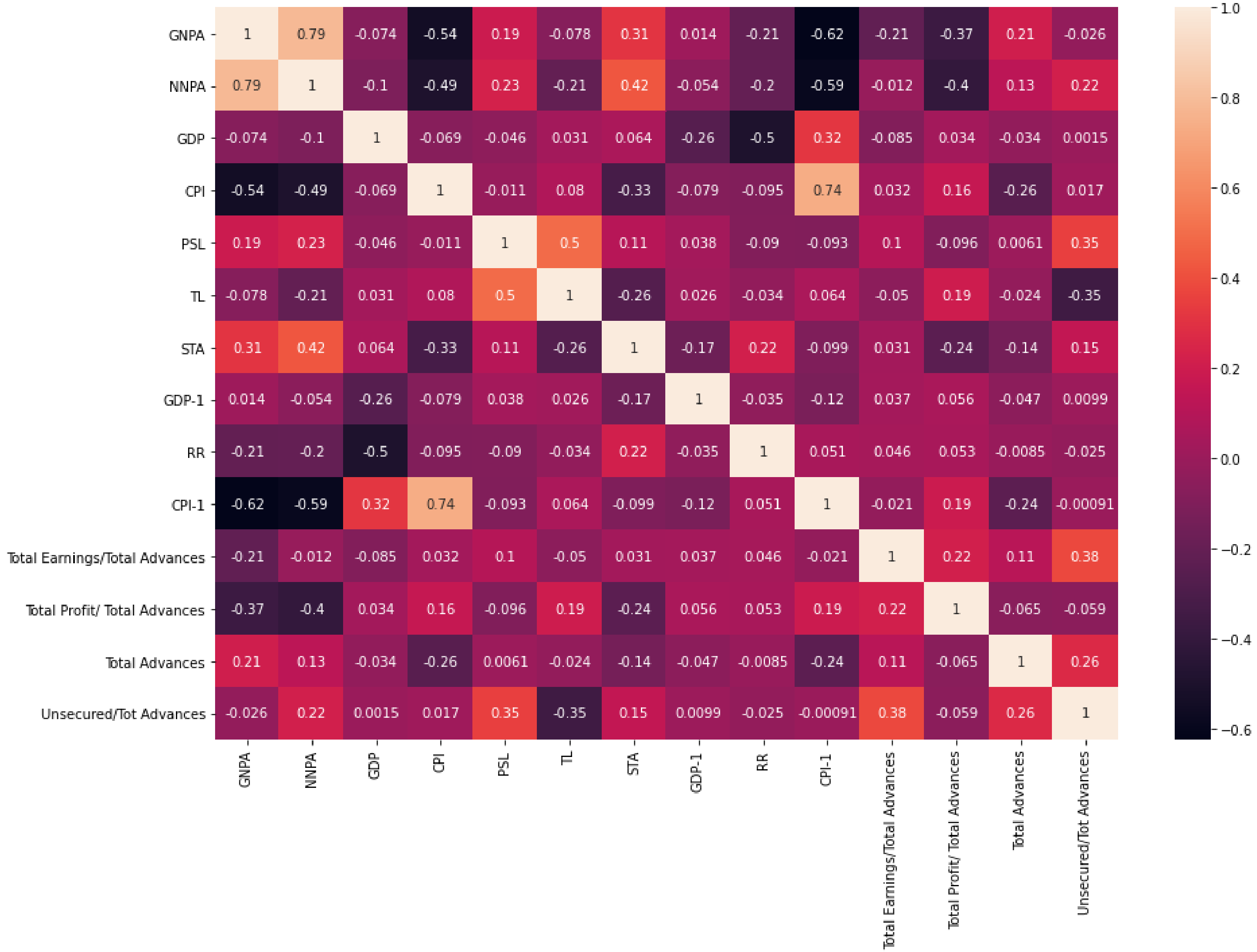
Table 4.5: Correlation matrix of Artificial neural network regression for GNPA



	GNPA	NNPA	GDP	CPI	PSL	TL	STA	GDP-1	RR	CPI-1	Total Earnings/Total Advances	Total Profit/ Total Advances	Total Advances	Unsecured/Tot Advances
GNPA	1	0.9	-0.11	-0.5	0.22	-0.11	0.36	-0.036	-0.22	-0.6	-0.16	-0.38	0.099	0.0085
NNPA	0.9	1	-0.1	-0.49	0.23	-0.21	0.42	-0.054	-0.2	-0.59	-0.012	-0.4	0.13	0.22
GDP	-0.11	-0.1	1	-0.069	-0.046	0.031	0.064	-0.26	-0.5	0.32	-0.085	0.034	-0.034	0.0015
CPI	-0.5	-0.49	-0.069	1	-0.011	0.08	-0.33	-0.079	-0.095	0.74	0.032	0.16	-0.26	0.017
PSL	0.22	0.23	-0.046	-0.011	1	0.5	0.11	0.038	-0.09	-0.093	0.1	-0.096	0.0061	0.35
TL	-0.11	-0.21	0.031	0.08	0.5	1	-0.26	0.026	-0.034	0.064	-0.05	0.19	-0.024	-0.35
STA	0.36	0.42	0.064	-0.33	0.11	-0.26	1	-0.17	0.22	-0.099	0.031	-0.24	-0.14	0.15
GDP-1	-0.036	-0.054	-0.26	-0.079	0.038	0.026	-0.17	1	-0.035	-0.12	0.037	0.056	-0.047	0.0099
RR	-0.22	-0.2	-0.5	-0.095	-0.09	-0.034	0.22	-0.035	1	0.051	0.046	0.053	-0.0085	-0.025
CPI-1	-0.6	-0.59	0.32	0.74	-0.093	0.064	-0.099	-0.12	0.051	1	-0.021	0.19	-0.24	-0.00091
Total Earnings/Total Advances	-0.16	-0.012	-0.085	0.032	0.1	-0.05	0.031	0.037	0.046	-0.021	1	0.22	0.11	0.38
Total Profit/ Total Advances	-0.38	-0.4	0.034	0.16	-0.096	0.19	-0.24	0.056	0.053	0.19	0.22	1	-0.065	-0.059
Total Advances	0.099	0.13	-0.034	-0.26	0.0061	-0.024	-0.14	-0.047	-0.0085	-0.24	0.11	-0.065	1	0.26
Unsecured/Tot Advances	0.0085	0.22	0.0015	0.017	0.35	-0.35	0.15	0.0099	-0.025	-0.00091	0.38	-0.059	0.26	1

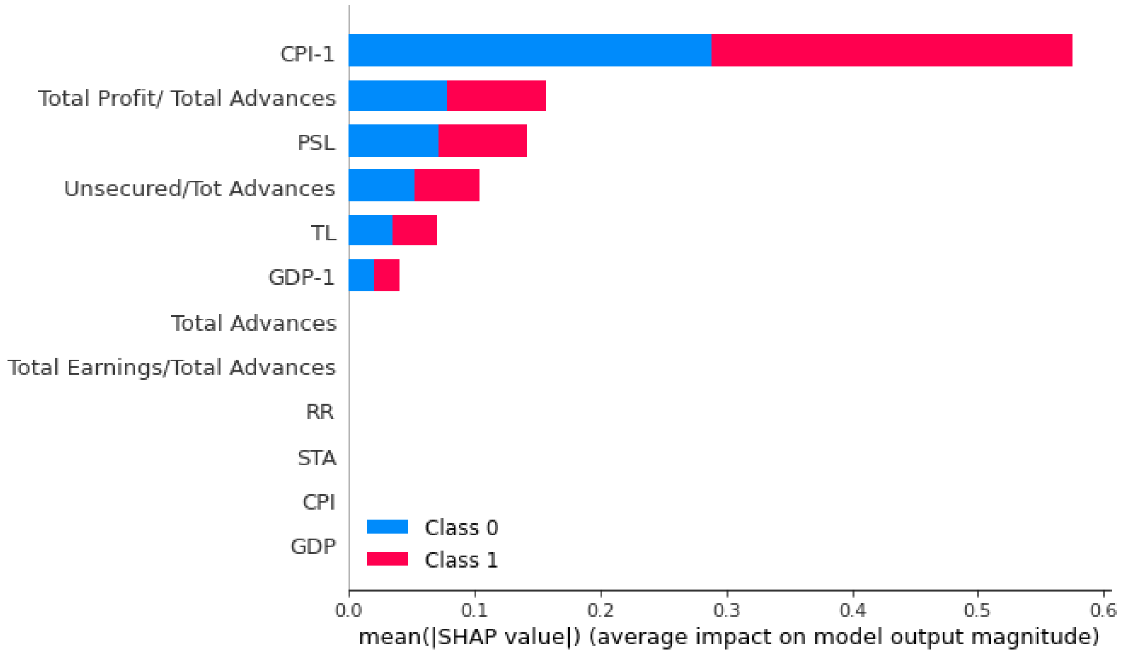
Table 4.5 evaluates the correlation matrix of ANN regression as per a dataset about different banks. It was predicted how the Gross NPA is getting affected by different variables. In this context, deployed an ANN to find out the relation and quantification with the dependent variable (GNPA) with other independent variables (GDP, CPI, PSL, TL, STA, GDP1, RR, CPI1, Total Earnings Total Advances, Total Profit Total Advances, Total Advances, Unsecured Total Advances). It obtained a MAPE of 2508944% using the ANN Regression, meaning that this data set is not suitable for Artificial Neural Networks. One of the possible reasons could be the meagre quantity of the data.

Table 4.6: Correlation matrix of Decision tree classification for GNPA



	GNPA	NNPA	GDP	CPI	PSL	TL	STA	GDP-1	RR	CPI-1	Total Earnings/Total Advances	Total Profit/ Total Advances	Total Advances	Unsecured/Tot Advances
GNPA	1	0.79	-0.074	-0.54	0.19	-0.078	0.31	0.014	-0.21	-0.62	-0.21	-0.37	0.21	-0.026
NNPA	0.79	1	-0.1	-0.49	0.23	-0.21	0.42	-0.054	-0.2	-0.59	-0.012	-0.4	0.13	0.22
GDP	-0.074	-0.1	1	-0.069	-0.046	0.031	0.064	-0.26	-0.5	0.32	-0.085	0.034	-0.034	0.0015
CPI	-0.54	-0.49	-0.069	1	-0.011	0.08	-0.33	-0.079	-0.095	0.74	0.032	0.16	-0.26	0.017
PSL	0.19	0.23	-0.046	-0.011	1	0.5	0.11	0.038	-0.09	-0.093	0.1	-0.096	0.0061	0.35
TL	-0.078	-0.21	0.031	0.08	0.5	1	-0.26	0.026	-0.034	0.064	-0.05	0.19	-0.024	-0.35
STA	0.31	0.42	0.064	-0.33	0.11	-0.26	1	-0.17	0.22	-0.099	0.031	-0.24	-0.14	0.15
GDP-1	0.014	-0.054	-0.26	-0.079	0.038	0.026	-0.17	1	-0.035	-0.12	0.037	0.056	-0.047	0.0099
RR	-0.21	-0.2	-0.5	-0.095	-0.09	-0.034	0.22	-0.035	1	0.051	0.046	0.053	-0.0085	-0.025
CPI-1	-0.62	-0.59	0.32	0.74	-0.093	0.064	-0.099	-0.12	0.051	1	-0.021	0.19	-0.24	-0.00091
Total Earnings/Total Advances	-0.21	-0.012	-0.085	0.032	0.1	-0.05	0.031	0.037	0.046	-0.021	1	0.22	0.11	0.38
Total Profit/ Total Advances	-0.37	-0.4	0.034	0.16	-0.096	0.19	-0.24	0.056	0.053	0.19	0.22	1	-0.065	-0.059
Total Advances	0.21	0.13	-0.034	-0.26	0.0061	-0.024	-0.14	-0.047	-0.0085	-0.24	0.11	-0.065	1	0.26
Unsecured/Tot Advances	-0.026	0.22	0.0015	0.017	0.35	-0.35	0.15	0.0099	-0.025	-0.00091	0.38	-0.059	0.26	1

Table 4.6 evaluates the correlation matrix of Decision tree classification as per a dataset about different banks. It was predicted how the Gross NPA is getting affected by different variables. In this context, deployed to find out the relation and quantification with the dependent variable (GNPA) would be 1 (High) or 0 (Low) based on independent variables (GDP, CPI, PSL, TL, STA, GDP1, RR, CPI1, Total Earnings Total Advances, Total Profit Total Advances, Total Advances, Unsecured Total Advances). We have used maximum depth till 3 to restrict overfitting of the model with Decision Tree Classification, which seems to be a more reasonable model than a 100% fit. The reason for overfitting is a dearth of data which has prevented us from segregating the data into training and testing. As per the feature importance, CPI-1 and Total Profit/ Total Advances are the 2 most important variables.

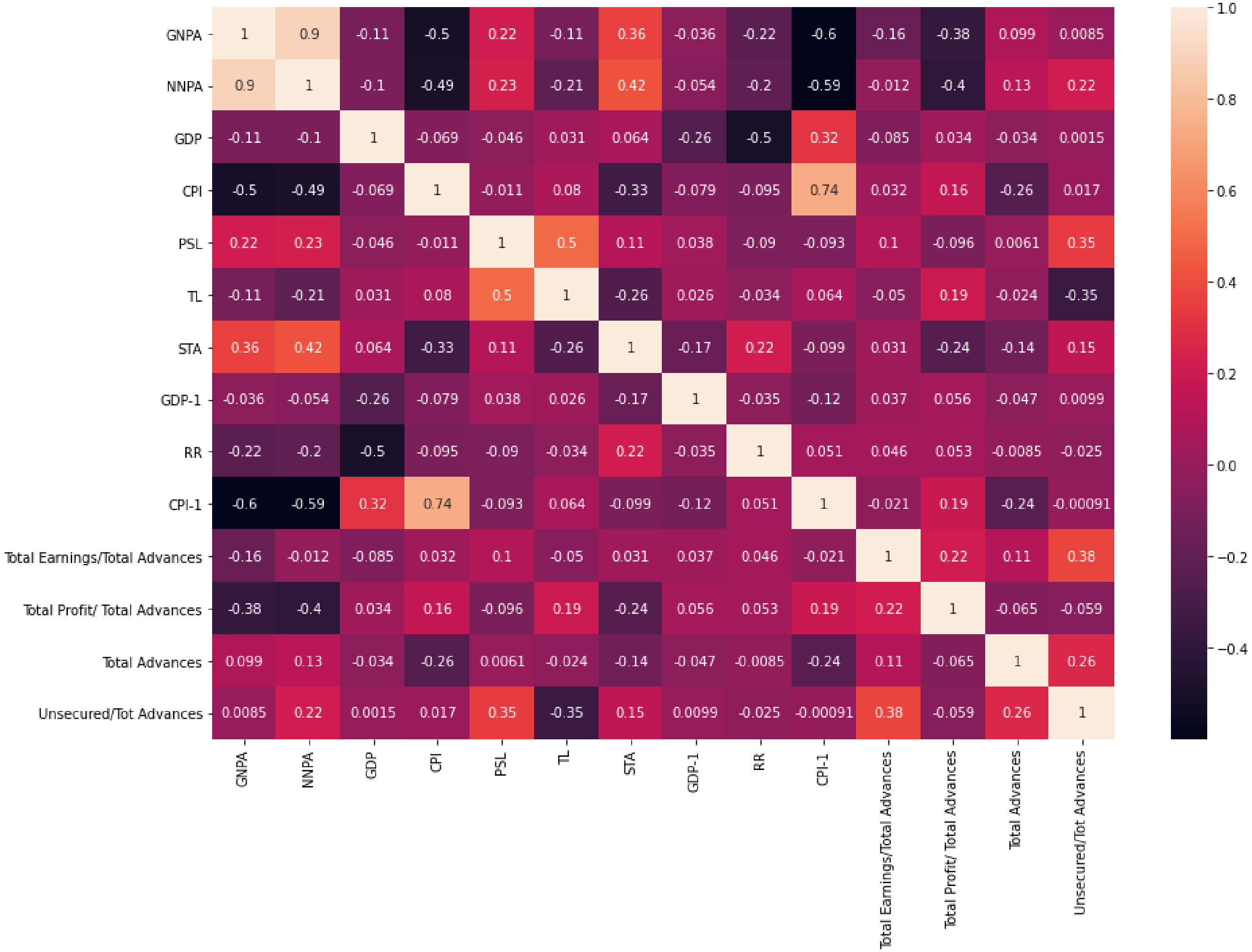


.

Figure 4.13: Graphical representation of SHAP values for independent variables

Features are arranged as per their importance; CPI-1 happens to be the most important feature followed by Total Profit/Total Advances Low value of CPI-1 (blue part) will push the output (GNPA) to be 1 (means chances to be NPA more). The high value of Total Profit/Total Advances (red part) will push the output (GNPA) to be 0 (means chances to be NPA less). Similarly, we can interpret the rest of the features (Fig 4.13).

Table 4.7: Correlation matrix of Random Forest Regression for GNPA



	GNPA	NNPA	GDP	CPI	PSL	TL	STA	GDP-1	RR	CPI-1	Total Earnings/Total Advances	Total Profit/ Total Advances	Total Advances	Unsecured/Tot Advances
GNPA	1	0.9	-0.11	-0.5	0.22	-0.11	0.36	-0.036	-0.22	-0.6	-0.16	-0.38	0.099	0.0085
NNPA	0.9	1	-0.1	-0.49	0.23	-0.21	0.42	-0.054	-0.2	-0.59	-0.012	-0.4	0.13	0.22
GDP	-0.11	-0.1	1	-0.069	-0.046	0.031	0.064	-0.26	-0.5	0.32	-0.085	0.034	-0.034	0.0015
CPI	-0.5	-0.49	-0.069	1	-0.011	0.08	-0.33	-0.079	-0.095	0.74	0.032	0.16	-0.26	0.017
PSL	0.22	0.23	-0.046	-0.011	1	0.5	0.11	0.038	-0.09	-0.093	0.1	-0.096	0.0061	0.35
TL	-0.11	-0.21	0.031	0.08	0.5	1	-0.26	0.026	-0.034	0.064	-0.05	0.19	-0.024	-0.35
STA	0.36	0.42	0.064	-0.33	0.11	-0.26	1	-0.17	0.22	-0.099	0.031	-0.24	-0.14	0.15
GDP-1	-0.036	-0.054	-0.26	-0.079	0.038	0.026	-0.17	1	-0.035	-0.12	0.037	0.056	-0.047	0.0099
RR	-0.22	-0.2	-0.5	-0.095	-0.09	-0.034	0.22	-0.035	1	0.051	0.046	0.053	-0.0085	-0.025
CPI-1	-0.6	-0.59	0.32	0.74	-0.093	0.064	-0.099	-0.12	0.051	1	-0.021	0.19	-0.24	-0.00091
Total Earnings/Total Advances	-0.16	-0.012	-0.085	0.032	0.1	-0.05	0.031	0.037	0.046	-0.021	1	0.22	0.11	0.38
Total Profit/ Total Advances	-0.38	-0.4	0.034	0.16	-0.096	0.19	-0.24	0.056	0.053	0.19	0.22	1	-0.065	-0.059
Total Advances	0.099	0.13	-0.034	-0.26	0.0061	-0.024	-0.14	-0.047	-0.0085	-0.24	0.11	-0.065	1	0.26
Unsecured/Tot Advances	0.0085	0.22	0.0015	0.017	0.35	-0.35	0.15	0.0099	-0.025	-0.00091	0.38	-0.059	0.26	1

Table 4.7 evaluates the correlation matrix of Random Forest regression as per a dataset about different banks. It was predicted how the Gross NPA is affected by different variables. In this context, deployed to find out the relation and quantification with the dependent variable (GNPA) with independent variables (GDP, CPI, PSL, TL, STA, GDP1, RR, CPI1, Total Earnings Total Advances, Total Profit Total Advances, Total Advances, Unsecured Total Advances). We have obtained a MAPE of 11% using the Random Forest Regression, meaning accuracy is about 89%.

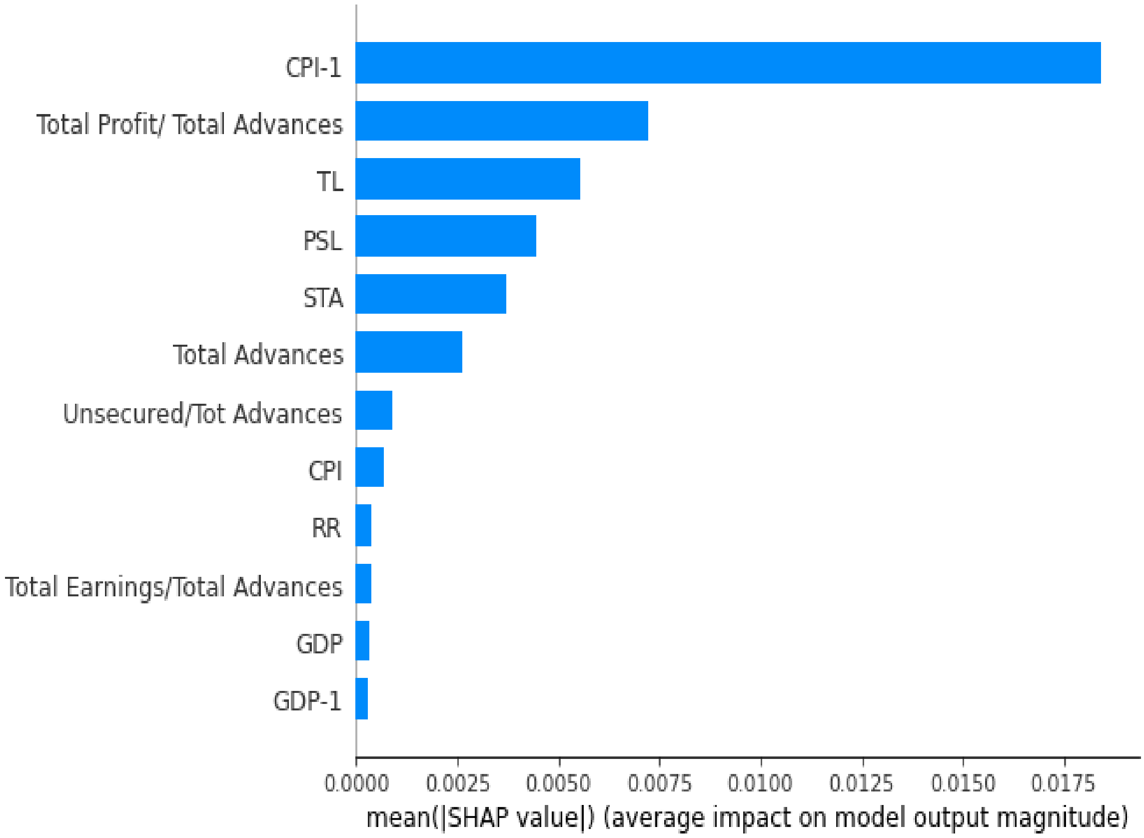
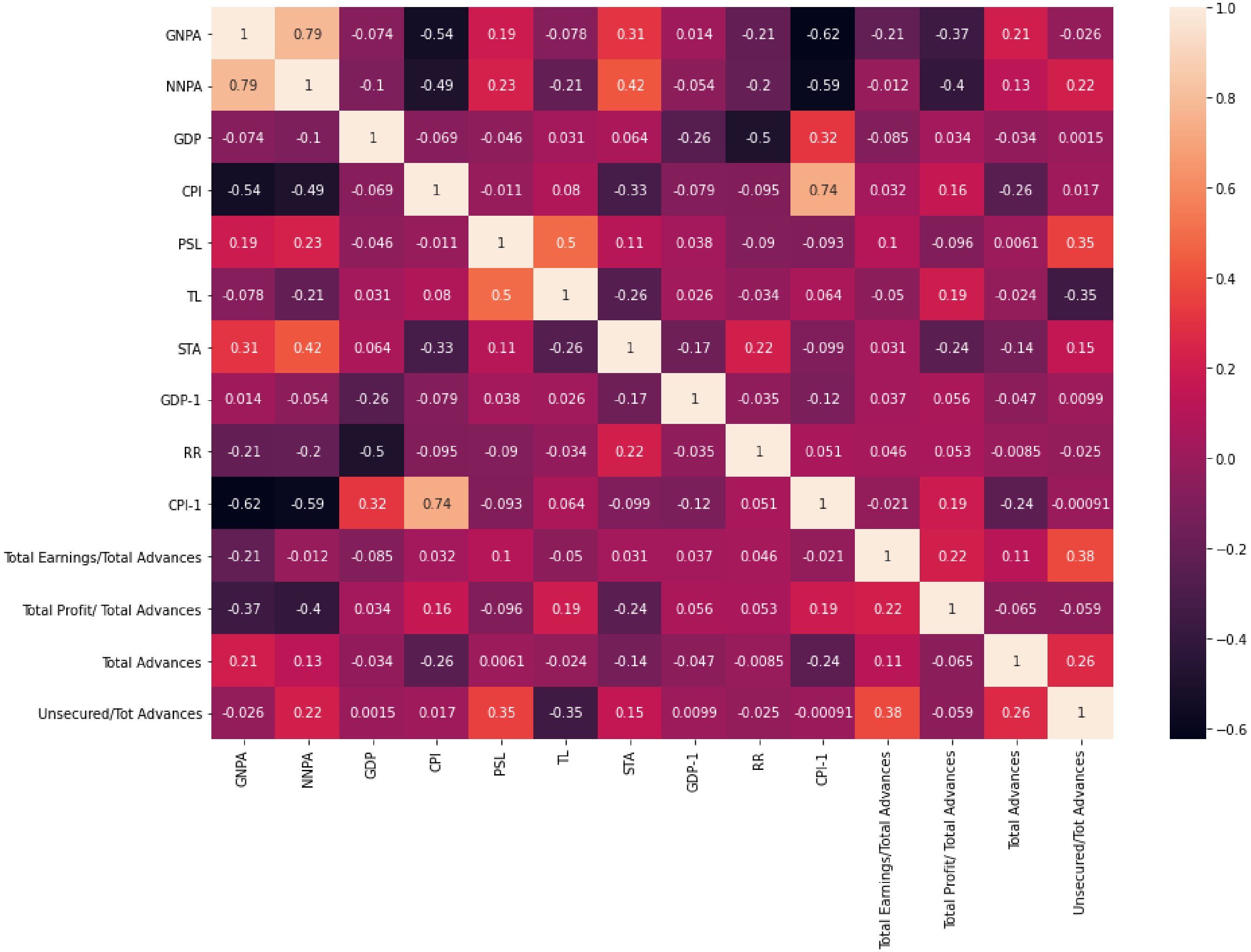


Figure 4.14: Graphical representation of SHAP values for independent variables

Features are arranged as per their importance; CPI-1 happens to be the most important feature followed by Total Profit/Total Advances. The low value of CPI-1 (blue part) will push the output (GNPA) to be more. The high value of Total Profit/Total Advances (red part) will push the output (GNPA) to be less. Similarly, we can interpret the rest of the features (Fig 4.14).

Table 4.8: Correlation matrix of Naïve Bayes Classification for GNPA



	GNPA	NNPA	GDP	CPI	PSL	TL	STA	GDP-1	RR	CPI-1	Total Earnings/Total Advances	Total Profit/ Total Advances	Total Advances	Unsecured/Tot Advances
GNPA	1	0.79	-0.074	-0.54	0.19	-0.078	0.31	0.014	-0.21	-0.62	-0.21	-0.37	0.21	-0.026
NNPA	0.79	1	-0.1	-0.49	0.23	-0.21	0.42	-0.054	-0.2	-0.59	-0.012	-0.4	0.13	0.22
GDP	-0.074	-0.1	1	-0.069	-0.046	0.031	0.064	-0.26	-0.5	0.32	-0.085	0.034	-0.034	0.0015
CPI	-0.54	-0.49	-0.069	1	-0.011	0.08	-0.33	-0.079	-0.095	0.74	0.032	0.16	-0.26	0.017
PSL	0.19	0.23	-0.046	-0.011	1	0.5	0.11	0.038	-0.09	-0.093	0.1	-0.096	0.0061	0.35
TL	-0.078	-0.21	0.031	0.08	0.5	1	-0.26	0.026	-0.034	0.064	-0.05	0.19	-0.024	-0.35
STA	0.31	0.42	0.064	-0.33	0.11	-0.26	1	-0.17	0.22	-0.099	0.031	-0.24	-0.14	0.15
GDP-1	0.014	-0.054	-0.26	-0.079	0.038	0.026	-0.17	1	-0.035	-0.12	0.037	0.056	-0.047	0.0099
RR	-0.21	-0.2	-0.5	-0.095	-0.09	-0.034	0.22	-0.035	1	0.051	0.046	0.053	-0.0085	-0.025
CPI-1	-0.62	-0.59	0.32	0.74	-0.093	0.064	-0.099	-0.12	0.051	1	-0.021	0.19	-0.24	-0.00091
Total Earnings/Total Advances	-0.21	-0.012	-0.085	0.032	0.1	-0.05	0.031	0.037	0.046	-0.021	1	0.22	0.11	0.38
Total Profit/ Total Advances	-0.37	-0.4	0.034	0.16	-0.096	0.19	-0.24	0.056	0.053	0.19	0.22	1	-0.065	-0.059
Total Advances	0.21	0.13	-0.034	-0.26	0.0061	-0.024	-0.14	-0.047	-0.0085	-0.24	0.11	-0.065	1	0.26
Unsecured/Tot Advances	-0.026	0.22	0.0015	0.017	0.35	-0.35	0.15	0.0099	-0.025	-0.00091	0.38	-0.059	0.26	1

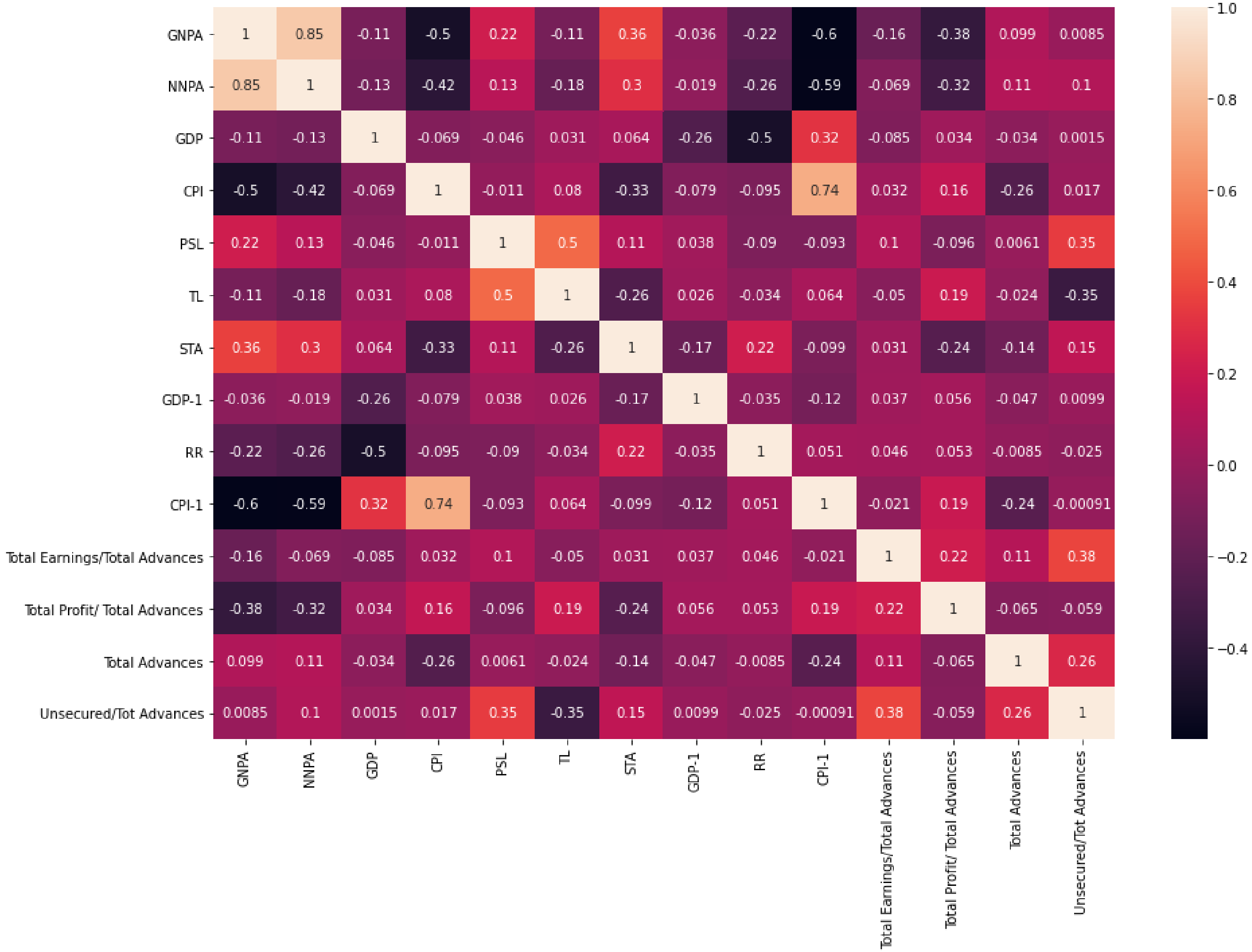
Table 4.8 evaluates the correlation matrix of Naïve Bayes Classification as per a dataset about different banks. It was predicted how the Gross NPA is getting affected by different variables. In this context, deployed to find out the relation and quantification with the dependent variable (GNPA) with independent variables (GDP, CPI, PSL, TL, STA, GDP1, RR, CPI1, Total Earnings Total Advances, Total Profit Total Advances, Total Advances, Unsecured-Total Advances). We have obtained an accuracy of about 86% using the Naive Bayes Classification. This seems like a more realistic model.

Table 4.9: Correlation Matrix of Random Forest Classification for GNPA

	NNPA	GDP	CPI	PSL	TL	STA	GDP-1	RR	CPI-1	Total Earnings/Total Advances	Total Profit/ Total Advances	Total Advances	Unsecured/Tot Advances
NNPA	1	-0.13	-0.42	0.13	-0.18	0.3	-0.019	-0.26	-0.59	-0.069	-0.32	0.1	0.1
GDP	-0.13	1	-0.069	-0.046	0.031	0.064	-0.26	-0.5	0.32	-0.085	0.034	-0.031	0.0015
CPI	-0.42	-0.069	1	-0.011	0.08	-0.33	-0.079	-0.095	0.74	0.032	0.16	-0.23	0.017
PSL	0.13	-0.046	-0.011	1	0.5	0.11	0.038	-0.09	-0.093	0.1	-0.096	0.0058	0.35
TL	-0.18	0.031	0.08	0.5	1	-0.26	0.026	-0.034	0.064	-0.05	0.19	-0.021	-0.35
STA	0.3	0.064	-0.33	0.11	-0.26	1	-0.17	0.22	-0.099	0.031	-0.24	-0.13	0.15
GDP-1	-0.019	-0.26	-0.079	0.038	0.026	-0.17	1	-0.035	-0.12	0.037	0.056	-0.043	0.0099
RR	-0.26	-0.5	-0.095	-0.09	-0.034	0.22	-0.035	1	0.051	0.046	0.053	-0.0077	-0.025
CPI-1	-0.59	0.32	0.74	-0.093	0.064	-0.099	-0.12	0.051	1	-0.021	0.19	-0.22	-0.00091
Total Earnings/Total Advances	-0.069	-0.085	0.032	0.1	-0.05	0.031	0.037	0.046	-0.021	1	0.22	0.088	0.38
Total Profit/ Total Advances	-0.32	0.034	0.16	-0.096	0.19	-0.24	0.056	0.053	0.19	0.22	1	-0.058	-0.059
Total Advances	0.1	-0.031	-0.23	0.0058	-0.021	-0.13	-0.043	-0.0077	-0.22	0.088	-0.058	1	0.25
Unsecured/Tot Advances	0.1	0.0015	0.017	0.35	-0.35	0.15	0.0099	-0.025	-0.00091	0.38	-0.059	0.25	1

Table 4.9 evaluates the correlation matrix of SVM Classification as per a dataset about different banks. It was predicted how the Gross NPA is getting affected by different variables. In this context, deployed to find out the relation and quantification with the dependent variable (GNPA) with independent variables (GDP, CPI, PSL, TL, STA, GDP1, RR, CPI1, Total Earnings Total Advances, Total Profit Total Advances, Total Advances, Unsecured Total Advances). We have obtained an accuracy of about 100% using the Random Forest Classification. This is a realistic model. The reason for overfitting is a dearth of data which has prevented us from segregating the data into training and testing. As per the feature importance, Total Profit/ Total Advances and CPI-1 are the 2 most important variables.

Table 4.10: Correlation matrix of Support Vector Classification for GNPA



4.13.1 Correlation matrix to check interdependency among variables by using Random Forest Regression

Table 4.11 evaluates the correlation matrix of RF regression as per a dataset about different banks regarding GNPA. We have a dataset about different banks. We are trying to predict how the Gross NPA is getting effected by different variables. We deployed a Random Forest regression to find out the relation and quantification with our dependent variable - GNPA with other independent variables.

In this case, Decision Tree is giving 0% error; clear case of over fitting; hence ignored. XG Boost, ANN, SVM were predicted with too high error rate, clear case of under fitted, hence not suitable for the given dataset. Random Forest seems the only realistic error, and best suited algorithm to predict the NNPA. Hence, this is used to predict the results. We got a MAPE of 11.5% using the Random Forest Regression with an accuracy of 86%.

Table 4.11: Correlation matrix of Random Forest regression for GNPA

	GNPA	NNPA	GDP	CPI	PSL	TL	STA	GDP-1	RR	CPI-1	Total Earnings/Total Advances	Total Profit/ Total Advances	Total Advances	Unsecured/Tot Advances
GNPA	1	0.83	-0.073	0.24	0.31	-0.15	0.42	-0.11	-0.3	0.22	-0.099	-0.38	0.0071	-0.09
NNPA	0.83	1	-0.11	-0.016	0.22	-0.25	0.46	-0.072	-0.13	-0.084	-0.015	-0.43	-0.032	0.21
GDP	-0.073	-0.11	1	0.12	-0.0056	0.045	0.057	-0.23	-0.42	0.082	-0.069	0.047	0.015	-0.024
CPI	0.24	-0.016	0.12	1	0.21	0.11	-0.016	-0.37	-0.65	0.77	0.13	0.032	0.26	-0.25
PSL	0.31	0.22	-0.0056	0.21	1	0.44	0.17	0.012	-0.22	0.32	0.14	-0.071	-0.028	0.18
TL	-0.15	-0.25	0.045	0.11	0.44	1	-0.3	0.0089	-0.1	0.14	-0.028	0.23	0.13	-0.36
STA	0.42	0.46	0.057	-0.016	0.17	-0.3	1	-0.15	0.16	-0.07	0.031	-0.28	-0.27	0.11
GDP-1	-0.11	-0.072	-0.23	-0.37	0.012	0.0089	-0.15	1	0.2	-0.13	-0.0057	0.07	-0.058	0.058
RR	-0.3	-0.13	-0.42	-0.65	-0.22	-0.1	0.16	0.2	1	-0.55	-0.055	0.0089	-0.21	0.16
CPI-1	0.22	-0.084	0.082	0.77	0.32	0.14	-0.07	-0.13	-0.55	1	0.13	0.1	0.38	-0.31
Total Earnings/Total Advances	-0.099	-0.015	-0.069	0.13	0.14	-0.028	0.031	-0.0057	-0.055	0.13	1	0.22	0.16	0.31
Total Profit/ Total Advances	-0.38	-0.43	0.047	0.032	-0.071	0.23	-0.28	0.07	0.0089	0.1	0.22	1	0.048	-0.071
Total Advances	0.0071	-0.032	0.015	0.26	-0.028	0.13	-0.27	-0.058	-0.21	0.38	0.16	0.048	1	0.046
Unsecured/Tot Advances	-0.09	0.21	-0.024	-0.25	0.18	-0.36	0.11	0.058	0.16	-0.31	0.31	-0.071	0.046	1

Predicting GNPA using regression machine learning model

The results of the Regression and Classification Models showed different degree of Accuracy measured as 1-MAPE (Mean Absolute Percentage Error) when the testing the relationship of independent variables with Gross Non-Performing Assets (GNPA). Random Forest Model showed the most reasonable MAPE of 11.5% while Decision Tree showed MAPE of 0%, which could be attributed to over-fitting of data set. Other models such as XG Boost, SVM and ANN have shown high MAPE values, which have not been considered due to high error values. However, given that the data set used in research is limited, and machine learning models results are most optimal with large data sets, it’s recommended that these models should be considered by banks for setting up risk appetite framework.

Table 4.12

ML Algorithm	MAPE (Error Rate)
Random Forest	11.5%
Decision Tree	0%
XG Boost	75%
ANN	NM
SVM	NM

NM: High error rates plausibly due to data limitations

Predicting NNPA using regression machine learning model

Machine learning regression model results to predict the net non-performing assets (NNPA) were similar with Random Forest model giving MAPE of 22.7% (Accuracy = 77.3%), which was the most reasonable fit amongst all the Models that were tested. Decision Tree results showed the lowest MAPE of 0% (Accuracy = 100%), as it was observed in GNPA predicted values. This can be attributed to over-fitting of data set. Other models such as XG Boost, SVM and ANN have shown high MAPE values, which have not been considered due to high error values. However, given that the data set used in research is limited, and machine learning models results are most optimal with large data sets, it's recommended that these models should be considered by banks for setting up risk appetite framework.

Table 4.13

ML Algorithm	**MAPE (Error Rate)**
Random Forest	22.70%
Decision Tree	0%
XG Boost	229%
ANN	NM
SVM	NM

NM: High error rates plausibly due to data limitations

Cohort Assessment: Predicting GNPA using Regression Random Forest Model

Machine learning regression Random Forest model was trained to further assess whether the predicted value of GNPA further improves by segregating data set based on risk-tiering (cohorts). For this purpose, data set comprising of 8 public sector banks was segregated into 4 Cohorts. Cohorts are designed in terms of size of banks' credit advances to ensure that similar sized banks are clubbed together. This Cohort Approach has been used to test whether different risk limits be designed by type of assets in risk appetite framework. Cohort 1 comprises of State Bank of India, Bank of Baroda and Canara Bank. Cohort 2 comprises of Bank of India, Andhra Bank. Cohort 3 is comprised of Central Bank of India and Allahabad Bank. Cohort 4 comprised of Bank of Maharashtra. GNPA predicted values have varying degree of MAPE, with Cohort 1 MAPE of 8.45% (Accuracy = 93.55%) to Cohort 4 with MAPE of 21.33% (Accuracy = 79.67%).

Table 4.14

Cohorts	MAPE (Error Rate)
1	8.45%
2	13.25%
3	9.91%
4	21.33%
Aggregate	**11.5%**

Cohort Assessment: Predicting NNPA using Regression Random Forest Model

Machine learning regression Random Forest model was trained to further assess whether the predicted value of NNPA further improves by segregating data set based on risk-tiering (cohorts). For this purpose, data set comprising of 8 public sector banks was segregated into 4 cohorts. NNPA predicted values have varying degree of MAPE, with Cohort 1 MAPE of 13.07% (Accuracy = 76.93%) to Cohort 2 with MAPE of 39.5% (Accuracy = 60.5%).

Table 4.15

Cohorts	MAPE (Error Rate)
1	13.07%
2	39.5%
3	17.36%
4	32.91%
Aggregate	**22.7%**

Predicting accuracy of GNPA classification using classification machine learning model

Machine learning classification models were tested to compute the accuracy of whether the dependent variable (GNPA) was correctly classified as GNPA, based on relationship of the variable with independent variables. The accuracy rates, which is defined by Precision value were found to be highest in case of Random Forest (100%) followed by Decision Tree and XG Boost (97.2%), SVM (95%), KNN (93%), Naïve Bayes (86%) and ANN (65%). It can be concluded that XG Boost and Decision Tree have yielded optimum accuracy rates and can be considered as good fit models. Random Forest with precision of 100% indicates model over-fitment. SVM, KNN and Naïve Bayes models should be considered in risk appetite framework development, as these models have high precision values and given the limited data sets available in this research, should be considered.

Table 4.16

ML Algorithm	**Accuracy**
Random Forest	100.00%
Decision Tree	97.20%
XG Boost	97.20%
Naive Bayes	86%
SVM	95%
ANN	65%
KNN	93%

Decision Tree : Key Metrics

Decision Tree

```
from sklearn.tree import DecisionTreeClassifier
classifier = DecisionTreeClassifier(max_depth=3,random_state=0)
classifier.fit(X, y)
#predicted values
y_pred = classifier.predict(X)
print(classification_report(y,y_pred))
```

```
              precision    recall  f1-score   support

           0       0.98      0.98      0.98        94
           1       0.96      0.96      0.96        50

    accuracy                           0.97       144
   macro avg       0.97      0.97      0.97       144
weighted avg       0.97      0.97      0.97       144
```

Predicting accuracy of NNPA classification using classification machine learning model

Machine learning classification models were tested to compute the accuracy of whether the dependent variable (NNPA) was correctly classified as NNPA, based on relationship of the variable with independent variables. The accuracy rates, which is defined by Precision value were found to be highest in case of Random Forest (100%) followed by Decision Tree and XG Boost (96.5%), SVM (95%), KNN (92%), Naïve Bayes (88%) and ANN (78%). It can be concluded that XG Boost and Decision Tree have yielded optimum accuracy rates and can be considered as good fit models. Random Forest with precision of 100% indicates model over-fitment. SVM, KNN and Naïve Bayes models should be considered in risk appetite framework development, as these models have high precision values and given the limited data sets available in this research, should be considered.

Table 4.17

ML Algorithm	**Accuracy**
Random Forest	100.00%
Decision Tree	96.50%
XG Boost	96.50%
Naive Bayes	88%
SVM	95%
ANN	78%
KNN	92%

Cohort assessment: Predicting accuracy of GNPA classification using Decision Tree machine learning model

Machine learning classification Decision Tree model was trained to further assess whether the classification of GNPA further improves by segregating data set based on risk-tiering (cohorts). For this purpose, data set comprising of 8 public sector banks was segregated into 4 cohorts. GNPA classification further improved by using the cohort approach with each cohort estimating 100% precision value.

Table 4.18

Cohorts	**Accuracy**
1	100%
2	100%
3	100%
4	100%
Aggregate	**96.5%**

Cohort assessment: Predicting accuracy of NNPA classification using Decision Tree machine learning model

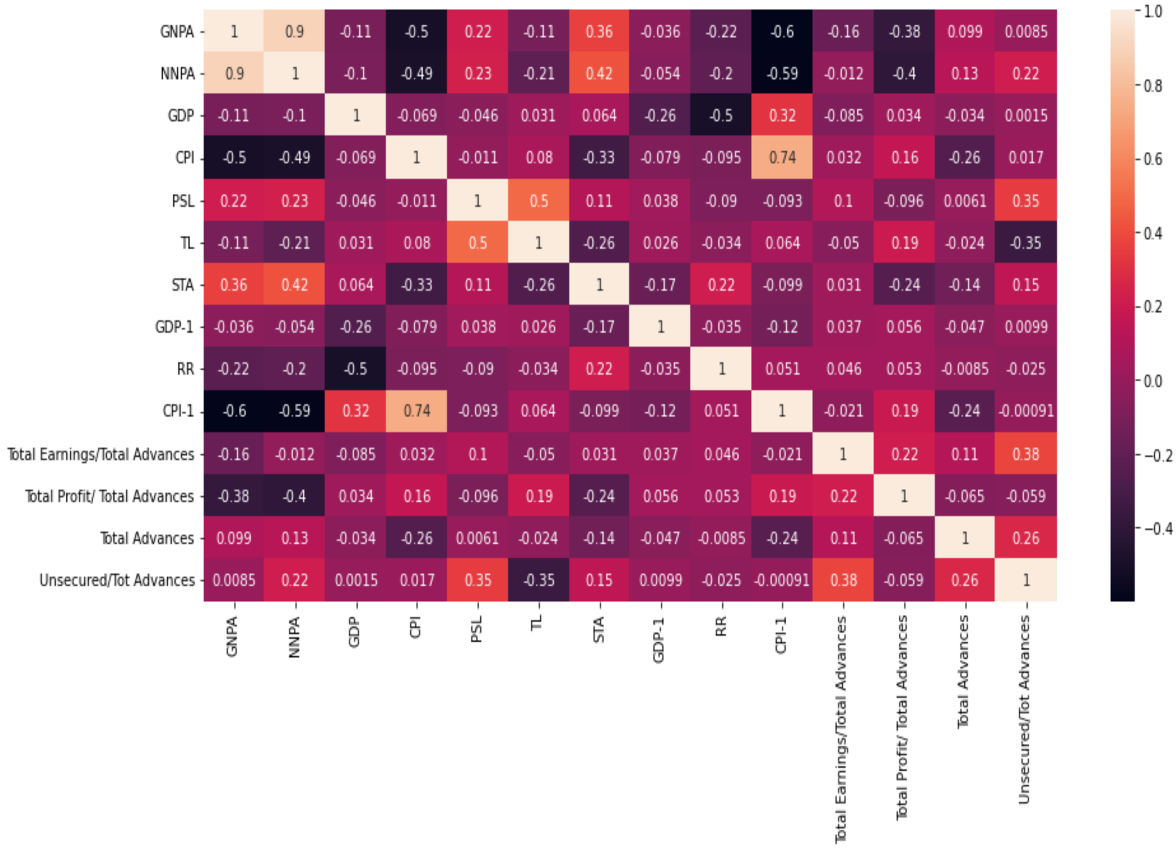
Machine learning classification Decision Tree model was trained to further assess whether the classification of NNPA further improves by segregating data set based on risk-tiering (cohorts). For this purpose, data set comprising of 8 public sector banks was segregated into 4 cohorts. GNPA classification further improved by using the cohort approach with each cohort estimating 100% precision value, except Cohort 1 which estimated precision value of 97.2%.

Table 4.19

Cohorts	**Accuracy**
1	97.2%
2	100%
3	100%
4	100%
Aggregate	**96.5%**

Table 4.20 evaluates the correlation matrix of RF regression as per a dataset about different banks regarding NNPA.

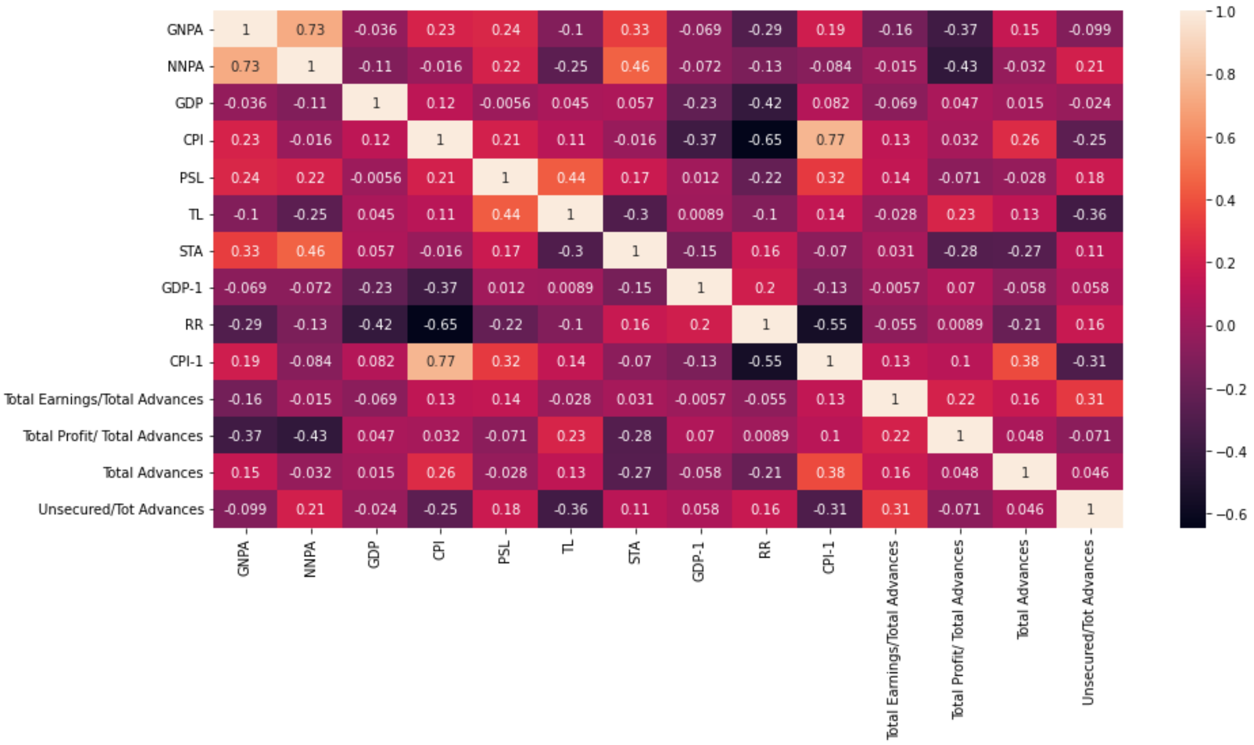
Table 4.20: Correlation matrix of Random Forest regression for NNPA



4.13.2 Correlation matrix to check interdependency among variables by using Random Forest Classification

Table 4.21 evaluates the correlation matrix of RF classification as per a dataset about different banks regarding GNPA.

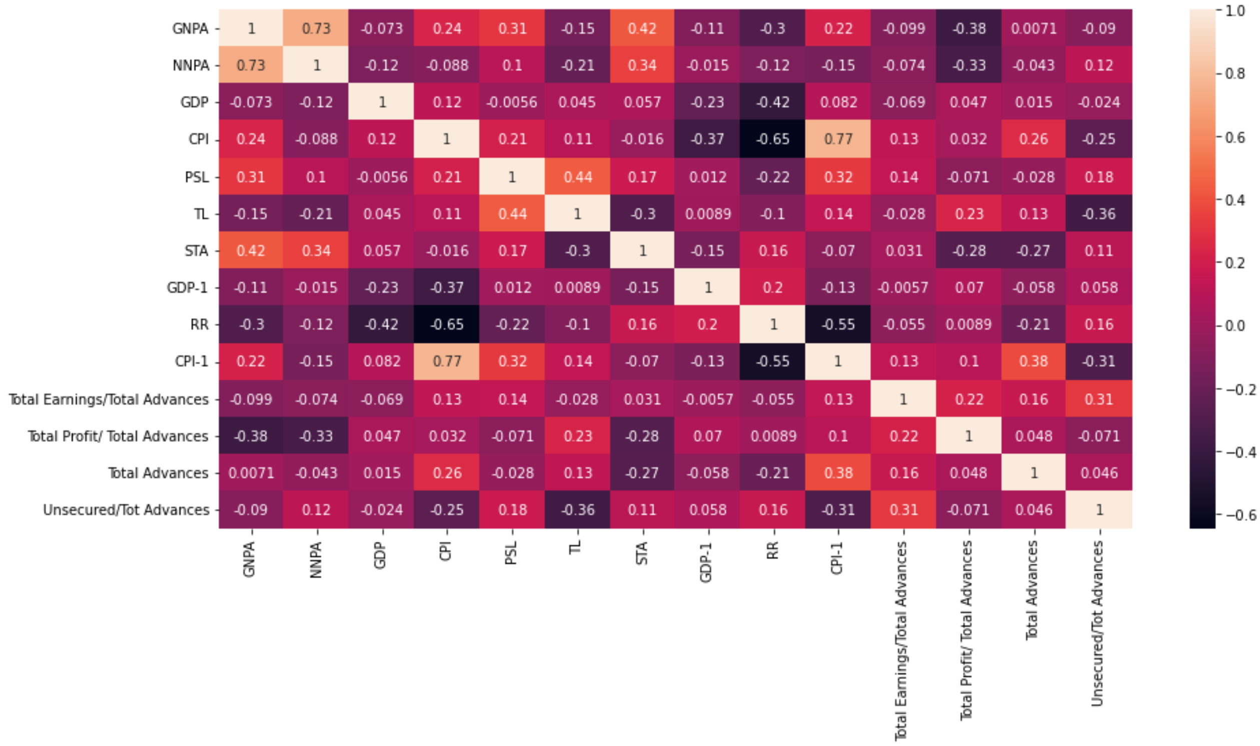
Table 4.21: Correlation matrix of Random Forest classification for GNPA



	GNPA	NNPA	GDP	CPI	PSL	TL	STA	GDP-1	RR	CPI-1	Total Earnings/Total Advances	Total Profit/ Total Advances	Total Advances	Unsecured/Tot Advances
GNPA	1	0.73	-0.036	0.23	0.24	-0.1	0.33	-0.069	-0.29	0.19	-0.16	-0.37	0.15	-0.099
NNPA	0.73	1	-0.11	-0.016	0.22	-0.25	0.46	-0.072	-0.13	-0.084	-0.015	-0.43	-0.032	0.21
GDP	-0.036	-0.11	1	0.12	-0.0056	0.045	0.057	-0.23	-0.42	0.082	-0.069	0.047	0.015	-0.024
CPI	0.23	-0.016	0.12	1	0.21	0.11	-0.016	-0.37	-0.65	0.77	0.13	0.032	0.26	-0.25
PSL	0.24	0.22	-0.0056	0.21	1	0.44	0.17	0.012	-0.22	0.32	0.14	-0.071	-0.028	0.18
TL	-0.1	-0.25	0.045	0.11	0.44	1	-0.3	0.0089	-0.1	0.14	-0.028	0.23	0.13	-0.36
STA	0.33	0.46	0.057	-0.016	0.17	-0.3	1	-0.15	0.16	-0.07	0.031	-0.28	-0.27	0.11
GDP-1	-0.069	-0.072	-0.23	-0.37	0.012	0.0089	-0.15	1	0.2	-0.13	-0.0057	0.07	-0.058	0.058
RR	-0.29	-0.13	-0.42	-0.65	-0.22	-0.1	0.16	0.2	1	-0.55	-0.055	0.0089	-0.21	0.16
CPI-1	0.19	-0.084	0.082	0.77	0.32	0.14	-0.07	-0.13	-0.55	1	0.13	0.1	0.38	-0.31
Total Earnings/Total Advances	-0.16	-0.015	-0.069	0.13	0.14	-0.028	0.031	-0.0057	-0.055	0.13	1	0.22	0.16	0.31
Total Profit/ Total Advances	-0.37	-0.43	0.047	0.032	-0.071	0.23	-0.28	0.07	0.0089	0.1	0.22	1	0.048	-0.071
Total Advances	0.15	-0.032	0.015	0.26	-0.028	0.13	-0.27	-0.058	-0.21	0.38	0.16	0.048	1	0.046
Unsecured/Tot Advances	-0.099	0.21	-0.024	-0.25	0.18	-0.36	0.11	0.058	0.16	-0.31	0.31	-0.071	0.046	1

Table 4.22 evaluates the correlation matrix of RF classification as per a dataset about different banks regarding GNPA.

Table 4.22: Correlation matrix of Random Forest classification for NNPA



4.13.3 Predicting GNPA and NNPA by using WEKA tool

Table 4.23 evaluates the summary results of studied ML algorithm models such as Bayes Net (BN), Naïve Bayes (NB), logistic regression (LgR), Sequential minimal optimization of Support Vector Machine regression (SMOreg), Linear Logistic Regression (SL), Classification via Regression (CR), LogitBoost (LB); Pruned and unpruned decision tree C4 (J48), Logistic Model Tree (LMT), Random Forest (RF), Random tree (RT), and Class implementing minimal cost-complexity pruning (Cart) related to 15 attributes viz. GNPA, NNPA, GDP, CPI, PSL, TL, STA, GDP-1, RR, CPI-1, TE, TP, and USTA as numeric as well as banks, Year, GNPA>6, and GNPA>7, as nominal of the dataset to know the overall performance accuracy of banking sectors. The performance of model accuracy of the above-mentioned ML algorithm classifications as per correctly and incorrectly classified instances, Kappa (K) statistics, mean absolute error (MAE) and root mean squared error (RMSE) were studied as per the 10-fold cross-validation test. In the case of algorithm model classification, the highest values were observed in LB (78.47%) and Cart (74.30%) followed by J48 (73.61%), CR (72.91%), and LMT (69.44%), and the lowest value in SMO (34.72%) as per 10-fold cross-validation test.

Table 4.23: Results on different classified instances and statistical values for different algorithm models

Classifier model	Correctly classified instances	Incorrectly classified instances	KS	MAE	RMSE
BN	58.33	41.67	0.54	0.07	0.22
NB	64.58	35.42	0.61	0.06	0.22
LgR	47.92	52.08	0.40	0.09	0.29
SMO	34.72	65.28	0.29	0.14	0.26
SL	61.11	38.89	0.56	0.07	0.22
CR	72.91	27.08	0.70	0.08	0.19
LB	78.47	21.53	0.76	0.04	0.16
J48	73.61	26.39	0.71	0.05	0.19
LMT	69.44	30.55	0.67	0.06	0.21
RF	50.69	49.30	0.46	0.11	0.23
RT	60.42	39.58	0.57	0.07	0.25
Cart	74.30	25.69	0.72	0.05	0.18

BN = Bayes Network; NB = NaiveBayes; LgR = Logistic Regression; SMOreg = Sequential minimal optimization of Support Vector Machine regression; SL = Linear Logistic Regression; CR = Classification via Regression; LB = LogitBoost; J48 = Pruned and unpruned decision tree C4; LMT = Logistic Model Tree; RF = Random Forest; RT = Random tree; Cart = Class implementing minimal cost-complexity pruning; KS = Kappa Statistics; MAE = Mean Absolute Error; RMSE = Root Mean Squared Error

Table 4.24 describes the representation of the detailed accuracy of studied models for the studied dataset. In the case of the accuracy of a class of values of TP, FP, precision, MCC, ROC, and PRC, the better performances were observed in LB and Cart followed by J48, CR, and LMT, and the lowest performance in the case of SMO. To determine the correctness of the classification, the rate of TP and FP are important statistical parameters of a dataset. In the present study, a higher rate of TP and a lower rate of FP was observed in the classification algorithms viz. LB and Cart followed by J48, CR, and LMT. Herein, the ROC curve is observed closely related to TP and FP rates and higher values were obtained in the above-mentioned models. The MCC is also an important statistical parameter to determine the good score in the prediction result and higher MCC values are observed in the above-mentioned models. In binary classification modelling, a higher PRC value is also determined by the performance accuracy in the dataset. The present study found better PRC (82% and 71%) values in the dataset, which are accepted regarding two ML algorithm models.

Table 4.24: Statistical data for prediction accuracy of studied algorithms

Classifier model	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC area	PRC area
BN	0.583	0.038	0.601	0.583	0.588	0.553	0.942	0.626
NB	0.646	0.032	0.642	0.646	0.637	0.609	0.952	0.662
LgR	0.583	0.038	0.601	0.583	0.588	0.553	0.942	0.626
SMO	0.347	0.059	0.345	0.347	0.338	0.284	0.912	0.444
SL	0.611	0.035	0.613	0.611	0.610	0.576	0.917	0.646
CR	0.729	0.025	0.717	0.729	0.713	0.695	0.966	0.771
LB	0.785	0.020	0.793	0.785	0.785	0.768	0.974	0.817
J48	0.736	0.024	0.746	0.736	0.732	0.714	0.912	0.654
LMT	0.694	0.028	0.689	0.694	0.689	0.663	0.917	0.679
RF	0.507	0.045	0.508	0.507	0.507	0.463	0.928	0.558
RT	0.604	0.036	0.620	0.604	0.607	0.574	0.787	0.464
Cart	0.743	0.023	0.743	0.743	0.738	0.718	0.933	0.706

BN = Bayes Network; NB = Naive Bayes; LgR = Logistic Regression; SMOreg = Sequential minimal optimization of Support Vector Machine regression; SL = Linear Logistic Regression; CR = Classification via Regression; LB = LogitBoost; J48 = Pruned and unpruned decision tree C4; LMT = Logistic Model Tree; RF = Random Forest; RT = Random tree; Cart = Class implementing minimal cost-complexity pruning; TP = True positive; FP = False positive; MCC = Matthews correlation coefficient; ROC = Receiver operating characteristic; PRC = Precision-recall curve

Conclusions from Weka tool analyses provides useful insights in understanding causal relationship between select independent and dependent variables of banks. A total of 12 classifier models were used to analyse causal relationship. Precision value which determines the accuracy of causal relationship is found to be most significant while using Logit Boost, J48, Cart and Logistic Model Tree classifier models with precision (accuracy) > 70%. These precision values are however lower than accuracy rates observed using Python based machine learning models. Weka's limitation is that it's not a widely used tool for loss forecasting in banks as on date and its usage is still in research. However, it provides useful insights for cross-validation of precision rates and other key indicators such as PRC, ROC etc. In future, as more research analyses is carried out, banks will consider adopting alternate tools to cross validate probability of credit default to further refine accuracy of loss forecasting in their forecasting models.

4.14 Kruskal Wallis test

Table 4.25: Kruskal Wallis test between years and bank related parameters

Test Statistics[a,b]

	GNPA	NNPA	GDP	CPI	PSL	TL	STA	GDP-1	RR	CPI-1	Total Earnings/Total Advances	Total Profit/ Total Advances	Total Advances	Unsecured/ Tot Advances
Chi-Square	59.29	57.03	143.00	143.00	3.25	1.11	38.31	143.00	143.00	143.00	2.30	24.69	27.62	8.58
Df	11	11	11	11	11	11	11	11	11	11	11	11	11	11
Asymp. Sig.	<0.001	<0.001	<0.001	<0.001	0.99	1.00	<0.001	<0.001	<0.001	<0.001	0.996	<0.001	<0.001	0.66

a. Kruskal Wallis Test, b. Grouping Variable: Year

Table 4.26: Kruskal Wallis test between banks and bank-related parameters

Test Statistics,[b]

	GNPA	NNPA	GDP	CPI	PSL	TL	STA	GDP-1	RR	CPI-1	Total Earnings/Total Advances	Total Profit/ Total Advances	Total Advances	Unsecured/Tot Advances
Chi-Square	44.30	47.92	0.00	0.00	123.43	122.10	47.24	0.00	0.00	0.00	128.84	62.42	83.81	119.16
Df	11	11	11	11	11	11	11	11	11	11	11	11	9	11
Asymp. Sig.	<0.001	<0.001	1.00	1.00	<0.001	<0.001	<0.001	1.00	1.00	1.00	<0.001	<0.001	<0.001	<0.001

a. Kruskal Wallis Test, b. Grouping Variable: Bank

Table 4.25 evaluates the Kruskal Wallis test between the years (2008-2019) and banking-related parameters. The parameters such as GNPA, NNPA, GDP, CPI, STA, GDP-1, RR, CPI-1, Total Profit/Total Advances, and Total Advances were observed with varying degree of significance (Chi-square value at 59.29, 57.03, 143.00, 143.00, 38.31, 143.00, 143.00, 143.00, 24.69 and 27.62; p<0.001). These results indicate that there is a significant difference in these parameters over the years. PSL, TL, Total Earnings/Total Advances and Unsecured/Tot Advances have p values >0.05 which show relatively low significance.

Table 4.26 evaluates the Kruskal Wallis test between banks and various selected variables. The parameters such as GNPA, NNPA, PSL, TL, STA, Total Earnings/Total Advances, Total Profit/Total Advances, Total Advances and Unsecured/Tot Advances were observed highly significant (Chi-square = 44.30, 47.92, 123.43, 122.10, 47.24, 128.84, 62.42, 83.81, and 119.16; p<0.001) This showed that there is a significant difference in these parameters across the banks. But for GDP, CPI, GDP-1, RR, and CPI-1 the p values are >0.05. We can conclude that GDP, CPI, GDP-1, RR, and CPI-1 are not statistically significantly different across the banks.

4.15 Tests of normality

Table 4.27: Kolmogorov-Smirnov test and Shapiro-Wilk test for bank-related parameters

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	**df**	**Sig.**	**Statistic**	**df**	**Sig.**
GNPA	0.168	120	<0.001	0.846	120	<0.001
NNPA	0.177	120	<0.001	0.845	120	<0.001
GDP	0.128	120	<0.001	0.949	120	<0.001
CPI	0.242	120	<0.001	0.865	120	<0.001
PSL	0.093	120	<0.001	0.963	120	<0.001
TL	0.118	120	<0.001	0.935	120	<0.001
STA	0.049	120	0.200	0.988	120	0.343
GDP-1	0.142	120	<0.001	0.940	120	<0.001
RR	0.226	120	<0.001	0.910	120	<0.001
CPI-1	0.255	120	<0.001	0.857	120	<0.001
Total Earnings/Total Advances	0.276	120	<0.001	0.742	120	<0.001
Total Profit/ Total Advances	0.350	120	<0.001	0.510	120	<0.001
Total Advances	0.220	120	<0.001	0.629	120	<0.001
Unsecured/Tot Advances	0.338	120	<0.001	0.705	120	<0.001

*. This is a lower bound of the true significance.

Table 4.19 evaluates tests of normality by using the Kolmogorov-Smirnov test and Shapiro-Wilk test for bank-related parameters such as GNPA and NNPA = Gross and Net non-performing assets; PSL = Priority sector lending; TL = Term loan; STA = Secured to total asset, RR = repo rate; CPI = Consumer price index; TE = Total earnings; TP = Total profits; TA = Total advance; USTA = Unsecured/Tot Advances*100; GDP = Gross domestic products. All the parameters observed highly significant levels ($p = <0.001$) except STA. This indicates that most of the parameters do not follow normal distribution. Fig 4.15 to 4.28 represents a histogram of the individual bank-related parameter to determine normal distribution.

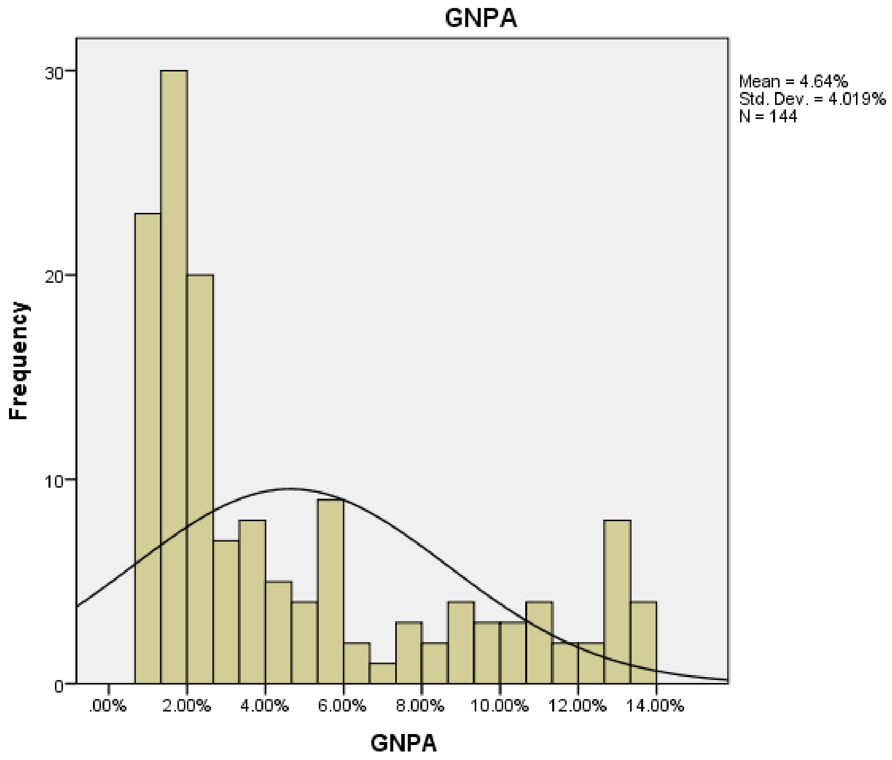


Figure 4.15: Frequency (%) of GNPA

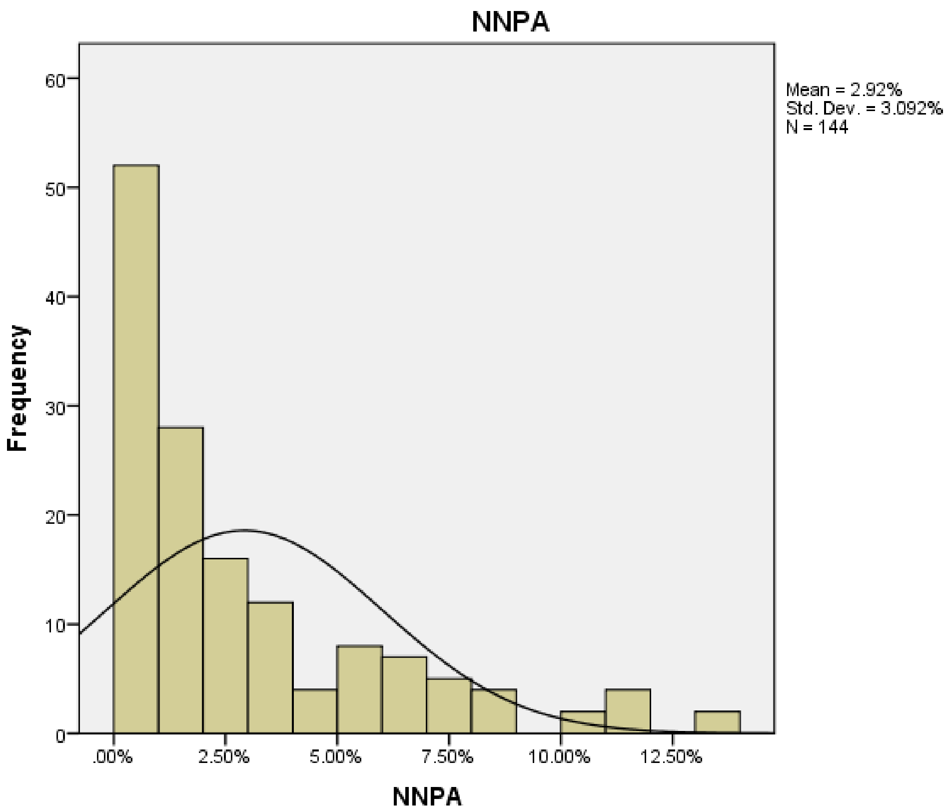


Figure 4.16: Frequency (%) of NNPA

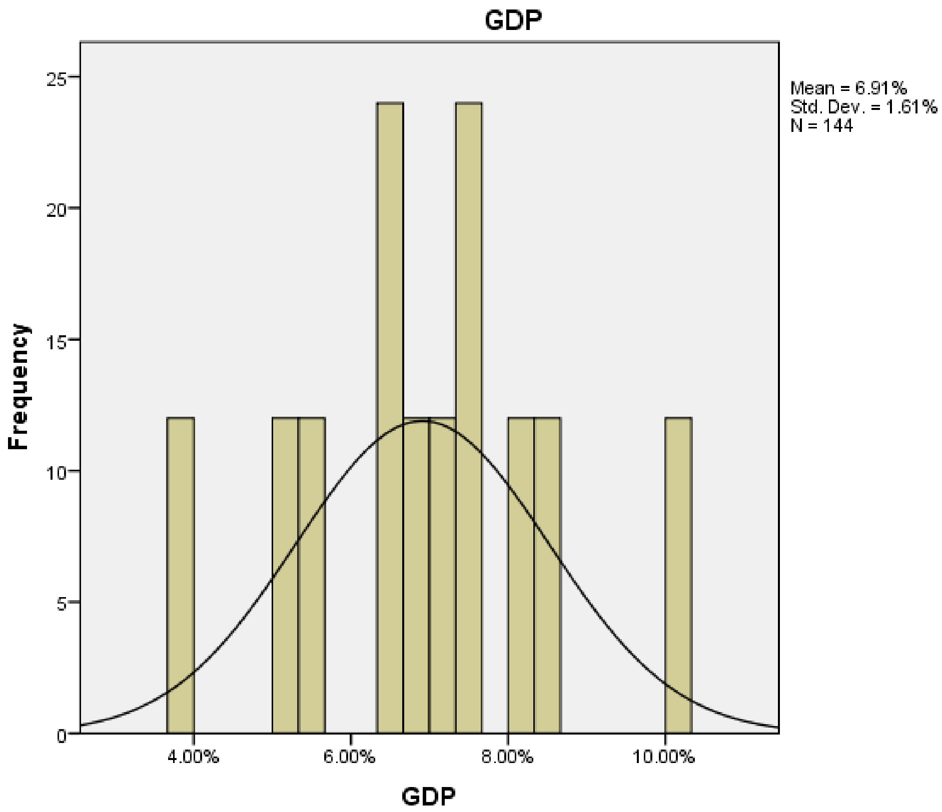


Figure 4.17: Frequency (%) of GDP

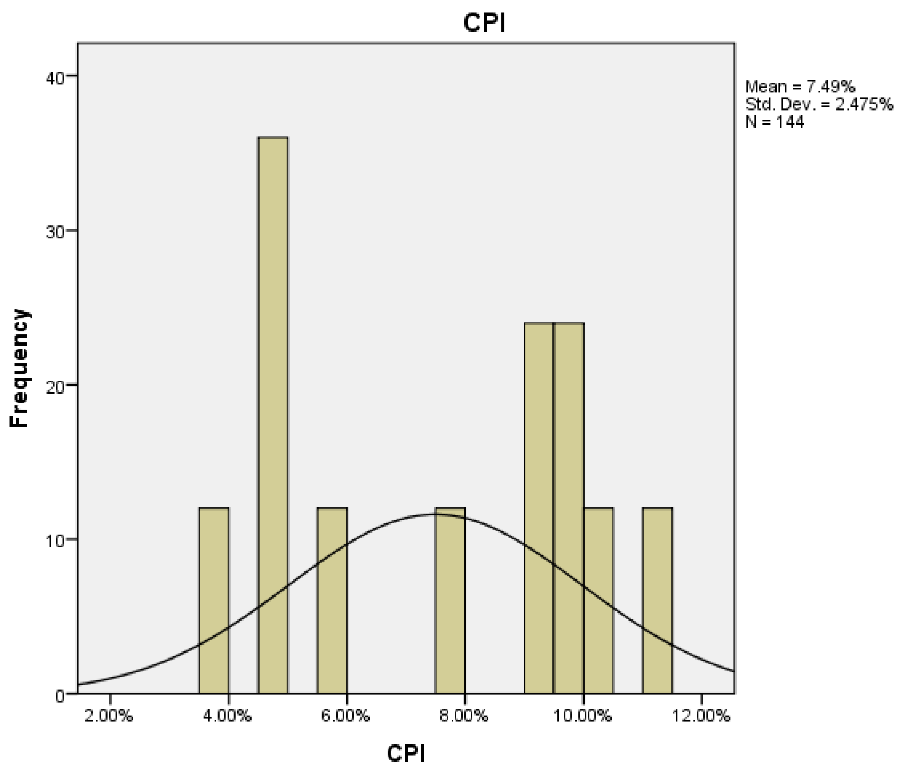


Figure 4.18: Frequency (%) of CPI

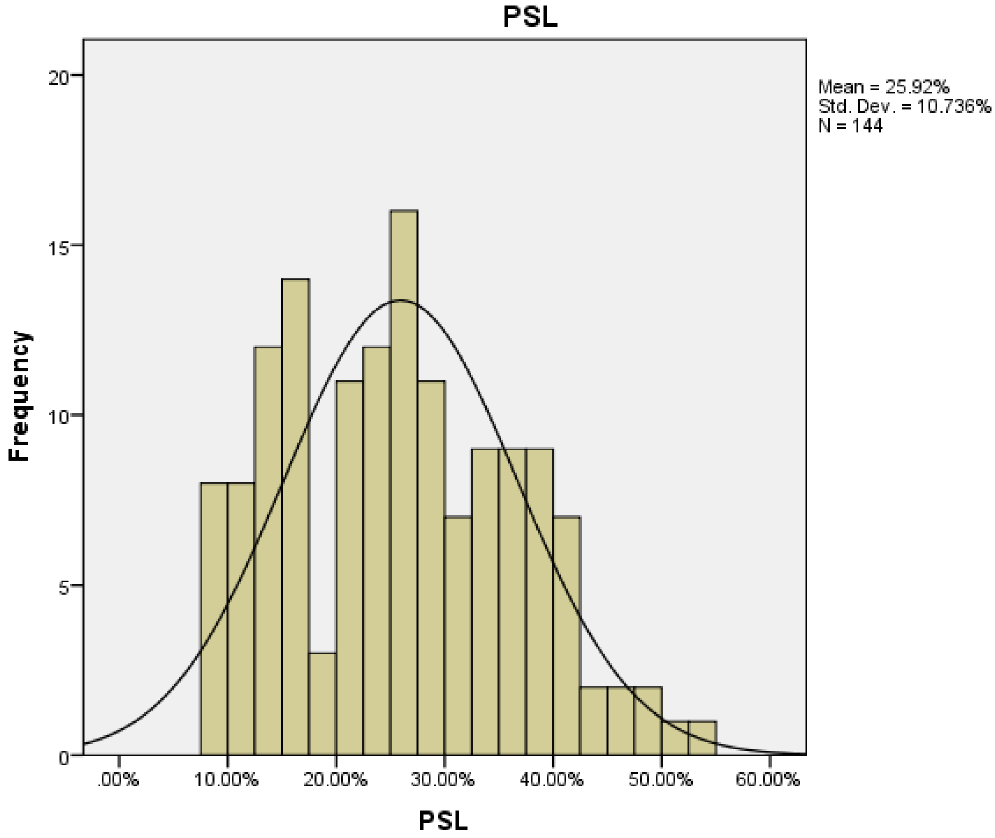


Figure 4.19: Frequency (%) of PSL

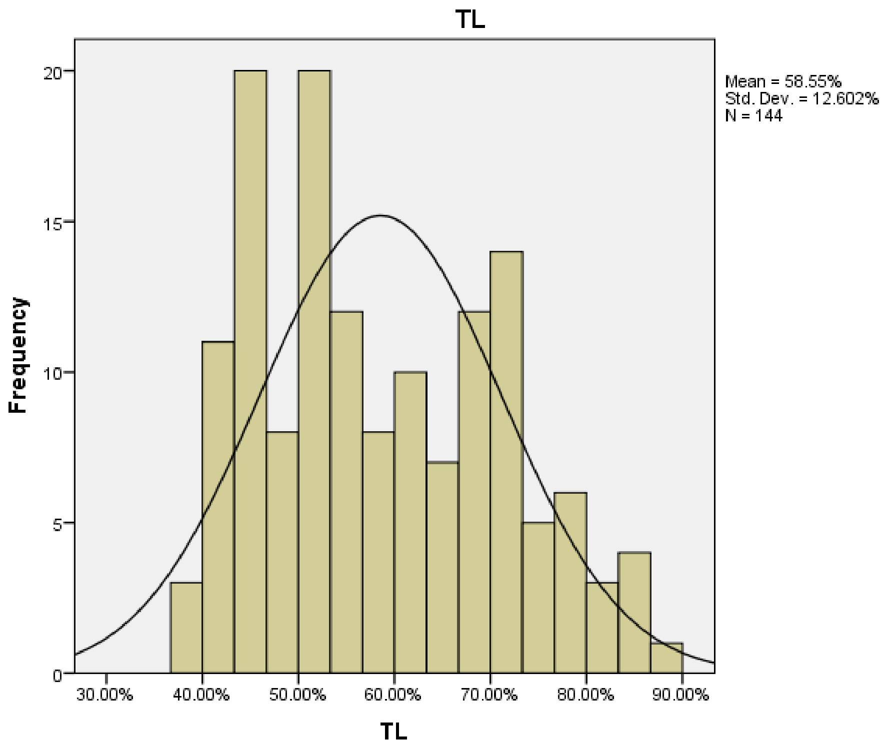


Figure 4.20: Frequency (%) of TL

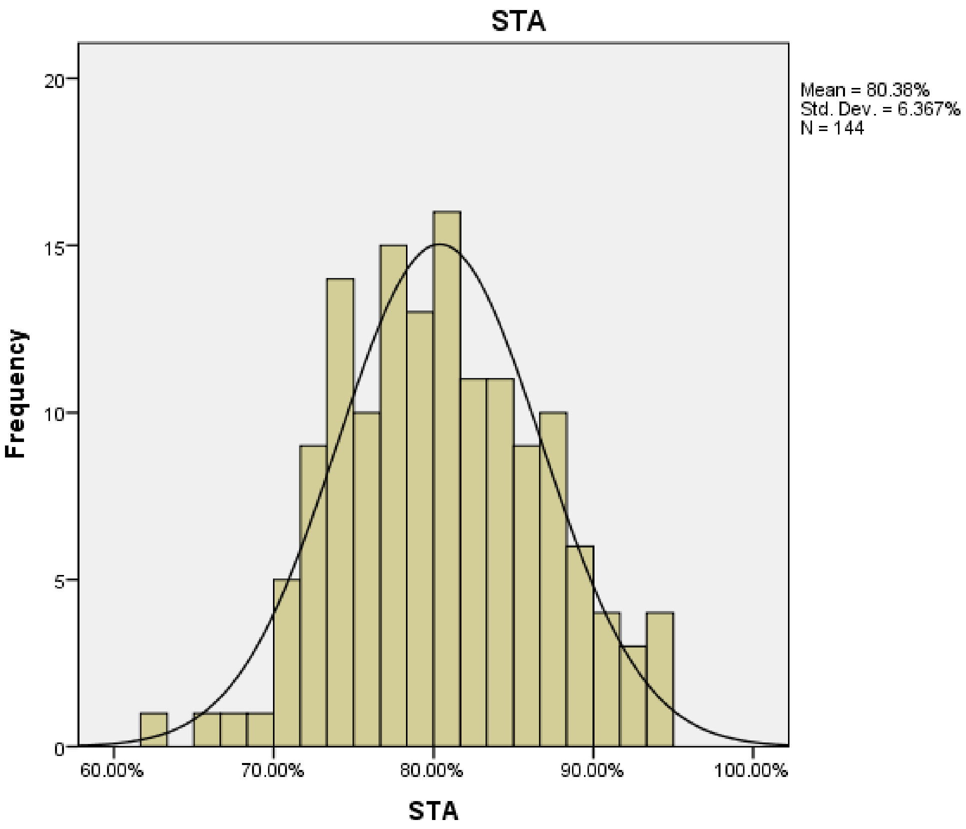


Figure 4.21: Frequency (%) of STA

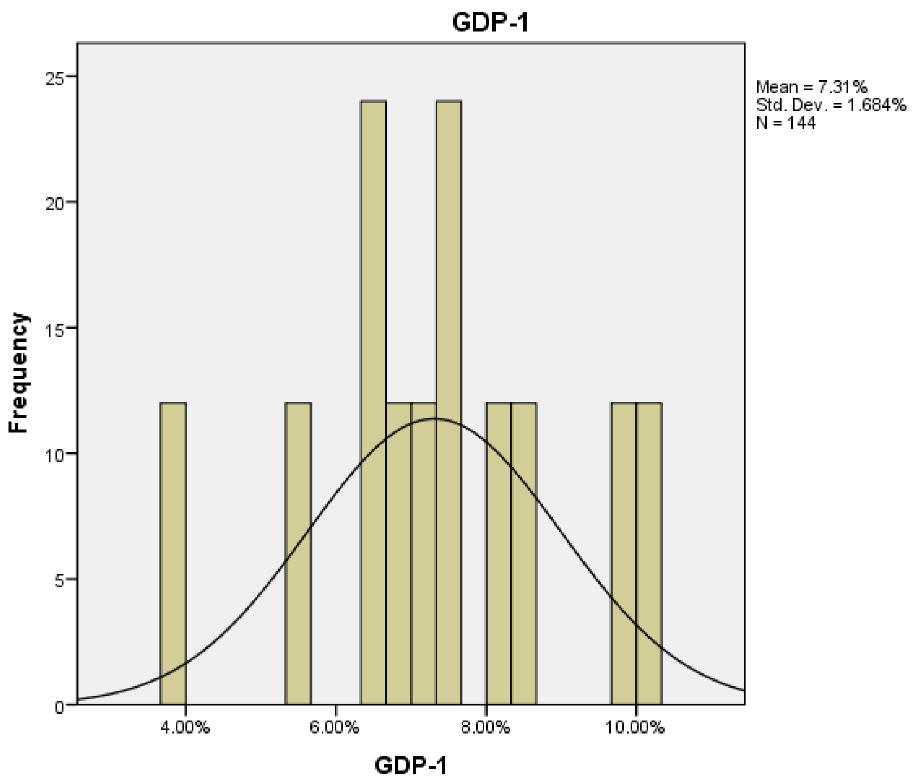


Figure 4.22: Frequency (%) of STA

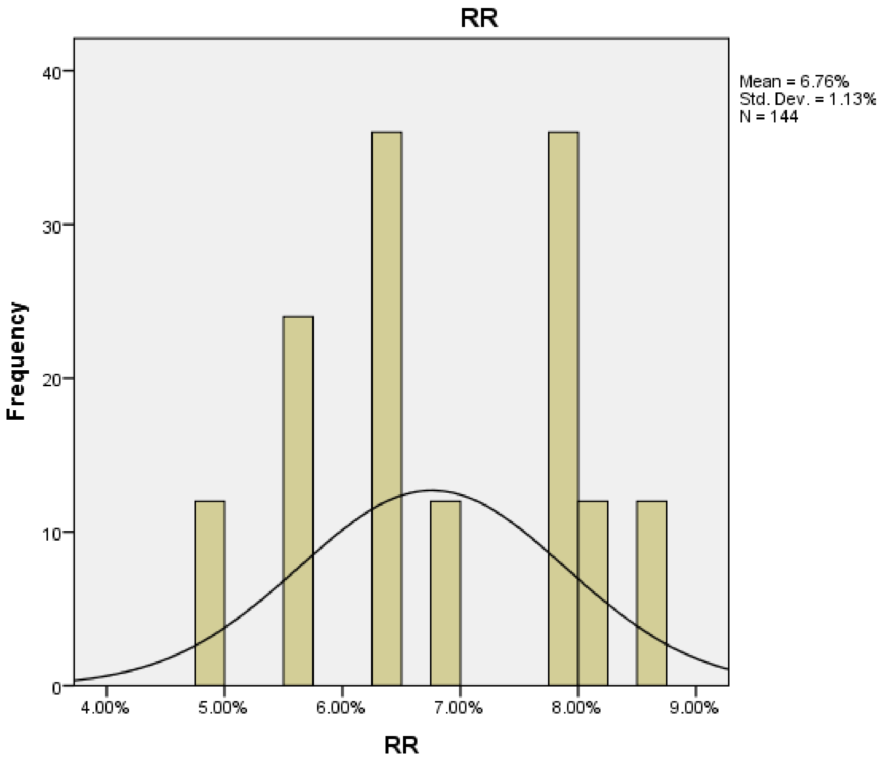


Figure 4.23: Frequency (%) of RR

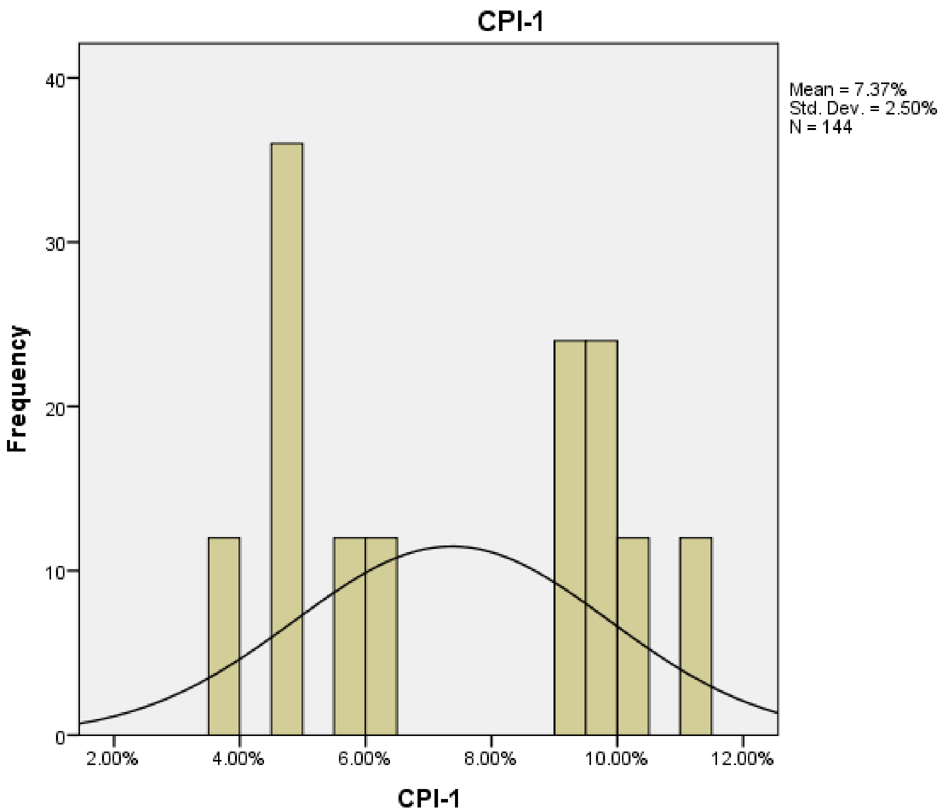


Figure 4.24: Frequency (%) of CPI-1

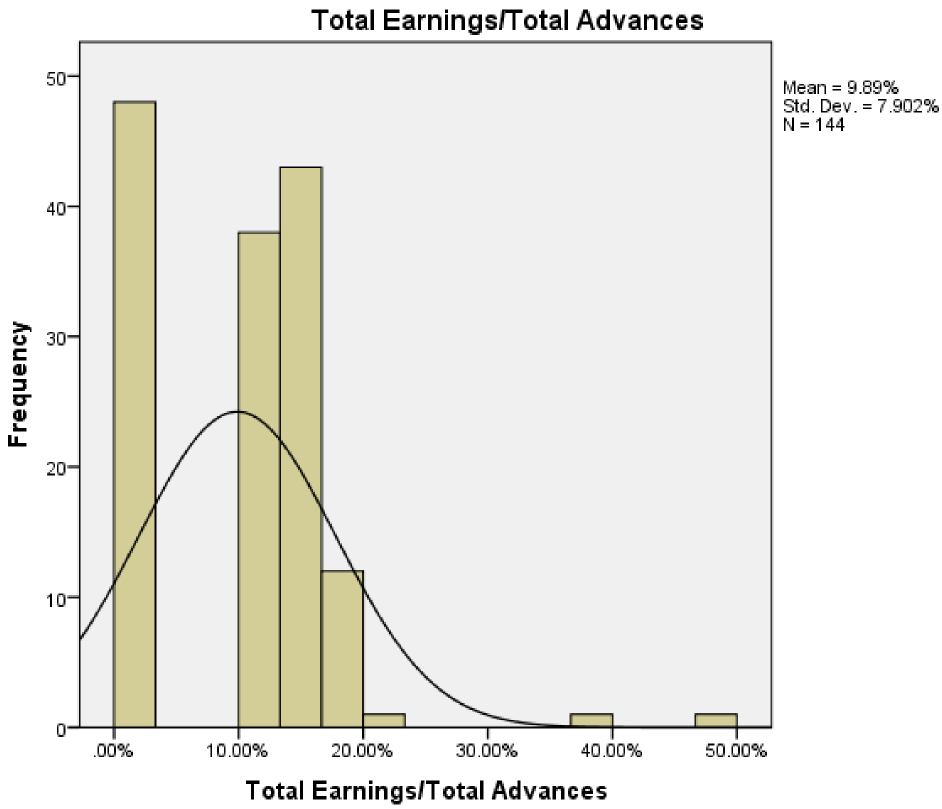


Figure 4.25: Frequency (%) of Total Earnings/Total Advances

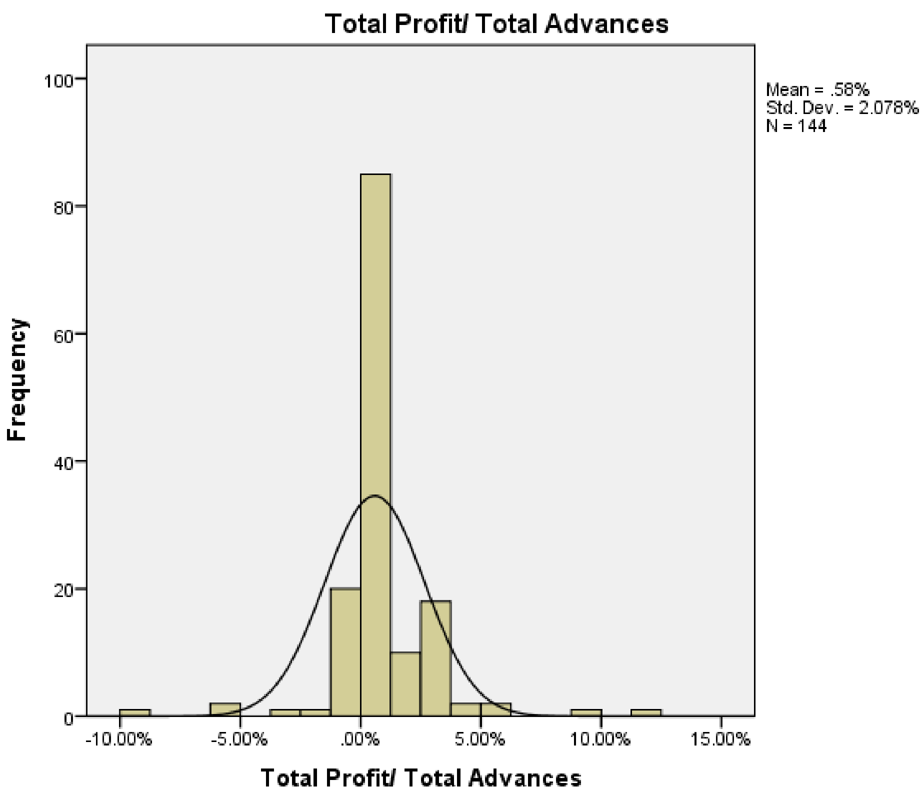


Figure 4.26: Frequency (%) of Total Profit/Total Advances

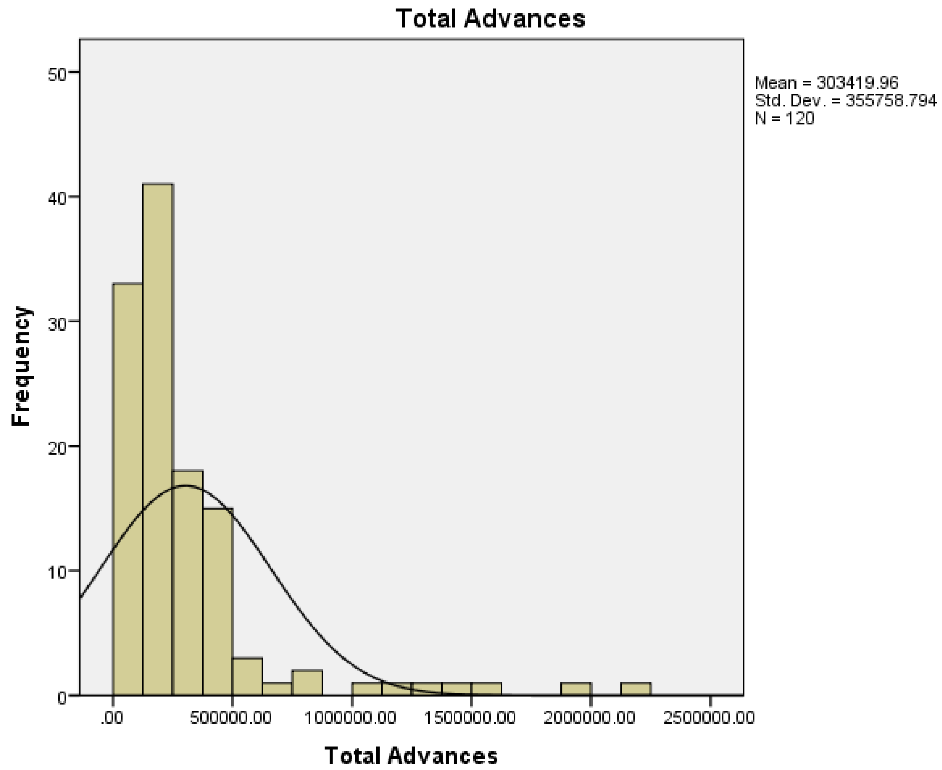


Figure 4.27: Frequency (%) of Total Advances

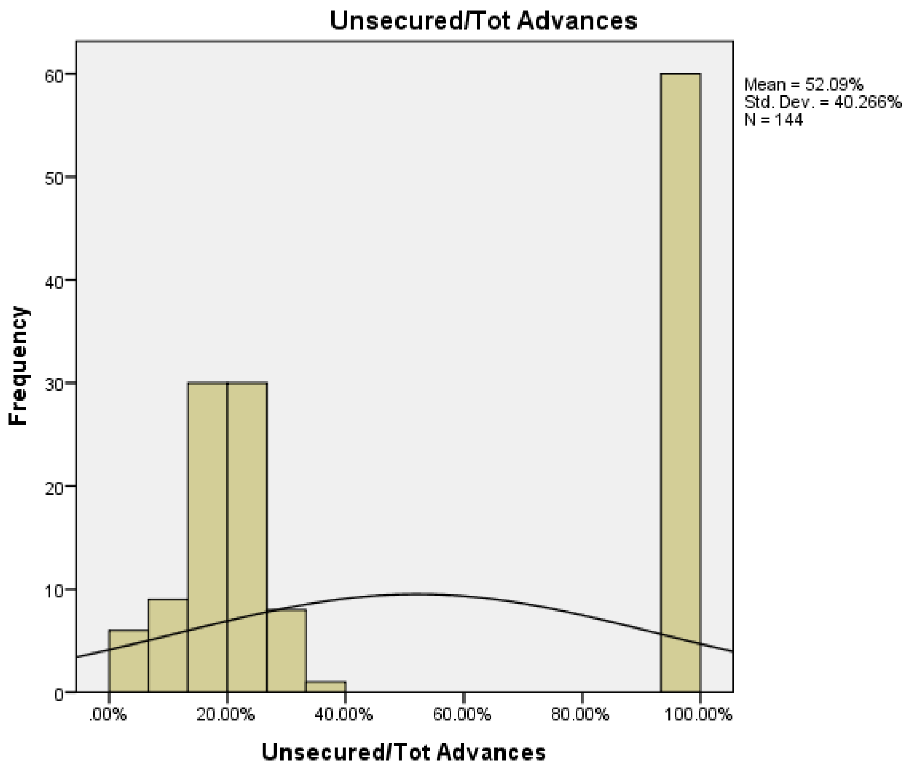


Figure 4.28: Frequency (%) of Unsecured/Total Advances

4.16 Analysis of correlation coefficient between GNPA and NNPA and bank related parameters

Table 4.28: Correlation coefficient analysis between GNPA and NNPA and bank related parameters

Correlations														
			GDP	CPI	PSL	TL	STA	GDP-1	RR	CPI-1	Total Earnings/ Total Advances	Total Profit/ Total Advances	Total Advances	Unsecured /Total Advances
Spearman's rho	GNPA	CC	-0.047	-0.515	0.190	-0.095	0.391	-0.056	-0.093	-0.537	-0.170	-0.509	0.391	-0.074
		p Value	0.573	<0.001	0.023	0.255	<0.001	0.509	0.270	<0.001	0.042	<0.001	<0.001	0.377
	NNPA	CC	-0.026	-0.518	0.161	-0.157	0.412	-0.072	-0.071	-0.534	-0.131	-0.523	0.358	0.041
		p Value	0.753	<0.001	0.054	0.060	<0.001	0.392	0.397	<0.001	0.118	<0.001	<0.001	0.629

CC = Correlation coefficient

Table 4.28 evaluates correlation coefficient between GNPA and NNPA and bank related parameters such as PSL = Priority sector lending; TL = Term loan; STA = Secured to total asset, RR = repo rate; CPI = Consumer price index; TE = Total earnings; TP = Total profits; TA = Total advance; USTA = Unsecured/Total Advances; GDP = Gross domestic products. GNPA and NNPA observed significantly (p = <0.001) negative correlation with CPI (-0.515 and -0.518), CPI-1 (-0.537 and-0.534), and Total Profit/Total Advances (-0.509 and-0.523) while significantly (p = 0.000) positive correlation with STA (0.391and 0.412) and Total Advances (0.391and0.358). It was also observed that GNPA significantly (p = 0.023) positively correlated with PSL (0.190) and significantly (p = 0.042) negatively correlated with Total Earnings/Total Advances (-0.170).

4.17 Regression analysis between GNPA and bank-related parameters

Table 4.29: Regression between GNPA and bank-related parameters

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1417.463	12	118.122	19.521	<0.001[b]
	Residual	647.443	107	6.051		
	Total	2064.906	119			

a. Dependent Variable: GNPA
b. Predictors: (Constant), Unsecured/Tot Advances, GDP, CPI, Total Profit/ Total Advances, TL, GDP-1, Total Advances, RR, STA, Total Earnings/Total Advances, PSL, CPI-1

Table 4.29 evaluates regression analysis to determine overall model fit related to F value. In the present study, the banking performance as per the GNPA as the dependent variable and bank-related factors such as Unsecured/Tot Advances, GDP, CPI, Total Profit/ Total Advances, TL, GDP-1, Total Advances, RR, STA, Total Earnings/Total Advances, PSL, CPI-1 as independent variables were studied. Thus, the total sum of squares (2064.906) with the F value (19.521) predicted that the overall model was significant ($p<0.001$) and predictive variables had a significant relationship with GNPA regarding banking performance.

4.18 Regression analysis between NNPA and bank-related parameters

Table 4.30: Regression analysis between NNPA and bank-related parameters

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	845.600	12	70.467	19.960	<0.000[b]
	Residual	377.757	107	3.530		
	Total	1223.357	119			

a. Dependent Variable: GNPA
b. Predictors: (Constant), Unsecured/Tot Advances, GDP, CPI, Total Profit/ Total Advances, TL, GDP-1, Total Advances, RR, STA, Total Earnings/Total Advances, PSL, CPI-1

Table 4.30 evaluates regression analysis to determine overall model fit related to F value. In the present study, the banking performance as per the NNPA as the dependent variable, and bank-related factors such as Unsecured/Tot Advances, GDP, CPI, Total Profit/ Total Advances, TL, GDP-1, Total Advances, RR, STA, Total Earnings/Total Advances, PSL, CPI-1 as independent variables were studied. Thus, the total sum of squares (1223.357) with the F value (19.960) predicted that the overall model was significant (p<0.001) and predictive variables had a significant relationship with NNPA regarding banking performance.

4.19 Overall model summary results for coefficient and regression analysis

Table 4.31: Regression analysis for overall banking performance as per GNPA

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.829[a]	.686	.651	2.45985%

a. Predictors: (Constant), Unsecured/Tot Advances, GDP, CPI, Total Profit/ Total Advances, TL, GDP-1, Total Advances, RR, STA, Total Earnings/Total Advances, PSL, CPI-1

Table 4.32: Correlation coefficient for overall banking performance as per GNPA

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	29.342	6.614		4.436	<0.001		
	GDP	-1.219	.278	-.471	-4.382	<0.001	.253	3.951
	CPI	-.898	.227	-.534	-3.963	<0.001	.161	6.200
	PSL	.033	.038	.089	.859	0.393	.275	3.642
	TL	-.002	.037	-.005	-.045	0.964	.267	3.750
	STA	.124	.051	.186	2.424	0.017	.498	2.007
	GDP-1	-.490	.157	-.198	-3.131	0.002	.730	1.370
	RR	-2.079	.328	-.564	-6.337	<0.001	.369	2.707
	CPI-1	-.166	.219	-.100	-.757	0.451	.169	5.916
	Total Earnings/Total Advances	-.027	.065	-.042	-.409	0.683	.276	3.619
	Total Profit/ Total Advances	-.114	.115	-.056	-.988	0.325	.898	1.113
	Total Advances	**-4.468E-07**	**.000**	**-.038**	**-.583**	**0.561**	**.683**	**1.464**
	Unsecured/Tot Advances	-.010	.010	-.101	-.976	0.331	.276	3.626

a. Dependent Variable: GNPA

Table 4.32 evaluates the multiple regressions in which the predictors of GNPA for banking performance of studied banks of India. The result on the predictive variable R, given the value of multiple correlation coefficients between the predictors and the outcome. The value of R is 0.829 indicates that the predictor variable has a significant impact on the outcome of overall banking performance related to GNPA. Thus, the predictors are good indicators of overall performance, and the model is a reasonable fit for the data sample.

Model – This is known to specify multiple models in a single regression command. This tells you the number of the model being reported.

R – R is the square root of R-squared and is the correlation between the observed and predicted values of the dependent variable.

R-Square – R-Square is the proportion of variance in the dependent variable (**GNPA**) which can be predicted from the 11 independent variables. This value indicates that 68.6% of the variance. Note that this is an overall measure of the strength of association and does not reflect the extent to which any particular independent variable is associated with the dependent variable. R-Square is also called the coefficient of determination.

Adjusted R-square – As predictors are added to the model, each predictor explained some of the variance in the dependent variable simply due to chance. One could continue to add predictors to the model which would continue to improve the ability of the predictors to explain the dependent variable, although some of this increase in R-square would be simply due to chance variation in that particular sample. The adjusted R-square attempts to yield a more honest value to estimate the R-squared for the population. The value of R-square was .686, while the value of Adjusted R-square was .651 Adjusted R-squared is computed using the formula 1 – ((1 – Rsq)(N – 1)/ (N – k – 1)). From this formula, we can see that when the number of observations is small and the number of predictors is large, there will be a much greater difference between R-square and adjusted R-square (because the ratio of (N – 1) / (N – k – 1) will be much greater than 1). By contrast, when the number of observations is very large compared to the number of predictors, the value of R-square and adjusted R-square will be much closer because the ratio of (N – 1)/(N – k – 1) will approach 1.

Std. Error of the Estimate – The standard error of the estimate, also called the root mean square error, is the standard deviation of the error term, and is the square root of the Mean Square Residual (or Error).

Table 4.32 evaluates the regression coefficient between GNPA (dependent variable) and bank-related parameters (independent variables). The β value observed about the relationship between the independent variables such as Unsecured/Tot Advances, GDP, CPI, Total

Profit/Total Advances, TL, GDP-1, Total Advances, RR, STA, Total Earnings/Total Advances, PSL, CPI-1 on overall banking performance. In this result, out of 12 independent variables, β value was observed positive for PSL without significant (β = 0.089; p = 0.393) change and STA with significant (β = 0.186; p = 0.017) change while negative for rest parameters such as CPI (β = -.534; $p < 0.001$), GDP-1 (β = -.198; p = 0.002), RR (β = -.564; $p < 0.001$) with significant value but TL (β = -.005; P=0.964), CPI-1 (β = -.100; P=0.451), Total Earnings/Total Advances (β = -.042; P=0.683), Total Profit/ Total Advances (β = -.056; P=0.325), Total advances (β = -.038; P=0.561) and Unsecured/Tot Advances (β = -.101; P=0.331) were negative, which indicates that these parameters did not contribute as the risk factors for the banking performance with P>0.05. As observed, this value does not contribute significantly to the model since the sig value is higher than 0.05. Since two tests provide contradictory results, it is not possible to confirm the relationship between the profitability and efficiency of the bank.

4.20 Overall model summary results for coefficient and regression analysis

Table 4.33: Regression analysis for overall banking performance as per NNPA

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.831[a]	.691	.657	1.87895%

a. Predictors: (Constant), Unsecured/Tot Advances, GDP, CPI, Total Profit/ Total Advances, TL, GDP-1, Total Advances, RR, STA, Total Earnings/Total Advances, PSL, CPI-1

Table 4.34: Regression coefficient for overall banking performance as per NNPA

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	16.892	5.052		3.344	0.001		
	GDP	-.861	.212	-.433	-4.055	<0.001	.253	3.951
	CPI	-.580	.173	-.448	-3.347	0.001	.161	6.200
	PSL	-.009	.029	-.033	-.323	0.747	.275	3.642
	TL	.014	.028	.051	.493	0.623	.267	3.750
	STA	.127	.039	.247	3.251	0.002	.498	2.007
	GDP-1	-.371	.120	-.195	-3.098	0.002	.730	1.370
	RR	-1.514	.251	-.534	-6.042	<0.001	.369	2.707
	CPI-1	-.221	.167	-.173	-1.322	0.189	.169	5.916
	Total Earnings/ Total Advances	-.004	.050	-.008	-.083	0.934	.276	3.619
	Total Profit/ Total Advances	-.179	.088	-.115	-2.030	0.045	.898	1.113
	Total Advances	**-6.458E-07**	**.000**	**-.072**	**-1.102**	**0.273**	**.683**	**1.464**
	Unsecured /Tot Advances	.013	.008	.174	1.701	0.092	.276	3.626

a. Dependent Variable: GNPA

Table 4.34 evaluates the multiple regressions in which the predictors of NNPA for banking performance of studied banks of India. The result on the predictive variable R, given the value of multiple correlation coefficients between the predictors and the outcome. The value of R is 0.831 indicating that the predictor variable has a significant impact on the outcome of overall banking performance related to NNPA. Thus, the predictors are good indicators of overall performance, and the model is a reasonable fit for the data sample.

Model – This is known to specify multiple models in a single regression command. This tells you the number of the model being reported.

R – R is the square root of R-squared and is the correlation between the observed and predicted values of the dependent variable.

R-Square – R-Square is the proportion of variance in the dependent variable (**NNPA**) which can be predicted from the 11 independent variables. This value indicates that 69.1% of the variance. Note that this is an overall measure of the strength of association and does not reflect the extent to which any particular independent variable is associated with the dependent variable. R-squared is also called the coefficient of determination.

Adjusted R-square – As predictors are added to the model, each predictor explained some of the variance in the dependent variable simply due to chance. One could continue to add predictors to the model which would continue to improve the ability of the predictors to explain the dependent variable, although some of this increase in R-square would be simply due to chance variation in that particular sample. The adjusted R-square attempts to yield a more honest value to estimate the R-squared for the population. The value of R-square was .691, while the value of Adjusted R-square was .675 Adjusted R-squared is computed using the formula $1 - ((1 - Rsq)(N - 1)/ (N - k - 1))$. From this formula, we can see that when the number of observations is small and the number of predictors is large, there will be a much greater difference between R-square and adjusted R-square (because the ratio of $(N - 1) / (N - k - 1)$ will be much greater than 1). By contrast, when the number of observations is very large compared to the number of predictors, the value of R-square and adjusted R-square will be much closer because the ratio of $(N - 1)/(N - k - 1)$ will approach 1.

Std. Error of the Estimate – The standard error of the estimate, also called the root mean square error, is the standard deviation of the error term, and is the square root of the Mean Square Residual (or Error).

Table 4.26 evaluates the regression coefficient between NNPA (dependent variable) and bank-related parameters (independent variables). The β value observed about the relationship between the independent variables such as Unsecured/Tot Advances, GDP, CPI, Total Profit/ Total Advances, TL, GDP-1, Total Advances, RR, STA, Total Earnings/Total Advances, PSL, CPI-1 on overall banking performance. In this result, out of 12 independent variables, β value was observed positive for STA (β = 0.247; p = 0.002) with significant change while TL (β = 0.051; p = 0.623) and Unsecured/Tot Advances (β = 0.174; p = 0.092) without significant change and STA with significant (p = 0.002) change while negative for rest parameters such as GDP (β = -0.433; p = 0.001), CPI (β = -0.448; p = 0.001), GDP-1 (β = -0.195; p = 0.002), RR (β = -0.534; p = 0.002) and Total Profit/Total Advances (β = -.115; P=0.045) with significant value while PSL (β = -0.033; p = 0.747), CPI-1 (β = -0.173; p = 0.189), Total Earnings/Total Advances (β = -.0008; P=0.934) and Total advances (β = -.072; P=0.273), which indicates that these parameters did not contribute as the risk factors for the banking performance with P>0.05.

4.21 Logistic regression analysis for GNPA

Table 4.35: Logistic regression for overall banking performance as per GNPA

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1[a]	GDP	-.492	.547	.806	1	0.369	.612	.209	1.788
	CPI	-.057	.461	.015	1	0.902	.945	.383	2.331
	PSL	.037	.077	.234	1	0.629	1.038	.893	1.207
	TL	.039	.083	.226	1	0.634	1.040	.884	1.224
	STA	.104	.095	1.198	1	0.274	1.110	.921	1.337
	GDP-1	.225	.362	.386	1	0.534	1.252	.616	2.547
	RR	-2.628	.744	12.462	1	<0.001	.072	.017	.311
	CPI-1	-.949	.612	2.408	1	0.121	.387	.117	1.284
	Total Earnings/Total Advances	.017	.131	.017	1	0.896	1.017	.787	1.315
	Total Profit/TotalAdvances	-.210	.362	.338	1	0.561	.810	.399	1.647
	Total Advances	.000	.000	1.158	1	0.282	1.000	1.000	1.000
	Unsecured Tot Advances	-.017	.019	.796	1	0.372	.983	.947	1.021
	Constant	13.931	10.654	1.710	1	0.191	1122487.221		

a. Variable(s) entered on step 1: GDP, CPI, PSL, TL, STA, GDP-1, RR, CPI-1, Total Earnings/Total Advances, Total Profit/Total Advances, Total Advances, Unsecured/Tot Advances.

Table 4.35 evaluates the results of the logistic regression. Since the P-value of factor 7 (RR) was observed significant ($p<0.001$) . In the case of RR, the highest odds ratio is 12.462 with 95% CI (0.017 – 0.311). Rest of the variables such as GDP, CPI, PSL, TL, STA, GDP-1, CPI-1, Total Earnings Total Advances, Total Profit Total Advances, Total Advances, and Unsecured Tot Advances did not show high significance as $P>0.05$. These variables are however extremely relevant to be used in designing the new generation agile risk appetite framework and for setting up risk limits for banks.

Step 1 - This is the first step (or model) with predictors in it. In this case, it is the full model that we expressed in the logistic regression command for GNPA.

Chi-square and **Sig.** - This is the chi-square statistic and its significance level. The value given in the Sig. column is the probability of obtaining the chi-square statistic assuming that the null hypothesis is true. In other words, this is the probability of obtaining this chi-square statistic if there is in fact no effect of the independent variables, taken together, on the dependent variable except RR. This is, of course, the p-value, which is compared to a critical value, perhaps .05 or .01 to determine if the overall model is statistically significant. In this case, the model is statistically significant because the p-value is less than .000.

df - This is the number of degrees of freedom for the model. There is one degree of freedom for each predictor in the model. In this example, we have 12 predictors.

-2 Log likelihood - This is the -2 log-likelihood for the final model. By itself, this number is not very informative. However, it can be used to compare nested (reduced) models.

Cox and Snell R Square and **Nagelkerke R Square** - These are pseudo R-squares. Logistic regression does not have an equivalent to the R-squared that is found in OLS regression; however, many people have tried to come up with one. There are a wide variety of pseudo-R-square statistics. Because this statistic does not mean what R-squared means in OLS regression (the proportion of variance explained by the predictors), we suggested interpreting this statistic with great caution.

Observed - This indicates the number of 0's and 1's that are observed in the dependent variable.

Predicted - These are the predicted values of the dependent variable based on the full logistic regression model. This table shows how many cases are correctly predicted, and how many cases are not correctly predicted.

Overall Percentage - This gives the overall percentage of cases that are correctly predicted by the model (in this case, the full model that we specified).

B - These are the values for the logistic regression equation for predicting the dependent variable from the independent variable. These estimates the relationship between the independent variables and the dependent variable, where the dependent variable is on the logit scale. These estimates tell the amount of increase (or decrease, if the sign of the coefficient is negative) in the predicted log odds that would be predicted by a 1 unit increase (or decrease) in the predictor, holding all other predictors constant. Note: For the independent variables that are not significant, the coefficients are not significantly different from 0, which should be taken into account when interpreting the coefficients.

S.E. - These are the standard errors associated with the coefficients. The standard error is used for testing whether the parameter is significantly different from 0; by dividing the parameter estimate by the standard error you obtain a t-value. The standard errors can also be used to form a confidence interval for the parameter.

Wald and **Sig.** - These columns provide the Wald chi-square value and 2-tailed p-value used in testing the null hypothesis that the coefficient (parameter) is 0. If you use a 2-tailed test, then you would compare each p-value to your preselected value of alpha. Coefficients having p-values less than alpha are statistically significant. For example, if you chose alpha to be 0.05, coefficients having a p-value of 0.05 or less would be statistically significant (i.e., you can reject the null hypothesis and say that the coefficient is significantly different from 0). If you use a 1-tailed test (i.e., you predict that the parameter will go in a particular direction), then you can divide the p-value by 2 before comparing it to your preselected alpha level. For the variable **read**, the p-value is .000, so the null hypothesis that the coefficient equals 0 would be rejected.

df - This column lists the degrees of freedom for each of the tests of the coefficients.

Exp(B) - These are the odds ratios for the predictors. They are the exponentiation of the coefficients. There is no odds ratio for the variables (as a variable with 1 degree of freedom) was not entered into the logistic regression equation.

4.22 Logistic regression analysis for NNPA

Table 4.36: Logistic regression for overall banking performance as per NNPA

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	GDP	-.034	.648	.003	1	`	.966	.271	3.444
	CPI	.732	.490	2.227	1	0.136	2.079	.795	5.437
	PSL	-.137	.085	2.582	1	0.108	.872	.738	1.031
	TL	.162	.106	2.316	1	0.128	1.176	.954	1.448
	STA	.272	.123	4.894	1	0.027	1.313	1.032	1.671
	GDP1	.116	.625	.035	1	0.853	1.123	.330	3.820
	RR	-2.339	1.017	5.287	1	0.021	.096	.013	.708
	CPI-1	-2.389	.870	7.532	1	0.006	.092	.017	.505
	Total Earnings/Total Advances	-.128	.135	.897	1	0.343	.880	.676	1.146
	Total Profit/Total Advances	-.245	.363	.454	1	0.501	.783	.384	1.596
	Total Advances	.000	.000	2.289	1	0.130	1.000	1.000	1.000
	Unsecured/Tot Advances	.044	.023	3.645	1	0.056	1.045	.999	1.093
	Constant	-5.501	11.320	.236	1	0.627	.004		
a. Variable(s) entered on step 1: GDP, CPI, PSL, TL, STA, GDP-1, RR, CPI-1, Total Earnings/Total Advances, Total Profit/Total Advances, Total Advances, Unsecured/Tot Advances.									

Table 4.36 evaluates the results of the logistic regression. Since the P-value of factor 8 (CPI-1) and 7 (RR) were observed significant (p=0.006 and p=0.021) so, the mitigation of NNPA as per cut off value of values of <5% and >5% is based on CPI-1 and RR, which will have a higher propensity to regulate overall banking performance of Indian banks. In the case of CPI-1 and RR, the highest odds ratio (7.532 and 5.287) with 95% CI (0.017 – 0.505 and 0.013 – 0.708). Rest variables such as GDP, CPI, PSL, TL, STA, GDP-1, Total Earnings Total Advances, Total Profit Total Advances, Total Advances, and Unsecured To Advances did not show significant change. These variables are however extremely relevant to be used in designing the new generation agile risk appetite framework and for setting up risk limits for banks.

Step 1 - This is the first step (or model) with predictors in it. In this case, it is the full model that we expressed in the **logistic regression** command for NNPA.

Chi-square and **Sig.** - This is the chi-square statistic and its significance level. The value given in the Sig. column is the probability of obtaining the chi-square statistic assuming that the null hypothesis is true. In other words, this is the probability of obtaining this chi-square statistic if there is in fact no effect of the independent variables, taken together, on the dependent variable except RR and CPI-1. This is, of course, the p-value, which is compared to a critical value, perhaps .05 or .01 to determine if the overall model is statistically significant. In this case, the model is statistically significant because the p-value is less than .000.

df - This is the number of degrees of freedom for the model. There is one degree of freedom for each predictor in the model. In this example, we have 12 predictors.

-2 Log likelihood - This is the -2 log-likelihood for the final model. By itself, this number is not very informative. However, it can be used to compare nested (reduced) models.

Cox and Snell R Square and **Nagelkerke R Square** - These are pseudo R-squares. Logistic regression does not have an equivalent to the R-squared that is found in OLS regression; however, many people have tried to come up with one. There are a wide variety of pseudo-R-square statistics. Because this statistic does not mean what R-squared means in OLS regression (the proportion of variance explained by the predictors), we suggested interpreting this statistic with great caution.

Observed - This indicates the number of 0's and 1's that are observed in the dependent variable.

Predicted - These are the predicted values of the dependent variable based on the full logistic regression model. This table shows how many cases are correctly predicted, and how many cases are not correctly predicted.

Overall Percentage - This gives the overall percentage of cases that are correctly predicted by the model (in this case, the full model that we specified).

B - These are the values for the logistic regression equation for predicting the dependent variable from the independent variable. These estimates the relationship between the independent variables and the dependent variable, where the dependent variable is on the logit scale. These estimates tell the amount of increase (or decrease, if the sign of the coefficient is negative) in the predicted log odds that would be predicted by a 1 unit increase (or decrease) in the predictor, holding all other predictors constant. Note: For the independent variables that are not significant, the coefficients are not significantly different from 0, which should be taken into account when interpreting the coefficients.

S.E. - These are the standard errors associated with the coefficients. The standard error is used for testing whether the parameter is significantly different from 0; by dividing the parameter estimate by the standard error you obtain a t-value. The standard errors can also be used to form a confidence interval for the parameter.

Wald and **Sig.** - These columns provide the Wald chi-square value and 2-tailed p-value used in testing the null hypothesis that the coefficient (parameter) is 0. If you use a 2-tailed test, then you would compare each p-value to your preselected value of alpha. Coefficients having p-values less than alpha are statistically significant. For example, if you chose alpha to be 0.05, coefficients having a p-value of 0.05 or less would be statistically significant (i.e., you can reject the null hypothesis and say that the coefficient is significantly different from 0). If you use a 1-tailed test (i.e., you predict that the parameter will go in a particular direction), then you can divide the p-value by 2 before comparing it to your preselected alpha level. For the variable **read**, the p-value is .000, so the null hypothesis that the coefficient equals 0 would be rejected.

df - This column lists the degrees of freedom for each of the tests of the coefficients.

Exp(B) - These are the odds ratios for the predictors. They are the exponentiation of the coefficients. There is no odds ratio for the variables (as a variable with 1 degree of freedom) was not entered into the logistic regression equation.

CHAPTER 5

SUMMARY OF FINDINGS AND CONCLUSIONS

5.1 Introduction

The present results derived from the study based on the reliability of statistical methods and validity through ML algorithms regarding Indian banking performance related to credit risk parameters especially GNPA and NNPA along with other bank-related parameters and macroeconomic variables for the period of 2008-2019.

This chapter takes up objective wise discussion on the investigations of the study done in the present research and compares it with earlier research on “Design and development of robust credit risk management and risk appetite framework for the banking sector” so as to find out the extent to which it supports or contradicts the present findings of the similar studies. The chapter concluded as per the results obtained to measure the parameters for improving the banking performance of public and private banking sectors of India. It also discusses regarding limitations of the present study and finally identifies the scope for future research in this research area.

5.2 Objectives and Conclusions

Use Machine learning algorithms models to predict non-performing assets (GNPA/NNPA) and test its accuracy and embed it in risk appetite framework for risk tolerance limits

Machine learning models and traditional multi-variate statistical techniques have been used to quantify the impact of bank-related factors and macroeconomic variables on the credit risk. The selected variables have varying degree of impact on credit risk. Thus, the selected variables are reasonable indicators which can be used to assess and measure credit risk.

Regression coefficient between GNPA/NNPA (dependent variable) and macro-economic and key internal variables (independent variables) revealed varying degrees of significance between the independent variables such as Unsecured/Tot Advances, GDP, CPI, Total

Profit/Total Advances, TL, GDP-1, Total Advances, RR, STA, Total Earnings/Total Advances, PSL, CPI-1. In the Logistic Regression analysis, GNPA/NNPA is highly correlated to RR.

Accuracy and prediction of dependent variable (GNPA/NNPA) through advanced machine learning algorithms were assessed by computing linkages between macroeconomic variables and key internal variables. Various machine learning regression models were used such as Random Forest, Decision Tree, XG Boost, Artificial Neural Network (ANN) and Support Vector Machine (SVM). Machine learning regression models were used to predict the non-performing asset percentage (GNPA/NNPA) and based on significance of the independent variables, use these variables in risk appetite framework to set risk concentration limits.

Machine learning classification models such as Random Forest, Decision Tree, XG Boost, Naïve Bayes, Artificial Neural Network (ANN), Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) were used to compute the accuracy of relationship between independent variables and dependent variable.

The results of the regression and classification models showed different degree of accuracy measured as 1-MAPE (Mean Absolute Percentage Error) when the testing the relationship of independent variables with gross non-performing assets (GNPA). Random Forest model showed the most reasonable MAPE of 11.5% while Decision Tree showed MAPE of 0%, which could be attributed to over-fitting of data set. Other models such as XG Boost, SVM and ANN have shown high MAPE values, which have not been considered due to high error values. However, given that the data set used in research is limited, and machine learning models results are most optimal with large data sets, it’s recommended that these models should be considered by banks for setting up risk appetite framework.

Machine learning regression model results to predict the net non-performing assets (NNPA) were similar with Random Forest model giving MAPE of 22.7% (Accuracy = 77.3%), which was the most reasonable fit amongst all the models that were tested. Decision Tree results showed the lowest MAPE of 0% (Accuracy = 100%), as it was observed in GNPA predicted values. This can be attributed to over-fitting of data set. Other models such as XG Boost, SVM and ANN have shown high MAPE values, which have not been considered due to high error values. However, given that the data set used in research is limited, and machine learning models results are most optimal with large data sets, it's recommended that these models should be considered by banks for setting up risk appetite framework.

Machine learning regression Random Forest model was trained to further assess whether the predicted value of GNPA further improves by segregating data set based on risk-tiering (cohorts). For this purpose, data set comprising of 8 public sector banks was segregated into 4 cohorts. Cohorts are designed in terms of size of banks' credit advances to ensure that similar sized banks are clubbed together. This cohort approach has been used to test whether different risk limits be designed by type of assets in risk appetite framework. Cohort 1 comprises of State Bank of India, Bank of Baroda and Canara Bank. Cohort 2 comprises of Bank of India, Andhra Bank. Cohort 3 is comprised of Central Bank of India and Allahabad Bank. Cohort 4 comprised of Bank of Maharashtra. GNPA predicted values have varying degree of MAPE, with Cohort 1 MAPE of 8.45% (Accuracy = 93.55%) to Cohort 4 with MAPE of 21.33% (Accuracy = 79.67%).

Machine learning regression Random Forest model was trained to further assess whether the predicted value of NNPA further improves by segregating data set based on risk-tiering (cohorts). For this purpose, data set comprising of 8 public sector banks was segregated into 4

Cohorts. NNPA predicted values have varying degree of MAPE, with Cohort 1 MAPE of 13.07% (Accuracy = 76.93%) to Cohort 2 with MAPE of 39.5% (Accuracy = 60.5%).

Machine learning classification models were tested to compute the accuracy of whether the dependent variable (GNPA) was correctly classified as GNPA, based on relationship of the variable with independent variables. The accuracy rates, which is defined by precision value were found to be highest in case of Random Forest (100%) followed by Decision Tree and XG Boost (97.2%), SVM (95%), KNN (93%), Naïve Bayes (86%) and ANN (65%). It can be concluded that XG Boost and Decision Tree have yielded optimum accuracy rates and can be considered as good fit models. Random Forest with precision of 100% indicates model over-fitment. SVM, KNN and Naïve Bayes models should be considered in risk appetite framework development, as these models have high precision values and given the limited data sets available in this research, should be considered.

Machine learning classification models were tested to compute the accuracy of whether the dependent variable (NNPA) was correctly classified as NNPA, based on relationship of the variable with independent variables. The accuracy rates, which is defined by precision value were found to be highest in case of Random Forest (100%) followed by Decision Tree and XG Boost (96.5%), SVM (95%), KNN (92%), Naïve Bayes (88%) and ANN (78%). It can be concluded that XG Boost and Decision Tree have yielded optimum accuracy rates and can be considered as good fit models. Random Forest with precision of 100% indicates model over-fitment. SVM, KNN and Naïve Bayes models should be considered in risk appetite framework development, as these models have high precision values and given the limited data sets available in this research, should be considered.

Machine learning classification Decision Tree model was trained to further assess whether the classification of GNPA further improves by segregating data set based on risk-tiering (cohorts). For this purpose, data set comprising of 8 public sector banks was segregated into 4 cohorts. GNPA classification further improved by using the cohort approach with each cohort estimating 100% precision value.

Machine learning classification Decision Tree model was trained to further assess whether the classification of NNPA further improves by segregating data set based on risk-tiering (cohorts). For this purpose, data set comprising of 8 public sector banks was segregated into 4 cohorts. GNPA classification further improved by using the cohort approach with each cohort estimating 100% precision value, except cohort 1 which estimated precision value of 97.2%.

Besides the empirical analysis, machine learning algorithm models have provided better insights in predicting the dependent variable values and its accuracy. ML algorithm models such as Linear Regression, Logistic Regression, Multivariate Regression, Naïve Bayes, Decision tree, Random Forest, SVM, ANN, KNN and hybrid models (simple hybrid models and class-wise classifier) have been successfully used in credit scoring, operational efficiency, but the predictive analysis of understanding the relationship of macro-economic variables and key internal variables with non-performing assets (GNPA/NNPA) is still not widely used in Indian banks in its the risk appetite framework.

Machine learning models, both classification and regression models accuracy was tested by training the datasets during the Out of Time period of 2020 to 2022. Testing was performed using only a select models that had provided optimum Predicted Values of GNPA and NNPA as well as Accuracy Rates.

Random Forest Classification Model was trained to test the accuracy of classification of NNPA. Accuracy rate of 100% was established. These results indicate over-fitting of data or lack of adequate size of data sets. However, the results directionally are in line with accuracy results observed during the In-Time Validation period of 2008-2019. Testing also revealed Total Profits to Advances and CPI (Lag) to have the highest contribution in estimating the Accuracy.

Random Forest Classification Model was trained to test the accuracy of classification of NNPA. Accuracy rate of 100% was established. These results indicate over-fitting of data or lack of adequate size of data sets. However, the results directionally are in line with accuracy results observed during the In-Time Validation period of 2008-2019. Testing also revealed Total Profits to Advances and CPI (Lag) to have the highest contribution in estimating the Accuracy.

Random Forest Regression Model was trained to test the accuracy of predicted values of GNPA. Model results show MAPE of 14%, thus demonstrating Accuracy of predicted values of GNPA to be 86% (1-MAPE). The results are in line with accuracy results observed during the In-Time Validation period of 2008-2019.

Random Forest Regression Model was trained to test the accuracy of predicted values of NNPA. Model results show MAPE of 23%, thus demonstrating Accuracy of predicted values of GNPA to be 77% (1-MAPE). The results are in line with accuracy results observed during the In-Time Validation period of 2008-2019.

Table 5.1 Risk appetite model framework key indicators selection

Variables Type	Variables	Significance	Type of model	Inclusion in risk appetite framework	Rationale for inclusion in model	Usage type in model
Macro-Economic	CPI (Lag)	Significant	Classification ML Model	Yes		Quantitative Based
Macro-Economic	GDP (Lag)	Moderately Significant	Classification ML Model	Yes	Critical Variable	Judgement based Overlay
Macro-Economic	Repo Rate / Int Rate	Not Significant	Classification ML Model	Yes	Critical Variable	Judgement based Overlay
QunMacro-Economic	CPI (No Lag)	Not Significant	Classification ML Model	Yes	Critical Variable	Judgement based Overlay
Macro-Economic	GDP (No Lag)	Not Significant	Classification ML Model	Yes	Critical Variable	Judgement based Overlay
Key Internal Variables	Profits/Tot Advance	Significant	Classification ML Model	Yes		Quantitative Based
Key Internal Variables	Priority Sector Advances	Significant	Classification ML Model	Yes		Quantitative Based
Key Internal Variables	Unsecured Advances	Moderately Significant	Classification ML Model	Yes		Quantitative Based
Key Internal Variables	Term Loans	Moderately Significant	Classification ML Model	Yes		Quantitative Based
Key Internal Variables	Secured Advances	Not Significant	Classification ML Model	Yes	Critical Variable	Judgement based Overlay
Key Internal Variables	Total Advances	Not Significant	Classification ML Model	No	Included through Secured/Uns ecured	

Variables Type	Variables	Significance	Type of model	Inclusion in risk appetite framework	Rationale for inclusion in model	Usage type in model
Key Internal Variables	Total Earnings/Advance	Not Significant	Classification ML Model	No	Included through Total Profit/Advance	
Macro-Economic	CPI (Lag)	Not Significant	Kruskal Wallis Test	Yes	Critical Variable	Judgement based Overlay
Macro-Economic	GDP (Lag)	Not Significant	Kruskal Wallis Test	Yes	Critical Variable	Judgement based Overlay
Macro-Economic	Repo Rate / Int Rate	Not Significant	Kruskal Wallis Test	Yes	Critical Variable	Judgement based Overlay
Macro-Economic	CPI (No Lag)	Not Significant	Kruskal Wallis Test	Yes	Critical Variable	Judgement based Overlay
Macro-Economic	GDP (No Lag)	Not Significant	Kruskal Wallis Test	Yes	Critical Variable	Judgement based Overlay
Key Internal Variables	Profits/Tot Advance	Significant	Kruskal Wallis Test	Yes		Quantitative Based
Key Internal Variables	Priority Sector Advances	Significant	Kruskal Wallis Test	Yes		Quantitative Based
Key Internal Variables	Unsecured Advances	Significant	Kruskal Wallis Test	Yes		Quantitative Based
Key Internal Variables	Term Loans	Significant	Kruskal Wallis Test	Yes		Quantitative Based
Key Internal Variables	Secured Advances	Significant	Kruskal Wallis Test	Yes		Quantitative Based
Key Internal Variables	Total Advances	Significant	Kruskal Wallis Test	No	Included through Secured/Unsecured	

Variables Type	Variables	Significance	Type of model	Inclusion in risk appetite framework	Rationale for inclusion in model	Usage type in model
Key Internal Variables	Total Earnings/Advance	Significant	Kruskal Wallis Test	No	Included through Total Profit/Advance	
Macro-Economic	CPI (Lag)	Significant	Correlation Coefficient Test	Yes		Quantitative Based
Macro-Economic	GDP (Lag)	Not Significant	Correlation Coefficient Test	Yes	Critical Variable	Judgement based Overlay
Macro-Economic	Repo Rate / Int Rate	Not Significant	Correlation Coefficient Test	Yes	Critical Variable	Judgement based Overlay
Macro-Economic	CPI (No Lag)	Significant	Correlation Coefficient Test	Yes		Quantitative Based
Macro-Economic	GDP (No Lag)	Not Significant	Correlation Coefficient Test	Yes	Critical Variable	Judgement based Overlay
Key Internal Variables	Profits/Tot Advance	Significant	Correlation Coefficient Test	Yes		Quantitative Based
Key Internal Variables	Priority Sector Advances	Moderately Significant	Correlation Coefficient Test	Yes		Quantitative Based
Key Internal Variables	Unsecured Advances	Not Significant	Correlation Coefficient Test	Yes	Critical Variable	Judgement based Overlay
Key Internal Variables	Term Loans	Not Significant	Correlation Coefficient Test	Yes	Critical Variable	Judgement based Overlay
Key Internal Variables	Secured Advances	Significant	Correlation Coefficient Test	Yes		Quantitative Based
Key Internal Variables	Total Advances	Significant	Correlation Coefficient Test	No		

Table 5.2 Risk appetite framework - model framework

Variables type	Independent variables	Dependent variable (GPNPA/NNPA) predicted value / accuracy						
		Random Forest	**Decision Tree**	**XG Boost**	**ANN**	**SVM**	**Derived Final Values (Aggregate)**	**Risk Appetite Threshold**
Macro-economic	CPI (Lag)	Use predicted value if predicted value - actual value ± 10%-25%	Use predicted value if predicted value - actual value ± 10%-25%	Use predicted value if predicted value - actual value ± 10%-25%	Use predicted value if predicted value - actual value ± 10%-25%	Use predicted value if predicted value - actual value ± 10%-25%	Review min and max values from each model output	Total earnings-Total expenses-Total expected loss (GNPA/NNPA) > X (bank risk appetite based)
Macro-economic	GDP (Lag)							
Macro-economic	Repo rate/Int rate							
Macro-economic	CPI (No Lag)							
Macro-economic	GDP (No Lag)							
Key internal variables	Profits/Tot advance							
Key internal variables	Priority sector advances							
Key internal variables	Unsecured advances							
Key internal variables	Term loans							
Key internal variables	Total earnings/advance							
Key internal variables	Secured advances							
Key internal variables	Total advances							
Key internal variables	Total earnings/ Advances							

Conclusions

The research work carried out and elaborated in the thesis document highlight the need to set up a robust and agile risk appetite framework. The research findings are pivoted towards identifying the right set of key bank indicators for setting up the risk tolerance limits which are forward looking and agile. The research focused on deriving relationship between independent variables and the most critical dependent variable and its usage in setting up a robust credit risk appetite framework. Analysis was focused on using latest machine learning intuitive models in addition to traditional statistical models, for better predictability. The new generation risk appetite framework is essential for banking system to remain robust and to proactively identify emerging risks and thus maintain an optimal level of credit defaults & non-performing assets (GNPA/NNPA) for sustainable profitability of returns on equity. The research findings will also provide important insights to regulators who can incorporate and leverage the recommendations on enhancing risk appetite frameworks, in the context of changing landscape where banks are operating in a challenging environment of artificial intelligence, cyber risks, frauds and most importantly increasing risk of bank failures which was recently witnessed in 2023 that resulted in collapse of number of regional banks in United States (Silicon Valley Bank, First Republic Bank, Signature Bank) and a leading European bank, Credit Suisse.

5.3 Limitations of Research

The major limitations of this research are presented below:

- Increased uncertainty on future events, constantly changing drivers, and a complex combination of economic factors require that banks should run scenarios that incorporate numerous external factors. The more factors and its combinations that they can model, the more robust the outcome can be, and it will be able to better identify and scope potential impacts on portfolios and overall performance of banks.

- However, in order to support more meaningful modelling results, scenarios and approaches need to be further enhanced, and not just base it on standardized macroeconomic inputs. In a period of increased complexity, scenario generation requires more forward-looking and extraneous factors, incorporating both economic and broader uncertainties (geopolitical risks, supply chain shocks, commodity cycles risk etc).

- These factors, combined with agile forecasting capabilities would enable rapid calculation of potential portfolio income and losses.

- To further develop insights on portfolio and customer level based on scenarios, new forward-looking credit assessment are not considered in this research work, for example, food prices and utility bill inflation or rent increases and customers' ability to pay. Such approaches, if modelled, would enable banks to identify microsegments that may be vulnerable to specific scenarios.

- This research work is based on selecting the most optimal macro-economic and internal variables and factors that impact Bank's performance and are relevant in today's context. Some of the factors that are highlighted above like geo-political risk, supply chain shocks and other extraneous risks that can have a material impact on unexpected losses of the Bank are not part of the chosen variables in the Research.

- Research work is also limited by the quantity of data available for the period of study. Granular level of data is required at account / customer level to build robust models and framework, which was a challenge given the strict confidentiality around customer data which banks would not provide. Instead, as a best proxy, RBI published data was used which is an aggregated data-wise. Machine learning models that are used in this Research generally tend to give highest accuracy results based on large sets of data. Thus, to that extent, research results can be impacted, however, directionally, they provide a concrete view of how the linkages are between different parameters and its impact on banks' performance.

5.4 Scope for Future Research

The research analysis and conclusions derived for design of risk appetite framework will provide insights to carry out further research on enhancing usage of machine learning models to identify key variables and risk indicators which have higher loss forecasting and probability of default predictability. The research work will provide regulators and banks useful insights of creating forward looking risk management dashboards for effective governance.

In the future studies, research should focus on further enhancing use of advanced artificial intelligence (AI) and machine learning models to develop new generation of risk appetite framework, which will be highly intuitive and will have very high prediction capability to estimate the risk a bank can take to earn optimized returns from its shareholders. The new generation risk appetite framework will be developed based on various parameters that have been used in this research as well as non-traditional dataset and variables such as customer spends, social profile, etc.

REFERENCES

Aburime, T. (2008). Determinants of bank profitability: Company-level evidence from Nigeria. *SSRN Electronic Journal*. Doi: http://dx.doi.org/10.2139/ssrn.1106825

Agarwal, M. K., and Preeti, M. (2022). Non-performing assets in Indian banking sector: An analytical and comparative study of selected public and private sector banks. *Journal Global Values*, *8*(1), 91-104.

Agarwala, V., and Agarwala, N. 2019. A critical review of non-performing assets in the Indian banking industry. Rajagiri Management Journal, vol. 13, no. 2, pp. 12-23.

Ahsan, M. (2018). Measuring financial performance based on CAMEL: a study on selected Islamic banks in Bangladesh. *Asian Business Review*, *6*(34).

Akinci, S., Kaynak, E., Atilgan, E., and Aksoy, S. (2007). Where does the logistic regression analysis stand in marketing literature?: A comparison of the market positioning of prominent marketing journals. *European Journal of Marketing*, *41*, 537-567.

Andersen, E. B. (1980). Discrete Statistical Models with Social Science Applications. North Holland.

Appiahene, P., Missah, Y. M.,and Najim, U. (2020). Predicting bank operational efficiency using machine learning algorithm: Comparative study of decision tree, random forest, and neural networks. *Advances in Fuzzy Systems*, 2020, Article ID 8581202. https://doi.org/10.1155/2020/8581202

Arora, S. (2013). Credit risk analysis in Indian commercial banks -An empirical investigation. *Asia Pacific Finance and Accounting Review*, *1*(2), 25-32.

Arunkumar, R., and Kotreshwar, G. (2005). Risk management in commercial banks. A case study of public and private sector banks. Market Conference Indian Institute of Capital Market, Mumbai, December 19-20.

Bagadi, S. M. B. (2020). Role of RBI in Management of NPAs in Corporate Banking. *IJRAR-International Journal of Research and Analytical Reviews (IJRAR)*, *7*(1), 268-276.

Bakshi, C. (2020). Random forest regression. Retrieved from: https://levelup.gitconnected.com/random-forest-regression-209c0f354c84

Baldan, C., Geretto, E., and Zen, F. (2016). A quantitative model to articulate the banking risk appetite framework. *Journal of Risk Management in Financial Institutions*, *9*(2).

Baruah, A. 2020. AI Applications in the Top 4 Indian Banks. February 27. Available from: Emerj: https://emerj.com/ai-sector-overviews/ai-applications-in-the-top-4-indian-banks/

Berger, A.N.,and DeYoung, R. (1997). Problem loans and cost efficiency in commercial banks. *Journal of Banking and Finance*, *21*, 849-870.

Bhaskar, P. J. (2014). Credit risk management in Indian banks. *International Journal of Advance Research in Computer Science and Management Studies*, *2*(1), 33.

Bhatia, A., Mahajan, P., and Chander, S. (2012). Determinants of profitability of private sector banks in India. *Journal of Commerce and Accounting Research*, *1*(2).

Bhatia, S., Sharma, P., Burman, R., Hazari, S., and Hande, R. (2017). Credit scoring using machine learning techniques. *International Journal of Computer Applications*, *161*(11), 1-4.

Bhattacharya, H. (2001). Banking strategy, credit appraisal and sending deviations. Oxford University press, New Delhi.

Bhullar, P. S., and Gupta, P. K. (2017). Empirical analysis of determinants of profitability of public sector banks. *International Journal of Accounting and financial reporting*, *7*(2), 404-416.

Bhuvaneswari, E., and Sarma Dhulipala, V. R. (2013). The study and analysis of classification algorithm for animal kingdom dataset. *Information Engineering*,*2*(1), 6-13.

Bittu, S.,and Dwivedi, A. K. (2012). Determinants of credit risk in Indian banking sector: some panel results. *International Journal of Business Continuity and Risk Management*, *3*(2),178-185.

Boateng, K. (2020). Credit risk management and profitability in select savings and loans companies in Ghana. PhD Thesis, Canara Bank School of Management Studies, Bangalore University, Bangalore. Published in International Journal of Advanced Research.

Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A.,and Scuse, D. (2020). WEKA manual for version 3-8-5. University of Waikato, Hamilton, New Zealand, December 21.

Brahmaiah, B. (2022). Credit risk management practices of indian banking industry: An empirical study. *International Journal of Economics and Financial Issues*,*12*(2),67-71.

Breeden, J. L. 2020. Survey of machine learning in credit risk. May 30. Available from: http://dx.doi.org/10.2139/ssrn.3616342

Brownlee, J. (2017). Difference between classification and regression in machine learning. Retrieved from: https://machinelearningmastery. com/ classification-versus-regression-in-machine-learning/

Carey, A. (2001). Effective risk management in financial institutions: the turnbull approach. *Balance Sheet*, *9*(3), 24-27.

CEBS (2010) ‘High level principles for risk management’, 16th February, Committee of European Banking Supervisors, London, available at: http://www.eba. europa.eu/documents/10180/16094/ HighLevelprinciplesonriskmanagement.pdf (accessed 5th April, 2022).

Chandani, A., Divekar, R., Salam, A., and Mehta, M. (2019, February). A Study To Analyse Impact Of Insolvency And Bankruptcy Code 2016 On NPA's Of Commercial Banks With Reference To Iron And Steel Sector. In *SIMSARC 2018: Proceedings of the 9th Annual International Conference on 4C's-Communication, Commerce, Connectivity, Culture, SIMSARC 2018, 17-19 December 2018, Pune, MH, India* (p. 477). European Alliance for Innovation.

Chao-Ying, J. P., and Tack-Shing, H. S.(2002). Logistic regression analysis and reporting: A primer. *Understanding Statistics*, *1*(1), 31-70.

Chao-Ying, J. P., Kuk, L. L., and Gary, M. I. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, *96*(1), 3-14.

Chijoriga, M. M. 2000. The interrelationship between bank failure and political interventions in Tanzania in the pre-liberalization period. African Journal of Finance and Management, vol. 9, no. 1, pp. 14-30.

Chitra, B., and Vani, U. (2014). Credit risk management for banking. *International Journal of Science and Research*, *3*(3), 135-137.

Corbellini, M. (2013). Risk appetite e pianificazione strategica in chiave ICAAP: una metodologia per le piccole banche', Bancaria, No. 3, pp. 70-76.

Cortez, A. (2011). Winning at risk. Wiley Finance, Hoboken, NJ

Coyle, C. (2022, May 16). *Russia's 1998 currency crisis: what lessons for today?* Economics Observatory. https://www.economicsobservatory.com/russias-1998-currency-crisis-what-lessons-for-today

Cucinelli, D., Di Battista, M. L., Marchese, M., and Nieri, L. (2018). Credit risk in European banks: The bright side of the internal ratings based approach. *Journal of Banking and Finance*, *93*, 213-229.

Dao, B.T.T., Nguyen, D.P., Pienkhuntod, A., Amornbunchornvei, C., Nantharath, P., Yooyanyong, P., and Tangjitprom, N. (2020). Determinants of profitability in commercial banks in Vietnam, Malaysia and Thailand. *Journal of Asian Finance, Economics and Business*, *7*(4), 133-143.

Das, K. K., and Palit, S. (2019). Credit risk management in the Indian banking sector: An insight study. *Journal of Emerging Technologies and Innovative Research*, *6*(6), 611-621.

Das,A.,and Ghosh,S. (2003). Determinants of credit risk. paper presented at the Conference on Money, Risk and Investment held at Nottingham Trent University, November.

Dash, D., Yang, Z., andLiang, L. (2006). Using DEA-neural network approach to evaluate branch efficiency of a large Canadian bank. *Expert Systems with Applications*, 31(1), 108-115.

Dash, M. K., and Kabra, G. 2010. The determinants of non-performing assets in Indian commercial banks: An econometric study. Middle Eastern Finance and Economics, vol. 7, pp. 94-106.

Dastidar, S., and Sarkar, S. C. (2017). An empirical analysis of the determinants of profitability of private sector banks. *International Journal of Advanced Research*, *5*(9), 770-780.

De Jongh, E., De Jongh, D., De Jongh, R., and Van Vuuren, G. (2013). A review of operational risk in banks and its role in the financial crisis. *South African Journal of Economic and Management Sciences*, *16*(4), 364-382.

Dogra, A., and Dogra, M. (2019). Sustainability and Ethical Banking: A Case Study of Punjab National Bank. *Amity Journal of Corporate Governance*, *4*(1), 15-27.

Espinoza, R. A., and Prasad, A. (2010). Non-performing loans in the GCC banking system and their macroeconomic effects. Working Paper No. 10/224, International Monetary Fund, Washington, DC.

Fatemi, A., and Fooladi, I. (2006). Credit risk management: a survey of practices. *Managerial finance*, *32*(3), 227-233.

Frank, E., Hall, M. A., and Witten, I. H. (2016). The WEKA workbench, Online appendix for data mining: Practical machine learning tools and techniques. Morgan Kaufmann, Fourth Edition.

FSB (Financial Stability Board). (2011). Intensity and effectiveness of SIFI supervision. Progress report on implementing the recommendations on enhanced supervision. 27th October, Financial Stability Board, Basel.Available from: http://www.fsb.org/wp-content/ uploads/r_101101.pdf? page_moved=1

FSB (Financial Stability Board). (2013a). Thematic review on risk governance. Peer Review report, 12th February, Financial Stability Board, Basel. available from: http:// www.fsb.org/wp-content/uploads/r_130212.pdf

FSB (Financial Stability Board). (2013b). Principles for an effective risk appetite framework. 18th November, Financial Stability Board, Basel. Available from: http:// www.fsb.org/wp-content/uploads/r_131118.pdf

Gandhi, R. (2018). Naive Bayes classifier. Retrieved from: https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c

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.

Gyasi, A. (2016). Effect of macroeconomic factors on credit risk of banks in developed and developing countries: Dynamic panel method. *International Journal of Economics and Financial Issues*, *6*(4), 1937-1944.

Global Non-Performing Loans Report. (2006). Ernst and Young Report of the Committee on Banking Sector Reforms (Chairman: Shri. M. Narasimham) Reserve Bank of India, Mumbai 1998.

Gonzalez-Hermosillo, B., Pazarbasioglu, C.and Billings, R.(1997). Determinants of banking system fragility: a case study of Mexico. IMF Staff Papers 3, 42-66.

Gopalakrishnan, T. V. 2005. Management of Mon performing Advances. 1st revised ed., Mumbai, Northern Book Centre publication, International Monetary Fund, (2012): Global Financial Stability Report, October IDFC Securities Research. Available from: www.idfc.com/capital/pdf/report/Asset-qualityAug11.pdf

Goyal, K. A. and Agrawal, S. (2010). Risk management in Indian banks: Some emerging issues. *Int. Eco. J. Res.*, 1(1), 102-109.

Guisse, M. L. (2012). Financial performance of the Malaysian banking industry: Domestic vs foreign banks. Master of science in Banking and Finance, Institute of Graduate Studies and Research, Eastern Mediterranean University, Gazimağusa, North Cyprus.

Gulati, R., Goswami, A.,and Kumar, S.(2019). What drives credit risk in the Indian banking industry? An empirical investigation. *Economic Systems*,*43*(1), 42-62.

Gupta, S. (2004). Statistical Methods. S. Chand Publications, New Delhi, 135-256.

Gurucharan, M. K. (2020). Machine learning basics: Decision tree regression. Retrieved from:https://towardsdatascience.com/machine-learning-basics-decision-tree-regression-1d73ea003fda

Hamid, A. J., and Ahmed, T. M. (2016). Developing prediction model of loan risk in banks using data mining. *Machine Learning and Applications: An International Journal,* 3(1), 1-9.

Hamid, A. J., and Ahmed, T. M. 2016. Developing prediction model of loan risk in banks using data mining. Machine Learning and Applications: An International Journal, vol. 3, no. 1, pp. 1-9.

Haneef, A. (2012). Domestic vs foreign : a comparison of financial performance of domestic and foreign banks in Pakistan. Master's thesis, Faculty of Economics and Social Science, University of Agder, Pakistan.

Hassan, M. M., and Tabasum, M. (2018). Predictive risk categorization of retail bank loans using data mining techniques.

Hoorunnisa and Reddy, B. B. (2019). Non-performing assets of IDBI banks in India. *International Journal of Advanced Scientific Research and Management*, *4*(12), 21-29.

Hulinsky, N. J. (2015). Risk balancing in the banking sector. M.Sc Thesis, Department of Agribusiness and Applied Economics, North Dakota State University.

Hyde, P., Liebert, T.,and Wackerbeck, P. (2009). A comprehensive risk appetite framework for banks. Leading Research, Booz and Company, New York.

Indeed. (2022). *Risk Appetite Framework (With Definition and Steps)*. Indeed Career Guide. https://www.indeed.com/career-advice/career-development/risk-appetite-framework

India. Committee on the Financial System; M. Narasimham (1992). Narasimham Committee report on the financial system, 1991. Standard Book Co. Retrieved from: https://books.google.co.in/books?id=h5qaAAAAIAAJandredir_esc=y

Jayadev, M. (2013). Basel III: Capital efficiency and challenges for Indian banks. *IIMB Management Review*, *25*(2), 68.

Karthik, L., Subramanyam, M., Shrivastava, A., and Joshi, A. R. (2018). Prediction of Wilful Defaults: An Empirical Study from Indian Corporate Loans. *International Journal of Intelligent Technologies and Applied Statistics*, *11*(1).

Karunakar, M., Vasuki, K., and Saravanan, S. (2008). Are non-performing assets gloomy or greedy from Indian perspective. *Research Journal of Social Sciences*, *13*, 6-7.

Kaur, M., and Singh, G.(2020). Calculation of client credit risk prediction in banking sector using data mining. *International Journal of Advance Research, Ideas and Innovations in Technology*, 6(5), 117-120.

Kaveri, V. S. (2001). Prevention of NPA suggested strategies *Vinimaya*, *23*(8), 7-9.

Kedia, N. (2016). Determinants of profitability of public sector banks. *International Journal of Management and Social Sciences*, *2*(3), 1-16.

Khasnobis, S. 2005. NPA emerging challenges. Indian Banker, IBA Journal, vol. 1, no. 11, pp. 17-20,

Kjosevski, J., and Petkovski, M. (2017). Non-performing loans in Baltic states: determinants and macroeconomic effects. *Baltic Journal of Economics*, *17*(1), 25-44.

Koju, L., Koju, R., and Wang, S. (2020). Macroeconomic determinants of credit risks: evidence from high-income countries. *European Journal of Management and Business Economics*, *29*(1), 41-53.

Konovalova, N., Kristovska, I., and Kudinska M. (2016). credit risk management in commercial banks. *Polish Journal of Management Studies*, *3*(2), 90-100.

Kumar, D., and Suresh, P. S. (2017). Financial performance of non-banking financial companies in India: An econometric study. *International Journal of Recent Scientific Research*, *8*(7), 18510-18517.

Kwan, S.,and Eisenbis, R.(1997). Bank risk, capitalisation and operating efficiency. *Journal of Financial Services Research*, *12*, 117-131.

Lam, J. (2014). Enterprise risk management: From incentives to controls.2nd Edition, Wiley.

Lam, J. (2015). Implementing an effective risk appetite. Statement on management accounting. Institute of Management Accountants, Montvale, New Jersey, USA.

Leo, M., Sharma, S., and Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, *7*(1), 29.

Levine, R. (1996). Financial development and economic growth. Policy Research Paper 1678, Development Research Group, World Bank.

Li, X-L., and Zhong, Y. (2012). An overview of personal credit scoring: techniques and future work. International Journal of Intelligence Science, 02(04),181-189.

Lokare, S. M. 2014. Re-emerging stress in the asset quality of Indian banks: Macro-financial linkages. RBI Working Paper Series. W P S (DEPR): 03/2014.

Madaan, M., Kumar, A., Keshri, C., Jain, R., and Nagrath, P. (2021). Loan default prediction using decision trees and random forest: A comparative study. *IOP Conf. Series: Materials Science and Engineering*,*1022*, 012042. doi:10.1088/1757-899X/1022/1/012042.

Mahato, J. K. (2022). Reserve Bank of India and Its Functions. *Supremo Amicus*, *29*, 95.

Makkar, A. (2016). Key determinants affecting the credit risk of Indian banks. International conference on recent innovations in science, management, education, and technology. retrieved from: www.conferenceworld.in

Malyadri, P., and Sirisha, S. (2011). A comparative study of non-performing assets in Indian banking industry. *International Journal of Economic Practices and Theories*, *1*(2), 77-87.

Malyadri, P., and Sirisha.S. (2003). Non-performing assets in commercial banks- An overview. *Banking Finance*,*16*(1), 9.

Misra, B. M., and Dhal, S. (2010). Pro-cyclical management of banks' non-performing loans by the Indian public sector banks. DBOD No. BP. BC. 6 /21.01.002/2009-10, July 2009, RBI/2009-2010/37, RBI (www.rbi.org.in).

Misra,T. P. (2003). Managing of NPAs – A professional approach. *IBA Bulletin*, *25*(1), 18.

Mittal, R. K., and Suneja, D. 2017.The problem of rising non-performing assets in banking sector in India: comparative analysis of public and private sector banks. International Journal of Management, IT and Engineering, vol. 7, no. 7, pp. 384-398.

Mohanty, P. K. and Patel, S. K. (2016). Basic Statistics. Scientific Publishers.

Muniappan,G. (2002). The NPAs overhang, magnitude, solutions and legal reforms. RBI Bulletin, May.

Muninarayanappa and Nirmala, (2004). Credit risk management in banks-key issues. *Journal of Accounting and Finance*, *18*(1),94-98.

Muniswaran,S. (2010). Nonperforming assets management of state bank groups.*MKU*,273.

Nag, A. K. (2015). Appraisal of non-performing assets in banking sector: An Indian perspective. *Indian Journal of Accounting*, *XLVII*(1), 133-143.

Naik, J. G. (2006). NPAs management challenges before banking sector. *The Management Accountant*, *41*(5), 360.

Pal, V. and Malik, N. S. (2007). A multivariate analysis of the financial characteristics of commercial banks in India. *The ICFAI Journal of Bank Management*, *VI*(3).

Pallart, J. (2005). SPSS survival manual: A step by step to data analysis using Spss version 15, 3rd ed. p. 146-165.

Pathak, B. V. 2009. The Indian Financial System – Markets. Institutions and Services, 2nd ed., Pearson Education.

Patnaik, P. (2019). Fifty Years after Bank Nationalization.

Patri. D. (2003). Statistical Methods. Kalyani Publishers, New Delhi, pp. 752-802.

Petkovski, M.,and Kjosevski, J. (2014). Does banking sector development promote economic growth? An empirical analysis for selected countries in Central and South Eastern Europe. *Economic Research-Ekonomska Istraživanja*, 27(1), 55-66.

Prakash, P. M. S. (2016). A study on the significance of risk management in banking sector. *International Journal of Research in IT and Management* , *6*(9), 135-139.

Prasanth, S., Nivetha, P., Ramapriya, M., and Sudhamathi, S. (2020). Factors affecting non-performing loan in India. *International Journal of Scientific and Technology Research*, *9*(1), 1654-1657.

Radelet, S., and Sachs, J. (1999). What have we learned, so far, from the Asian financial crisis?. *Harvard Institute for International Development, mimeo*.

RAF guidelines (2018). Guidelines on Risk Appetite Practices for Banks. European bank for Reconstruction and Development (EBRD). Available at: ENG%20Smernice%20v%20zvezi%20s%20praksami%20na%20področju%20nagnjenosti%20k%20prevzemanju%20tveganj.pdf

Raghavan, R.S. (2003). Risk management in banks. *Chartered Accountant*, 841-851.

Raj, A. (2020). Unlocking the true power of support vector regression. Retrieved from: https://towardsdatascience.com/unlocking-the-true-power-of-support-vector-regression-847fd123a4a0

Rajadhyaksha, N. (2004). The rise of financial conglomerates. *Business World,* 28-33.

Rajaraman, I., Bhaumik, S., and Bhatia, N.(1999). NPA variations across Indian commercial banks: some findings. *Economic and Political Weekly, 34*, 16-23.

Rajeswari, M. (2014). A study on credit risk management in scheduled banks. *Journal of Management*, *5*(12), 79-89.

Rajput, N., Anu Priya Arora, A. P., and Baljeet Kaur, B. (2011). Non-performing assets in the Indian public sector banks: an analytical study. *Banks and Bank Systems*, *6*(4), 84-89.

Rajput,N. and Goyal, A. K. (2019). Indian banking sector a major contributor to economy: Constancy major concern. *International Journal of Recent Technology and Engineering*, *8*(4), 11596-11608.

Rawlin, R., Shanmugam, R., and Bhat, V. (2014). A comparison of key determinants of profitability of India's largest public and private sector banks. *European Journal of Business and Management*, *6*(34), 62-68.

RBI (2020). Discussion paper on Governance in Commercial Banks in India. Available at: https://www.rbi.org.in/Scripts/PublicationsView.aspx?id=19613#:~:text=%E2%80%9CRisk %20Appetite%20Statement%20(RAS),other%20relevant%20measures%20as%20appropriate

Reddy, P. K. (2002). A comparative study of non-performing assets in India in the global context-similarities, dissimilarities and remedial measures. *SSRN Electronic Journal*, doi: 10.2139/ssrn.361322

Reddy, R. G. and Bhargavi, S. T. (2004). An appraisal of Indian banking from NPAs perspective. *The Journal of Accounting and Finance*, *18*(1), 53.

Reddy,Y.V. (2004). Credit policy systems and culture. RBI Bulletin, March.

Rehman, Z.U., Muhammad, N., Sarwar, B., and Raz, M. A. (2019). Impact of risk management strategies on the credit risk faced by commercial banks of Balochistan. *Financial Innovation*, *5*, 44. Doi: https://doi.org/10.1186/s40854-019-0159-8

Rehman, Z.U., Muhammad, N., Sarwar, B., and Raz, M.A. (2019). Impact of risk management strategies on the credit risk faced by commercial banks of Balochistan. *Financial Innovation*, *5*(1), 44.

Roberts, D. (2015). Exploring risk perception and management in U.K banks. Ph.D Thesis, University of Essex, UK.

Roland, C. (2006, January). Banking sector liberalization in India. In *Indian Institute of Capital Markets 9th Capital Markets Conference Paper*.

Salas, V.and Saurina, J.(2002). Credit risk in two institutional regimes: Spanish commercial and savings banks. *Journal of Financial Services Research*, *22*, 203-224.

Sankareswari, S. 2012. A study of credit risk management in commercial banks. Ph.D Thesis, Department of Commerce, Madurai Kamraj University, Madurai, Tamil Nadu.

Santoni, A., Ricci, E., and Kelshiker, A. 2010. US banks, causes of bank failures in 2009 and early warning indicators. Studi e Note di Economia, Anno XV, no.1-2010, pp. 161-176.

Sarstedt, M. and Mooi, E. (2014). A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistics. Springer, p.194.

Sarwar, B., Muhammad, N., and Zaman, N.U. (2020). Diversification, industry concentration, and bank margins: Empirical evidence from an emerging south Asian economy. *Journal of Asian Finance Economics and Business*, *7*(7), 349-360.

Schmitt, K. R. (2023). *Credit risk: Definition, Role of Ratings, and Examples*. Investopedia. https://www.investopedia.com/terms/c/creditrisk.asp#:~:text=%25%2025%25%200%25-

Sharifi, S., Haldar, A., and Nageswara Rao, S.V.D. (2019). The relationship between credit risk management and non-performing assets of commercial banks in India. *Managerial Finance*, *45*(3), 399-412.

Shrey, B., Tanmayee, B., Mohak, C., Animesh, D., and Geetha, I. (2018). Role of payment banks in India: opportunities and challenges. *Int. J. Adv. Manag. Econ*.

Sikdar, P., and Makkad, M. 2013. Role of non-performing assets in the risk framework of commercial banks: A study of select Indian commercial banks. AIMA Journal of Management and Research, vol. 7, no. 2/4.

Singh, A. (2013). Credit risk management practices of Indian commercial banks. *International Journal in Management and Social Science*, *3*(1), 89-94.

Singh, A., and Prasad, S. (2020). Artificial intelligence in digital credit in India. Dvara Research, April 13, 2020. Available from: https://www.dvara.com/blog/2020/04/13/artificial-intelligence-in-digital-credit-in-india/

Singh, A., and Sharma, A. K. (2016). An empirical analysis of macroeconomic and bank-specific factorsaffecting liquidity of Indian banks. *Future Business Journal*,*2*, 40-53.

Singh, C., Pattanayak, D., Dixit, D., Antony, K., Agarwala, M., Kant, R., ... and Mathur, V. (2016). Frauds in the Indian banking industry. *IIM Bangalore Research Paper*, (505).

Singh, C.,and Mohammad Farook Khan, M. F. (2005). Effectiveness of DRTs in recovery of banks dues. *IBA Bulletin*, *27*(10), 30.

Singh, M. (2023, March 19). *The 2007-2008 Financial Crisis in Review*. Investopedia. https://www.investopedia.com/articles/economics/09/financial-crisis review.asp#:~:text=The%202008%20financial%20crisis%20began

Singh, S. (2015). The measurement and management of risks in banks. *Academike*, *16*, 38.

Sudhakar, M., and Krishna Reddy, C. V. (2014). Credit evaluation model of loan proposals for banks using data mining. *International Journal of Latest Research in Science and Technology*, 126-131.

Swamy, B.N.A. (2001). New competition, deregulation and emerging changes in Indian banking. *Bank Quest The Journal of Indian Institute of Bankers*, *729*(3), 3-22.

Taori, K.J. (2000). Management of NPAs in public sector banks. *Banking Finance*,13(8), 11.

Tiwari, R., Chauhan, A. S., and Singh, P. (2022). Role Of Leadership For Sustainability In Indian Banking: A Case Of HDFC Bank. *Tiwari, R., Chauhan, AS, and Praveen Singh, H.(2022). Role Of Leadership For Sustainability In Indian Banking: A Case Of HDFC Bank. Journal of Positive School Psychology*, *6*, 67-74.

Tuo, M. (2016). An empirical analysis of Chinese commercial banks' efficiency and influencing factors – under the constraint of non-performing loans. *American Journal of Industrial and Business Management*, *6*, 455-466.

Uppal, R.K. (2009). Priority sector advance-trends, issues and strategies. *Journal of Accounting and Taxation, 1*, 86.

Valasmma, A. (2004). A menace to the banking industry. *Southern Economist,42*(17), 23.

Vallabh G., Bhatia A., and Mishra, S. 2007. Non-performing assets of Indian public, private and foreign sector banks: An empirical assessment. *The IUP Journal of Bank Management*, VI, 7-28.

Van Greuning, H., and Bratanovic, S. S. B. (2020). Market Risk Management.

Vidyashree D.V., and Rathod, P. (2015). *International Journal of Research in Finance and Marketing*, *5*(7), 44-47.

Wang, K. (2019). Comparative analysis of business management of Chinese and foreign commercial banks—based on the perspective of non-performing loans of commercial banks. *Modern Economy, 10*, 108-119.

World Bank Development Report (1988). An assessment of trends in poverty in Ghana 1988-1992. New York: Oxford University Press for the World Bank.

Zhang, D., Huang, H., Chen, Q., and Jiang, Y. (2007). Comparison of credit scoring models. Third International Conference of Natural Computation.

Zhang, D., Huang, H., Chen, Q., and Jiang, Y. (2007). Comparison of credit scoring models. Third International Conference of Natural Computation.

Zheng, C., Sarker, N., and Nahar, S. (2018). Factors affecting bank credit risk: An empirical insight. *Journal of Applied Finance and Banking*, *8*(2), 45-67.

Zhu, L., Qiu, D., Ergu, D., Ying, C.,and Liu, K.(2019). A study on predicting loan default based on the random forest algorithm *Procedia Computer Science*, *162*, 503-513.

Zou, Y., and Li, F. (2014). The impact of credit risk management on profitability of commercial banks: A study of Europe.

List of publications and working papers

Papers Published

- Ankur Joshi, NVM Rao and A K Vaish., (2022), " Usage of machine learning algorithm models to predict operational efficiency performance of select banks in India", International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 06, June 2022)
- Ankur Joshi, NVM Rao and A K Vaish (2022), " An empirical analysis of credit risk based on non-performing assets of select banks in India: data validation by using machine learning algorithms, International Journal of Emerging Technology and Advanced Engineering, (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Accepted.

Working Papers

- Three (3) Research Papers drawn from the work are communicated to Journals and they are under review, which are as follows:
  - Ankur Joshi, and NVM Rao (2023). Use of machine learning methodologies to create an econometric model to predict probability of default of banks. Springer Open, (Communicated).
  - Ankur Joshi, and NVM Rao (2023). Validating machine learning model results by out-of-time validation to test the accuracy. Springer Open, (Communicated).
  - Ankur Joshi, and NVM Rao (2022). Digital lending practices in Asian banks.

Brief Biography of Research Supervisor

Prof. N V Muralidhar Rao is a Senior Professor at the Department of Economics and Finance, Birla Institute of Technology and Science (BITS), Pilani – Pilani Campus. He has been working here since 1994 with involvement in all the four-fold activities of the institute including teaching, research, consultancy and institutional development. He is currently Senior Professor and Chief of Centralized Purchases at BITS Pilani. He has taught various courses and guided several PhD students. His research interests include Applied Econometrics, Economic Development, Economics of Strategy, Financial Economics, Behavioural Economics and Finance, Risk Management, Health Economics and Healthcare management, Economics of Education. e-governance and Digital revolution,

He is a lifetime member of Indian Economic Association, The Indian Econometric Society, Indian Society of Labour Economics and Association of Management Scholars International. He has eighty research articles published in national and international journals of repute and attended diverse national and international conferences. He has also been appointed as "Country Expert" by World Intellectual Property Organization, Japan and represented India on Multi-Country WIPO-UNU joint major research project on "Impact of Intellectual Property System on Economic Growth" and has addressed the UN University, Japan in 2007.

Brief Biography of Candidate

Ankur Joshi is a research scholar at the Department of Economics and Finance, Birla Institute of Technology and Science, Pilani – Pilani Campus.

He is currently working as Director, Regional Risk Officer, Asia Pacific at Citibank NA Singapore. Ankur has 19 years of experience in Citibank across credit risk, credit operations and credit risk analytics. He has worked in Citibank in different capacities holding senior leadership roles in Singapore, Malaysia and India.

Some of the recent positions that he has held include being Sr. Vice President, Global Risk Review, Consumer Business, Asia Pacific at Citibank, Kuala Lumpur, Malaysia. He was Sr. Vice President, Portfolio Credit Risk Head, Secured Lending at Citibank NA, India. He has held position of Risk Head, Mortgages business when he was working for Citibank India.

He is a subject matter expert in risk management with expertise of credit risk management for Retail lending portfolios in Asia Pacific & EMEA markets. He has rich experience of India lending business and this research work is motivated by his vast experience of India market.

Prior to Citibank, he has worked as mergers & acquisitions and venture capital analyst with Cadila Pharmaceuticals and Commonwealth Development Corporation funded venture capital fund in India.