"A STUDY OF NON-PERFORMING ASSETS ANALYSIS IN INDIAN BANKS USING DATA ANALYTIC TECHNIQUES WITH PARETO ANALYSIS"

Pramuk.R.Bhargav,
M.E (CS) student, HICET, Coimbatore.
Mrs.Uma Priya,
Asst Professor, Department of Computer Science and Engineering,
HICET, Coimbatore.
Dr.S.Uma,
Professor, Department of computer science and Engineering,
Hindustan College of Engineering and Technology, Coimbatore.

Abstract:
The research paper analyses the Non-Performing Assets analysis in Indian banks using data analytic Techniques with Pareto analysis." The journal paper will analyze NPA of Indian Banks Using Data Analytic Techniques with pareto technique segregation of data . The python analytical tool is used for analyzing the vast data and visualizing the data with required formats and outcomes required to draw conclusion useful for decision makers in the top positions in banks and the government policy makers to take decisions . The pareto analysis of segregating the data w.r.t the 70:30 formula with an extended pareto technique of 80 :20 or 90:10 to find which seventy percent of NPA is held by 30 percent of even less number of companies having highest NPA holding for speedy actions and reducing of NPA burden on banks as well as nation.

This will help all stake holders of NPA including tax payers to take proper steps to reduce the NPA , by prioritising their financial help and redressal process aimed towards more sick sector (among manufacturing, service and mixed sectors) and specifically pin pointing actions to red alert NPA specific industry (hold Maximum NPA of banks) and customer set.

Keywords : NPA , Indian Banks,

1. Introduction:

1.1 : Banking and NPA : The banking is process in which a banking organization , which collects deposits from public and lends to other people or organizations mostly which are business entities like proprietary concerns , partnership firms or companies. Banks operate and make profits by collecting deposits at a lower rate of interest than the lending rate leading to public and business. The lending business of banks drive these companies to two major categories – Performing companies and Non-Performing companies –financially because of many reasons. They are said in banking terms as – Performing assets and Non-Performing assets (– NPA-)for the banks and the whole country.

1.2 : Project : The project has selected three major banks and analyze the NPA of these banks for a decade and find "sector"(power sector, service sector, manufacturing sector etc..) wise and "loss making reason" wise (management reasons, market forces driving companies to losses, global reasons like recession, wars etc..and other reasons like firm owner's death, division in partnership , division of companies ,amalgamation of companies etc..). pareto analysis of banks will go ahead with nationalized, scheduled and private banks – three selected sample banks on a comparative basis – the NPA analysis and bring out suggestions and conclusions useful for the policy makers of India to reduce the NPA of banks to find specific NPA alert company lists itself on this basis .

The use pareto diagram techniques and analysis to go into micro level surgical analysis to find – 70 : 30 , 80 : 20 and 90 : 10 analysis process to find the list of companies which hold 70 percent (80 & 90 respectively) NPA total amount of list for speedy action and steps for reducing of NPA amounts.

2. Literature survey :

The literature survey on the following aspects mentioned below have be exhaustively done to and arrive at new strategies with pareto analysis , data analytics and python programming for Indian banking NPA data analysis. The bird view of the literature survey conducted is given below:

2.1 Data analytics be used to address the common problems of NPA management:

When we analyze the root cause of the issue with NPAs, we find that as far as borrowing entity is concerned, the issue relates to the mis-utilization of cash and loan, suboptimal project management and oversight, invisible and unclear corporate structures, excessive leverage on the balance sheet and creative accounting practices.As far as banks are concerned, highly divergent loan appraisal standards, absence of a common database, lack of timely intervention on red flags, excessive use of discretion, non-perfection of security interests, lack of loan covenants enforcement and not informing the statutory authorities are the common issues with NPA management.

Some of the primary benefits of Data Analytics for non-performing assets are as follows:

Forecasting of NPAs : Data analytics can help in the forecasting of the NPAs as NPAs are going to pose a massive challenge after Covid-19 considering a lot of businesses are not doing well. It can help in predicting loans that are going to turn NPAs in the near future.

Devising right strategies for reduction of NPA : Data analytics can help the business recovery teams to come up with the right strategy to deal with the account in terms of measures and strategies.

Prevention of cartelization: By using analytics, cartelization that is resorted by the borrowers to prevent banks from recovering the fair value of security can be thwarted.

Enhancement of the process and reliable data base automation : Data analytics will cause the enhancement of the process and reliable database automation, which will further help in improving the transparency in decision making. It will also help the bankers not to fear later accountability, which is one of the significant obstacles in the rebooting of the system.

2.2 : The research paper "Performing Loans Financial Risk: A Study And Analytics Though Data Mining" : by Rashmi Bisht and , Pooja Dixit at -Symbiosis Institute, pune state as follows in its abstract that "The principle focus of this paper is to recognize and analysis the risk in giving loans of financial association. For analysis, the risk giving loans, data mining techniques are used. The technique includes process and analysis of data that collect from various resources and summaries that data into meaningful information. This paper, examine the risk of Non Performing loan (NPL) in India. this research use the data mining Regression technique by utilizing multiple linear regression (MLR) and Classification techniques by utilizing Naïve Bayes and Support Vector Machine (SVM).

The data predicted by using regression model that help to predict the NPL, Which can help to produce better prediction results. Which is apply on Datasets, predict the GDP, Gross NPA, Repo rate of banks, and give better financial system for customer retention? The banks and numerous business organizations are pioneers in exploiting Data Mining. Subsequently, this research talks about that, how NPL of commercial and public banks connected with the financial consideration."

The paper introduces NPL – non performing loan / NPA as follows: At present both developed and developing country governments removing the consideration the issue of NPL. NPL is significant since they indicate the credit nature of the loan portfolio of banks. A comprehension of the components that affect the degree of NPL is essential for the risk management capacity of banks and for national bank managers answerable for banking security. The determinants of nonperforming loans inside and across nations is a significance theme in the non-performing loan scholastic writing [Skarica,2013].

2.3 Pareto analysis is a formal technique useful where many possible courses of action are competing for attention. In essence, the problem-solver estimates the benefit delivered by each action, then selects a number of the most effective actions that deliver a total benefit reasonably close to the maximal possible one.

Ref: Wikipedia.

Pareto analysis is premised on the idea that 80% of a project's benefit can be achieved by doing 20% of the work—or, conversely, 80% of problems can be traced to 20% of the causes. Pareto analysis is a powerful quality and decision-making tool.

Ref :Pareto Analysis Definition – Investopedia www. investopedia . com

Pareto Charts & 80-20 Rule

The Pareto Chart is a very powerful tool for showing the relative importance of problems.

It contains both bars and lines, where individual values are represented in descending order by bars, and the cumulative total of the sample is represented by the curved line. An 80% cut off line is also included to indicate where the 80/20 rule applies i.e. the vital few factors that warrant the most attention sit under the 80% cut off line.

80/20 Rule – The Pareto Principle

The 80/20 Rule (also known as the **Pareto principle** or the law of the vital few & trivial many) states that, for many events, roughly 80% of the effects come from 20% of the causes. Joseph Juran (a well regarded Quality Management consultant) suggested the principle and named it after the Italian economist Vilfredo Pareto, who noted the 80/20 connection in 1896. Vilfredo Pareto showed that approximately 80% of the land in Italy was owned by 20% of the population. Pareto also observed that 20% of the peapods in his garden contained 80% of the peas. According to the Pareto Principle, in any group of things that contribute to a common effect, a relatively few contributors account for the majority of the effect. Commonly, it is found that:

- 80% of complaints come from 20% of customers
- 80% of sales come from 20% of clients
- 80% of computer crashes come from 20% of IT bugs

The ordering in a Pareto Chart helps identify the 'vital few' (the factors that warrant the most attention i.e. factors whose cumulative per cent (dots) fall under the 80% cut off line) from the 'trivial many' (factors that, while useful to know about, have a relatively smaller effect i.e. cumulative per cent dots that fall above the 80% cut off line).

Using a Pareto diagram helps a team concentrate its efforts on the factors that have the greatest impact. It also helps a team communicate the rationale for focusing on certain areas.

The example in Figure 1 (above) shows a Pareto Chart of types of medication errors. An audit of 430 medication errors was conducted to determine the categories (types) of errors and their frequency. The results were collected initially in a Tally Sheet then the data was placed in descending order of frequency in a Pareto Chart Template in Excel.

The types of errors that fall under the 80% cut off line indicate the 'vital few' types of medication error that should be addressed as a priority as they contribute most to the problem ie:

- Dose missed
- Wrong time
- Wrong drug
- Over dose

The types of medication errors that fall above the 80% cut off line are known as the 'trivial many' and are generally seen as not a high priority to address when compared to the 'vital few' factors. However, some of the 'trivial many' factors may be simple to address (low hanging fruit) and therefore may be acted upon earlier rather than later.

2.4 The article “Evaluating the Performance of Indian Domestic Banks Through the Lens of Pareto–Koopmans Efficiency “ published in December 2020 Global Business Review by Authors: Karan Singh Khati and Deep Mukherjee state as under :

In this article, we use data envelopment analysis to obtain Pareto–Koopmans (PK) measures of technical efficiency (TE) of India’s domestic commercial banks for the period between the global financial crisis and merger of the State Bank of India and its associates. This article aims to contribute to the growing body of literature on the efficiency of Indian banks by adopting the concept of PK efficiency to overcome the restrictive nature of radial and orientation-specific TE measures. To the best of our knowledge, this article is the first of its kind where one can disaggregate overall TE into two separate components by measuring input and output efficiency in the Indian banking sector. We assume a three-input three-output technology for both groups and utilize a balanced panel of 26 public sector banks (PSBs) and 19 private banks (PVBs) from 2010–2011 to 2016–2017. The mean PK efficiencies across the study period are 0.86 and 0.72 for PSBs and PVBs, respectively. Hence, there is considerable scope of improvement in the productive performance of PVBs. The disaggregation of PK efficiencies into input- and output-specific components reveals that for PSBs, the inefficiencies primarily result from physical assets, while for PVBs, they emerge mainly from other incomes. Hence, the management should specifically target these aspects of banking operations to improve their performance. Second-stage regression analysis reveals that PK-TE has a non-linear relationship with the size of a bank.

Deposit to liability ratio and management quality negatively impact PK efficiency, while priority sector lending positively influences it.

3. System analysis :

The proposed system of analysis **of project** will use python language for data analytics of volume of NPA data of Indian banks (sample data selected from Indian banks) and gives better data graphical representations for better visualisation of NPA problem of the banks and to find the reason behind such NPA generation . The project also brings out consolidation statements of NPA amts in each sector wise/ bank wise /reason wise/ group wise etc.. The proposed system of analysis of will use python language for data analytics of volume of NPA data of Indian banks and effectively uses Pareto techniques to go into pin pointed "small " list of companies which generate the maximum NPA in banks.

4. System design :

To analyze the NPA data of banks. the dataset is imported in python, some packages are imported to plot graph, segregating , reordering the data set on required attributes for decision making. This analyze the to find the main reasons for generation of NPA etc.. sector wise , reason wise, banks wise etc..

The project designs an algorithm to match "Pareto technique analysis (70:30) which states that 70 percent of Total NPA amount of list falls in 30 percent of the NPA list of companies " of accumulating the NPA amounts into a counter of 70%, 80% and 90% of NPA amount of the "total NPA" amount of the list of all NPA companies .
The project makes a NPA counters with 70%, 80% and 90% tag with reference to the total NPA amount of the whole NPA list of banks and checks and gives out a brief list of specific companies which are responsible for accumulating the 70% of NPA amount , 80 % of NPA amount and 90% of the NPA amount out of total NPA of list respectively (and finds whether these lists fall with in 10 or 20 or 30 percent of total list of NPA companies) which are responsible for generating – "focussed corrective actions list" of few specific companies and leaves the rest of list for later actions and greater NPA recovery on NPA rich "specific" groups.

5. Research methodology :

The project will go ahead with nationalized, scheduled and private banks on a comparative basis – the NPA analysis for a decade and bring out suggestions and conclusions useful for the policy makers of India to reduce the NPA of banks which are becoming a major burden on the shoulders of central and state governments. Includes macro category of sectors on the sample data provided from three major banks in India year wise will be analysed. This phase ii will use pareto diagram techniques and analysis to go into micro level surgical analysis to find – 70 : 30 , 80 : 20 and 90 : 10 analysis process to find

I) On the basis of NAP amount: from total list of companies :

(program algorithm shown in pages 1-2 of this file)

- To find whether 70 percent of NPA total amount is held in 30 percent of companies or
- To find whether 80 percent of NPA total amount is held in 20 percent of companies or
- To find whether 90 percent of NPA total amount is held in 10 percent of companies .

II) On the basis of sector: to find top list in each sector :

- To find whether 70 percent of NPA total amount is held in 30 percent of companies in each sector or
- To find whether 80 percent of NPA total amount is held in 20 percent of companies in each sector or
- To find whether 90 percent of NPA total amount is held in 10 percent of companies in each sector.

iii) On the basis of Individual banks: to find top list in each Bank (Canara / SBI / ICICI) :

The same program is run for each bank (Canara , SBI and ICICI) to find:

- To find whether 70 percent of NPA total amount is held in 30 percent of companies in each individual bank or
- To find whether 80 percent of NPA total amount is held in 20 percent of companies in each individual bank or
- To find whether 90 percent of NPA total amount is held in 10 percent of companies in each individual bank .

iV) which specific "reason" category group is producing height NPA:

- To find whether 70 percent of NPA total amount is held in 30 percent of companies in each "reason" or
- To find whether 80 percent of NPA total amount is held in 20 percent of companies in each "reason" or
- To find whether 90 percent of NPA total amount is held in 10 percent of companies in each "reason" .

v) which specific industrial group (list five NPA top companies) is producing height NPA: (under each bank and each sector) :which 10 percent of **specific industrial group is responsible for** producing 40-60 percent of NPA amount of each bank / sector .

6. The conclusions and suggestions:

The project was implanted using python language on the basis of the above planning and has come with useful outputs. The project has made successful "result analysis" of the NPA data to find three dimensional results by combining more than two attributes of data set with all categories of cluster diagrams and consolidated and summary reports –on the one hand And also the project has made successful Pareto techniques implementation – categorisation of the whole list by classification of " highest NPA holding companies " from the rest of NPA list to give out specific hand full of company list for speedy and focussed action and better results to reduce NPA total burden on the banks on the other hand.

Both these results – either graphical representation of the red alert areas in each sector/ bank/ reason (with consolidated results) and the specific pin pointed names (NPA companies) from the total list (to drastically reduce NPA burden by 70% / 80% / 90% etc.. based on the selection of NPA high alert companies from the list.) This will help for better understanding of the NPA data from banks for decision making purpose from policy makers , bank management , reserve bank of India and government of India – finance directorates .

Bibliography :

a. https://enterslice.com/learning/online-data-analytics-for-non-performing-assets-npa/ . research author profile : Ashish M. Shaji has done his graduation in law (BA. LLB) from CCS University. He has keen interests in doing extensive research and writing on legal subjects especially on criminal and corporate law. He is a creative thinker and has a great interest in exploring legal subjects.
b. Non Performing Loans Financial Risk: A Study And Analytics Though Data Mining Rashmi Bisht1 Assistant Professor , Pooja Dixit3 , Research Scholar at - Symbiosis Institute of Management Studies, Pune, India Symbiosis Institute of Management Studies, Pune, India tushar.rathore@rediffmail.com JOURNAL OF CRITICAL REVIEWS ISSN- 2394-5125 VOL 7, ISSUE 15, 2020 3905.
c. Advantages and Disadvantages of Pareto Chart – Geeks for Geeks www. geeksforgeeks.org › advantages-and-disadva..
d. "Evaluating the Performance of Indian Domestic Banks Through the Lens of Pareto–Koopmans Efficiency " published in December 2020 Global Business Review by Authors: Karan Singh Khati and Deep Mukherjee