"THE NON-PERFORMING ASSETS ANALYSIS IN INDIAN BANKS USING DATA ANALYTIC TECHNIQUES AND PYTHON PROGRAMMING."

Pramuk .R.Bhrgav, student, M.E-CSE, HICET, Coimbatore**.**

Mrs.UMA PRIYA, Assistant Professor, Guide, Department of CSE,
Hindustan College of EngineeringandTechnology,Coimbatore–32.

Abstract:

The paper is analyzing "**THE NON-PERFORMING ASSETS ANALYSIS IN INDIAN BANKS USING DATA ANALYTIC TECHNIQUES AND PYTHON PROGRAMMING."** of Indian banks. 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 w.r.t the priority they may have give to type of bank, type of sector, type of industry to reduce the NPA of banks which are becoming burden to the country and tax payers.

The paper will try to bring out programming strategy to segregate and stratify - list of NPA data of banks , to help management of banks, reserve bank of India, governmental financial policy makers to take proper steps to reduce the NPA , by prioritizing 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 and customer set.

The conclusions of the paper will come handy in the hand of policy makers to take immediate actions as they know the main reason of NPA generation and which are of high value area, and in which bank , sector, type of specific industry needs much attention and calls for aids, subsidies , incentives and governmental financial supports like load wavers , rate of interest reduction & rescheduling of loans for more than a decades to see that the industries wake up and help the countries progress.

Key words used: NPA – non-performing assets, BANKS- Indian banks,

1. Introduction: 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 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 drives 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.

1.1 Introduction to research paper : Phase –I:

The paper first phase will select five major nationalized 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..). In this paper phase - python is used as backend and jupyter notebook as frontend for better analysis and visualizing the result and analyzing red alert zones of NPA high density to find which sector, sub-sector, banks, banks category, reason category is responsible for NPA increase so that policy of subsidy, help , incentives , rescheduling of loans, loan period from 4 yrs to 12 yrs to bring back industrial environment back into life and

NPA data analysis using bar charts, scatter diagrams and data re-re-ordering and data segregation on the basis of class and categories. The regression is applied on the basis of following pairs. NPA Amount v/s sector, NPA Amount v/s banks,NPA Amount v/s reasons, Sector v/s banks v/s NPA ,Sector v/s reason v/s NPA And classification and ordered data sets on sector, sub-sector, bank wise, reason wise etc.. With NPA amounts and total amounts in each category, sub category and classifications. **RESULT** analysis of the NPA data to find three dimensional results by combining more than two attributes of data set for better understanding of the data for decision making purpose from policy makers.

2. MAIN OBJECTIVE AND AIM OF PAPER : The main aim of the paper is to analyze NPA data on different categories and . The main objective is to find the strategy for NPA reduction by pin pointed sector/ sub sector / reason / bank/ customer group /. Python is used for analyzing and visualizing the data. First the data is collected and the data is pre-processed by removing the duplicate votes and removing null values. NPA data is used for 2d and 3d Visualization on the basis of paired attributes is that visualizing the NPA reducing strategy by finding red alert zones in these attributes. The paper Uses the machine learning algorithm to find the red zones by comparison and summation and reordering the data based on attribute categories. The algorithm that is used for analyzing the data are Logistic Regression, Polynomial Regression and Linear Regression. NPA analyzed and the results are also visualized using the pair plot.

3. REVIEW OF LITERATURE

3.1 Introduction: Non-Performing Assets (NPA) is a loan or advance for which the principal or interest payment has remained overdue for a period of 90 days. A loan may be classified as a non-performing asset when it has not being repaid by the borrower.

2.2 data analytics be the answer for the management of NPA : Digitization has been implemented in many industries, and banking is no different. However, investment and learning's have been limited to the retail landscape. Corporate banking and **risk management** applications have taken a back seat, or it started late. Lack of implementation or late implementation of digital technology has impacted the loans set up, and it has become one of the significant issues concerning the Indian Banking set up.It is noteworthy to mention that data analytics can be used to improve the deficiencies of the banking industry and control the trickery of the borrowers. Established used cases from retail, as well as consumer banking, can be used without a massive investment and IT change.

3.3 ***Utility of data analytics for finding the major factors :that are responsible for the surge of NPAs are as follows:*** Improper Due Diligence, Lenient Credit terms, Loose Credit Monitoring, Collateral free loans, Frauds,Wilful default by customers It may be noted that these factors can be addressed and controlled by AI and Data Analytics. **2.5 the benefits of using data analytics for Non-performing assets:** As stated earlier, **data analytics** provide a number of solutions for the management of non-performing assets. Therefore it is beneficial.

3.4.Some of the primary benefits of Data Analytics for non-performing assets are as follows: forecasting of NPA's, devising right strategies, prevention of cartelization, enhancement of process and reliable data automation, no need to fear about later accountability.

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

Ref: enterslice.com / learning / online-data-analytics-for-non-performing-assets-npa . research author profile : Ashish M. Shaji

3.5 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.

References : Non Performing Loans Financial Risk: A Study And Analytics Though Data Mining Rashmi Bisht and Pooja Dixit , from -Symbiosis Institute of Management Studies, Pune, JOURNAL OF CRITICAL REVIEWS ISSN- 2394-5125 VOL 7, ISSUE 15, 2020 3905.

4. System analysis:

Present system of NPA analysis and proposed data analytic tools using python : Presently No data analytic tools are effectively used in India fro NPA data analysis and this paper would try to bring before the use of python for analyzing the NPA large data for the purpose of bringing reasonably good inferences on the data and better understanding of the NPA state in India and which would be useful for the management of banks as well as the government of India.

5. Significance: The python language and data analytic tools taken take upon with machine learning algorithm to analyse and bring specific and accurate outcomes which may be useful in better understanding of NPA huge data of banks. The Outcomes , inferences are going to be useful for management of Indian banks, reserve bank of India and government of India to take corrective actions for reduction of NPA in India. So it clearly shows the significance of analysis and paper and it's utility to policy makers at banks as well as government.

6. SYSTEM SPECIFICATION

6.1 HARDWARE REQUIRMENTS -Device name- WINDOES-74PPV22**,**Processor - Intel(R) Core(TM) i3-7020U CPU@ 2.30GHz 2.30 GHz**,** Installed RAM- 4.00 GB (3.43 GB usable) Device ID-AEF4C744-BF5D-433D-B41A-93AF49E3C1D7**,**Product ID-00326-10000-00000-AA591**,**System type- 64-bit Operating system, x64-based processor**,** Pen and touch-No pen or touch input is available for this display

6. SOFTWARE and HARD ware REQUIRMENTS -Operating systems: Windows* 7 or later, macOS, and Linux**,**Python* version: 3.8.5**,**Operating System: Windows 10 Home/1909 **,**Front-end Tool: Jupyter Notebook 6.1.4

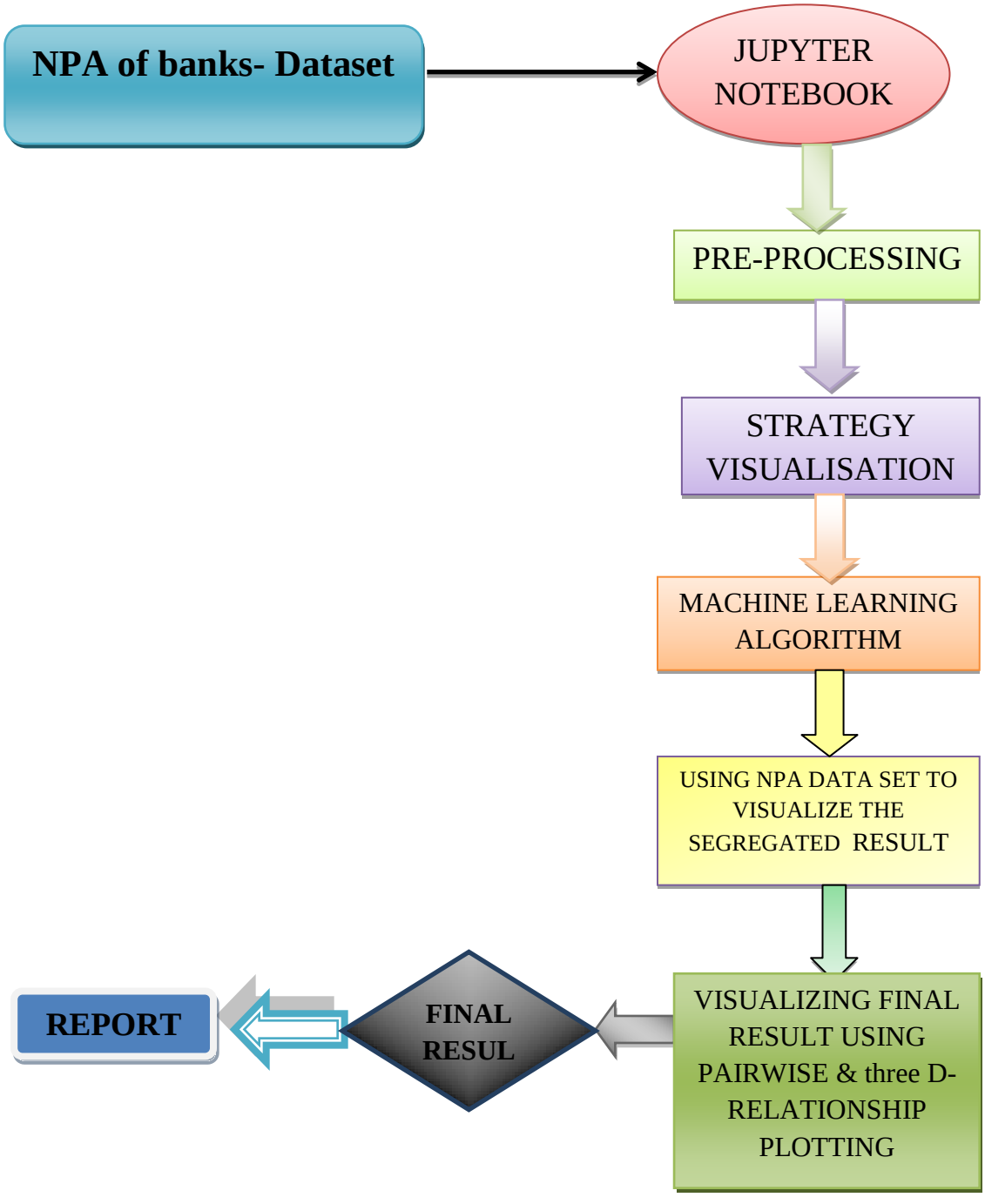
ABOUT THE SOFTAWRE: Python Programming Language: Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding; make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together.

7. 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.
- Then the dataset is pre-processed by removing null values and to know the information about the dataset.
- After finding red alert areas of NPA, by using Machine learning Algorithm researchcan analyze the to find the main reasons for generation of NPA etc.. sector wise , reason wise, banks wise etc..
- Here Logistic Regression, Polynomial Regression and Linear Regression is used to get accurate results. Logistic Regression compares and find sector wise , bank wise , reason wise results w.r.t NPA of banks are used. . Polynomial Regression is used to find the most NPA concentration between paired attributes.
- Paper used scatter plot to visualize the dataset of NPA data better understanding of NPA trend and concentration .
- Linear Regression shows that (bank – NPA amount v/s sector , bank– NPA amount v/s reason, bank– NPA amount v/s specific industry).
- Three dimension representation will also be used like – amounts (y axis), sector (x- axis) and reason in colours which in the dots which represent the particular data value by comparing two attribute and also comparing three attributes are analyzed and the final useful paper results are also visualized using the pair plot.

8. DATA FLOW DIAGRAM:

The flow chart explains about the DATA FLOW about the paper. Connectivity



9 DATABASE DESIGN

TABLE NAME: NPA data set NPA_DATA

Field name	Data type
bank name	object
Customer-Name	int64
year	int64
Bank type	object
Company_type	object
sector_type	object
Sub_sector_type	object
Reason_type	object
NPA-Amt-cat	object
NPA_amount	int64
Year	int64

THE CLASSIFICATIONS OF THE NPA DATA AND ANALYSIS TECHNIQUES PLANNED :

The NPA data collected from banks for the year 2020-21 are tabulateas follows:

SL no	Bank name	Name of customer/ company	Type of bank	Sector Category (major)	Specific Sector (macro)	NPA Amount pending	NPA Pending amt category	Reasons For NPA. NPA category	

Table contents :NPA TABLE

1. Sl-no :
2. Name of the customer (loaner / defaulter Of the bank which are considered by bank under NPA) or the company
3. Name of the bank.
4. Type of the company.
5. sector wise (sector or type of industry –under main sectors)
6. Specific sector under main sector.
7. NPA – pending amount from customers
8. Amount – NPA value - (value wise classification code)
9. Reasons (wise classification code)
10. Special info : (Special information regarding NPA customers if any)

Type of the bank: B1 - public sector banks, B2 - Schedules banks**,** B3 – private banks **and** B4 – foreign banks

Type of the company: C1 - public sector, C2 – private sector**,** C3 – partnerships / proprietary etc **and** C4 – foreign companies and firms

Type of the industry or sector / company's activity : Manufacturing (under that 10 sub-classifications) **,**Service (under that 9 sub-classifications) **and** Mixed sector - M1 – many sector based industrial groups

Amount – NPA value - (value wise classification code) classifications: A1-, A2-A3- between A4 –and **A5** – on the basis of pending NPA amounts

Reasons: (wise classification code) **(Reasons behind the NPA generation in companies)**

R1- Management Reasons (Ill Management),

R2- Market Forces Driving Companies To Losses,

R3- Global Reasons Like Recession, Wars, Diseases Etc..

R4- Other Reasons Like Firm Owner's Death, Division In Partnership, Division Of Companies, Amalgamation Of Companies etc.

10 .Research PAPER MODULES -DETAILS AND INDIVIDUAL MODULE WORKING DESCRIPTION

1. Module 1 :Identifying and pre-processing NPA data on the type of bank, main sector, sub sector, reason category, NPA pending amt category with codes given in the schedule –I of paper and customizing them and removing null values.

2. Module 2:Taking the past NPA data set into account classifying the data on the basis of bank type, banks, main sector , sub sector, reason wise on the order of priority and analysing them and visualising the comparison therefore, which will help In next mosules of predicting the top slot –red alert NPA zones under them.

3. Module 3: then data is entered w.r.t the (type of bank, main sector, sub sector, reason category, NPA pending amt category) NPA amounts and banks with other details in the date fields with the four main NPA analysis strategies and analysing with python programs for visualising Final Result using pair wise relationship plotting and calculate the accuracy score for logistic , polynomial and linear regression for NPA amt V/s sector (manufacturing , service and mixed) , NPA amt v/s sub sector, NPA amt v/s type of bank , NPA amt v/s banks, NPA amt v/s "reason wise" category analysis.

4. Module 4: Then the final data is filtered and the data is interpreted with visualization with the help of data analytic and embedded statistical keywords available in python language with bar diagrams , segregation of data on the basis of above categories with 3 D visualisation with colour of NPA amt category, x- axis (sector or sub sector or "reason") and Y-axis as banks .

Or y- axis NPA amount pending, x- axis – sector and colour of dots (bank category)

5. Module 5 : inferences and conclusions on graphical representations , segregated data outputs with w.r.t each classification attributes parameters (sector/ sub-sector/ bank/ reason/) to the help the banks top management s and government financial departments to understand NPA red alert zones and decisions based to reduce the NPA amounts of bansk with proper fiscal policy changes in their financial strategies

10.1. Research Paper outcomes :

I. Graphical 2d, 3d Visualizations:

II. The Graphs And Stratification On The Basis Of Regression Analysis

III. Program Codes And Outcomes For Consolidated Reports Of Npa Pending On Different Classifications

11.1 CLASSIFICATION OF DATA ON THE BASIS OF FOLLOWING CATAGORIES

I) SECTORS - npa PENDING

SUB-SECTOR WISE – NPA PENDING

REASON WISE – NPA PENDING

NPA PENDING CLASSIFICATION WISE -

II) BANK WISE - NPA PENDING

COMPANY CLASSIFCATION WISE – NPA AMOUNTS

BANK WISE - NPA AMOUNTS

III) TWO PAIRED , THREE PAIRED AND MULTI- PAIRED CONSOLIDATED LIST GENERATIONS -IN DECENDING ORDER OF NPA AMOUNTS PENDING:

A) - BANK WISE, SECTOR WISE, SUB SECTOR WISE – NPA PEDING TOTALS

B) - BANK WISE, SECTOR WISE, REASON WISE – NPA PEDING TOTALS

C) SECTOR, REASON, NPA CLASSIFICATION CODE WISE, NPA AMOUNTS TOTALS

10.2 Other outcomes of the project:

- count of bank, number of npa cases, amounts and amount category
- count of bank wise nap number of companies and amounts
- predicted value 1 - to find high npa values
- to analyse red alert npa zones in data base
- increase or decrease of npa – sector wise, bank wise, reason wise using polynomial regression
- to create an attribute which is npa value of banks for finding high –low – medium npa zones and compare with sector , reason and amount zone wise comparison with result, then find linear regression -npa data analysis ,amount v/s sector, amount v/s banks and amount v/s reasons

11 .Results, discussions and conclusions :

11.1 RESULT: The analysis the NPA data to find three dimensional results by combining more than two attributes of data set for better understanding of the data for decision making purpose from policy makers like banks, governments and other corporate sectors – who are the main stake holders of NPA . Then above data is Plotted of NPA data y by mentioning x label as sector and y label as amounts – amount zones and title as Plotting of NPA –Indian banks.

11.2 DISCUSSIONS : This paper recognized to analyze the NPA data And got following results N offer R analysis Here researchuse Logistic Regression and Polynomial Regression to analyze the npa data of Indian banks Using Linear Regression compare total NPA value in each category which shows us better understanding of NPA increase , red alert area of NPA , customer zones, etc..(in all classifications and categories)

The results are as follows:_The programs will be able to pin-point NPA red alert zones in each banks, type of banks, in each sector, in each sub-sector, and also will be able pin point the "main reason " for such NPA high value generations. (if the same program techniques are used on the real –time data of the banks which will run for lakhs of records).

11. 3 CONCLUSION

Inferences and conclusion on based on the following comparison of attributes and classification - In this paper - python is used as backend and jupyter notebook as frontend for better analysis and visualizing the result and analyzing red alert zones of NPA high density to find which sector, sub-sector, banks, banks category, reason category is responsible for NPA increase so that policy of subsidy, help , incentives , rescheduling of loans, loan period from 4 yrs to 12 yrs to bring back industrial environment back into life and save employments in the society.

The paper will bring out the specific category ,sector, customer wise and NPA amount wise for understanding the trend and alert zones of the NPA pending .

(NPA data analysis used bar charts, scatter diagrams and data re-re-ordering and data segregation on the basis of class and categories. The regression is applied on the basis of following pairs. NPA Amount v/s sector, NPA Amount v/s banks, NPA Amount v/s reasons, Sector v/s banks v/s NPA ,Sector v/s reason v/s NPA And classification and ordered data sets on sector, sub-sector, bank wise, reason wise etc.. With NPA amounts and total amounts in each category , sub category and classifications.)

RESULT analysis of the NPA data to find three dimensional results by combining more than two attributes of data set for better understanding of the data for decision making purpose from policy makers .

11.4 Scope for further study : There is a big scope for future improvement and modifications of the paper by way of inroducing furthur modules of analysis of NPA data of all banks in india (including private and foreign banks besides – public sector, nationaised and scheduled banks) analysis of NPA data , company wise cataory , major sector wise mapping and international level analysis for better understanding .

REFERENCES:

- Ref: enterslice.com / learning / online-data-analytics-for-non-performing-assets-npa . research author profile : Ashish M. Shaji
- References : Non Performing Loans Financial Risk: A Study And Analytics Though Data Mining Rashmi Bisht and Pooja Dixit , from -Symbiosis Institute of Management Studies, Pune, JOURNAL OF CRITICAL REVIEWS ISSN- 2394-5125 VOL 7, ISSUE 15, 2020 3905.