Grenze International Journal of Engineering and Technology, Jan Issue

Bank Reliability and Stability: An Important Concern

A S Punith[1], Namana J Jain[2], Nagavishnu H[3] and Bhaskar G[4]
[1-3]Student, Dept. Of CSE, Siddaganga Institute of Technology, Tumakuru, Karnataka, India.
{punithsujatha, namanajjain, nagavishnuh}@gmail.com
[4]Assistant Professor, Dept. Of CSE, Siddaganga Institute of Technology, Tumakuru, Karnataka, India.
bhaskar_gopal@sit.ac.in

***Abstract*—In modern era, use of technology has made the life of mankind simple. Also, it is vital to see that the technology must help the people in securing their lifestyle. Money is a key factor for any person's living. It's important to secure this for an individual's life too. People had believed that the most secured source of investing the money is in banks and co-operative societies. In this paper, we have discussed the challenges and difficulties that one faces if an individual invests money in any of the bank or co-operative sector without having the prior knowledge. This paper also discusses the cons of the existing solution and the pros of the approached solution. Further, the important parameters to be considered along with the prepared dataset has been discussed which leads to the application of Machine Learning Algorithms on solving the problem. Analysis of the performance of bank based on valid parameters is the concerned part of paper.**

***Index Terms*— CRISIL, Stability, Reliability, Performance, Machine Learning, Classification.**

I. INTRODUCTION

In the modern days, securing money for the future use is the most vital part of any individual's lifestyle. It is important for everyone to keep their money safe, and more importantly the place where the money is kept safe becomes utmost crucial. A proper source for the right investment of our money must be found. Banks and co-operative societies are one such promising sectors who can provide an extent of security on our money. But, blindly believing anyone (or anything) is quite not good. Hence, there are some cons in every sector.

The 2020 pandemic had made a drastic change in main two sources: one is cinema and another one is cricket. But these are the two main visible affected sectors. But the undiscovered sector that got affected by the pandemic is the Banking System of India. Before the verge of the pandemic, there was already an easily visible increase in non-performing assets, commercial bank failures, financial frauds and many more. There was also a rise in the loan moratoriums which is not good for any financial sector. Let us see this in detail and in a clearer way.

In recent times banks and co-operatives have no longer been that secure and it turns out that they are volatile whenever there is situation of crisis or bankruptcy. More than 80% Indians nowadays have a bank account which is the same portion that has a smartphone, and they have to consider the fact that no bank is going to assure them about the safety of their money and recovery if they face any crisis or bankruptcy.

In the past few years due to the negligence of the government in the banking sector has led to huge losses and degraded the trust of customers. One such case is of the Laxmi Vilas Bank, which reported a loss of Rs 650 crores in March 2018. Since then, the bank has been in losses quarterly. They decided to merge with India Bulls Housing Finance, as they were not able to attract investors, and to regain the trust amongst their investors. India

Grenze ID: 01.GIJET.8.1.19

Bulls aimed for a full-fledged banking license. RBI refused to grant permission to this merger. Before rejecting the deal, RBI had placed Lakshmi Vilas Bank under prompt corrective action.
Another recent crisis which was bought to spotlight was of Guru Raghavendra Co-operative Bank. This case is similar to PMC Bank (Punjab and Maharashtra co-operative bank). In PMC Bank, loans were provided to same individuals through numerous fake accounts. It is yet to be known, whether the so-called '62 customers' are all verified accounts or a pre-planned scheme to fake. Guru Raghavendra Co-operative Bank is facing debt from 62 customers. RBI has imposed a cap on withdrawals of Rs.35000/- per account, and a permanent ban on accepting fresh deposits, providing new loans or making new investments. The bank has eight of its branches spread across Bengaluru, and a major proportion of its depositors are senior citizens. The bank officials claim that all major advances are supported by collaterals, and appropriate security measures will help them realize the proceedings and meet all important commitments within March 31, 2020.
Another peculiar case is of YES Bank [1] on which the RBI had placed a 30 days moratorium (temporary prohibition). Their loans nearly had grown to four times taking it to around 2.25 trillion on September 2019.
Adding to their wounds, their deposits growth failed to keep pace and asset quality worsened and came to a scanner.
It is therefore very much important for us to be aware of the banks in which we invest or deposit our money as the volatile nature of the banks is such that there is no guarantee whether our money is completely secured or not. Making use of emerging technologies like Machine Learning Algorithms, a novel approach is proposed in the following sections of this paper which gives a picture of analysis that must be done before proceeding with the banks and co-operative sectors. In the rest of the paper, banks and co-operative societies are together addressed as *Banks*.
The rest of the paper is organized as, Section II describes the Problem Statement in brief. Section III contains Literature Survey. Section IV discusses about the proposed solution to the problem in detail. Conclusion is presented in Section V.

II. Problem Statement

To analyse and arrive at a conclusion regarding the stability and reliability of banking sectors through the use of Machine Learning Algorithms based on the factors like Capital Investment, FDs, Bonds verses NPAs.

III. Literature Survey

Proper research was made to check whether a solution would really exist for this type of problem. One such solution was found but it was not completely solving the chosen problem. That solution is CRISIL Rating.
Research paper [2] gives us a complete picture about CRISIL rating. CRISIL is the abbreviation for "Credit Rating Information Services of India Limited". CRISIL is a global rating agency which does the debt assessment for the mutual fund companies. CRISIL gives the rating based on several parameters like Risk analysis of the company, project funding, cash flows of the company and many more. The top rating given by this agency is 'AAA' rating which indicates *prime company*. In contrast, the poor rating is named as 'C' rating which indicates company as *highly risky*. And a default rating 'D' is also given by CRISIL. But assessing this agency is also an important aspect.
Research paper [3] attempts to assess the quality of credit rating function performed by CRISIL (Credit Rating and Information Services of India Ltd.). With this objective, the paper addresses two important questions that can be stated as: a) Are CRISIL's rating standards match with the international standards? (b) Are CRISIL's ratings stable? Based on the assessment of CIRIL, companies which are rated AAA on the S&P standards, the authors come to the conclusion that CRISIL's credit rating standards are below international standards, and they are inconsistent in certain aspects.
Research paper [4-5] evaluates the different market responses to credit rating amendments in the pre-global and post-global financial crisis period using relevant data of India. By analysing and evaluating the stock price, to the announcement of long-term rating changes during the period from 1996 to 2015, the study reveals evidence that the stock price reacted less to rating announcements after the global financial crisis of 2008. However, the difference in the cumulative unusual returns before the global financial crisis and after the global financial crisis is not that significant.
Based on these many research works, there is no exact solution to our problem. CRISIL just attempts to give rating for the shares of a company. But many people depend on the banks for keeping their deposits too.

	Year	Bank_Name	RBI License	Total Income	Total Expenditure	NPA	Dividend	Net Profit	Capitals and Liabilities	Threshold NPA(%)	Dividend Class	Profit Percentage	NPA Classes
0	2011	State Bank of India	Yes	97218.96	88954.44	12346.90	30.0	8264.52	1223736.20	12.700095	1	8.500934	3
1	2012	State Bank of India	Yes	120872.90	109165.61	15818.85	35.0	11707.29	1335519.23	13.087177	1	9.685620	3
2	2013	State Bank of India	Yes	135691.94	121586.96	21956.48	41.0	14104.98	1566261.04	16.181123	1	10.394855	4
3	2014	State Bank of India	Yes	154903.72	144012.55	31096.07	3.0	10891.17	1792748.29	20.074450	1	7.030929	4
4	2015	State Bank of India	Yes	174972.97	161871.39	27590.58	3.5	13101.58	2048079.80	15.768481	1	7.487774	4

Deposits must be given a good interest rate. Having less interest rate is also not good [6]. This was addressed in focus along with the shares issued by bank in proposed approach.

IV. PROPOSED SOLUTION

As the existing solution won't solve our problem completely, it is important to develop a proper solution which will address our problem completely. A novel approach was proposed in this paper to solve the problem in all possible aspects. Let's see it in detail.

To say whether a bank is stable or not, we need to consider the annual report of that bank for at least 2-3 years. But a corner case would be that there are chances of having annual report of that present year. So, while deciding the parameters that need to be evaluated, this corner case must also be kept in conscious. Parameter like cash flows of a bank cannot be considered to say whether a bank is reliable or not as cash flows may vary largely in consecutive years. Hence choosing the right parameters to evaluate the stability and reliability of a bank is the first and foremost thing that need to be considered. Also, the parameters which will be chosen must be that type of information which each and every person can get it free of cost. After all the research, a few parameters are listed and some other parameters are calculated. This shall be seen in detail now.

A. Dataset Details

For any Machine Learning project, the important and foremost required thing is the *Data*. Only if we have the proper legitimate data, we can proceed with the project further. Hence, in this case there is no existing dataset. Now, it is major task to create a proper dataset with the right parameters. After a considerable amount of research on several parameters present in the balance sheet of the bank, few parameters were decided to be chosen [7] and other parameters were derived from these parameters. A snap of the dataset is shown in the Fig.

1. So, let's see each of the parameters one by one: -
   1) *Year*: - It is the year of record of a particular bank.
   2) *Bank Name*: - This is required to provide a further extension of the proposed solution
   3) *RBI License*: - A Bank is said to be completely dependable (to an extent) if it is validated by RBI.
   4) *Total Income*: - The excess revenue that is generated from the interest earned on assets over the interest paid out on deposits is the net interest income.
   5) *Total Expenditure*: - This means all expenditures including those of an operating and capital nature that was spent by the bank in that particular *Year*.
   6) *NPA*: - A Non-Performing Asset (NPA) refers to a classification for loans or advances that are in default or in arrears [8].
   7) *Dividend*: - A dividend is the distribution of some of a company's earnings to a class of its shareholders, as determined by the company's board of directors [9].
   8) *Capital & Liabilities*: - Capital is the value of the investment in the business by the owner(s). Liabilities are the debts owed by the firm.

These are the basic parameters that are taken from the user as input either as static or as file input. Based on these, other parameters that are designed are: -

   9) *Net Profit*: - It represents the financial profit of the bank after all the expenditures that are paid from revenue. However, profit is much dependent on NPA [10]. Equation (1) gives the formula to calculate net profit.

Net Profit = Total Income – Total Expenditure (1)

10) *Threshold NPA*: - This is the NPA percentage of the bank in that year. Equation (2) gives the formula to calculate Threshold NPA.

Threshold NPA (%) = (NPA/Total Income) * 100 (2)

11) *Dividend Class*: - This indicates if the customers were given dividend in that year or not. Value 1 indicates dividend is given and value 0 indicates dividend is not given.

12) *Profit Percentage*: - This indicates the percentage of profit earned by the bank in that year. Equation (3) gives formula for profit percentage.

Profit Percentage = (Net Profit/Total Income)*100 (3)

13) *NPA Class*: - Banks that have a threshold value of:
   i) <9% will fall in class 1 indicating as *Highly Secured & Trustable Bank.*
   ii) 9-12% will fall in class 2 indicating as *Trustable Bank.*
   iii) 12-15% will fall in class 3 indicating as *Average Trustable Bank*
   iv) >15% will fall in class 4 indicating as *Bank Not valid for trust.*

So, these are the main parameters of the dataset. Hence using these parameters, two new target variables are predicted in future. That shall be seen in later sections. Hence, now with these parameters, machine learning model can be created to predict the required performance.

Data-flow Diagram

A Data-flow diagram (DFD) gives us the idea of the flow of a process from start to end. A DFD can be represented in different levels. Lower levels give the outlook of the process where higher levels give the idea of process flow in depth. So, with the designed parameters, a process flow is created which is represented as a DFD of two levels.

Level-0 DFD: - Level-0 DFD is created as shown in Fig. 2. This level just indicates the entire context of the solution. User gives the inputs and a process is done at the back-end and then the result is displayed to the user. This is the simplest way of representing the entire solution.

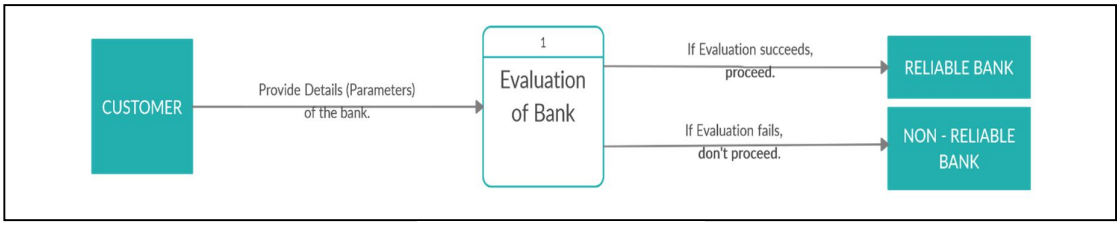

Fig 2. Level-0 Data-flow Diagram.

Level-1 DFD: - Fig. 3 shows the level-1 DFD. This level gives us a better idea about the entire process in detail. Here it is observable that based on complete dataset, two target variables are predicted. Based on the prepared dataset, at first, the target variable let say *target_1* is predicted. *target_1* is the attribute which is used to know whether the customer can proceed with keeping deposits or to buy a share from the bank or both or none. This newly predicted variable is a categorical data with various values:

   i) *'d'* - > indicates customer to proceed with buying share as the bank provides dividend.
   ii) *'p'* -> indicates bank provides interest on the deposits for sure.
   iii) *'dp'* -> indicates bank is good to go with both buying share and keeping deposits.
   iv) *ndp'* -> indicates the contrast meaning of 'dp'.

Then, including the newly predicted *target_1* along with other parameters, a final target variable is predicted saying that as the *Result.* This is the actual and final target variable.

This variable can take value from 0-5 where 0 indicates *very poor dependent* and 5 indicates *highly dependent* bank. Based on the score value, the bank can be judged by how much it is stable. Hence, this is the overall idea of the proposed solution of the problem. But it is clearly understandable in the Implementation section i.e., in the next section.

Implementation: -

So, all set to go for the prediction of the actual result. But, no, it is not possible to directly apply an ML algorithm to the raw dataset. A sample dataset is used for the entire implementation. The chosen dataset has 51 tuples. A very important phase before prediction is the phase of exploring data which is also called as

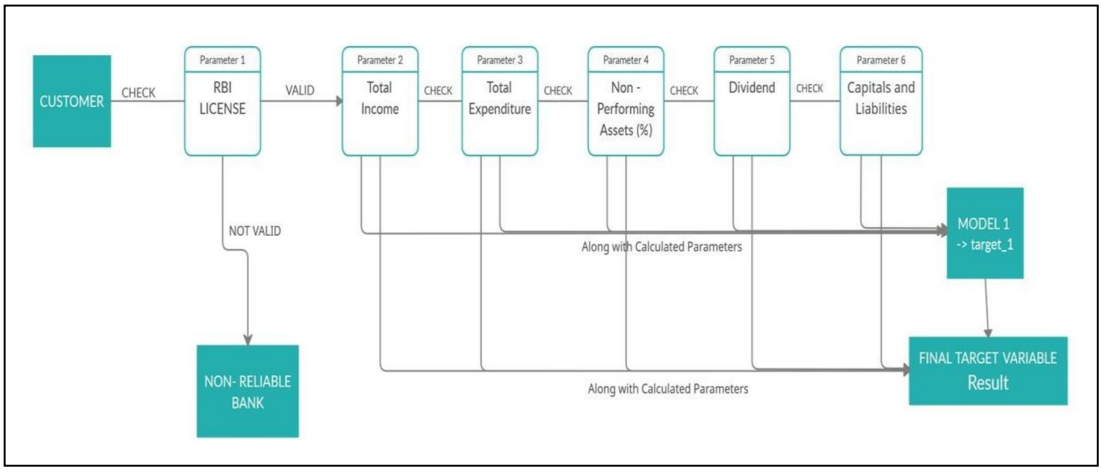

Fig. 3: Level-1 Data-flow Diagram

Exploratory Data Analysis. So, a considerable amount of analysis must be done before proceeding with application of algorithm. Analysis is seen in the following steps: -

1) Result Analysis: - The final target variable is analysed in the chosen dataset. The analysis is shown in the Fig. 4. It can be seen that the count of all the possible cases is not approximately equal too. Hence if this happens with the dataset which has high number of tuples, it is also important to balance the data using several algorithms.

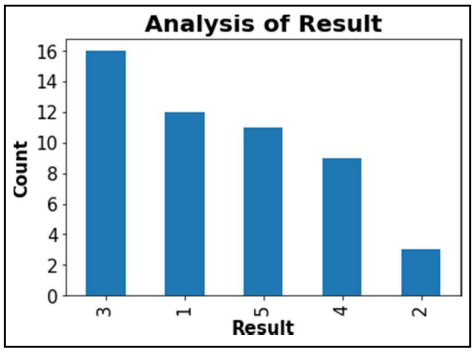

Fig. 4: Result Value Analysis

2) *Correlation Matrix*: -This matrix gives the idea of interdependency of the attributes inside the dataset. Also, categorical data attribute is neglected in this matrix. Based on the correlation matrix shown in the Fig. 5, it is decided to consider all the attributes to train the model.

3) Filling Missing Values: - It is an important aspect of data pre-processing phase to take care of the missing values that are present in dataset. There are many ways to fill the missing values. One can use mean of all the other values in that particular attribute and fill the missing values with it. Similarly, one can use median and constant value to fill empty values. Here using *SimpleImputer*, empty values are filled with the constant 0.

4) *Encoding Categorical Values*: - One important thing to do with categorical data is that it should be encoded to numerical data. If it is not encoded to numerical data, it is not able to train the model for that attribute and there will be a greater chance of missing an important attribute. There are many types of encoding algorithms. Here, *LabelEncoder* has been used and the categorical values is turned into numerical numbers of range 0 to 3(incl. of both ends). The encoded table is shown in the Table I.

5) *Feature Scaling*: - This is the most important phase of this part. Feature Scaling, in general, refers to equalising entire data within a small interval. Generally, interval is considered to be between -3 to +3. This is important because a parameter which takes values in millions and another parameter which

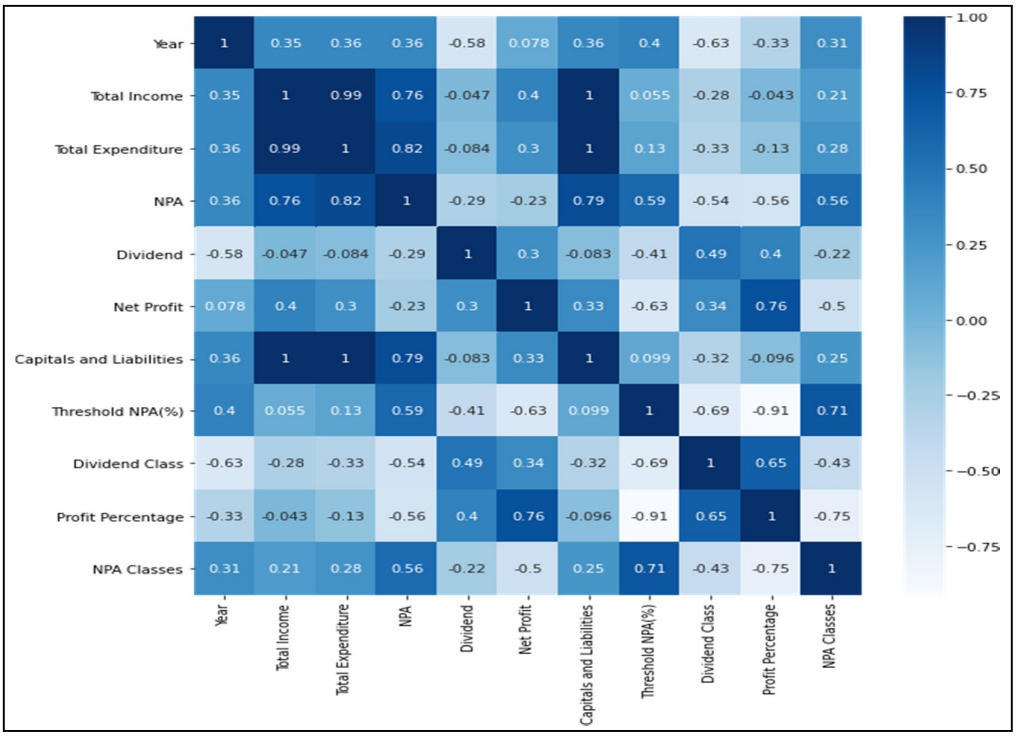

Fig. 5: Correlation Matrix

TABLE I. ENCODING CATEGORICAL TO NUMERICAL VALUES

Categorical Value	Numerical Values
p	3
d	0
dp	1
ndp	2

may take value in range of hundreds need to be correlated with each other. Else, model will neglect the second parameter and will consider the first parameter particularly. A simple example of two parameters of this kind is 'Total Income' and 'Dividend Class'. Both are important equally, but the range of values that those take has huge difference. Hence, feature scaling is important. There are many methods to scale the data. Here, *StandardScaler* is used as it the convenient way of standardising data and it uses mean, variance and standard deviation to scale data.

6) *Splitting Data into Train and Test Set*: - This process is done to test the model against the trained data. This process of testing the model helps to know the performance of the model. Knowing the performance will help to choose the right algorithm that can be applied to model for the current problem statement. Here, 20% of the data is assumed as the test data and remaining 80% of the data is used for training model.
7) *Applying Various Algorithms*: - There are various number of algorithms in Machine Learning. The cache in the proposed solution is that, there are two models that need to be used to get the final target variable. One model is to predict the attribute 'target_1'. Only if this is predicted first, it is possible to predict the final 'Result' variable. Hence, both the models were applied with different algorithms and separate analysis is done for both. As both the target variable have fixed number of classes in them, using classification algorithm is suitable. There are many classification algorithms in ML [11]. Only a few were applied and tested here. Those are:
   a) *K-Nearest Neighbours*: - In this algorithm, the new data is classified based on the similarity that the data share with the k-nearest tuples of old data [12]. This is considerably most simple approach

among other classification algorithms. A simple working of this algorithm is said like: considering there are 5 members in a group with 3 yes and 2 no. A new member X want to join the group. The value of k is to be 3. Hence when the similarity is calculated, the three nearest neighbours to X has been grouped as 2 yes and 1 no. Hence, based on the majority voting, X belongs to yes. KNN is majorly preferred for the small datasets.

b) *Kernel-SVM*: - Kernel SVM is the extended version of SVM algorithm. SVM stands for Support Vector Machine. There are various transformations that can be used in SVM [13]. One such transformation is Kernel Transformation. Normally two features can be separated using a line. But there are few peculiar cases where two features are surrounded in a circular fashion. At that time, a single line cannot separate two features. But a plane can definitely separate them. Hence, this is the working of Kernel-SVM.

Decision Tree Classification: - As the name itself says the meaning, a tree is constructed based on several calculations, in which leaf nodes corresponds to the final result and intermediate nodes represents attributes [14]. An attribute is selected as the root of the tree based on the concept of Information Gain (IG) that is earned by that attribute. An attribute which has high IG becomes root node. Subsequently, subtrees are created based on the remaining attributes and information gain from sub-table. At last, leaf node gives the target value of an instance.

c) *Random Forest Classification*: - This is the extension of Decision Tree algorithm. The name of the algorithm itself is the key point of the concept. A forest is created using several trees. Now, when a new instance enters, based on the results of all the trees, according to majority voting, the result of new instance is classified. The advantage of Random Forest over Decision Tree is that this algorithm reduces the overfitting of the model, as many instances are created [15].

Hence, these are some algorithms that were applied on the model and evaluated based on the performance. Apart from these algorithms, many other algorithms can also be used in the same approach.

8) *Performance and Analysis*: - After the model is tested and trained for various algorithms, it is important to check the performance of each algorithm. The basic performance metrics *time taken* and *accuracy* are considered for the analysis. A model is said to be a good model if it takes less time to produce more accurate results. This is the general knowledge required to select the right model. Fig. 6 and Fig. 7 gives a scatter plot of time taken by the above discussed algorithms along with the accuracy for Model 1 and Model 2 respectively.

Hence, based on this observation we can easily choose the right algorithm. It is found that Decision Tree is found to have an 100% accuracy in both the models. However, if one chooses decision tree based on this observation, there are possibilities of model overfitting. Hence it is advised to choose the correct model to obtain high performance.

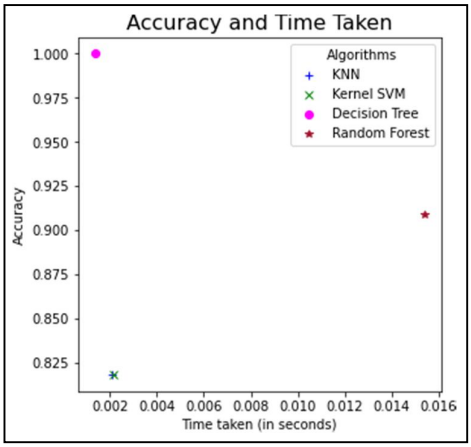

Fig.6: Analysis of Accuracy and Time Taken by various algorithms for Model-1

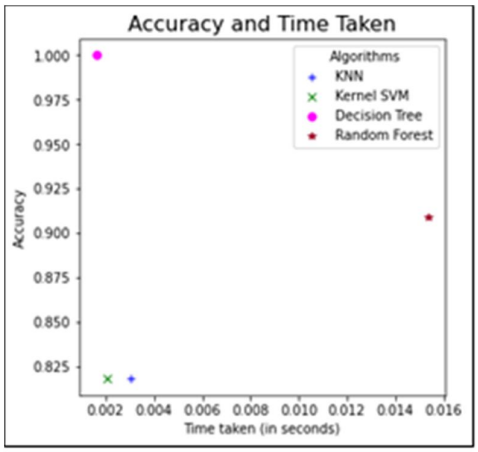

Fig.7: Analysis of Accuracy and Time Taken by various algorithms for Model-2

So, this is the proposed solution in detail. An approach like this covers all the aspects of the problem statement and gives a clear picture of solution.

V. RESULTS

The above given solution is implemented and has been hosted as a website. Some result snapshots are as shown in the following figures. Figure 8 gives the performance and analysis of a bank which has been performing moderately. Also, figure 9 gives the performance and analysis of a bank which is performing so well. The performance matrix is designed with scores like 0, 20, 40, 60, 80 and 100. The analysis parameters i.e., NPA and Profit are measured in percentage. Class is assigned a score between 0-5. This analysis is general in nature which means that this analysis can be done to any bank, no matter how big or how small or how old the bank is to the market.

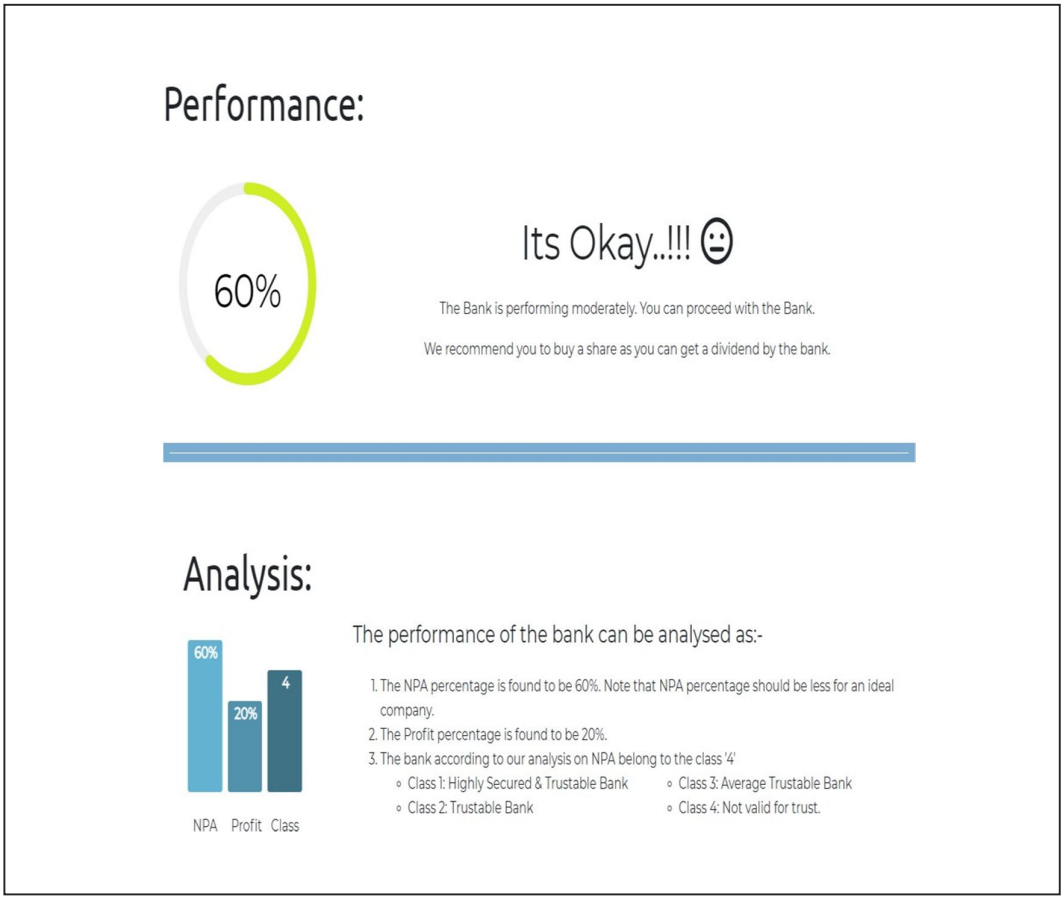

Fig 8. A snap of the performance and analysis of a moderately performing bank

VI. CONCLUSION

The main objective of the entire problem statement is to save money for future in secure hands. Evaluating the performance of a bank must be done based on the valid parameters. So, a novel approach to the problem of evaluating performance of the bank along with a proposed solution is discussed so far. The way of implementation of solution using many algorithms are also discussed.

Many a times, there are chances of missing certain aspects of a parameter when evaluation is done manually. Use of modern technology like Machine Learning Algorithms can help to get rid of these mistakes. This paper proves that use of Machine Learning Algorithms helps in solving society-related problems. ML is the future scope for many innovations which helps to achieve a better mankind.

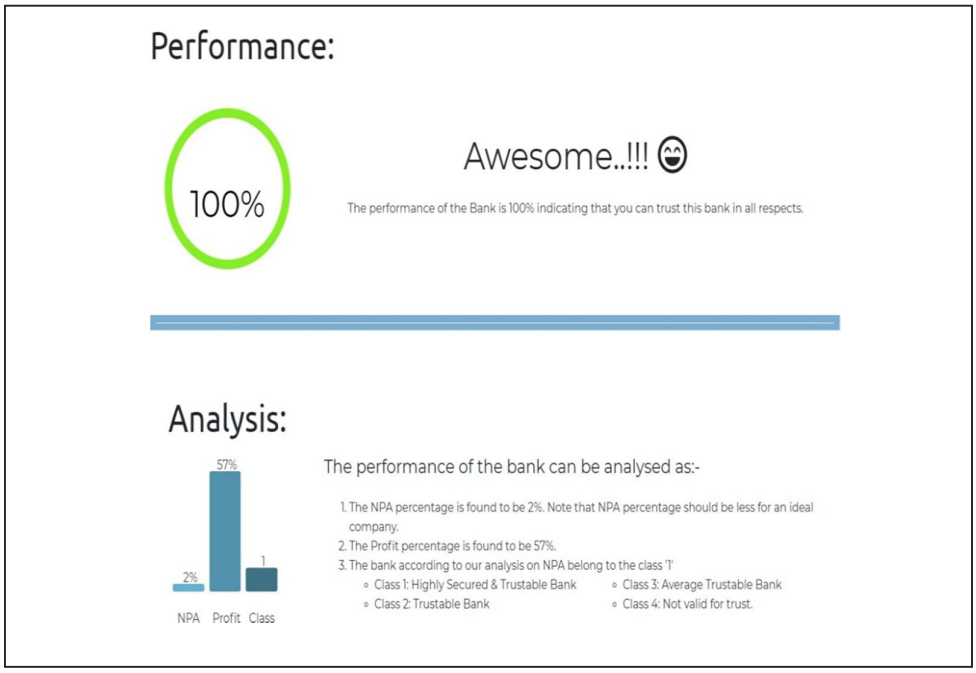

Fig 9. A snap of the performance and analysis of a highly performing bank

REFERENCES

[1] Barik, T. R. (2021). Yes Bank Crisis-A Critical Analysis on Causes, Effects & Recommendations. International Journal of Research and Analysis in Commerce and Management, 1(1), 17-17.

[2] Kumar, K. V., & Rao, S. H. (2012). Credit Rating-Role In Modern Financial System. International Journal of Marketing, Financial Services & Management Research, 1(8), 126-138.

[3] Raghunathan, V., & Varma, J. R. (1992). CRISIL Rating: When Does AAA Mean B?. Vikalpa, 17(2), 35-42.

[4] Thampy, A. (2020). Kaveri Krishnan1 Sankarshan Basu2. Journal of Emerging Market Finance, 19(1), 7-32.

[5] Krishnan, K., Basu, S., & Thampy, A. (2020). Has the Global Financial Crisis Changed the Market Response to Credit Ratings? Evidence from an Emerging Market. Journal of Emerging Market Finance, 19(1), 7-32.

[6] Lopez, J. A., Rose, A. K., & Spiegel, M. M. (2020). Why have negative nominal interest rates had such a small effect on bank performance? Cross country evidence. European Economic Review, 124, 103402.

[7] Pinto, P., & Joseph, N. R. (2017). Capital structure and financial performance of banks. International Journal of Applied Business and Economic Research, 15(23), 303-312.

[8] Ali, A., Singh, D., & Ali, S. (2020). A Systematic Analysis of Commercial Banks in India–A Study of Nonperforming Assets (Npa). International Journal of Management, 11(9).

[9] Sharma, K. (2018). Impact of dividend policy on the value of Indian banks. SCMS Journal of Indian Management, 15(3), 14-19.

[10] Chakraborty, S. A. (2017). Effect of NPA on banks profitability. International Journal of Engineering Technology, Management and Applied Sciences, 5(5), 201-210.

[11] Mahesh, B. (2020). Machine Learning Algorithms-A Review. International Journal of Science and Research (IJSR).[Internet], 9, 381-386.

[12] Cunningham, P., & Delany, S. J. (2020). k-Nearest neighbour classifiers: (with Python examples). arXiv preprint arXiv:2004.04523.

[13] Altan, A., & Karasu, S. (2019). The effect of kernel values in support vector machine to forecasting performance of financial time series. The Journal of Cognitive Systems, 4(1), 17-21.

[14] Chen, C., Geng, L., & Zhou, S. (2021). Design and implementation of bank CRM system based on decision tree algorithm. Neural Computing and Applications, 33(14), 8237-8247.

[15] Jaiswal, J. K., & Samikannu, R. (2017, February). Application of random forest algorithm on feature subset selection and classification and regression. In 2017 World Congress on Computing and Communication Technologies (WCCCT) (pp. 65-68). IEEE.