Predicting Delinquency and Loan Default with Advanced Analytics, Machine Learning and analyzing the performance of Banks in MSME Sector

Jai Kushwaha
Great Learning
Pune, India
jaikushwaha7@gmail.com

Richa Agarwal
Great Learning
Mumbai, India
richaagarwal1902@gmail.com

Abstract--The concept of Non-Performing Assets (NPAs) was introduced in the Narasimhan Committee on Financial System Reforms that was tabled in Parliament on December 17th, 1991. The Committee studied the prevailing financial system, identified its shortcomings, and made valuable recommendations in line with internationally accepted norms. Many of those recommendations are in place, yet we continue to observe a steep rise in NPAs especially with MSME (Micro, Small, and Medium Enterprises) sector. Even recent reforms by the government indicate the dire need to reduce the MSME defaults. However, why do MSME loan default, and is there a pattern which can detect early sign to ensure banks take the precautionary steps. In this paper, we study patterns of erratic EMI payment leading to delinquency and explore factors that make the MSME loan go default. A database of 32191 MSME borrowers from 2017 to 2019 was constituted and borrower loan with even one missed EMI payment was identified as delinquent while borrower loan for which interest or principal repayment has not been received for more than 90 days was considered as Loan Default. Using VIF scores, the independent variables were identified and a Logistic Regression model was developed for delinquency prediction. The model was successfully able to spot 97 of the 100 delinquent accounts. Different ML techniques, like Random Forest, Extreme gradient boosting, were applied for Loan default model development. The best model was able to identify 75 of 100 loan defaults. Finally, a user interface was developed to allow financial institutions to consume the proposed model for identifying delinquency and potential loan default.
Keywords—PSU, NPA, MSME, KYC, Engg, Corr, EMI

I. Introduction (Banking and Analytics)

The Banking industry generates a huge volume of data on a day-to-day basis. To differentiate themselves from the competition, banks are increasingly adopting big data analytics as part of their core strategy. Analytics will be the critical game-changer for the banks. Adopting it has become necessary in almost all sectors that banks deal with. One such sector that we are going into depth is the MSME sector.

The Micro, Small, and Medium Enterprises (MSME) sector has emerged as a highly vibrant and dynamic sector of the Indian economy over the last five decades. It contributes significantly to the economic and social development of the country by fostering entrepreneurship and generating the largest employment opportunities at comparatively lower capital costs, next only to agriculture. MSMEs are complementary to large industries as ancillary units and this sector contributes significantly to the inclusive industrial development of the country. The MSMEs are widening their domain across sectors of the economy, producing a diverse range of products and services to meet the demand of domestic as well as global markets

II. Objective

As a part of our project, we are proposing the use of Advanced Analytics and Machine learning to help financial institutes like banks understand their MSME borrowers and predict possible delinquency. It is then that adequate steps could be taken to understand the reasons behind payment defaults and thereby prevent NPAs. Through this project, we propose to address interdependence, reduce financial risk and improve financial stability.

In this paper, we present the advanced Machine Learning techniques for predicting the MSME loan going delinquent and also what are the factors that lead to advance becoming a default.

III. Literature Review

As the government has increased its emphasis on the MSME sector as this sector will be the driving force of the Indian Economy. These are the growth accelerators as well growth engines [1] as many economists.

So, identify the important features of the customer which drive this sector and which hampers this sector in the form of loan default is an important thing especially for the financial firms.

As a part of the Atma Nirbhar Bharat, the Indian Government is taking many initiatives to revive the Indian economy by investing Rs 3 Lakh crore especially for the MSME sector distributed via Banks.[2]

The percentage share of the MSME sector has increased in the GDP contribution over the years so has the NPAs of the public sector banks.

Table 1: NPA Comparison and % Share of 3 main Banking Firms

NPA across Indian Banks Year wise Comparison			
%wise Comparison Across Years			
Year	**FOREIGN BANKS**	**PRIVATE SECTOR BANKS**	**PUBLIC SECTOR BANKS**
2011	3.14%	10.60%	86.26%
2012	2.17%	6.75%	91.08%
2013	2.70%	6.07%	91.23%
2014	2.21%	6.21%	91.57%
2015	3.89%	31.17%	64.95%
2016	0.79%	7.63%	91.58%
2017	0.49%	11.03%	88.47%
2018	0.30%	12.37%	87.33%
2019	0.58%	18.99%	80.43%
2020	0.72%	19.31%	79.97%

Indian Government is targeting the investment in MSME to go up to 50% by 2022 [3].

Table 2: MSME contribution to Indian GDP

Year	Share of MSME in GDP (in %)
2011-12	30
2012-13	30.4
2013-14	30.2
2014-15	29.7
2015-16	29.2
2016-17	28.9
2017-18	28.9
2018-19	29
2019-20	30

NPA in MSME and its ache the banks' faces as we can see the rise of NPA in the Indian Banking sector (figure 2) from 2010 till 2020. Especially in the public sector banks.[4]

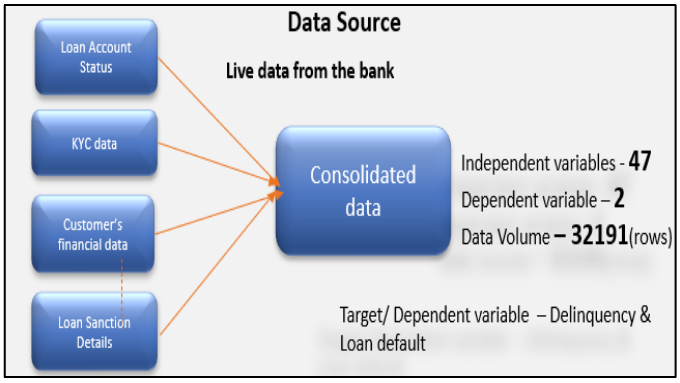
As we can see the rise of gross NPA across the year and major concerning thing is the continuously increasing NPA in public sector banks. If we think it concerning MSME sector which forms around 30% of GDP, there major concern for this sector especially as it will be playing a major part in resolving this problem.
So, keeping the importance of this and banks concerns we are analyzing the factors which impact MSME advance and how we can help banks overcome this problem using machine learning techniques.

IV. Data Ingestion and Data Encapsulation

The live data from a PSU bank consist of some masked fields for which we are looking to create a predictive model to predict delinquency and loan default of the customers. The data set received is from various departments and we have collated it as per our analytical problem statement. We also derived a few new columns from the current data for better learning of the model.
We collected more data around the details of the customer, their KYC details, payment schedules, etc. which in turn will help us in accurately creating a model for delinquency and loan default. The key data sets are:

1. Loan Account Status
2. KYC data (masked) like mobile number, type of documents submitted as a part of KYC, employment condition, etc.
3. Customer's financial data (liabilities and credit scores)
4. Loan Sanction Details like asset cost, disbursement amount.

Figure 1: Summary of Data Gathering



We merged the data and created two master files containing all the variables from all the above files based on customer ID as unique key as:

- **Data_MSME_merged_Loan_default:** The target variable present is whether the customer does loan default or not.
- **Data_MSME_merged_Delinquent:** The target variable present is Delinquent – whether the customer has paid EMI on the respective dates or missed the EMI payment date.

V. Exploratory Data Analysis

Basic Structure of the final data merged:

Descriptions <chr>	Value <chr>
Sample size (nrow)	32191
No. of variables (ncol)	47
No. of numeric/interger variables	31
No. of factor variables	0
No. of text variables	16
No. of logical variables	0
No. of identifier variables	1
No. of date variables	0
No. of zero variance variables (uniform)	1
%. of variables having complete cases	100% (47)

Table 3: Summary of the MSME final Dataset

Table 4: **Summary of Variables**

Customer Financial Data	Customer Loan Account Details
CUSTOMERID	CUSTOMERID
PRI.ACTIVE.ACCTS	AC_NO
PRI.OVERDUE.ACCTS	AC_TITLE
PRI.CURRENT.BALANCE	OPEN_DT
PRI.SANCTIONED.AMOUNT	SME_SECTOR
PRI.DISBURSED.AMOUNT	SECTOR_NAME
SEC.NO.OF.ACCTS	SME_CATEGORY
SEC.ACTIVE.ACCTS	MODULE_ID
SEC.OVERDUE.ACCTS	SCHEME_CD
SEC.CURRENT.BALANCE	BSR_ACTIVITY_CD
SEC.SANCTIONED.AMOUNT	INDUSTRY_NAME
SEC.DISBURSED.AMOUNT	BSR_ORG_CD
PRIMARY.INSTAL.AMT	BRANCH_NAME
SEC.INSTAL.AMT	DISTRICT_NAME
NEW.ACCTS.IN.LAST.SIX.MONTHS	REGION_NAME
DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS	STATE_NAME
PERFORM_CNS.SCORE	
PERFORM_CNS.SCORE.DESCRIPTION	

Customer KYC Data	Customer MSME Loan Ac Status	Customer MSME Loan Ac Sanction Details
CUSTOMERID	CUSTOMERID	CUSTOMERID
DATE.OF.BIRTH	AC_NO	AC_NO
EMPLOYMENT.TYPE	DELINQUENT	DISBURSED_AMOUNT
MOBILENO_AVL_FLAG	LOAN_DEFAULT	ASSET_COST
AADHAR_FLAG		LTV
PAN_FLAG		
VOTERID_FLAG		
DRIVING_FLAG		
PASSPORT_FLAG		

VI. Missing Values

On drilling down further, we found that NA is considered as a factor and some blank spaces are present in the data so initially, data was showing no missing, so we modified the code to induce NA in the data. On further converting the missing string in the data as NA. **The total missing values in the master file is 6707.**

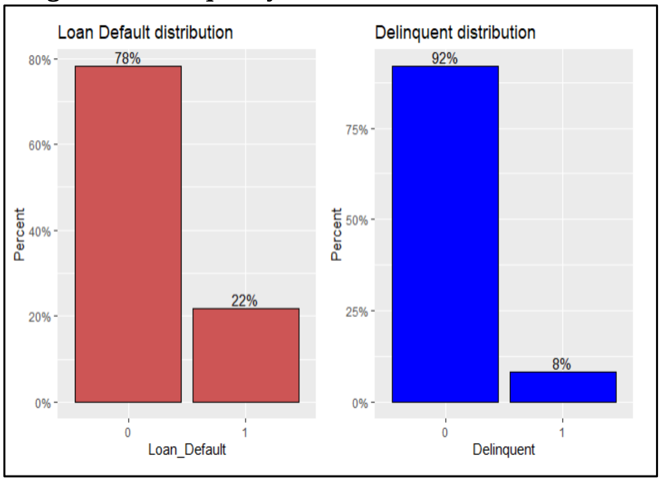
VII. Data cleaning and variable datatype modification of the variables

As most of the fields are read as integers while importing so we converted them to factor as per business understanding.

- We started with the CNS score which is **a credit score/rating for an individual customer or a firm.**
- As per the business understanding, we categorized CNS score into 6 different categories based on the severity of risk with a score ranging from 0 to 900 (900 is the lowest risk).
- Variable **perform_CNS.score.The description** was having multiple categories that can be converted to simple levels so we feature engineered the variable.
- Cleaning up SME category field as in some places category 'Medium' was showing as "Mediu'
- Converted the rest of the variables as a factor.
- Converted the date field [open_dt, Date of Birth] in date format ('DD-MM-YYYY') as initially it was read as a character then converted to factor then to a date format.
- We further omitted the NA's and created the data set having proper date columns and total rows now reduced to 28696.

VIII. Target Variable Distribution

Figure 2: **Delinquency and Loan Default Distribution**



Defaulters are **22 %** of the customers while **delinquent customers** are **8%**. So, there is an imbalance of the target variable.

IX. Feature importance Understanding

We further deep dive into the variable with the impact on our target variable. Using decision and random forest selection, we found out the importance of a variable.

Table 5: **Top Identified Features based on decision tree**

	Features	Score
5	ltv	483.194196
44	PERFORM_CNS.SCORE.DESCRIPTION_Simplified_Very ...	254.518240
25	PERFORM_CNS.SCORE	240.649305
3	disbursed_amount	167.771414
42	PERFORM_CNS.SCORE.DESCRIPTION_Simplified_Not S...	130.686215
6	PRI.ACTIVE.ACCTS	108.099196
26	Age_of_Customer	61.363191
43	PERFORM_CNS.SCORE.DESCRIPTION_Simplified_Very ...	60.476066
22	VoterID_flag	55.937090
20	Aadhar_flag	50.019011

Observation:

- From above it is evident the credit score of the customer reflects most of the aspect of the type of customer.
- LTV is the loan to value of asset ratio that describes the intention of the customer whether he is in practical need of a loan and an honest customer
- Several active accounts talk about the credits portfolio and future indications can be sorted out while giving the in practical in banks.
- Age, VoterID, and Aadhaar give relevant emphasis of KYC play an important role in advance. Better to you know the customer the better chances that you can predict the delinquency and loan default of a loan account.
- The top 2 features playing an important role in determining the delinquency and default in the MSME sector are LTV and disbursed amount.
- As business understands disbursed amount and the sanctioned amount will be correlated and data correlation matrix proves it. The same goes for sec sanction and disbursal as it was evident from the data.

X. Feature Engineering

- Converting relevant numeric fields to factor
- Count frequency encoding for the variable with factor levels greater than 10.
- Age of Customer and Age of loan by converting the date of Birth and Date of opening account respectively.
- Count frequency encoding for the variable with factor levels greater than 10.

XI. Modeling for Delinquency

a) For modeling we started working with Logistic regression where we received the following scores on the data were obtained:

Table 6: Delinquency Model Performance

	precision	recall	f1-score	support
0	0.92	1.00	0.96	23677
1	0.41	0.03	0.05	2075
accuracy			0.92	25752
macro avg	0.67	0.51	0.50	25752
weighted avg	0.88	0.92	0.88	25752

Observation: As we can see the model is performing well for predicting whether the customer will go delinquent or not pretty well.
When we performed logistic regression on the unscaled data accuracy for delinquent customers was coming very low.

b) On further enhancing the model and scaling the data we were able to get good results.

Table 7: Delinquency model performance for scaled data

	precision	recall	f1-score	support
0	0.99	1.00	1.00	23677
1	1.00	0.90	0.94	2075
accuracy			0.99	25752
macro avg	1.00	0.95	0.97	25752
weighted avg	0.99	0.99	0.99	25752

Using tuned parameter and scaled data (min-max scaling used to normalize the data) in logistic regression prediction power for delinquent customers increased to **0.94.**
Final Model inferred as below

Table 8: **Model iteration and output of test data for Delinquency**

Model	F1 Score	AUC/ROC	Comments
Logistic Regression Iteration-3	0.97	0.91	Using VIF removed the Multicollinear columns and we are left with 42 independent variables and finally scaled the data and ran Logistic Regression.
Random Forest Iteration1	1.0	1.0	Data with 42 variables removed the high collinear variables

Table 9: **Top coefficient of Logistic Regression**

Variables	Model Coef Value
PRI.OVERDUE.ACCTS	13.222337
PRI.ACTIVE.ACCTS	6.542507
SEC.OVERDUE.ACCTS	4.404274
Passport_flag	3.917191
PRI.DISBURSED.AMOUNT	1.842276

As we can infer that Random forest is overfitting the data. However, we settle with the logistic regression with Coefficient as per the Logistic Regression.

XII. Modeling for Loan Default

We tried different models and analyzed the accuracy to obtain the best prediction model and compare the results.

After running the model with different modeling techniques starting with logistic regression, then decision tree, random forest, SVM, KNN, light gradient boost, extreme gradient boost, and Deep Neural Network we ran more than 250 iterations on the train, test, and validation data.

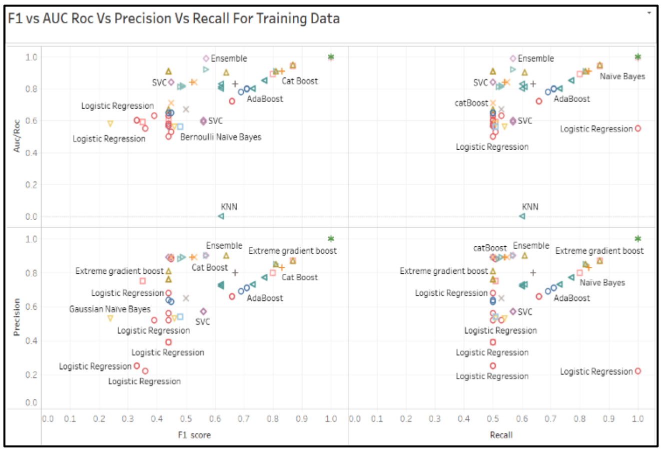
Here is the main summary of all the models and best scores across different models and best metric scores on test data and different kinds of data we created.

Table 10: Model comparison on multiple datasets and the best score on test data for different models

Model	Data File Used	F1 Score	AUC ROC	Model Iteration
Ada Boost	Feature Engineered, Smote	0.71	0.79	5 (Normal, Feature Engg, PCA, Smote, PCA Top30, Corr Removed)
Cat Boost	Feature Engineered, Smote	**0.77**	**0.86**	5 (Normal, Feature Engg, PCA, Smote, PCA Top30, Corr Removed)
Decision Tree	Feature Engineered Data. Smote and Cross-Validation	**0.99**	**0.99**	Feature Engineered Data. Smote and Cross-Validation
Ensemble	Feature Engineered Data and Mean of All Probability of All models	0.45	0.61	2 (Normal and Feature Engg. Data with Threshold Optimization)
Extreme gradient boost	Feature Engineered, Smote	0.76	0.84	Feature Engineered and Used Smote
KNN	Feature Engineered and Scaled data for Outlier treatment	0.63	0.69	5 (Normal, Feature Engg, PCA, Smote, PCA Top30, Corr Removed)
Light gradient boost	Feature Engineered and Used Smote	0.76	0.85	5 (Normal, Feature Engg, PCA, Smote, PCA Top30, Corr Removed)
Logistic Regression	Feature Engineered, PCA 30 component, Scaled data	0.83	0.97	5 (Normal, Feature Engg, PCA, Smote, PCA Top30, Corr Removed)
Naïve Bayes	Feature Engineered, Smote	0.76	0.85	5 (Normal, Feature Engg, PCA, Smote, PCA Top30, Corr Removed)
Neural Network	Feature Engineered, Scaled	0.51	0.56	2 (Normal, Feature Engg)
Random Forest	Feature Engineered Scaled, PCA 30 Components	0.91	0.98	5 (Normal, Feature Engg, PCA, Smote, PCA Top30, Corr Removed)
SVC	Feature Engineered, PCA Top 30 components	0.57	0.59	5 (Normal, Feature Engg, PCA, Smote, PCA Top30, Corr Removed)

a) Models Comparison and Selection for Training Data

Figure 3: Metric Comparison on Training data using Multiple ML Algorithms

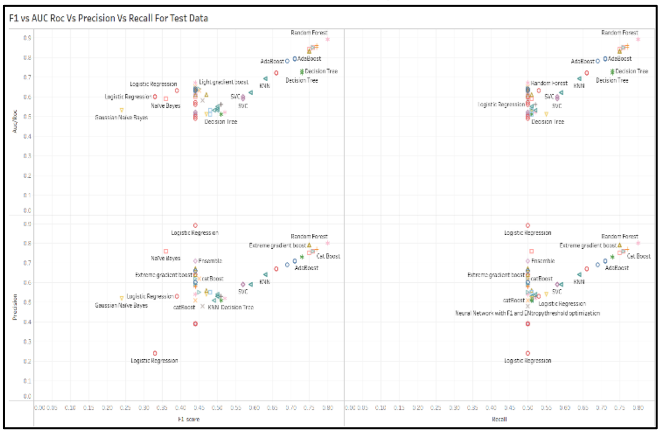


Key Highlights

- Extreme Gradient Boosting creates a balance between Precision and Recall.
- Random Forest and Decision Tree with High Variance overfit the training data.
- The ensemble of All model lies in the center for all the metrics.
- KNN performs the worst

b) Models Comparison and Selection for Testing Data

Figure 4: **Metric Comparison on Testing data using Multiple ML Algorithm**

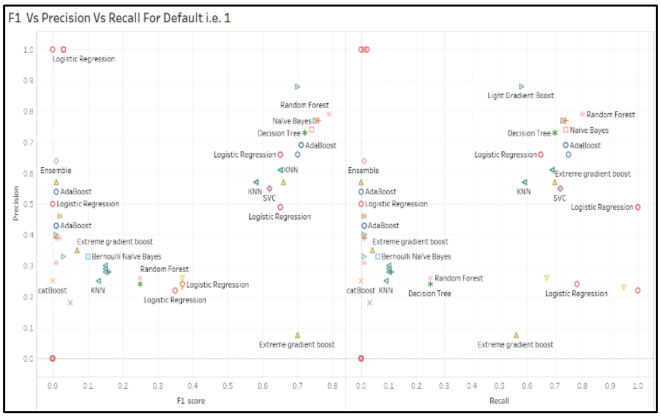


Key Highlights

- Best model RF and XGB
- Bias induced while using improvising techniques like SMOTE
- XGB for final deployment gives importance to weak learners i.e., for Default cases.
- Random forest and Decision Tree giving high accuracy and AUC ROC score on test data but it's mostly due to bias created by SMOTE.

c) Models Comparison – Default Capture

Figure 5: **Model Comparison comparing metrics for only the results for Loan Default i.e., "1"**



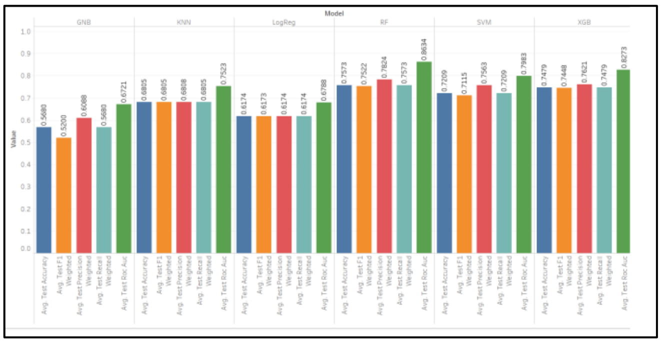
Key Highlights

- The random forest provides a balance between F1 recall and Precision for Default cases predicted.
- Extreme Gradient Boosting performs well with general feature-engineered data.
- Able to increase score in XGB using Hyper Parameter tuning.
- Random Forest performed exceptionally well with smote-induced data but it is creating a high bias.

d) Model Comparison and Final Inference

Now, to decide with which model(s) we will proceed further, we used the confusion matrix outputs of these models to evaluate their accuracy. A confusion matrix is a table that allows us to visualize the performance of a classification model. One can also use the information in it to calculate measures that can help us determine the usefulness of the model. As our primary focus here is to predict the 'loan defaults', the most important measure that we focused on was the f1 score. As we can see in the image below, **the Random Forest and XGB model are the top two performers.**

Figure 6: **Accuracy Comparison across multiple ML Models**



XIII. Application and Conclusion

B. Deployment

The most important part while deploying an analytical solution is to bridge the gap between the IT team and the data science team. The main objective is to ensure that the final solution is ready to be used within the operational environment and that end users have all the required tools to act upon the analytical insights discovered during the development phases of the project.

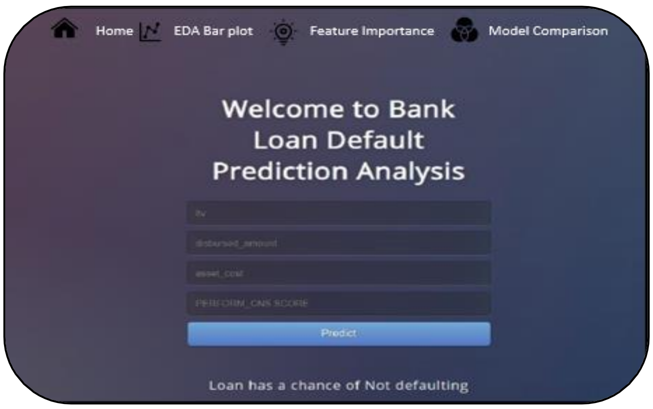
a) User Interfaces for a banking enterprise

We have created an end-to-end user-friendly interface front-end application that will be deployed as a web application. Banks will have to provide the top 4 mentioned details of the customer to predict if the customer will default or not. Based on which the bank can decide if further details are required of the customer and if the loan can be given.

We plan to improvise the application by including more features and data which will help the banks to understand the customers' financial status and to better predict loan defaults to reduce NPAs.

In the future enhancements, we also plan to create an application where the end users can themselves input their data in the required and can ensure their eligibility for the loan.

Figure 7: App created for Banking Official to early detect and target Loan before becoming default/NPA.



b) Achievements

We were able to achieve an accuracy of **94%** for predicting the delinquency of the customer and around **75%** for predicting the defaults much early in the lending cycle. There was also some analytics taken during the study which helped the bank to identify various other issues and how the bank can handle those.

The key achievements which can help the bank to cater to the current challenges are as below:

Table 11: Delinquency Variable Explanation and Impact:

Variable	y* = ln(p/(1-p))	p = exp(y*)/(exp(y*) +1)	% Change	Explanation
Primary Overdue Account 0	-3.28	0.04		Change in the probability account being Delinquent if Primary Overdue Account change from 0 to 1
Primary Overdue Account 1	9.95	1.00	96%	
Primary Overdue Account 2	23.17	1.00	0%	No change if the no. of Overdue Accounts increased further.
Primary Active Account 0	-3.28	0.04		93 % Change in the probability of an account being Delinquent if Primary Active Account change from 0 to 1
Primary Active Account 1	3.27	0.96	93%	
Passport Flag 0	-3.28	0.04		62 % Change in the probability of an account being Delinquent if Passport Flag change from 0 to 1
Passport Flag 1	0.64	0.66	62%	
Credit Score with medium risk flag 0	-3.28	0.04		3 % Change in the probability of an account being Delinquent if Credit Risk Medium Flag change from 0 to 1
Credit Score with medium risk flag 0	-2.63	0.07	3%	
Credit Risk Very High flag 0	-3.28	0.04		-3 % Change in the probability of an account being Delinquent if Credit Risk Medium Flag change from 0 to 1
Credit Risk Very High flag 1	-5.78	0.00	-3%	
LTV is 0	-3.28	0.04		-4 % Change in the probability of an account being Delinquent if LTV change from 0 to 80% which is an ideal condition
LTV ratio 80%	-17.51	0.00	-4%	

- The bank needs to give importance to the LTV ratio for better security of loans as from the data we can see that LTV is an important factor.

Table 12: LTV deep analysis on Loan Default

LTV	Non-Defaulter	Defaulter	Ratio Defaulter/Non-Defaulter
[0,25]	15	2	13%
(25,50]	992	137	14%
(50,60]	1904	280	15%
(60,70]	4274	826	19%
(70,80]	9008	2467	27%
(80,90]	8892	3310	37%
(90,100]	66	18	27%

- Banks should maintain a good amount of gap between the loan amount and actual value for which a loan is given as we can see from that more than 80% of the default happens between 70% to 100% LTV values.
- As recommended bank should maintain at most LTV of 70-75% for preventing a loan from going to default.
- The amount of loan disbursed is also a crucial factor as the overvaluation of asset and over credit lead to improper functioning of advances and add additional pressure on the borrower to meet the EMI deadlines which eventually leading to the loan going defunct.

Table 13: LTV and Disbursed Amount - Top 10 bins impacting Loan Default

LTV	Disbursed Amount	No Default	Default	Ratio Default/No-Default
(80,90]	(513030,570130]	3818	1490	39%
(80,90]	(570130,9905720]	2704	1040	38%
(80,90]	(139900,448490]	270	98	36%
(70,80]	(570130,9905720]	1962	657	33%
(25,50]	(570130,9905720]	45	15	33%
(80,90]	(448490,513030]	2100	682	32%
(90,100]	(570130,9905720]	46	14	30%
(60,70]	(570130,9905720]	1008	285	28%
(70,80]	(513030,570130]	1525	430	28%
(70,80]	(139900,448490]	2026	508	25%

- With LTV 80 to 90 and the disbursed amount ranging from 5 lakhs to 6 lakhs probability of the MSME loan going into goes up to **39%** which can be seen from the above analysis.
- MSME is a core participant in the Indian GDP so government and bank can drive the flow of credit to the general public but banks need to be carefully evaluated a portfolio as sometimes an honest entrepreneur can miss out on the EMI payment and become a delinquent customer as uncertain times like the emergence of COVID also the depression created by the pandemic. So additional margin should be considered for safeguarding the banks' interests.

- Model run using the XGBOOST gave a balance between the training data and test data using the F1 and AUC ROC metrics.
- Constant Model updating and monitoring are required with new data as the MSME sector is updating as per new government of India guidelines but the customer remains the same and more parameters will give more insights into the customers and loan performance.
- Consolidation of data spread across departments to achieve the analytical needs of the bank.
- Deep dive into various aspects of loans given by the banks to MSME sectors and the important measures that are required while giving the loan
- Constantly monitoring
- Sector-wise planning

ACKNOWLEDGMENT

We wish to extend special appreciation and gratitude to my colleagues Sanjeev Sidharth, Amol Pal, and Anu Rathi who help me through the process and valuable suggestions.

REFERENCES

[1] https://www.livemint.com/news/india/msmes-the-growth-engines-of-the-indian-economy-11597923225239.html

[2] https://www.lendingkart.com/msme-loan/what-is-msme/

[3] https://m.economictimes.com/small-biz/sme-sector/vision-is-to-increase-msmes-contribution-to-gdp-to-50-nitin-gadkari/articleshow/69978436.cms#:~:text=Union%20Minister%20for%20MSME%20Nitin,MSME%20sector%20will%20be%20crucial.

[4] https://dbie.rbi.org.in/DBIE/dbie.rbi?site=publications#!4

[5] RBI Circular: RBI/2018-19/100 DBR.No.BP.BC.18/21.04.048/2018-19 Dated: 1st January 2020.

[6] Banerjee, R., Verma, D., and Jaiswal, B. (2018), "Non-performing assets: a comparative study of the

[7] Indian commercial banks", International Journal of Social Relevance and Concern, Vol. 6 No. 2

[8] Dey, S. (2018), "Recovery mechanisms of non-performing assets in Indian commercial banks: an

[9] empirical study", NSOU Open Journal, Vol. 1 No. 2.

[10] Kumar, S. (2018), "A study on Non-Performing assets of Indian banks: Trend and recovery",

[11] International Journal of Electronics, Electrical, and Computational System, Vol. 7 No. 3

[12] Kumar, R.B., Subba Rao, B.V.S.S. and Kusuma, G.D.V. (2018), "Genesis for the increase of NPAs in Indian

[13] banks – an empirical analysis", Journal of Banking and Finance Management, Vol. 1 No. 1 Mishra, M.K. and Pawaskar, J.R. (2017), "A study of non-performing assets and its impact on banking sector", Journal for Research, Vol. 3 No. 1

[14] Sharma, G. (2018), "Unmasking the ordeal of growing NPAs in the Indian banking sector", International

[15] Journal of Business and Management Invention, Vol. 7 No. 7

[16] https://www.thehindubusinessline.com/opinion/columns/s-murlidharan/npa-crisis-prevention-is-the-cure/article30551370.ece

[17] https://www.ndtv.com/business/rbi-npa-norms-rbi-issues-revised-circular-for-resolution-of-stressed-assets-2049736

[18] http://sudeep.co/data-science/Understanding-the-Data-Science-Lifecycle/