Available online at [*http://www.advancedscientificjournal.com*](http://www.advancedscientificjournal.com)
[*http://www.krishmapublication.com*](http://www.krishmapublication.com)
IJMASRI, Vol. 2, issue 12, pp.805- 816, Dec. -2022
https://doi.org/10.53633/ijmasri

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY ADVANCED SCIENTIFIC RESEARCH AND INNOVATION (IJMASRI)

ISSN: 2582-9130 **IBI IMPACT FACTOR 1.5** **DOI: 10.53633/IJMASRI**

RESEARCH ARTICLE

LOAN FINANCIAL RISK ANALYSIS AND VISUALISATION

Jhalak Kumar[1] and Tarun Pangotra[2]

[1,2] Department of Information Technology, Maharaja Agrasen Institute of Technology affiliated to Guru Gobind Singh Indraprastha University, Rohini, Delhi
Email: tarunpangotra11@gmail.com , jhalakkumar456@gmail.com

Abstract

Credit risk is associated with the possibility of a client failing to meet contractual obligations, such as mortgages, credit card debts, and other types of loans. Minimising the risk of default is a major concern for financial institutions. For this reason, commercial and investment banks, venture capital funds, asset management companies and insurance firms, to name a few, are increasingly relying on technology to predict which clients are more prone to stop honouring their debts. By this project, we are trying to compare three gradient boosting ensemble machine learning algorithms and find which one can work better based on some performance metrics like Precision, Recall and F1 score. After successfully comparing the three algorithms we can draw our decisions by communicating with the officials with which algorithm they wanted to proceed with. People who are exploring machine learning as a field to shift their careers or people who are just curious about why there is such a buzz of machine learning all over the place often have one burning question in their mind – what all possible things can one achieve with machine learning. Well, the short answer is – that the possibility is endless and one's creativity is the only limit. The main motive of this project was to build machine learning algorithms that would be able to identify potential defaulters and therefore reduce company loss. In this fast-pacing world, it has become very difficult to trust anyone due to which person in need don't get proper help he/she is seeking for and in order to minimise the chances of fraud there is a need for the companies to have a proper methods, algorithms and data driven approach to find the fraudsters. Machine learning is the best possible way for companies in order to separate out these people and get proper insights whether it is beneficial for them to provide them credit support or not. Being from a computer science background, our main motive is to devise such ways and algorithms by which we can reduce human efforts for managing things and increase automation. In order to do so, we must check which method we are implementing is best suitable and has best performance.

Keywords: ML, Unsupervised Learning, Simulation, Smart Recommendations, Data-driven approach

Introduction

The primary goal of this project is to create machine learning algorithms capable of identifying probable defaulters and therefore reducing corporate damage. In today's fast-paced world, it is impossible to trust anyone, which prevents those in need from receiving the assistance they require. Additionally, in order to reduce the likelihood of fraud, firms must have correct procedures, algorithms, and data-driven approaches in place to identify fraudsters. Machine learning is the greatest approach for businesses to filter out these people and gain correct insights into whether it is worthwhile to give them credit help or not. As computer scientists, our fundamental goal is to create methods and algorithms that decrease human effort in managing things and boost automation. To accomplish so, we must determine which strategy is most appropriate and has the best performance. We took the data set with 45,000 rows and 43 columns. First, we will analyse the data and attempt to transform it into a usable format so that we may undertake more analysis on it. Following an examination, we will develop three gradient boosting algorithms, namely XG Boost, Light GBM, and Cat Boost, to compare them and see which one performs best for this sort of issue and dataset. Precision and recall will be the parameters used to compare these methods. As previously mentioned, one of the study papers compares the Artificial Neural Network method and the gradient boosting technique for loan risk analysis. The analysis clearly shows that the gradient boosting algorithm outperforms the Artificial Neural Network algorithm, which is understandable given that the gradient boosting algorithm is a type of ensemble learning and is a combination of more than one algorithm or an algorithm that is applied multiple times to achieve better results. So, in our study, we attempted to broaden that analysis since we know that there are several gradient boosting methods, and it would be preferable if we could determine which of them can be applied in practice and produce better outcomes which we came to finally understand after doing literature survey completely.

Literature Survey

Table 1 - Literature Survey table for Loan Financial Risk Analysis and Visualisation

Study	Main Objective	Main Conclusion	Main Contribution
Galih Kurniawan (2021)	Analyse and design scoring of default status and fines computation processes in Islamic bank.	The fines will be imposed if and only if the customer cannot prove that the delay was due to inadvertence.	conducted to determine whether the customer's financing application is acceptable or not.
Yuejin Zhang (2016)	Construct a credit scoring model by fusing social media information based on a decision tree through credit risk	Borrowed credit score, the number of successes, prestige, the number of failures, repayment period, forum currency are the most important attributes for predicting	The main contribution of this paper is to add social behaviour factors into the traditional credit scoring model, which are not considered in prior

	analysis.	the default.	research of credit scoring models.
Beibei Niu (2019)	Test the reliability of social network information in predicting loan default.	Results show that there is a statistically significant correlation between social network information and loan default. Social network information can improve loan default prediction performance significantly. Social network information is valuable for credit scoring.	This model shows the variables and uses a logistic regression model to identify the relationship between borrowers' social network information and loan default.
Arya (2013)	Estimates credit scores using an online FICO estimator, based on information reported by the subjects. These estimated credit scores are compared with incentivized measures of risk attitudes, trustworthiness, and time preference, and a survey measure of impulsivity. Main purpose is to determine the behavioural correlates of credit behaviour reflected by credit scores.	This model suggests that , in addition to the important effect of income, there are certain behavioural factors, such as impulsivity, time preference (or future orientation), and trustworthiness that are correlates of credit scores.	This project demonstrates how the behavioural factors processes work in the analysis of credit scores.

Rory P. Bunker (2018)	Investigate the extent to which features derived from bank statements provided by loan applicants, and which are not declared on an application form, can enhance a credit scoring model.	Data from bank statements obtained through applications like Credit Sense could be used in investigating possible fraud cases, for example, by using methods of outlier detection.	Results in this model show that a combined feature model performs better than both of the two baseline models, and that a number of the bank statement derived features have value in improving the credit scoring model.
Thakar Shivam (2021)	Analyse the gold loan facility provided by Muthoot Fincorp limited and the rate of customer satisfaction.	There are more men who benefit from Muthoot Fincorp's services. Most of the respondents are satisfied with the services which are provided by the bank because they will prefer others to try these services. Low-income earners take more loans as compared to high income earners.	This model tends towards surveys of people taking gold loans throughout India at lower rates at Muthoot FinServ.
Mohammad Ahmad Sheikh (2020)	Predict the safety of the loan.	The model concludes that a bank should not only target the rich customers for granting loans but it should assess the other attributes of a customer as well which play a very important part in credit granting decisions and predicting the loan defaulters.	The models are compared based on the performance measures such as sensitivity and specificity.

I O Eweoya (2019)	Predict fraud in bank loan administration and subsequently avoid loan default that manual scrutiny by a credit officer would not have discovered.	It has been revealed that false positives can be reduced with an employment of decision trees, thereby getting a highly reliable accuracy that financial institutions can depend on while scrutinising loan applications.	The machine learning approach uses the decision tree method to predict fraud and it delivers an accuracy of 75.9 percent ,also reducing human labour by introducing machine learning softwares.
Omprakash Yadav (2019)	Effectively predict the credibility of customers who have applied for a loan.	In this paper, a loan prediction system has been introduced that helps the organisations in making the right decision to approve or reject the loan request of the customers. This will also definitely help the banking industry to open efficient delivery channels.	Saving the result of the prediction system which will help the bank employee to next time process the application of the person applying for loan and the system would produce an even better and more accurate result.
B. Yamuna (2022)	Build an ML application which can reduce the time required to approve a loan using an ML based prediction model to approve the loan with minimal human intervention by filtering a huge number of applications and forwarding very few applications for human verification.	As per the prediction model uses several attributes of the applicant which also include non-financial attributes, we can obtain a highly reliable model when compared to the ones which include only financial attributes with 80% accuracy.	The model concludes that a bank must not only focus on wealthy consumers when giving loans to applicants, but we should also take consideration of a customer's remaining attributes, which play an essential factor in credit repayment and forecasting loan escapers.

Credit scoring system is a classic problem which is still interesting to study. There are many studies on credit scoring. But, most of them only discuss feasibility analysis. In fact, the credit scoring system should accommodate all processes from feasibility analysis until the end of contract. This study is aimed to analyze and design scoring of default status and fine computation processes in Islamic banks. BPMN 2.0 was used to model their processes. Besides that, this study proposed new mechanisms and algorithms using Interval Type-2 Fuzzy Sets for maintaining Sharia rules and fairness guarantee. The results showed that the new methods offer more fairness and compliance with sharia than existing methods as discussed in (Arya *et al.*, 2011). In recent years, the online Peer-to-Peer (P2P) lending market is rapidly expanding in China. In this paper, we use a public dataset from PPDai, a leading online P2P platform in China to study loan default. We construct a credit scoring model by fusing social media information based on a decision tree. The experimental result shows that our model has good classification accuracy. From the credit scoring model and classification rules, we get a conclusion that the loan information, social media information, and credit information are the most important factors for predicting the default. However, the credit rating is not as important as the platform described as seen in. (Yuejin Zhang 2016)

Financial institutions use credit scoring to evaluate potential loan default risks. However, insufficient credit information limits the peer-to-peer (P2P) lending platform's capacity to build effective credit scoring. In recent years, many types of data are used for credit scoring to compensate for the lack of credit history data. Whether social network information can be used to strengthen financial institutions' predictive power has received much attention in the industry and academia. The aim of this study is to test the reliability of social network information in predicting loan default. We extract borrowers' social network information from mobile phones and then use logistic regression to test the relationship between social network information and loan default. Three machine learning algorithms—random forest, AdaBoost, and LightGBM—were constructed to demonstrate the predictive performance of social network information. The logistic regression results show that there is a statistically significant correlation between social network information and loan default. The machine learning algorithm results show that social network information can improve loan default prediction performance significantly. The experiment results suggest that social network information is valuable for credit scoring as discussed in (Bunker *et al.*, 2016). In (Beibei Niu. 2019) we will see that Credit scoring has become an increasingly popular topic in recent years—in the media, in business, and at the dinner table. In these days of easy access to information, a negative credit event such as a mortgage default or bankruptcy can haunt a consumer for a considerable period of time. A credit score is a number that represents an assessment of the creditworthiness of a person, or the likelihood that the person will repay his or her debts. Credit scores are generated based on the statistical analysis of a person's credit report; credit bureaus such as Experian, Equifax and TransUnion maintain a record of a person's borrowing and repaying activities. In addition, credit scores are also used to determine insurance rates and for pre-employment screening. Employers as well as lenders use credit reports and scores to gain insight into the records and tendencies of prospective employees, making the assumption that credit scores correlate with general trustworthiness. There is even a dating website, creditscoredating.com, that purports to match subscribers with high-score partners. With reports and scores available to the public for little or no cost, the FICO score has become a part of the dating and mating process: along with a criminal background check, a credit report reveals much about a person's personality and behavioral tendencies.

The repayment of debt is contingent upon two factors: the ability to pay the debt, and the borrower's willingness to pay. The first condition is largely determined by income, while the second is more psychological in nature. Debtors may choose to pay their balances and reduce funds available to spend on other items, or default on their loans and keep their current level of liquidity, accruing penalties and credit bruises in the process.

Credit scoring models take a vector of attributes for a loan applicant, and given these attributes, attempt to discriminate between goods and bads; that is, to discriminate between those that are not likely to default or be in arrears with their payments, and those that are. A lending company in New Zealand scores loan applicants based on 11 attributes obtained from an application form, which is filled out online when applying for a loan. The lending company had recently begun to incorporate data from an application called Credit Sense 1 into their data warehouse, which automatically extracts line-by-line bank statement data for a 90-day period of spending. The bank statement data is currently used for purposes such as income verification. The company was interested in exploring the potential of this data for credit scoring purposes. The company was firstly interested in seeing whether a sufficiently predictive scoring model could be created using only the bank statement derived features, in which case the questions in the online application form would then be unnecessary to ask. Loan applicants would then be able to complete the application process faster, and the lending company would potentially be able to write more loans. If this model was not found to be sufficiently predictive, then the lending company was interested in investigating the extent to which the bank statement derived features could be of use in developing an improved scoring model by supplementing existing scoring features as discussed in (Eweoya 2019)

From (Omprakash Yadav (2019) we can conclude that there are more men who benefit from Muthoot Fincorp's services. There are mostly youth who take visits to Muthoot Fincorp because mostly responses collected from that age group. Most of the respondents are satisfied with the services which are provided by the bank because they will prefer others to try these services. Low income earners take more loans as compared to high income earners.

By predicting the loan defaulters, the bank can reduce its Non- Performing Assets. This makes the study of this phenomenon very important. Previous research in this era has shown that there are so many methods to study the problem of controlling loan default. But as the right predictions are very important for the maximization of profits, it is essential to study the nature of the different methods and their comparison. A very important approach in predictive analytics is used to study the problem of predicting loan defaulters: The Logistic regression model. The data is collected from the Kaggle for studying and prediction. Logistic Regression models have been performed and the different measures of performances are computed. The models are compared on the basis of the performance measures such as sensitivity and specificity. The final results have shown that the model produces different results. Model is marginally better because it includes variables other than checking account information (which shows the wealth of a customer) that should be taken into account to calculate the probability of default on loan correctly. Therefore, by using a logistic regression approach, the right customers to be targeted for granting loan can be easily detected by evaluating their likelihood of default on loan as we have seen in (Mohammad Ahmad Sheikh. (2021). The rate at which banks lose funds to loan beneficiaries due to loan default is alarming. This trend has led to the closure of many banks, potential beneficiaries deprived of access to loan; and many workers losing their jobs in the banks and other sectors. This work uses past loan records based on the employment of machine learning to predict fraud in bank loan administration and subsequently avoid loan default that manual scrutiny by a credit officer would not have discovered. However, such hidden patterns are revealed by machine learning. Statistical and conventional approaches in this direction are restricted in their accuracy capabilities. With a large volume and variety of data, credit history judgement by man is inefficient; case-based, analogy-based reasoning and statistical approaches have been employed but the 21st century fraudulent attempts cannot be discovered by these approaches, hence; the machine learning approach using the decision tree method to predict fraud as published in (Shiva and Kumar 2008).

Nowadays, Banks are struggling a lot to get an upper edge over each other to enhance overall business due to tight competition. Most of the banks have now realized that retaining the customers and preventing fraud must be the strategy tool for a healthy competition. Availability of the huge quantity of data, creation of knowledge base and efficient utilization of the same have helped banks to open up efficient delivery channels. Business decisions can be optimized very well through data mining. Credit scoring, Customer segmentation, predicting payment from customers, marketing, detecting fraud transactions, cash management and forecasting operations, optimizing stock portfolios and ranking investments are some of the areas where data mining techniques can be very useful and can be used widely in the banking industry (Galih *et al.*, 2014). Finally, from (Keramati and Yousefi (2011). Banks have various products to sell in our banking system, but their major source of money is their credit lines. As a result, they can profit from the interest on the loans they credit. Loans play a key role in determining profit or loss of the bank, i.e., whether consumers repay the loan or fail. Any bank can avoid its Non-Profiting Assets by pre-identifying loan absconders. As the outcome, research into this process is crucial. Existing works now have proved that there are so many implementations for studying the topic of loan escape control. However, as perfect forecasts are critical for the profit maximization, it is crucial to research and compare the various methodologies. To explore the topic of predicting loan escapers, the XG Boost model is utilized, which is a much important procedure in predictive research. Kaggle data is used to research and predict. The several performance metrics were computed using XG Boost models. Performance indicators like sensitivity and specifications were used for comparing these models. The final results show that the model gives-out different outcomes. Model is slightly better as it also includes attributes (customer personal attributes such as age, dependents, education, background, employment, etc) other than just financial information (which indicates a customer's money background only) that should be taken when calculating the probability of loan escape correctly. As a result, by calculating the possibility of loan escape, the ideal customers to target for loan giving will be easily identified using a XG Boost model approach. The model concludes that a bank must not only focus on wealthy consumers when giving loans to applicants, but we should also take consideration of a customer's remaining attributes, which play an essential factor in credit repayment and forecasting loan escapers.

Methodology

The data set that we used has 45,000 rows and 43 columns. In order to undertake additional analysis on the data, we will first examine it and attempt to transform it into a form that is helpful. After reviewing it, we will put three gradient boosting algorithms, XG Boost, Light GBM, and Cat Boost, to the test to see which one can handle this kind of problem and dataset the best. Precision and recall are the metrics we'll use to compare these systems. As we've mentioned, one of the study papers compares the gradient boosting algorithm and the artificial neural network algorithm for loan risk analysis. The analysis's findings make it clear that the gradient boosting algorithm outperforms the artificial neural network algorithm, which is quite justified given that ensemble learning, which includes the gradient boosting algorithm, combines multiple algorithms or applies one algorithm repeatedly to produce better results. Since there are various gradient boosting techniques, we attempted to broaden that analysis in our project. It would be preferable if we could determine which of them may be applied practically and are producing the best results. The results of the analysis show that the gradient boosting algorithm performs better than the artificial neural network algorithm, which is quite justified given that ensemble learning, which includes the gradient boosting algorithm, combines different algorithms or repeatedly applies one algorithm to get better results. In our effort, we tried to widen that analysis because there are numerous gradient boosting methods. It would be better if we could figure out which of them is most applicable practically and yield the best outcomes.

The target variable we are attempting to forecast is target default, a True/False feature. We discovered after investigating the data set that some characteristics contained outliers and missing values. Additional factors that wouldn't improve the model were eliminated. We are currently working on the following features after cleaning the data set and addressing the missing values:

target_default, score_1, score_2, score_3, score_4, score_5, score_6, risk_rate, last_amount_borrowed, last_borrowed_in_months, credit_limit, income, facebook_profile, state, real_state, ok_since, n_bankruptcies, n_accounts, n_issues, application_time_in_funnel, external_data_provider_credit_checks_last_month, external_data_provider_credit_checks_last_year, external_data_provider_email_seen_before, external_data_provider_fraud_score, reported_income, shipping_state.

Preprocessing must be done before configuring the machine learning algorithms namely XGBoost, LightGBM and CATBoost. We'll preprocess our data using Scikit Learn's LabelEncoder for the binary variables and pandas' get dummies for the remaining category variables because most Machine Learning algorithms perform better with numerical inputs. We must separate the data into training and test sets before creating the models. After that, using Standard Scaler and Random Under Sampler, respectively, we will standardise and resample the training set because we are working with an unbalanced data set.

Dataset used - ['ids', 'target_default', 'score_1', 'score_2', 'score_3', 'score_4','score_5', 'score_6', 'risk_rate', 'last_amount_borrowed','last_borrowed_in_months', 'credit_limit', 'reason', 'income','facebook_profile', 'state', 'zip', 'channel', 'job_name', 'real_state','ok_since', 'n_bankruptcies', 'n_defaulted_loans', 'n_accounts','n_issues', 'application_time_applied', 'application_time_in_funnel','email', 'external_data_provider_credit_checks_last_2_year', 'external_data_provider_credit_checks_last_month','external_data_provider_credit_checks_last_year','external_data_provider_email_seen_before','external_data_provider_first_name','external_data_provider_fraud_score', 'lat_lon', 'marketing_channel','profile_phone_number', 'reported_income', 'shipping_state', 'shipping_zip_code', 'profile_tags', 'user_agent', 'target_fraud']

Result

With regards to the evaluation of the models, it's worth mentioning that we should consider Precision, Recall and F1 score as evaluation metrics, for the following reasons:

1. **Precision** will give us the proportion of positive identifications that were indeed correct. It can be defined as:
   Precision = (True Positives) / (True Positives + False Positives).
2. **Recall** will determine the proportion of real positives that were correctly identified, and it can be defined as: Recall = (True Positives) / (True Positives + False Negatives).
3. **F1 Score** is a metric that is useful when we need to seek a balance between precision and recall. The formula is defined as: F1 = 2 * (Precision * Recall) / (Precision + Recall)

Since our objective is to minimize company loss, predicting the risk of client default, a good recall rate is desirable because we want to identify the maximum number of clients that are indeed prone to stop paying their debts, thus, we are pursuing a small number of *False Negatives*.

Additionally, we also seek to minimize the number of False Positives because we don't want clients to be mistakenly identified as defaulters. Therefore, a good precision rate is also desirable.

However, there is always a tradeoff between precision and recall. For our sake, we chose to give more emphasis to recall, using it as our evaluation metric because we need to check the possibility of real positives being accurate.

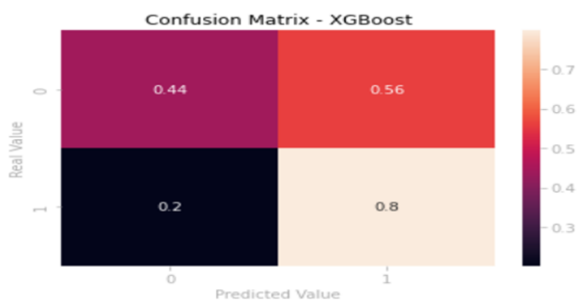
Confusion matrix:

The confusion matrix is a matrix used to determine the performance of the classification models for a given set of test data. It can only be determined if the true values for test data are known. The matrix itself can be easily understood, but the related terminologies may be confusing. Since it shows the errors in the model performance in the form of a matrix, hence also known as an error matrix. Some features of Confusion matrix are given below:

1. For the 2 prediction classes of classifiers, the matrix is of 2*2 table, for 3 classes, it is 3*3 tables, and so on.
2. The matrix is divided into two dimensions, that are predicted values and actual values along with the total number of predictions.
3. Predicted values are those values, which are predicted by the model, and actual values are the true values for the given observations.

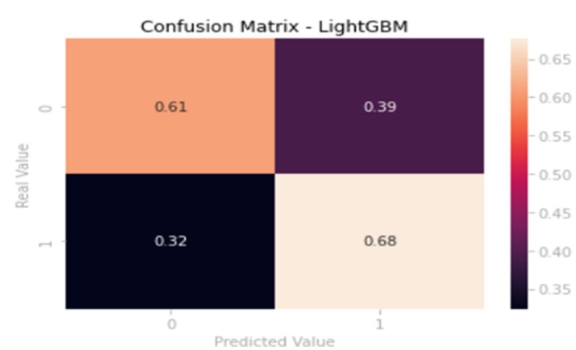
XGBOOST

	precision	recall	f1-score	support
0	0.92	0.44	0.60	8771
1	0.21	0.80	0.34	1665
accuracy			0.50	10436
macro avg	0.57	0.62	0.47	10436
weighted avg	0.81	0.50	0.56	10436



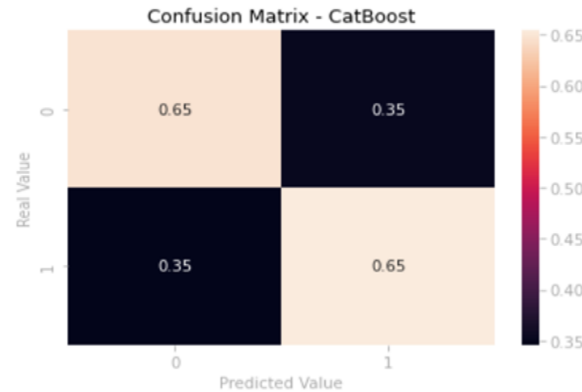
LIGHTGBM:

0	0.91	0.61	0.73	8771
1	0.25	0.68	0.36	1665
accuracy			0.62	10436
macro avg	0.58	0.64	0.55	10436
weighted avg	0.80	0.62	0.67	10436



CATBOOST:

	precision	recall	f1-score	support
0	0.91	0.65	0.76	8771
1	0.26	0.65	0.37	1665
accuracy			0.65	10436
macro avg	0.58	0.65	0.56	10436
weighted avg	0.80	0.65	0.69	10436



After completing and testing the project we have reached to the following results for each of three gradient boosting algorithms:

XGBoost - Best recall rate: 0.80
LightGBM - Best recall rate: 0.68
CatBoost - Best recall rate: 0.65

Conclusion

In these growing times of technology, it can be easily stated that there is a need to automate the tasks as much as possible. Also, everyone is living a fast-paced life, so it becomes very time consuming to do things manually. Before we used to work manually, then we shifted towards basic computation power where we just hard coded the program and got corresponding results but now with the help of Artificial Intelligence and Machine learning we have started giving the computers power to think, predict and act accordingly. So, after going through a lot of papers and research we tried to implement one of the famous machine learning algorithms which is gradient boosting. After completing the project, we get to understand the power of machine learning and get to know about the potential of it for the future and how it can shape the computational power in future. So, in our project we tried to implement the three gradient boosting algorithms and then compared them based on some parameters like precision, recall and F1 score for the loan risk analysis dataset of more than 45000 clients and concluded below the results that we have got. The best model possible would be the one that could minimise false negatives, identifying all defaulters among the client base, while also minimising false positives, preventing clients to be wrongly classified as defaulters. Meeting these requirements can be quite tricky as there is a trade-off between precision and recall, meaning that increasing the value of one of these metrics often decreases the value of the other. Considering the importance of minimising company loss, we decided to give more emphasis on reducing false positives, searching for the best hyperparameters that could increase the recall rate. Among the three Gradient Boosting Algorithms tested, XG Boost yielded the best results, with a recall rate of 81%, although it delivered an undesired 56% of false positives. On the other hand, Light GBM and Cat Boost delivered a better count of false positives, with 38% and 33% respectively, but their false negatives were substantially higher than that of XG Boost, resulting in a weaker recall rate. This Major project presents a classic evaluation metrics dilemma. In this case, it would be up to the company's decision-makers to analyse the big picture, with the aid of the machine learning algorithms, and decide the best plan to follow.

Some disadvantages of this project include Data misuse, Data manipulation, Lack of privacy, Security, etc. Some Barriers are also there in implementing them like adaptability, usability and sustainability.

Reference

1. Arya,S., Catherine, E and Wichman, C. (2011). “Anatomy of the Credit Score,” Journal of Economic Behavior & Organization 95 DOI:10.1016/j.jebo.2011.05.005.
2. Yuejin Zhang. (2016). "Research into credit scoring by merging social media information in online peer-to-peer lending", Procedia Computer Science, Vol. 91, 2016, pp. 168-174.
3. Bunker, R., Naeem, Aand Zhang, W. (2016). “Improving a Credit Scoring Model by Incorporating BankStatement Derived Features,” 1611.00252.pdf (arxiv.org).
4. Beibei Niu (2019). “Credit Scoring Using Machine Learning by Combining Social Network Information,” Information (Switzerland) 10(12):397 DOI:10.3390/info10120397.
5. Eweoya. I. O. (2019). "Predicting fraud in bank loan management using decision trees", Journal of Physics: Lecture Series, Volume 1299, 3rd International Conference on Science and Sustainable Development (ICSSD 2019) "Science, Technology and Research: Key to Sustainable Development” May 6-8, 2019, Center for Research, Innovation and Discovery, University of Covenant, Land of Canaan, State of Ogun, Nigeria and Ota.
6. Omprakash Yadav, (2019). “Loan Prediction System Using Decision Tree,” international journal of information and computing science ISSN NO: 0972-1347.
7. Mohammad Ahmad Sheikh. (2021) “Internal Credit Risk Models and Digital Transformation,” An Approach for Prediction of Loan Approval using Machine Learning Algorithm | IEEE Conference Publication | IEEE Xplore.
8. Shiva, T and Kumar, P. (2008). "A study of customer satisfaction with gold lending with special respect for Muthoot Fincorp Pt. Ltd", (PDF) Project Report summer internship on "A Gold Loan Customer Satisfaction Study for Muthoot Fincorp Pt. Ltd' (researchgate.ne).
9. Galih, K., Sidik, Djatna, T and Buono, A. (2014). “An IT2FS model forsharia credit scoring: analysis & design,” Jurnal Sistem Informasi 58(2) DOI:10.21609/jsi. v9i2.352.
10. Keramati, A and Yousefi, N. (2011). “Recommended Classification of Data Mining Techniques in Credit Scoring,” in Proceedings of the 2011 International Conference on Industrial Engineering and Operations Management, p. 22–24, Kuala Lumpur, Malaysia, Journal, January 2011.

11. F. Doko, S. Kalajdziski and I. Mishkovski, "Credit Risk Modeling Based on Central Bank Credit Registration Data", Journal of Financial and Risk Management, vol. 14, no. 3, tr. 138, 2021.

*****