Predictive Analytics for Loan Approval and Comparative Bank Interest Rate Evaluation

Ritika Sharma
MCA, Department of Computer Application & IT, IIS(Deemed to be University), Jaipur
Email: sharma.ritika.150705@gmail.com
Dr. Preeti Tiwari, Associate Professor, International School of Informatics & Management, Jaipur
Email: Preeti.tiwari@icfia.org

Abstract: The loan approval process is a vital process within the Banking industry. A variety of factors are taken into consideration when approving a loan to an applicant. These include income, age, experience in employment, credit score, loan amount requested, education, property ownership, and previous loan history. Machine learning can be used to analyse the data and predict whether an application will be approved or denied. This research contains two parts. Part A contains a dataset with 45,000 entries of loan applications. This dataset includes a variety of information about the applicant including income, employment, credit score, loan amount, reason for loan, etc. The target column indicates whether the loan has been approved or denied. Various machine learning classification models such as Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbours (KNN), and Support Vector Machine (SVM) can be applied to the dataset. These can then be compared using accuracy, precision, recall, and F1-score to see which machine learning model works best for predicting loan approvals. Part B of the research looks at comparing the lending rates of different banks using data provided by the Reserve Bank of India (RBI). The term loan rates for different banks can be compared. This can be simplified by looking at a set loan amount and time period. We can then calculate the estimated monthly EMI using the lending rates. This will help us understand how varying lending rates can impact the monthly repayment amount. This research combines machine learning to predict loan approvals and comparing lending rates. This can help us understand the key factors that relate to loan approvals and provides a simple way of determining which bank has the cheapest lending rates. Keywords— Loan Approval Prediction, Machine Learning, Random Forest, Credit Risk, MCLR, EMI, Data Science.

I. INTRODUCTION

Access to credit affects household consumption, education, property, and business. When banks receive loan applications they must determine whether they will be repaid and what rate to lend at. Traditional loan approval is time-consuming and can be inconsistent between officers. Credit scoring has therefore been used by lenders for decades. This uses statistical models to rank applicants based on their risk [12][13]. With the growth of digital lending there are now large amounts of data available. Machine learning (ML) can identify complex non-linear relationships between applicant attributes and previous lending decisions [7][8]. It is important to understand the factors influencing the predictions to make the predictions useful. Lenders need to know why loan applications are being declined and applicants need to know what will improve their chance of approval.

Only part of the lending process has been looked at when predicting loan approvals. Once an applicant has been identified as likely to be approved they will need to decide where to get their loan from. In India there are different lending rates between public sector, private and foreign banks. From April 2016, loans have been priced based on the Marginal Cost of Funds based Lending Rate (MCLR). MCLR is an internal benchmark that shows how much the bank is spending on funds [14]. The Equated Monthly Instalment (EMI) calculates the monthly repayment based on the interest rate, loan amount and term. Most borrowers do not have a structured method to compare their options and studies published often look at the prediction and pricing issues separately. A comprehensive view is needed with data to understand the lender and borrower issues.

This paper looks at both problems using two datasets. Part A uses the loan_data_new.csv file which contains 45,000 loan applications. This will allow us to understand the applicant, financial, credit and loan attributes of loan approvals and train ML models. Part B uses the bank details.csv file to compare the MCLR, term loan rates and estimate EMI of different banks to identify the relatively suitable options. There are 20 research questions throughout the investigation. These are split into 5 themes: Dataset & Descriptive Analysis, Factors related to loan approvals, Machine Learning Prediction, Bank Lending Rate Comparison, and EMI Based Bank Selection.

The main contributions of this paper are: Firstly, we have described and identified the key factors that affect the large dataset of loan applications. This includes how many loans were approved based on property ownership, previous loan history, and loan reason. Secondly, we have compared 4 classifiers and evaluated them using Accuracy, Precision, Recall, F1, ROC-AUC, and Confusion Matrix. As the dataset is imbalanced (22.22% were approved) only using accuracy may be misleading. Finally, we have compared the

different banks based on their benchmark rates and a set hypothetical loan. This has led to one clear recommendation for the most suitable bank. The rest of the paper is structured as follows: Section II looks at related work, Section III describes the experimental setup, Section IV outlines the experiments, Section V describes the results and discussion, and finally Section VI concludes.

II. LITERATURE REVIEW

Kruppa et al. [1] estimated the probability of default for individual consumers. They compared various machine learning models including Logistic Regression and Random Forests for credit risk prediction. This showed that data driven learners can reliably predict probabilities for individuals. However, this research only looks at the risk of the borrower and does not link the prediction to the lending rates of the banks.

Dastile et al. [2] performed a survey of 74 studies published between 2010 and 2018. Various statistical and machine learning models were reviewed including Logistic Regression, Discriminant Analysis, Naïve Bayes, K-Nearest Neighbours, and Ensemble models. Explainability and class imbalance were highlighted as an issue. This research only looked at models.

Uddin et al. [3] created a loan approval prediction model for banks using various classifiers including Logistic Regression, Decision Tree, Random Forest, Extra Trees, SVM, KNN, Gaussian Naïve Bayes, AdaBoost, and Gradient Boosting. The data was balanced using SMOTE. The models were combined using Voting Ensembles. A Smart Application was developed. This research focuses on automation and predicting defaulters. It does not incorporate the MCLR term loan rates or EMI.

Kurniawan [4] used the Random Forest algorithm to analyse credit risk. This used data understanding, feature extraction, preprocessing, exploratory data analysis, and modelling. The models were evaluated using recall. This research highlighted the key factors affecting credit risk. This research only looked at one algorithm and did not look at the benchmark rates of the banks.

Wang et al. [5] performed a comparison of credit risk models using bank loan data. Various classifiers were evaluated including Naïve Bayes, Logistic Regression, K-Nearest Neighbours, and Random Forest. This can help identify the most suitable algorithms for credit scoring. This research does not look at the MCLR or the lending rates of the banks.

Singh et al. [6] looked at automated loan approval prediction for banks. Various classification models were compared including Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine. This research focused on the prediction performance of the ML models. This research does not look at integrating the lending rates of the banks.

Khandani et al. [7] applied machine learning techniques to consumer credit data of a commercial bank to predict delinquency. They stated that the models helped improve out of sample predictions. This could lead to large cost savings. This research focuses on predicting losses for the lender. It does not look at the factors affecting approvals in the application data or the lending rates of the banks.

Lessmann et al. [8] performed an update to previous benchmarking work. 41 classification models were compared using multiple credit scoring datasets. They identified that advanced classifiers (particularly Ensemble models) performed the best. This has motivated the use of Random Forest in this research. This research only looks at benchmarking classifiers based on accuracy and profit. It does not look at interest rate benchmarking or comparing based on EMI.

No.	Author(s), Year	Methodology	Tools	Research Gap
1	Singh et al., 2025 [6]	Automated loan approval and algorithm comparison	Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), SVM	No integration of Lending Rates, MCLR, and EMI comparison.
2	Kurniawan, 2024 [4]	Data understanding, feature extraction, EDA, and modelling	Random Forest; recall-based evaluation	Single algorithm; focuses only on risk factors
3	Uddin et al., 2023 [3]	Loan approval prediction using SMOTE and voting ensembles	Logistic Regression, Decision Tree, Random Forest, Extra Trees, SVM, KNN, GNB, AdaBoost, GB.	No MCLR, lending-rate, or EMI analysis

4	Dastile et al., 2020 [2]	Systematic survey of 74 studies (2010–2018)	LR, LDA, Naïve Bayes, KNN, and ensemble methods	Discusses explainability and class imbalance; no bank-wise lending-rate comparison
5	Wang et al., 2020 [5]	Comparative evaluation of classifiers for credit scoring	Naïve Bayes, LR, KNN, Random Forest	No MCLR or bank lending-rate analysis
6	Lessmann et al., 2015 [8]	Benchmarking 41 classifiers on credit-scoring datasets	Single and ensemble classifiers	Focuses on classification benchmarks; no lending-rate or EMI comparison
7	Kruppa et al., 2013 [1]	Individual credit-risk probability estimation and comparison of machine-learning methods	Logistic Regression, Random Forest	Individual risk estimation only; no lending-rate or EMI comparison
8	Khandani et al., 2010 [7]	Delinquency forecasting using consumer credit data	Machine-learning forecasting models	Focuses on lender-side loss forecasting; no lending-rate comparison

Table 1: Comprehensive Summary of Literature Review

III. EXPERIMENTAL SETUP: HARDWARE, SOFTWARE (TOOLS AND TECHNIQUES)

All experiments were run in Python on a standard personal computer. No GPU was needed because the datasets are small and the models are classical ML algorithms. Table 2 lists the tools and techniques used at each stage.

Component	Tool / Technique	Purpose
Hardware	Standard personal computer (CPU only)	Running data analysis and model training
Language / environment	Python 3, Jupyter-style notebook	Scripting, analysis and reporting
Data handling	pandas [15], NumPy	Loading CSV files, grouping, aggregation, correlation
Visualisation	Matplotlib	Bar charts, box plot, ROC curve, confusion matrix
Machine learning	scikit-learn [10]	Preprocessing, model training, evaluation metrics
Classifiers	Logistic Regression, Decision Tree, Random Forest [9], KNN	Loan approval / rejection prediction
Evaluation	Accuracy, Precision, Recall, F1, ROC-AUC, confusion matrix	Model comparison beyond accuracy
Bank analysis	Sorting, group means, EMI formula	Comparing MCLR, term-loan rates and EMI

Table 2: Hardware, Software, Tools and Techniques

(a) Data Set

Two datasets, in the form of CSV files, were used in this study. **Dataset 1 (loan_data_new.csv)** has 45,000 loan applications with numerical and categorical applicant details and a two-class target, *Loan Status* (1 = approved, 0 = rejected); it is used in Part A. **Dataset 2 (bank details.csv)** has one row for each bank with its category (Public Sector, Private or Foreign), MCLR and term-loan interest rate; it is used in Part B. The two datasets describe different things (applicants and banks), so they are analysed separately. Table 3 describes the attributes and Fig. 1 shows a snapshot of Dataset 1.

Group	Attributes	Type	Role
Applicant	Age, Gender, Education	Numeric / Categorical	Applicant profile
Financial	Person Income, Loan Amount, Loan percentage (loan-to-income), Loan interest Rate	Numeric	Capacity and cost of loan
Employment	Employee Experience	Numeric	Stability of income
Property	Home Ownership (RENT, OWN, MORTGAGE, OTHER)	Categorical	Asset position
Credit	Credit History, Credit Score, Previous Loan (Yes/No)	Numeric / Categorical	Creditworthiness
Loan	Loan Intent (e.g., Personal, Education, Medical, Venture)	Categorical	Purpose of loan
Target	Loan Status	Binary (0/1)	Approved / Rejected
Bank data	Bank Type, Bank Name, MCLR, Term Loan Interest Rate	Categorical / Numeric	Bank comparison (Part B)

Table 3: Data Set Description (Dataset 1: 14 columns; Dataset 2: 4 columns)

Age	Gender	Education	Person Income	Employee Experience	Home Onwership	Loan Amount	Loan Intent	Loan interest Rate	Loan percentage	Credit History	Credit Score	Previous Loan	Loan Statu
22	female	Master	71948	0	RENT	35000	PERSONAL	16.02	0.49	3	561	No	1
21	female	High School	12282	0	OWN	1000	EDUCATION	11.14	0.08	2	504	Yes	0
25	female	High School	12438	3	MORTGAGE	5500	MEDICAL	12.87	0.44	3	635	No	1
23	female	Bachelor	79753	0	RENT	35000	MEDICAL	15.23	0.44	2	675	No	1
24	male	Master	66135	1	RENT	35000	MEDICAL	14.27	0.53	4	586	No	1
21	female	High School	12951	0	OWN	2500	VENTURE	7.14	0.19	2	532	No	1
26	female	Bachelor	93471	1	RENT	35000	EDUCATION	12.42	0.37	3	701	No	1
24	female	High School	95550	5	RENT	35000	MEDICAL	11.11	0.37	4	585	No	1
24	female	Associate	100684	3	RENT	35000	PERSONAL	8.9	0.35	2	544	No	1
21	female	High School	12739	0	OWN	1600	VENTURE	14.74	0.13	3	640	No	1
22	female	High School	102985	0	RENT	35000	VENTURE	10.37	0.34	4	621	No	1
21	female	Associate	13113	0	OWN	4500	HOMEIMPROV	8.63	0.34	2	651	No	1
23	male	Bachelor	114860	3	RENT	35000	VENTURE	7.9	0.3	2	573	No	1
26	male	Master	130713	0	RENT	35000	EDUCATION	18.39	0.27	4	708	No	1

Fig. 1: Snapshot of loan_data_new.csv

(b) Methodology:

The methodology is shown in Fig. 2. There are seven steps to the process. (1) Data Collection: The CSV files are imported. (2) Data Preprocessing: Missing values and duplicates are identified (None in Dataset 1). The data is split into numeric and categorical attributes. Categorical attributes are one-hot encoded. The data is split into training (80%) and testing (20%) sets (9000 test records). (3) Exploratory Data Analysis: The number of records, percentage of approved loans, averages, medians, and approval rates based on Education, Home Ownership, Previous Loan, and Loan Intent are calculated and plotted. (4) Feature Analysis: The correlation between numeric attributes and Loan Status are identified. The importance of each attribute is identified using Random Forest. (5) Machine Learning: Logistic Regression, Decision Tree, Random Forest, and KNN are trained. The

models are compared using F1-Score. The best model is evaluated using Accuracy, Precision, Recall, F1, ROC-AUC, and Confusion Matrix. (6) Bank Comparison: The banks are ranked based on MCLR and Term Loan Rates. The averages of each category are compared. The EMI is calculated using the formula: EMI = P.r.(1+r)n / ((1+r)n - 1). Where P is the loan amount, r is the monthly rate (Annual rate / 12), and n is the number of months. A hypothetical loan of ₹5,00,000 for 60 months was used for comparison purposes using the minimum term loan rate of each bank. (7) Result Analysis: The findings from both parts are analysed.

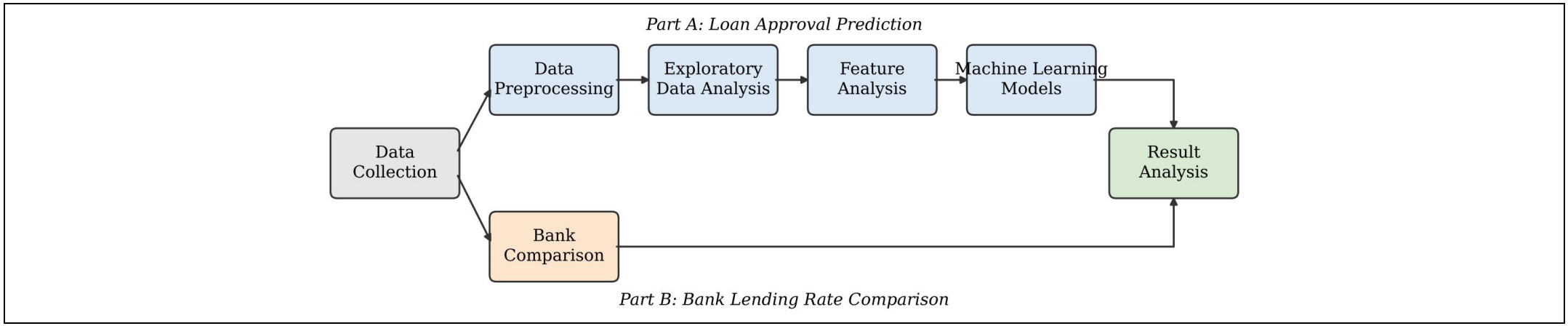


Fig. 2: Methodology of the proposed study

IV. EXPERIMENTS

The analysis is built around twenty research questions. Questions Q1–Q5 describe the data, Q6–Q10 look at factors linked to approval, Q11–Q15 build and test ML models, and Q16–Q20 compare banks. Table 4 lists each group, its questions and why it matters.

S.No	Section Header	Question	Significance
1	Dataset & Descriptive Analysis (Q1–Q5)	Q1: How many loan applications are present? Q2: What percentage of loans is approved or rejected? Q3: What is the average applicant income? Q4: What are the average and median loan amounts? Q5: Which education level has the most applications?	Helps understand the dataset and gives a basic summary of the applicants
2	Factors Associated with Loan Approval (Q6–Q10)	Q6: Does home ownership affect approval? Q7: Does previous loan history affect approval? Q8: Does credit score influence approval? Q9: Which loan purpose has the highest approval rate? Q10: Which features are most strongly associated with approval?	Finds the applicant and loan factors linked to approval
3	Machine Learning Prediction (Q11–Q15)	Q11: Can ML predict approval or rejection? Q12: Which algorithm performs best? Q13: What is the accuracy of the best model? Q14: Which features contribute most to prediction? Q15: How well does the model perform on Precision, Recall, F1 and ROC-AUC?	Builds, compares and tests ML models for approval prediction
4	Bank Lending Rate Comparison (Q16–Q18)	Q16: Which banks have the lowest MCLR? Q17: Which banks have the lowest term-loan interest rates? Q18: How do Public Sector, Private and Foreign banks differ?	Compares lending rates across banks and bank categories

S.No	Section Header	Question	Significance
5	EMI & Bank Selection (Q19–Q20)	Q19: Which banks have the lowest estimated EMI? Q20: Which bank is comparatively more suitable on rates and EMI?	Calculates EMI and finds the bank with better terms

Table 4: Table of Research Questions

V. RESULTS AND DISCUSSIONS

A. Dataset and Descriptive Analysis (Q1–Q5)

Q1–Q5. Table 5 summarises the descriptive results. The dataset is complete, with no missing values or duplicate rows, so it is suitable for statistical analysis and modelling. Only about one in five applications (22.22%) is approved (Fig. 3). This means the target is imbalanced, so accuracy alone is not enough to judge a model. The mean applicant income is ₹80,319.05. The mean loan amount (₹9,583.16) is higher than the median (₹8,000.00), which shows that a few large loans pull the average up. Bachelor-degree holders form the largest group of applicants.

Measure	Result
Total applications / columns	45,000 / 14 (9 numerical, 5 categorical)
Missing values / duplicate rows	0 / 0
Approved vs. rejected (Q2)	22.22% (10,000) vs. 77.78% (35,000)
Average applicant income (Q3)	₹80,319.05
Average / median loan amount (Q4)	₹9,583.16 / ₹8,000.00
Education with most applications (Q5)	Bachelor

Table 5: Descriptive Summary of Dataset 1 (Q1–Q5)

B. Factors Associated with Loan Approval (Q6–Q10)

Q6 – Home ownership. Approval rates differ a lot by home ownership (Fig. 4): OTHER (33.33%) and RENT (32.40%) are the highest, MORTGAGE is about 12%, and OWN is the lowest at 7.52%.

Q7 – Previous loan. Applicants with no previous loan have a 45.16% approval rate, but applicants with a previous loan have 0% in this dataset. This makes previous-loan history a very strong, rule-like factor.

Q8 – Credit score. The credit-score distributions of approved and rejected applicants are almost the same (Fig. 5), with medians close to 640. So credit score has only a weak link with approval here.

Q9 – Loan purpose. DEBTCONSOLIDATION has the highest approval rate (30.27%) and VENTURE has the lowest (14.43%). **Q10 – Strongest features.** Loan percentage (0.385) and loan interest rate (0.332) have the strongest positive correlations with Loan Status, while Person Income has a weak negative correlation (−0.136) (Fig. 6).

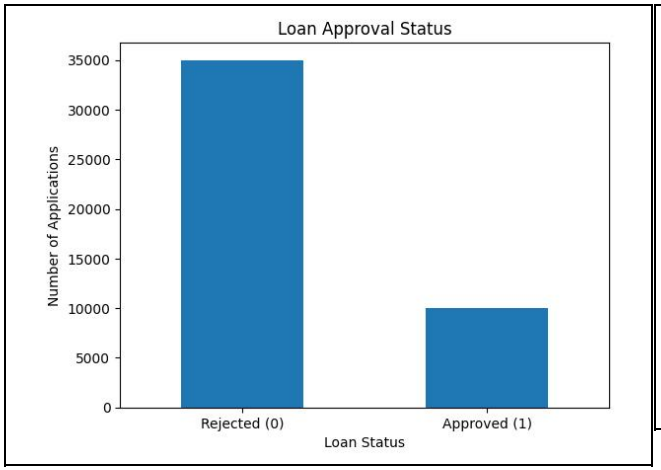


Fig. 3: Loan approval status (Q2)

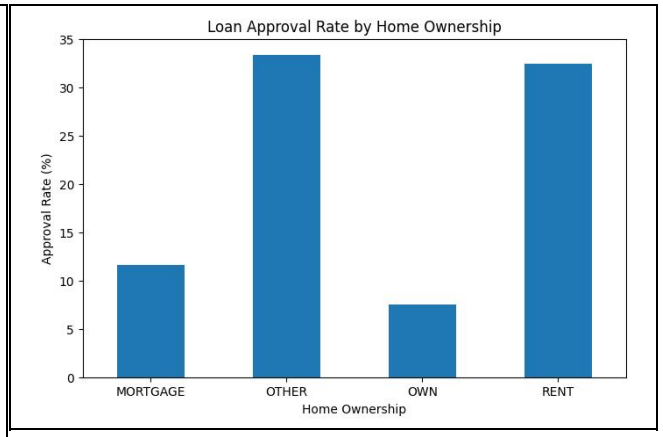


Fig. 4: Approval rate by home ownership (Q6)

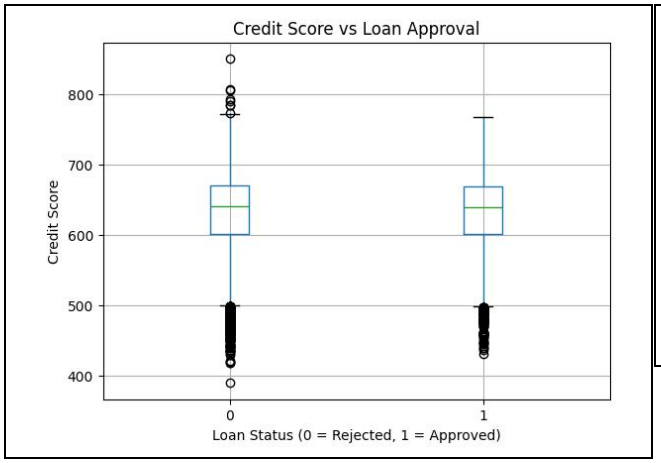


Fig. 5: Credit score vs. loan approval (Q8)

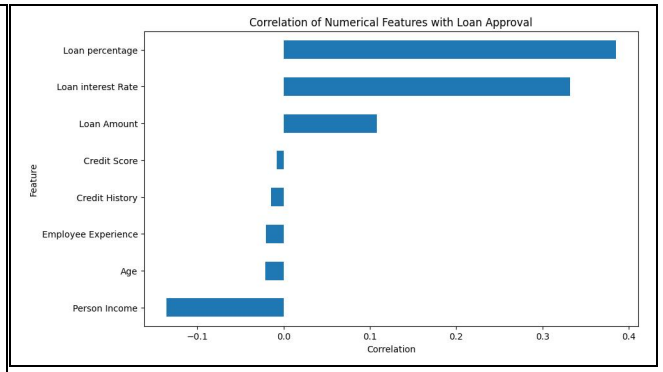


Fig. 6: Correlation of numerical features with approval (Q10)

C. Machine Learning Prediction (Q11–Q15)

Q11 – Can ML predict approval? Yes. A Logistic Regression baseline reached 90% accuracy. For rejected applications (Class 0), precision, recall and F1 were 93%, 94% and 94%, but for approved applications (Class 1) they were only 79%, 75% and 77%. The model works better on the majority class, so accuracy alone should not be used to judge it.

Q12 – Best algorithm. Random Forest achieved the highest F1-score among the four classifiers (Fig. 7). Decision Tree and Logistic Regression were close behind, and KNN was the lowest.

Q13 – Accuracy. Random Forest reached 92.74% accuracy (Table 6).

Q14 – Important features. Loan interest rate (15.04%), loan percentage (14.09%), previous loan (Yes 13.22%, No 12.48%, combined 25.70%) and person income (11.71%) are the top features in the importance ranking (Fig. 8). Loan amount (5.84%) and credit score (5.25%) matter less. This matches the correlation analysis in Q10 and the weak role of credit score in Q8. **Q15 – Detailed performance.** On the 9,000 test records, the confusion matrix (Fig. 9) shows 6,814 true negatives, 186 false positives, 467 false negatives and 1,533 true positives. Precision is 89.18% but recall is 76.65%, so the model is careful: it rarely approves a loan wrongly, but it misses about a quarter of the applications that were truly approved. The ROC-AUC of 97.41% (Fig. 10) shows that the model separates the two classes very well.

Metric	Logistic Regression (Class 1)	Random Forest (best model)
Accuracy	90%	92.74%
Precision	79%	89.18%
Recall	75%	76.65%
F1-score	77%	82.44%
ROC-AUC	Not reported	97.41%

Table 6: Performance Comparison of Logistic Regression and Random Forest

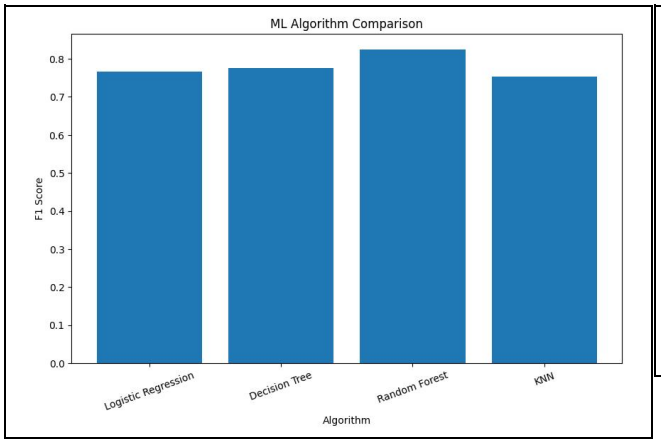


Fig. 7: F1-score of ML algorithms (Q12)

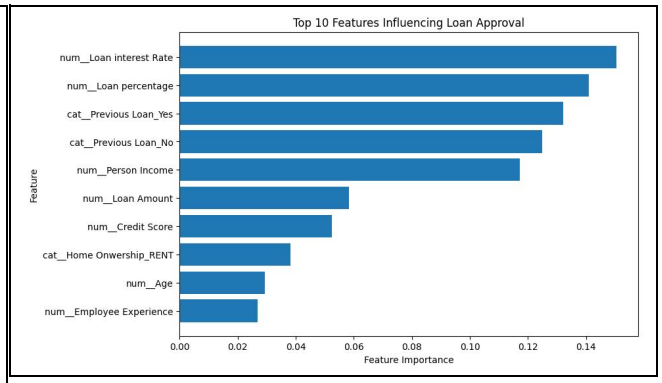


Fig. 8: Random Forest feature importance (Q14)

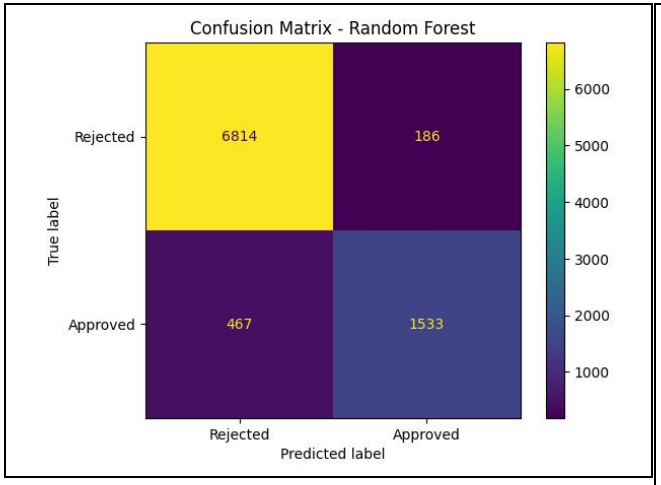


Fig. 9: Confusion matrix, Random Forest (Q15)

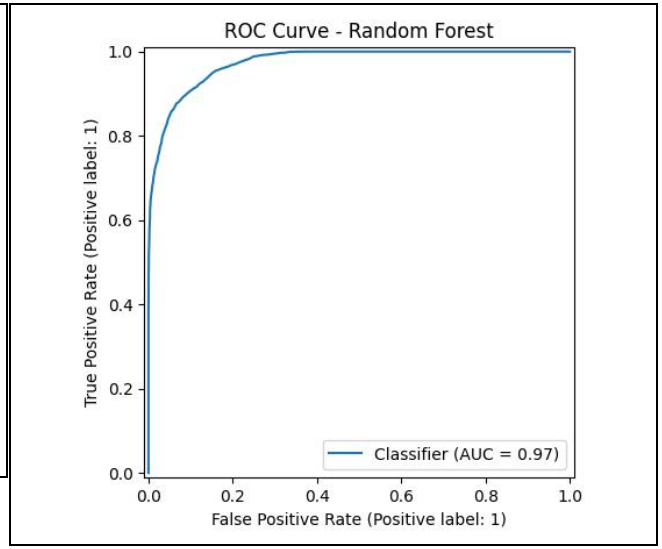


Fig. 10: ROC curve, Random Forest (Q15)

D. Bank Lending Rate Comparison (Q16–Q18)

Q16 – Lowest MCLR. American Express Banking Corp. has the lowest MCLR (4.78%), followed by UBS AG (4.87%) and Credit Agricole Corporate and Investment Bank (Fig. 11). Most banks in the lowest-MCLR list are foreign banks.

Q17 – Lowest term-loan rates. Standard Chartered Bank has the lowest minimum term-loan rate (1.00%), followed by RBL Bank, Doha Bank and IDFC First Bank (about 2.00%), Kotak Mahindra Bank (about 2.07%), HSBC, HDFC Bank and Federal Bank (2.50%), and IDBI Bank and Indian Bank (3.00%) (Fig. 12).

Q18 – Bank categories. On average, private banks have the lowest minimum term-loan rate (about 4%), public-sector banks are in the middle (about 5%) and foreign banks have the highest (about 6%) (Fig. 13). However, foreign banks appear at both ends of the ranking, so bank category alone is not a reliable guide and individual banks must be compared.

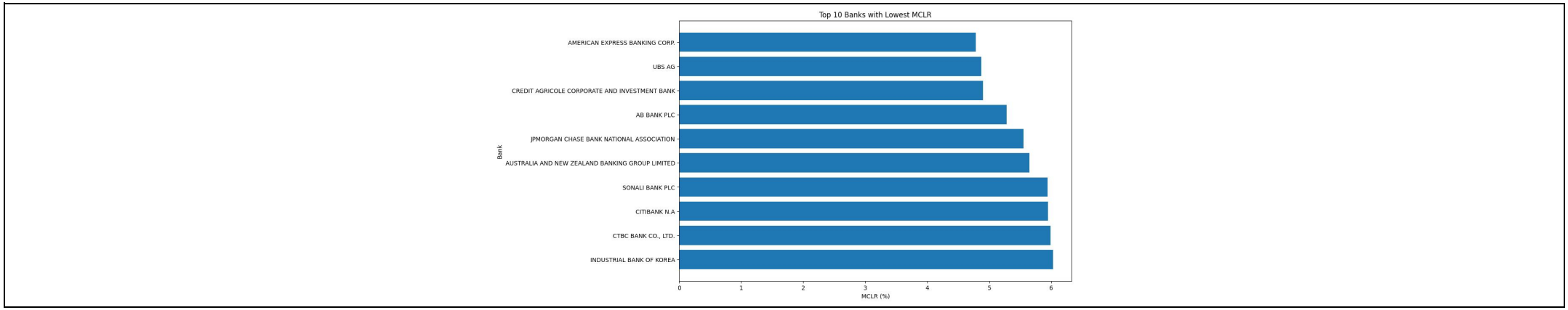


Fig. 11: Top 10 banks with lowest MCLR (Q16)

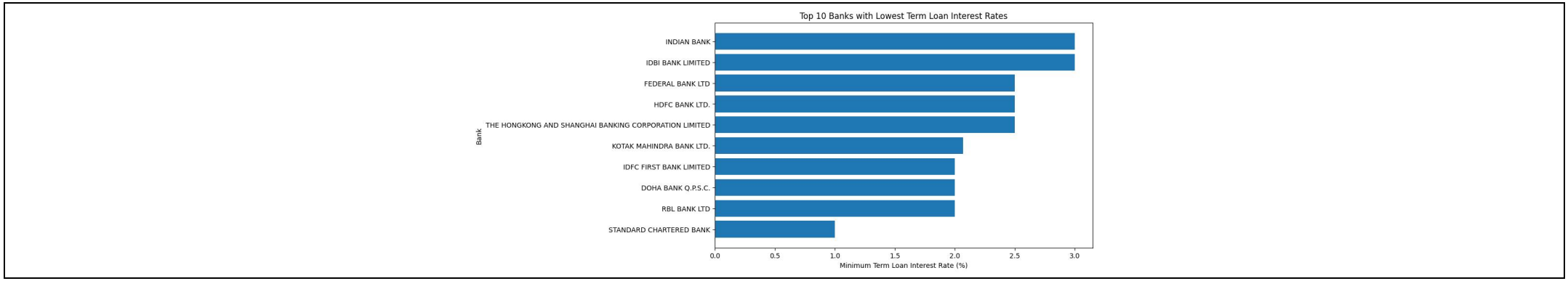


Fig. 12: Top 10 banks with lowest minimum term-loan interest rates (Q17)

E. EMI and Bank Selection (Q19–Q20)

Q19 – Lowest EMI. For the hypothetical loan of ₹5,00,000 over 60 months, Standard Chartered Bank has the lowest estimated EMI (₹8,546.87). It is followed by RBL Bank, Doha Bank, IDFC First Bank and Kotak Mahindra Bank, whose EMIs are close to one another (Fig. 14, Table 7).

Q20 – Most suitable bank. Standard Chartered Bank (Foreign) has both the lowest minimum term-loan rate (1.00%) and the lowest estimated EMI. Its MCLR is 7.8 and its suitability score is 88.04. On these measures, it is the most suitable bank among those compared. Its MCLR is not the lowest, which shows that the benchmark rate and the actual cost of borrowing can rank banks differently.

Rank	Bank	Min. term-loan rate (%)	Estimated EMI (₹)
1	Standard Chartered Bank	1.00	8,546.87
2	RBL Bank Ltd	2.00	8,763.88
2	Doha Bank Q.P.S.C.	2.00	8,763.88
2	IDFC First Bank Ltd	2.00	8,763.88
5	Kotak Mahindra Bank Ltd	≈2.07	≈8,779

Table 7: Banks with the Lowest Estimated EMI (₹5,00,000, 60 months)

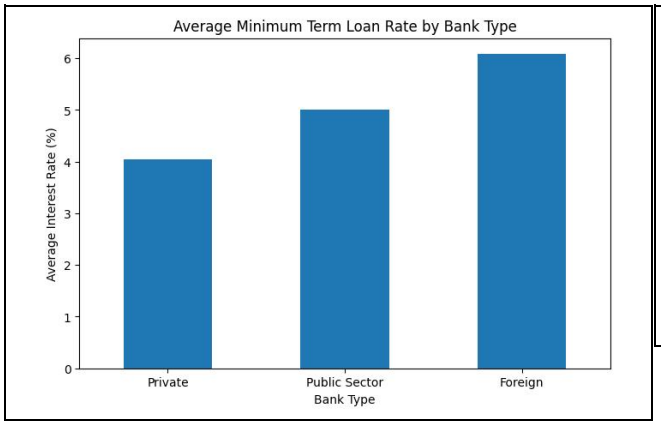


Fig. 13: Average minimum term-loan rate by bank type (Q18)

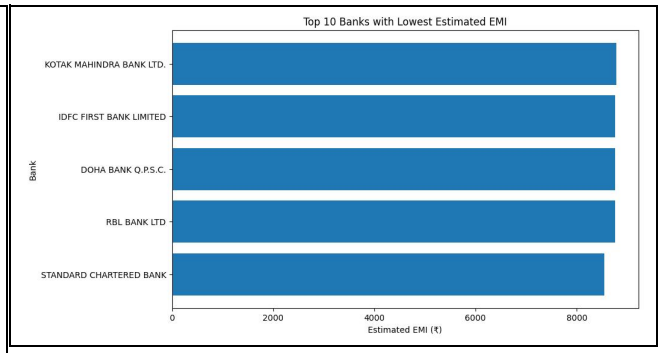


Fig. 14: Banks with lowest estimated EMI (Q19)

F. Discussion

There are three main observations made. Firstly, the loan approval of this dataset is determined by factors such as loan cost, loan size and income rather than credit score. It's clear that there is an issue with data labelling within the dataset considering that there is zero approval for anyone requesting loans with previous loan history. This means that the model is unlikely to be transferrable to different lenders. Secondly, the Random Forest model achieved better results than the other classifiers which is consistent with benchmark studies [8]. However, a low recall score of 76.65% indicates that further work could be done to balance the data (e.g. using SMOTE [11]), and/or adjusting the threshold. Finally, the bank ranking used the lowest advertised rates for term loans and only one hypothetical loan. The actual loan rate will vary depending on the applicant's profile, length of the loan, fees and loan type. Therefore, this recommendation is only a guideline.

VI. CONCLUSION

This paper studied loan approval patterns and bank lending rates using data science and machine learning. In Part A, 45,000 loan applications were studied and four classifiers were compared. Loan-to-income ratio, loan interest rate, previous-loan history and income were the most important factors, while credit score contributed little. Random Forest was the best model, with 92.74% accuracy, 89.18% precision, 76.65% recall, 82.44% F1-score and 97.41% ROC-AUC. In Part B, banks were compared on MCLR, minimum term-loan rate and estimated EMI. Standard Chartered Bank had the lowest minimum term-loan rate (1.00%) and the lowest estimated EMI (₹8,546.87), so it came out as the most suitable option among the banks compared. Overall, the study shows that loan approval prediction and rate comparison can be combined to help both banks and borrowers. The main limitations are the imbalanced target, the previous-loan pattern that is specific to this dataset, and the use of minimum rates and a single hypothetical loan. In future work, imbalance handling, hyper-parameter tuning and explainability methods can be added, applicant profiles can be linked with bank-specific rates, and the findings can be checked on more banks and recent data.

VII. REFERENCES

[1] J. Kruppa, A. Schwarz, G. Arminger and A. Ziegler, "Consumer credit risk: Individual probability estimates using machine learning," *Expert Systems with Applications*, vol. 40, no. 13, pp. 5125–5131, 2013.

[2] X. Dastile, T. Celik and M. Potsane, "Statistical and machine learning models in credit scoring: A systematic literature survey," *Applied Soft Computing*, vol. 91, Art. 106263, 2020.

[3] N. Uddin et al., "An ensemble machine learning based bank loan approval predictions system with a smart application," *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 327–339, 2023.

[4] R. Kurniawan, "Application of random forest algorithm on credit risk analysis," *Procedia Computer Science*, vol. 234, pp. 169–176, 2024.

[5] Y. Wang, Y. Zhang, Y. Lu and X. Yu, "A comparative assessment of credit risk model based on machine learning: A case study of bank loan data," *Procedia Computer Science*, vol. 174, 2020.

[6] D. P. Singh, A. Khare, N. Purusothaman, R. Lal, S. S. Chauhan and C. Kalaiarasan, "Predictive modeling for bank loan approval: From data to decisions," *Procedia Computer Science*, 2025.

[7] A. E. Khandani, A. J. Kim and A. W. Lo, "Consumer credit-risk models via machine-learning algorithms and risk management," *Journal of Banking & Finance*, vol. 34, no. 11, pp. 2767–2787, 2010.

[8] S. Lessmann, B. Baesens, H.-V. Seow and L. C. Thomas, "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research," *European Journal of Operational Research*, vol. 247, no. 1, pp. 124–136, 2015.

[9] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[10] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[11] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[12] D. J. Hand and W. E. Henley, "Statistical classification methods in consumer credit scoring: A review," *Journal of the Royal Statistical Society: Series A*, vol. 160, no. 3, pp. 523–541, 1997.

[13] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens and J. Vanthienen, "Benchmarking state-of-the-art classification algorithms for credit scoring," *Journal of the Operational Research Society*, vol. 54, no. 6, pp. 627–635, 2003.

[14] Reserve Bank of India, "Marginal Cost of Funds based Lending Rate (MCLR) – Guidelines," Mumbai, India, 2016.

[15] W. McKinney, "Data structures for statistical computing in Python," in *Proc. 9th Python in Science Conf.*, 2010, pp. 56–61.