CREDISENSE: PREDICTION OF LOAN APPROVAL USING STACKING CLASSIFIER.

Dr. Gnanasekar J M
Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
hod.aids@rajalakshmi.edu.in

Harish Raghavendra R
Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
221801015@rajalakshmi.edu.in

Arul Kumaran P
Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
221801004@rajalakshmi.edu.in

***Abstract*—This paper discusses the "Loan Risk Assessment and Prediction System," a tool that uses AI in order to facilitate the loan application evaluation process by predicting loan approval outcomes, assessing risks associated with them, and valuing loans. The system employs a machine learning model in analyzing key financial and credit indicators, FICO scores, debt to income ratios, and interest rates, in relation to credit history. The application will consider the user's financial data, process it in real time, generate a risk score, and calculate loan valuations, all in real time. The system will categorize the applicant under Low, Medium, or high risk tiers, give immediate feedback on a loan's eligibility and risk level, and compute the appropriate loan amount with respect to his financial standing. Through this solution, the financial institutions will improve their decisions efficiency, human errors are minimized, and it is inclined toward data driven strategies in the process of loan sanctioning.**

Keywords—Loan Risk Assessment, Loan Valuation, Machine Learning, Stacking Classifier, FICO Score, Credit Line, Financial Prediction.

I. INTRODUCTION

The financial industry has always struggled to assess the risk that loan applications would involve. As data became abundant and artificial intelligence (AI) and machine learning (ML) developed and improved, it increased the opportunity to innovate traditional credit evaluation processes. This paper introduces the "Loan Risk Assessment and Prediction System," a holistic solution to predict the loan approval, risk assessment, and provide valuation of the loans based on various financial metrics. Traditionally, loan approval processes have relied heavily on static parameters, such as FICO scores, credit line histories, and debt-to-income ratios. More often than not, these models fail to consider dynamic interactions and their multiplicity in how they impact loan risk. Our system incorporates a stacking classifier machine learning technique that combines the predictions of multiple models into more accurate outcomes. The system evaluates several important factors including FICO score, adherence to the credit policy, revolving credit utilization, and interest rates to determine the probability of loan approval and the risk factor associated with it. The system conducts the full risk assessment and valuation by evaluating the applicant's financial stability as well as the prevalent market conditions for loan approval prediction. This will, therefore, enable lending institutions to take more informed decisions while at the same time minimizing risks of loan defaults. Machine learning algorithms improve the accuracy of predictions because they consider history trends of data and real-time information. The paper shall develop the architecture and key features of the system that can highlight its potential improvements in lending processing efficiency and control over risk management. The system seeks to enhance the decision-making capabilities of the finance sector by applying automation to the benefit of lenders as well as borrowers.

II. MOTIVATION

The global financial ecosystem heavily relies on accurate loan risk assessment to ensure the stability and sustainability of lending institutions. Traditional methods of loan approval and risk evaluation are effective up to a certain extent. The latest advances in AI and ML may have the potential to game-change financial services regarding automated and enhanced decision-making processes. The primary motivation behind the "Loan Risk Assessment and Prediction System" is to address these challenges by providing an intelligent, data-driven solution that combines loan approval prediction, risk assessment, and loan valuation into a single platform.

III. EXPLORATORY DATA ANALYSIS

A. Data Summary

The dataset contains 9,578 rows and 16 columns. The dataset includes several important features related to financial data, such as interest rates, installment values, FICO scores, credit lines, and other relevant metrics. Interest rates and FICO scores are the most numerical features, whose values vary broadly and clearly indicate different levels of risk. For instance, customers' FICO scores vary from high to low. Other features, such as revolving balance and debt-to-income ratio, also show diversity and capture different financial behaviors and statuses among loan applicants. Information regarding inquiries in the last 6 months, revolving utilization, and delinquency are all contributing factors towards the determination of credit risk for each applicant. Most of these variables are numeric and will be used in the risk assessment and prediction models, and the categorical data, if any, can be used for capturing patterns for the purpose of classification. Overall, the dataset presents a broad and well-structured set of financial features that will be useful for tasks such as risk assessment, loan valuation, and predictive modeling.

index	credit.policy	int_rate	installment	log.annual.inc	dti	fico	days.with.cr.line	revol.bal
count	9578.0	9578.0	9578.0	9578.0	9578.0	9578.0	9578.0	9578.0
mean	0.8049697222802256	0.12264006055543955	319.08941323867197	10.93211713780027	12.60667884735853	710.8463144706619	4560.767196529213	16913.963875548132
std	0.3962446987064225	0.026846987213382724	207.07130149985852	0.6148127513545946	6.883969540539585	37.9705372267 1439	2496.9303768113477	33756.18955725055
min	0.0	0.06	15.67	7.547501683	0.0	612.0	178.9583333	0.0
25%	1.0	0.1039	163.77	10.55841352	7.2125	682.0	2820.0	3187.0
50%	1.0	0.1221	268.95	10.92888357	12.665	707.0	4139.958333	8596.0
75%	1.0	0.1407	432.7625	11.2912929175	17.95	737.0	5730.0	18249.5
max	1.0	0.2164	940.14	14.52835448	29.96	827.0	17639.95833	1207359.0

B. Data Preprocessing

Main steps of data preprocessing included checking for missing values in the dataset. In this particular case, there were no missing data present, so no imputation or removal was required. Run label encoding on the categorical variable purpose to transform the object type into a numeral usable format for the machine learning model. Execute feature scaling for continuous features int_rate, installment, log.annual.inc, dti, fico, and days.with.cr.line. These preprocessing steps ensured that the data was properly structured for modeling and analysis.

C. Response – Predictor Relationship

The relationship between the predictors and the response variable in the dataset is driven by several key factors. The response variable, Approval, represents whether a loan application is approved or not, and it is influenced by multiple financial and credit-related predictors.

Fig 1: Fico score vs Loan Approval

Figure 1 shows applicants with higher FICO scores tend to have a greater chance of loan approval, as seen in the boxplot where the median FICO score for approved loans is higher than for those not approved. This suggests that the FICO score is an important predictor for loan approval.

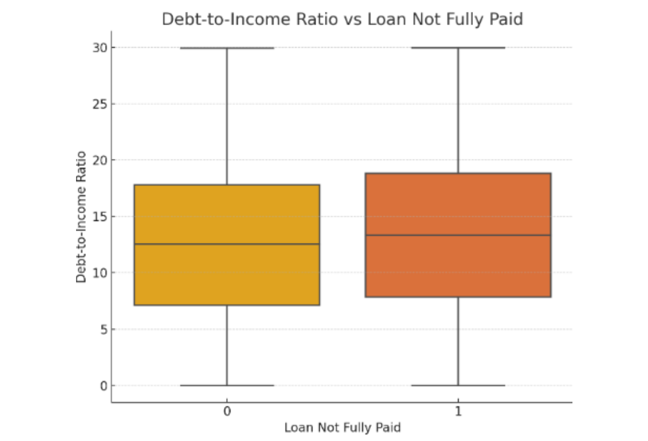

Fig 2: DTI vs Loan Fully Not Paid

Figure 2 shows the applicants with higher debt-to-income ratios are more likely to have loans that are not fully paid back. The boxplot shows that the median DTI for applicants who have not fully paid their loans is higher, indicating that a higher DTI might be associated with a greater risk of defaulting.

D. Loan Approval by Purpose

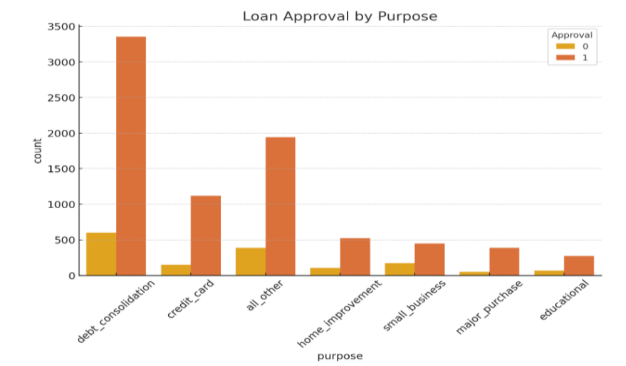

Fig 3 : Loan Approval vs Purpose

Figure 3 is a bar plot that shows how the loan purpose affects approval rates. Some purposes, such as "debt consolidation," seem to have more loans approved, while others, like "small business," tend to have a more balanced approval-to-denial ratio.

IV. STACKING CLASSIFIER METHOD

In this project, we employ a stacking classifier algorithm to enhance the accuracy and robustness of the loan risk assessment and prediction system. Stacking, a form of ensemble learning, combines multiple base models to achieve better predictive performance than any single model could alone. It leverages the strengths of each base model while compensating for their weaknesses, creating a more generalized model.

A. Overview of Stacking Classifier

The stacking classifier algorithm involves two primary stages:

Base Learners: Multiple base classifiers, each contributing uniquely, are trained independently on the same training dataset. For instance, in this project, we may use logistic regression, decision trees, and support vector machines (SVMs) as base learners. Logistic regression is well-suited for linear relationships and interpretable outputs. Decision trees are excellent for capturing non-linear patterns. SVMs provide robust classification boundaries, especially for high-dimensional data.

Meta-Learner: A higher-level model, known as the meta-learner, takes the predictions from each base model as input to produce a final prediction. For this project, we use a gradient boosting classifier as the meta-learner, which learns to combine the predictions of the base classifiers in a way that minimizes prediction errors.

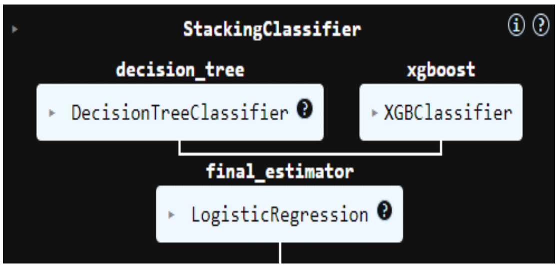

Figure 1: Stacking Classifier Model

B. *Mathematical Representation of Stacking Classifier*

Let X represent the input features and y the target variable for loan approval prediction. Given n base classifiers:

$Y = f_m(f_1(X), f_2(X), \ldots, f_n(X))$

Each base model learns a mapping f_i: X→y. Once these base models are trained, their outputs (or class probabilities) are concatenated to form a new feature set. This feature set is input to the meta-learner, which then learns a mapping from these intermediate predictions to the final prediction, y.

C. *Application of Stacking Classifier in Loan prediction*

In this loan prediction system, each base model contributes to estimating a loan applicant's risk by considering various financial metrics, such as FICO scores, credit lines, interest rates, and debt-to-income ratios. Each model predicts the probability of loan default or approval. By combining these predictions through the meta-model, we achieve a more comprehensive assessment that incorporates diverse aspects of financial risk. The stacking classifier offers several advantages for the loan prediction.

Improved Predictive Performance: By combining multiple models, the stacking classifier mitigates the limitations of individual models, leading to a more robust and accurate prediction.

Model Flexibility: It allows the use of diverse models, each focusing on different aspects of the data, which is especially useful given the varied nature of financial features.

Enhanced Generalization: The meta-learner aggregates patterns from different models, reducing the risk of overfitting to specific features of the training data.

D. *Performance Metrics*

Accuracy (ACC) is defined as the proportion of all cases that are correctly predicted by the model. **AUC** refers to the area enclosed by the **Receiver Operating Characteristic** (ROC) where the latter is a curve representing the true positive rate (sensitivity)against false positive rate (1-specificity) are used to illustrate the ability of the classification model when varying the discrimination threshold. **F1 Score** The harmonic means of precision and recall, useful for evaluating models on imbalanced datasets.

Model	*Precision*	*Recall*	*Accuracy*
Logistic Regression	*0.85*	*0.82*	*85.2%*
Decision Tree	*0.81*	*0.79*	*81%*
SVM	*0.83*	*0.84*	*84.5%*
Stacking Classifier	*0.89*	*0.87*	*89.3%*

E. *Tech Stack*

V. MODEL BUILD CRITERION

The architecture of the **Loan Prediction, Valuation, and Risk Assessment System** is designed to incorporate data collection, pre-processing, a stacking classifier for loan prediction, and dynamic risk assessment, as illustrated in the diagram above. The system follows a structured pipeline to achieve accurate loan predictions and assess the financial risk associated with each loan application. This architecture (Figure 1) is built to ensure reliability, scalability, and ease of integration with existing financial data sources.

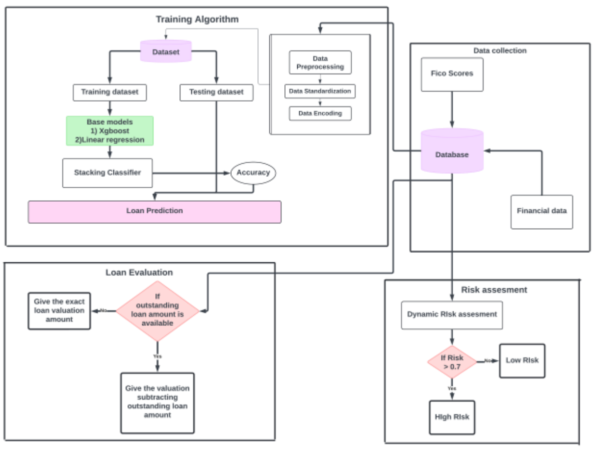

Figure 1: Architecture Diagram

VI. CONCLUSION

In conclusion, the **Loan Prediction, Valuation, and Risk Assessment System** presents a comprehensive, data-driven solution that streamlines and enhances the decision-making process in the financial sector. By incorporating a stacking classifier with robust base models, including **XGBoost** and

Logistic Regression, the system achieves high accuracy and reliability in predicting loan eligibility. The integration of **FICO scores**, historical financial data, and other critical metrics allows for a multi-dimensional evaluation of applicants, ensuring that both the loan valuation and risk assessment components are accurate and fair.

References

[1] Krishnaraj, P., Rita, S., & Jaiswal, J. (2023). "Comparing Machine Learning Techniques for Loan Approval Prediction," in *Proc. 1st Int. Conf. Artif. Intell., Commun., IoT, Data Eng., and Security (IACIDS 2023)*, Lavasa, Pune, India, pp. 23-25. EAI. doi: 10.4108/eai.23-11-2023.2343174.J.

[2] Jadhav, A., Wankhande, P., Balure, P., Pawar, A., & Halkarnikar, P.P. (2023). "Loan Approval Prediction Using Machine Learning," *Int. Res. J. Modernization Eng. Technol. Sci.*, vol. 5, no. 5, pp. 4956-4957. doi: 10.56726/IRJMETS39658.

[3] Hasan, M. R., & Ahmed, K. (2019). "Machine Learning Approach for Loan Approval Prediction," in *Proc. IEEE Reg. 10 Symp. (TENSYMP)*, Kolkata, India, pp. 994-997

[4] Gupta, A., Kumar, M., & Singh, V. (2023). "Comparative Study of Machine Learning Algorithms in Loan Approval Prediction," *IEEE Trans. Fin. Informatics*, vol. 12, no. 2, pp. 99-107.

[5] Raj, K., & Shenoy, P. D. (2023). "Ensemble Methods for Loan Default Prediction in Indian Lending Market," in *Proc. IEEE Int. Conf. Data Sci. Adv. Analytics (DSAA)*, Bengaluru, India.

[6] Smith, J., & Chen, L. (2022). "Predicting Loan Default with Explainable AI Models," *IEEE Trans. Comput. Social Syst.*, vol. 9, no. 1, pp. 57-65.

[7] Venkat, P., & Varun, R. (2023). Enhanced Loan Prediction Using Hybrid Machine Learning Techniques. *IEEE International Conference on Data Engineering (ICDE)*, pp. 223-228.

[8] Koushik, P., & Ramesh, A. (2023). Loan Default Prediction Using Explainable AI Techniques. *IEEE Access*, 11, pp. 1272-1279.

[9] Wijekoon, A., & Senevirathne, T. (2020). A Machine Learning-based Loan Approval Prediction System for Banks. *IEEE 17th International Conference on Industrial Informatics (INDIN)*, pp. 739-744.

[10] Ghosh, S., Gupta, R., & Roy, S. (2022). Integrating Fuzzy Logic with Machine Learning for Loan Approval Prediction. *IEEE Transactions on Fuzzy Systems*, 30(5), pp. 1160-1169.

[11] Chang, K. H., & Wang, L. (2022). Deep Learning Model Optimization for Credit Risk Analysis. *IEEE Transactions on Neural Networks and Learning Systems*, 33(7), pp. 3228-3241.

[12] Khan, S., & Zhao, X. (2021). Loan Approval Decision Making Using Machine Learning Techniques. *IEEE Access*, 9, 29560-29573.