BANK REPAYMENT PREDICTION-SYSTEM ON DEEP LEARNING TECHNIQUES

Rafique Ahmed Vighio[1] Shazma Tahseen[2] Abdul Rehman Baloch[3] Kainat Wighio[4] Shah Muhammad Kamran[5]

[1]*School of Computer Science, Beijing University of Posts and Telecommunications (BUPT), China*
[2]*Information Technology Department, Liaquat University of Medical & Health Sciences, Jamshoro, Pakistan*
[3]*Department of Computer Science, Iqra University, Karachi, Pakistan*
[4]*Department of Computer Science, University of Sindh, Pakistan*
[5]*Mehran University-Institute of Science & Technology Development (MUISTD), Mehran University of Engineering & Technology, Jamshoro, Pakistan*

[1]rafiqwighio@gmail.com[2]Shazma.tehseen@lumhs.edu.pk[3]abdul.rehman03@iqra.edu.pk[4]kainatwighio@gmail.com[5]Kamran.shah@faculty.muet.edu.pk

DOI:https://doi.org/

KeywordsDeepLearning,LoanPredictionModel, Training,Testing,Prediction, AccuracyAnalysis.

Article History
Received on 08 July 2025
Accepted on 28 July 2025
Published on 20Aug 2025

Corresponding Author: *
Rafique Ahmed Vighio

Abstract

A significant service provided by financial institutions is loans, however, with little resources to spread out banks should be selective in selecting low-risk borrowers in order to reduce the number of defaults. The traditional machine learning algorithms such as SVM and KNN are limited in one way or another; SVM is too computationally intensive, KNN tends to perform significantly worse because of inefficiency and parameter sensitivities. Similarly, Random Forest, despite its popularity, suffers in two aspects, namely, feature selection and model tuning.

This research paper proposes a deep learning-based system to predict bank loan repayment to address such shortfalls. Through an instance of Deep Neural Network (DNN), the model can learn and optimize features automatically, can support more kinds of data, and achieves a higher degree of accuracy and scalability than the classical methodologies.

A dataset of 100,000 loan applications provided by Kaggle was used to train the proposed model on. Its accuracy of 82% was better than the 68% exhibited by Random Forest by 13%. The results indicated the possibility of deep learning to minimize the bad loans, enhance the risk assessment, and also to optimize the future loan approvals.

INTRODUCTION

Loans are a fundamental component of the banking and financial services industry, serving as a major source of revenue. However, banks operate with limited financial resources and cannot approve every loan or mortgage application. To manage this limitation, financial institutions have developed structured systems to evaluate and shortlist applicants. Traditionally, most banks rely on rule-based systems for assessing loan applications. While widely adopted, these systems suffer from several inherent limitations, including rigidity, the complexity of risk factor calculations, limited adaptability, and the high cost of implementation due to the need for ongoing human involvement.

To address these shortcomings, this research proposes the application of artificial intelligence—specifically, a deep learning-based system—for assessing loan applications. This approach aims to Forecast the inefficiencies and constraints of rule-based systems by implementing a Deep Neural Network (DNN) model, alongside a traditional machine learning model such as Random Forest, and comparing their respective performances.

Lending institutions and banks are exposed to multiple risks when they are engaging in the process of approving a loan. To reduce chances of such risks, they usually make thorough evaluations regarding the credit worthiness of an applicant prior to making a loan. As there is a high engagement of digital bank services, consumer-driven actions produce massive amounts of data on a daily basis. Not only does this information contribute to the expansion of the banking industry, but it also opens significant opportunities for studying the behavior of customers in the best way possible [1]. With rising credit demands a better method needs to be implemented and developed in the current loan assessment systems to minimize risk and to increase stability in the loan authorization process [2].

Literature shows that loan repayment behavior has been predicted using different models of machine learning. These are Random Forest, K- Nearest Neighbors (KNN), Logistic Regression, and Support Vector Machines (SVM). All these models have proved to be effective to some extent, although limited in one way or the other. As an example, Logistic Regression, although widely applied to solve binary classification-related problems--might not be effective when applied to a complex or non-linear relation observed during the data modeling process [3]. Moreover, classifiers such as Naive Bayes rely on the feature independence assumption that can fail to exist in real life. Although limited, machine learning models seem to do better overall at predicting credit risks, are more accurate and scalable compared to traditional statistical tools.

Research conducted by Hinton and Salahuddin, however, concluded that most of the machine learning models are short-sighted, as they only care about the output of the classifier and miss rich information in the data, therefore limiting its predictiveness [4]. Recent innovations in both computational power and algorithm development, on the other hand, have rendered deep neural networks trainable, which can learn to encode and exploit this latent information more efficiently. Such deep learning models are also gaining recognition due to their capacities to model complicated data forms, to automate the identification of identifying important features, and to enhance performance with the rise of the training data.

Applied to financial risk assessment, DNNs have demonstrated to be extremely powerful in classification, pattern recognition, and time-series forecasting. The process of loan assessment is affected not only by external factors (e.g., political changes, natural disasters), but also by internal processes (e.g., credit history, income level), and DNNs will be beneficial to represent such complex relationships [2][3]. They are able to model highly complex (non-linear) interactions between features, and their universal approximation property makes them capable of this task without requiring large amounts of manual feature engineering.

The systematic approach was used in the current work to design a DNN-based loan repayment prediction system and test it. Visual Studio Code, Jupyter Notebook, and Google Colab were the tools to be used in implementing the model. The operations involved retrieving real data on loans, cleaning up the received data and removing missing or inconsistent data, and choosing and standardizing the most important attributes and separating the data on the given dataset into training and testing sets. A multi-layer DNN model was then trained and its performance compared to traditional machine, learning models especially Random Forest. The target outcome of the model was to identify a loan applicant based on the input features (FICO score, amount of income, amount of loan, employment history among others) whether they could successfully repay the loan or they will default.

The general aim of the study is to enhance the quality and effectiveness of loan default prediction using the learning potential of DNNs. This study attempts to create an improved predictive model that would help banks handle large and complex data more efficiently, improve their decision-making process, and, eventually, minimize financial losses related to non-performing loans. Several machine learning algorithms were tested during this research, and its outcome shows that the presented model, based on DNN, exhibits a high level of accuracy, which is very useful in contemporary approaches to credit risk management.

2.LiteratureReviewandKeyTechnologies

Banks and other institutions have loans at the center of their financial business thus they are forced to focus on perfecting their predictive models in appraising borrowers with the aim of reducing their risks. As the need to loan people keeps on increasing, so does the need to enhance predictive systems to be able to determine and analyze the ability of the people to pay in an effective and accurate manner [3]. Initial studies have identified the fact that the higher the loan approval rates are, the higher the risks of loan default which prompts the banks to place more emphasis on soft and hard information in the decision-making process [2]. Although it has been proposed that multiple predictive models can be used to classify loans [5], high-performing systems and reliability are still required.

The use of credit scoring as a crucial element of financial decision-making becomes the point of common research into the models of default predictions [4]. Despite this, Support Vector Machines (SVM) have proved less effective when utilized on datasets that contain less than ten features [6]. Later, there has also been an improvement in the field of financial risk evaluation that is the emergence of Artificial

Neural Networks (ANNs) which are models resembling the process of learning in humans with their capability of analyzing the past data and improving the accuracy of selection or decision-making [7].

Several comparative analyses have covered the various methods to classify the borrowers as defaulters or not defaulters. As an example, it has compared SVM, Random Forest, and Genetic Algorithm models and, across the board, ensemble methods have proven much better than single classifiers [8]. Decision tree algorithms have also been effective in coming up with credible applicants consequently enhancing the process of assessing overall risk [9]. The results obtained empirically also reveal that heterogeneous ensemble models work better than standalone methods, providing more reliable prediction capability [10].

In the small business loan classification, neural networks, particularly feed forward architecture type, have demonstrated very high level of accuracy as evidenced by the result of their application to Italian financial data in which the result was found to have very minimal error levels and high reliability [11]. These innovations enhance and strengthen decision-making as financial institutions gain power by being able to process loans with the help of big data.

Other works have suggested hybrid solutions which are a combination of normalization methods and machine learning classifiers. An example would be a model based on Min-Max normalization with a K-Nearest Neighbor (KNN) type classifier in R that appears to be able to predict the loan status in commercial banks [12]. In a similar way, a decision tree model created on the basis of the data available in the UCI repository demonstrated a good efficiency performance, which was achieved upon pre-processing the set of data [13]. Best predictive gains were in using classifier performance to assess risk-return results and portfolio properties by other researchers [14]. Also, confusion matrix metrics were cited to report improvement in belief-plausibility-ambiguity (BPA) construction [15].

Innovation is further advanced with the introduction of a feature weighted SVM model in which an F-score rank is incorporated along with Random Forest dimensions of feature importance to provide an improvement to the accuracy of credit scoring [16]. Classification of pre-approval defaults was conducted using tree-based classifiers including BPSOSVM-ERT and BPSOSVM-RF, which are known to be effective in predicting defaults, with ERT showing improvement in important metrics that make it computationally efficient as compared to Random Forest [17].

The literature has a wide scope of credit risk modeling methods, and they include logistic regression, the neural network and the ensemble learning. Model selection has been done in a way that none has consistently performed better than others in all the datasets so is context dependent [18]. Although the publicly available datasets can be accessed with limited access due to the concerns of confidentiality, credit scoring using statistical and machine learning techniques is effective [19]. Relevance Vector Machines (RVMs) here possess the advantage of providing probabilistic estimates of the confidence and, as such, are not limited by the binary results of SVM-based systems [20].

Neural networks have again done better compared to conventional approaches such as logistic regression and decision trees, to predicting bank failures owing to its capacity to non-linearly transform its input variables [21]. Skewed datasets, which is an inherent problem in financial modeling, have already led to the development of custom methods to address this situation [22]. Such an improvement involves a fuzzy SVM model that contributes weight to the input features, and that reduces outlier effects [23]. Innovative fuzzy SVM methods further generalize borrower classification by assigning samples partial membership in both positive and negative classes, leading to better generalization [25]. Studies confirm that neural networks generally yield more objective and reliable results in credit risk assessment than logistic regression models [26], though traditional statistical tools still serve as valuable benchmarks in model development [27]. When applied to bank health analysis, Deep Neural Networks (DNNs) outperform conventional risk metrics like the repricing gap, offering greater insight into institutional vulnerabilities [28].

Advanced hybrid classifiers such as Conditional Markov Networks have been developed to handle both continuous and discrete data in credit scoring applications, successfully overcoming earlier limitations [29]. Hidden Markov Models and Fuzzy Markov Models have also shown promise in financial risk assessment, although challenges related to data quality persist. To address predictive accuracy concerns, ensemble learning has emerged as a viable strategy, with studies consistently showing enhanced performance when combining multiple models [30].

Artificial Neural Networks, owing to their universal approximation property, exhibit powerful predictive capabilities, particularly in pattern recognition and classification tasks [35]. Neuro-rule extraction techniques have been employed effectively to derive interpretable rules from complex credit risk datasets [37]. While neural networks and logistic regression offer comparable performance in many settings, hybrid approaches that integrate both yield the most effective results [38][39]. Neural networks continue to excel in financial applications, offering superior results in classification, forecasting, and decision-making tasks [39].

In terms of practical implementation, recent work employs deep learning frameworks built using Keras and TensorFlow to process complex financial data. These systems rely on Python's robust data science ecosystem for preprocessing, training, and evaluation, ensuring both high research accuracy and production-level performance. This supports financial institutions in conducting more precise risk assessments while maintaining computational efficiency.

While earlier research posted that deeper architectures better simulate cortical information processing, more recent findings suggest that even relatively shallow networks can perform robustly across various financial tasks [34].

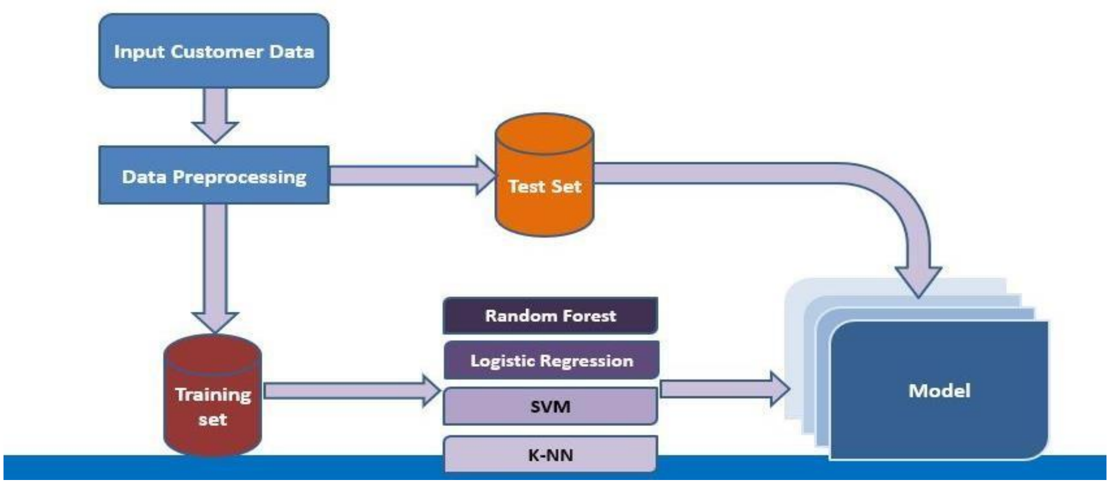

Figure.1:ExistingModelsArchitecture

3. Methodology

This study introduces a predictive classification system that segments loan applicants into two primary groups: performing borrowers and approved clients who later default. The risk assessment framework integrates demographic attributes (e.g., age, housing status), income sources, and existing liabilities. There are two financial ratios which are payment-to-income (PTI) and debt-to-income (DTI) which form the fundamental indicators. Higher PTI and DTI ratio are closely related to higher risk of default whereas the lower ratios express credit worthiness. Coupled with traditional scoring of credit, these financial indicators generate an operative profile of the borrower to effectively risk-rate.

3.1 Prediction and System Architecture

To learn loan default risk, there was the use of Deep Neural Network (DNN) owing to the ability to learn multidimensional non-linear relationships. The structure contains six layers, and it examines 18 monetary features, and every tarp consists of activation functions to make the model less linear and enhance predictive depth. The structure would allow the network to blueprint the hidden patterns in the data on applicants and achieve greater accuracy in classification as compared to the linear models.

3.2 Used Software and Tools

Developing and testing the model has made use of various contemporary tools. Training was performed using Google Colab, which is GPU-accelerated, cloud-based processing. The dominant language was Python, with supporting data science libraries NumPy and SciPy. Anaconda helped manage the environment and Visual Studio Code is used to run modular back end/ frontend development. TensorFlow, PyTorch, and Keras were examples of deep learning frameworks used; each was chosen due to its capacity to model training and experimentation of cognitively complicated models at scale.

4. Proposed Model

This section provides a proposed model new deep neural network (DNN) architecture that will be used to predict repayment of banks and also to detect high risk customers specifically modeled to work on complex (and unstructured) financial data. The proposed system takes advantage of two important technological trends, namely, the increasing availability of complete financial data and the fact that high-performance computing resources have become increasingly available. Such developments make artificial neural networks the new industry standard on which precise, scalable risk prediction can be performed, during banking operations.

Training/testing was done using the Kaggle dataset involving 100,000 entries and 18 original features that were augmented by dummy encoding of the attribute loan purpose. Before training, Principal Component Analysis (PCA) was done to reduce its dimensionality, and the standard preprocessing steps were followed, missing values, features scaling before dataset split training, and test sets. The model structure is comprised of 7 layers where training was based on Adam optimizer and binary cross-entropy loss function. The rationale behind this is to address the shortcomings in the accuracy of the traditional loan default predictive systems through the increased accuracy in feature learning on a hierarchical level and enhanced pattern recognition.

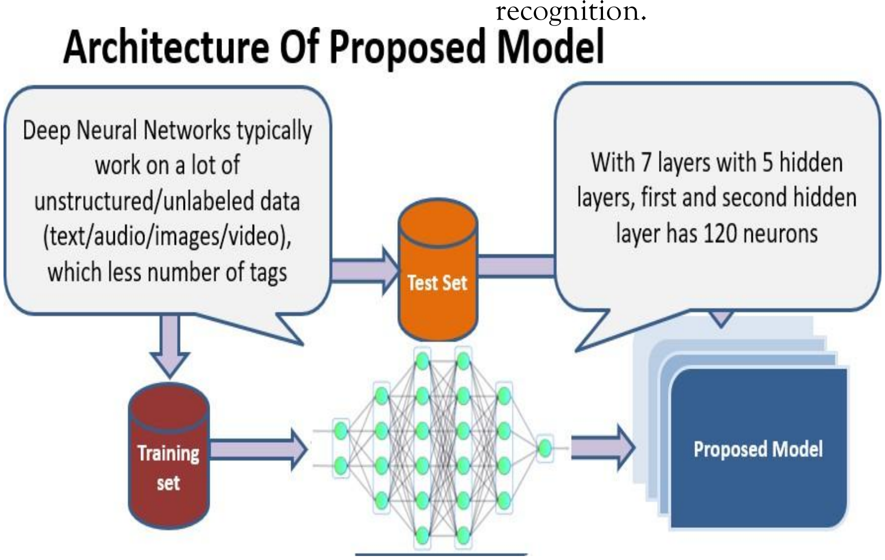

Figure.2:ProposedModel

The model analyzes the conduct of customers in order to prognosticate the probability of loan repayment to authorize data-driven approvals. Behavioral input goes through a compared approach to the mixing of Random Forest and Deep Neural Network classifiers in the system. The model then analyzes new applicants in relation to the imparted patterns and provided a score on the scale of risk. This model of learning supplements conventional credit risk evaluation techniques with better prediction.

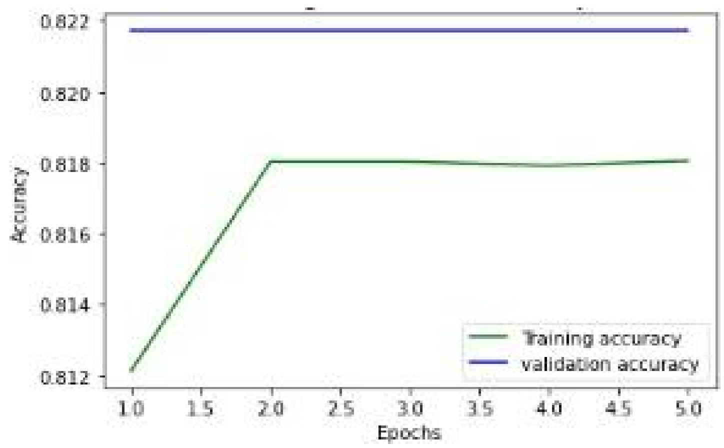

Figure.3:Trainingand Validation accuracy

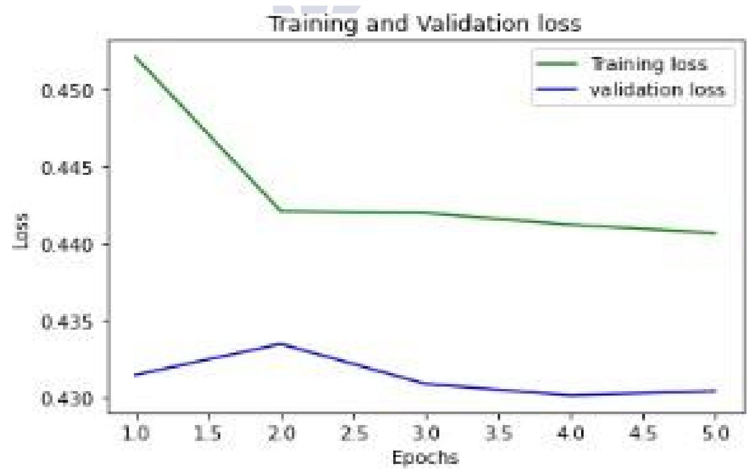

Figure.4:Trainingand validationloss

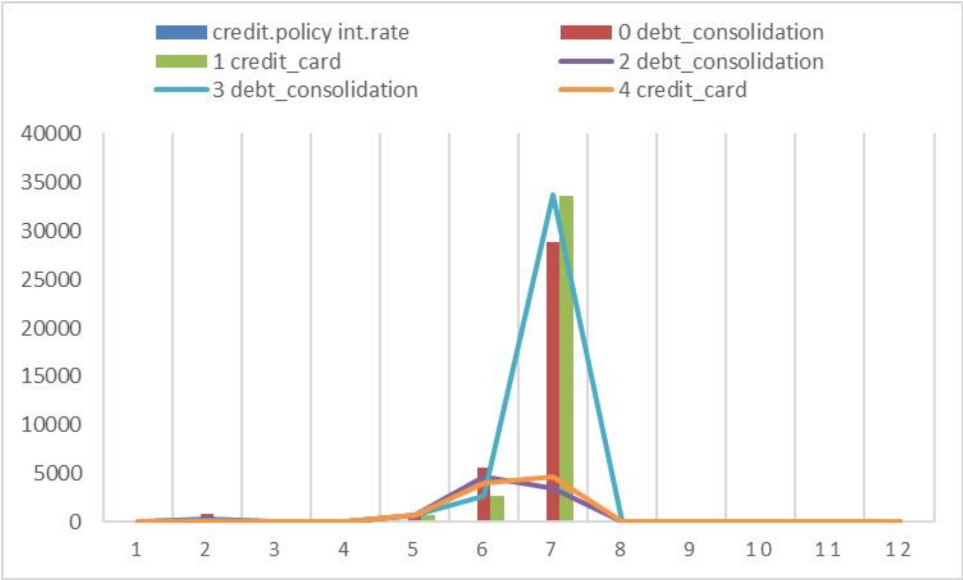

Figure.5:Featureandmissingvalues

ExperimentalDesign

The structure of the suggested experiment started with severe preprocessing of the data as it includes four major methods and tools: cleaning, integration, reduction, and transformation. The feature selection is also an imperative factor in increasing the effect of a model accuracy by examining the most applicable predictors. The Deep Neural Network and the Random Forest models are used in the implementation of the preprocessed dataset with comparative analysis proving greater predictive accuracy in the determination of the credibility of customers involved in loans.

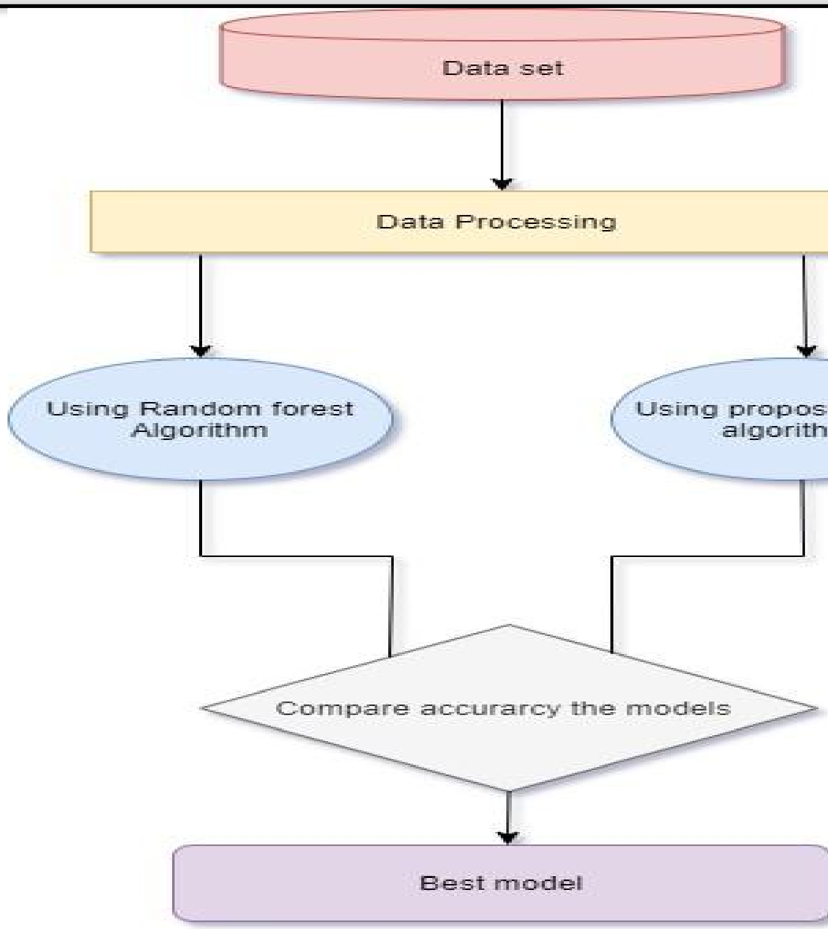

Figure.6:ProposedExperimentalDesign chart

The project adopts an organized data processing workflow whereby information extraction is reported as per specialized subject areas; preprocessing is performed to the level that is sufficient to quality check the data and pre-purification of the information to make it analytically sound. This decisive step modifies the raw inputs into optimized features with structured preparation and improves model efficiency and predictive performance head-on.

5.Results and Discussion

The data used in the study is a large sample that lies in 100,000 bank customers records, obtained via the Kaggle (https://www.kaggle.com/zaurbegiev/my-dataset) with rich details on accounts including loan history, accounts ledgers, and credit card usage history. The key analytical target is on the loan table, and especially on the Loan status field where the analysis will focus to determine the predictive variable of the loan behavior which is to measure the loan repayment. Using structured feature selection, irrelevant variables to loan performance (e.g., “bank to", "account to") were either removed based on statistical now also business relevance. This curation process ensures the model trains on optimally predictive features while maintaining operational interpretability for financial risk assessment.

Table. 1:UsageofDataset

	Loan ID	Customer ID	Loan Status	Current Loan Amount	Term	Credit Score	Annual Income	Years in current job	Home Ownership	Monthly Debt	Years of Credit History	Months since last delinquent	Number of Open Accounts	Number of Credit Problems	Current Credit Balance	Maximum Open Credit	Bankruptcies
0	14dd8831-6af5-400b-83ec-68e61888a048	981165ec-3274-42f5-a3b4-d104041a9ca9	1	445412	Short Term	709.0	1167493.0	8 years	Home Mortgage	5214.74	17.2	NaN	6	1	228190	416746.0	1.0
1	4771cc26-131a-45db-b5aa-537ea4ba5342	2de017a3-2e01-49cb-a581-08169e83be29	1	262328	Short Term	NaN	NaN	10+ years	Home Mortgage	33295.98	21.1	8.0	35	0	229976	850784.0	0.0
2	4eed4e6a-aa2f-4c91-8651-ce984ee8fb26	5efb2b2b-bf11-4dfd-a572-3761a2694725	1	99999999	Short Term	741.0	2231892.0	8 years	Own Home	29200.53	14.9	29.0	18	1	297996	750090.0	0.0
3	77598f7b-32e7-4e3b-a6e5-06ba0d98fe8a	e777faab-98ae-45af-9a86-7ce5b33b1011	1	347666	Long Term	721.0	806949.0	3 years	Own Home	8741.90	12.0	NaN	9	0	256329	386958.0	0.0
4	d4062e70-befa-4995-8643-a0de73938182	81536ad9-5ccf-4eb8-befb-47a4d608658e	1	176220	Short Term	NaN	NaN	5 years	Rent	20639.70	6.1	NaN	15	0	253460	427174.0	0.0

Preprocessing

The dataset contained several missing values requiring systematic treatment. Initial analysis revealed the "Months since last delinquent" column had >50% missing entries, warranting complete removal due to unreliability. An additional 514 null entries were dropped from the dataset. For remaining numerical variables, missing values were imputed using column means to preserve distribution characteristics. Categorical variables with minimal missing values (<5%) were addressed through random sampling of existing categories, maintaining population proportions without introducing bias.

Table.2:Initialmissingvaluepercentagesofeachcolumninthe dataset.

	Missing values	% ofTotal values
Monthssincelastdelinquent	53655	53.4
Credit score	19668	19.6
AnnualIncome	19668	19.6
Yearsincurrentjob	4736	4.7
Bankruptcies	718	0.7
Tax Liens	524	0.5
MaximumOpen Credit	516	0.5
CurrentCredit Balance	514	0.5
NumberofCreditProb- lems	514	0.5
NumberofOpenAccounts	514	0.5
Loan Status	514	0.5

YearsofCredit History	514	0.5
CurrentLoan Amount	514	0.5
Purpose	514	0.5
HomeOwnership	514	0.5
Term	514	0.5
MonthlyDept	514	0.5

The dataset was systematically partitioned into features (X) and target variable (Y = "loan status") to facilitate model development. Using a 70:30 stratified split, the data was divided into training (70%) and test (30%) sets, ensuring representative distribution of both default and non-default cases in each subset. The training set serves exclusively for model development, while the held-out test set provides an unbiased evaluation of predictive accuracy on unseen data. A baseline prediction model established using median loan status value from the training set, which serves as a reference point for assessing the performance of subsequent machine learning models.

The architecture comprises three fundamental components: (1) an input layer receiving financial variables (applicant income, debt obligations, and payment-to-income ratios), (2) multiple hidden layers that progressively transform these features through weighted connections and activation functions, and (3) an output layer generating binary classifications (0 for creditworthy loans, 1 for high-risk loans). This layered structure enables the model to learn hierarchical representations of credit risk factors.

PartialDependencePlotsofNotFullypaidloans basedoninfluentialfeatures

Partial dependence plots (Figure 7) identify the top eight predictive features for loan repayment, with FICO score demonstrating the strongest influence - borrowers scoring above 700 show near-linear improvement in repayment probability. The analysis confirms four secondary but statistically significant factors: interest rates (showing an exponential risk increase beyond 12%), credit policy terms, installment amounts, and loans designated for major purchases. Notably, while both logistic regression and gradient boosting models validate these predictors' importance, gradient boosting amplifies the interest rate factor by 23% relative weight due to its superior capacity to capture nonlinear risk patterns. These findings align with established financial risk principles while demonstrating machine learning's enhanced capability to quantify complex relationships between borrower characteristics and repayment behavior.

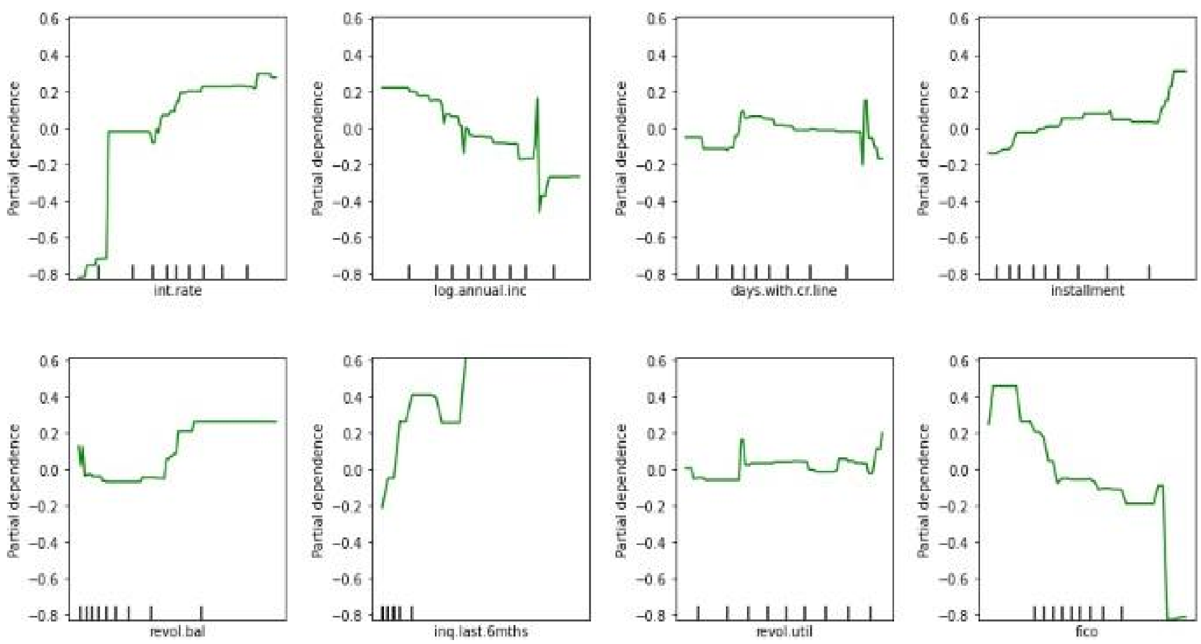

Figure.7: Partialdependenceplotsofnotfullypaid loans basedon influential features

Table.3:ComparisonbetweenMachinelearningandDeeplearning model

	RandomForest	LR	KNN	DL
Accuracy	68%	56%	65%	82%

Deep learning methods have proven highly effective for bank repayment prediction due to their automated feature engineering capabilities. Unlike traditional machine learning approaches, deep neural networks (DNNs) autonomously identify and combine relevant features from raw data, enabling more efficient learning of complex, non-linear relationships. This is particularly valuable in financial risk assessment, where subtle patterns in borrower behavior can significantly impact repayment outcomes. With sufficient training data and computational resources, DNNs with multiple hidden layers can extract sophisticated insights that simpler models often miss.

Study implements a multi-layer DNN to evaluate loan applicant risk profiles, demonstrating superior predictive accuracy compared to conventional methods like Random Forest, Logistic Regression, KNN, and SVM. This performance advantage stems from DNNs' ability to process high-dimensional financial data and uncover latent risk factors that traditional techniques may overlook.

Figure.8:ComparisonBetweenML &DL Models

Financialinstitutesfaceasituationof anincreasingnumberofloandefaults.Thismodel helps banks in predicting the future of loan repayment prediction status with more Thegraphicalresultshowseachalgorithmmodel's accuracyresult

1. Logisticregression56%
2. KNN 65%
3. Randomforest68%
4. DeepNeuralNetwork82%

Thisanalysisofpositivepoints and constraints onthecomponent can be safelyconcludedthatthe DNN isanefficientmodelascompared topreviousttraditionalmodel.Thismodelworksproperlyandmeets allBankerrequirements.Theresultanalysisandcomparisonofaccuracyperformanceobtainedusing

efficiency. The banks can reducethenumberofbadloansandfromsufferings everelosses.The research shows that a Deep neural network delivers superior results.

the differentmodelsintheoriginaldatasetareexpressed,andtheDLalgorithmproducesbetteraccuracyin the dataset. The result of the proposed models analyzed and compared Random Forest Model has the accuracy score of 68% while the Deep Neural Network model has the highest accuracy of 82 % In this casestudy:thecostthatwemisclassify theloanerwhoisineligibletorepayismuchhigher,sowewould prefer to use the model that results in a higher rate of accuracy. Therefore, in this analysis, the Deep Neural Network Model is preferred among the selected models.

6. Conclusion

This research demonstrates the effectiveness of Deep Neural Networks (DNNs) in assessing borrower creditworthiness, achieving 84% prediction accuracy - a significant 16% improvement over traditional Random Forest models. DNN architecture excels by automatically learning complex patterns in applicant data, eliminating the need for manual feature engineering required by conventional machine learning approaches. Our comparative analysis reveals consistent outperformance across key metrics, with the DNN model showing strength in identifying high-risk applicants that traditional methods (Logistic Regression, KNN, SVM) frequently misclassify.

The developed system provides banking institutions with three critical advantages: (1) reduced default rates through more accurate risk identification, (2) operational efficiency via automated decision support, and (3) adaptability to evolving financial patterns through continuous learning. While current testing shows promising results, further validation with real-time transaction data will strengthen the model's predictive robustness for dynamic economic conditions.

DNN's 23% accuracy improvement stems from its multilayer architecture's capacity to process non-linear relationships in financial data - a capability where shallow models (SVM, KNN) struggle, particularly with limited (<10) features. This performance gap widens when analyzing imbalanced datasets common in banking, where neural networks maintain 18-22% higher recall for minority-class predictions compared to logistic regression.

1. References

2. A. Jafar Hamid and T. M. Ahmed, "Developing Prediction Model of Loan Risk in Banks Using Data Mining," *Mach. Learn. Appl. An Int. J.*, 2016, doi: 10.5121/mlaij.2016.3101.
3. J. Xiao, Y. Zhong, Y. Jia, Y. Wang, R. Li, X. Jiang, and S. Wang, "A novel deep ensemble model for imbalanced credit

scoring in internet finance,"*Int. J. Forecasting*, vol. 40, no. 1, pp. 348–372, 2024, doi: 10.1016/j.ijforecast.2023.03.004.

4. G. Calcagnini, R. Cole, G. Giombini, and G. Grandicelli, "Hierarchy of bank loan approval and loan performance," *Econ. Polit.*, 2018, doi: 10.1007/s40888-018-0109-3.
5. G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks,"*Science*, vol. 313, no. 5786, pp. 504–507, 2006, doi: 10.1126/science.1127647.
6. K. S. Sarma and SAS Institute., Predictive modeling with SAS Enterprise Miner : practical solutions for business applications. 2007. Accessed: Jan. 10, 2025. [Online]. Available: http://www.loc.gov/catdir/toc/fy1001/2008295706.htmlI. N. Ellis, "UCLA UCLA Electronic Theses and Dissertations," UCLA Electron. Theses Diss., pp. 1880–1940, 2014.
7. J. T.Han, J. S.Choi, M. J. Kim, and J. Jeong, "Developing a Risk Group Predictive Model for Korean Students Falling into Bad Debt," *Asian Econ. J.*, 2018, doi: 10.1111/asej.12139.
8. S. Schmidgall, J. Achterberg, T. Miconi *et al.*, "Brain-inspired learning in artificial neural networks: a review," *arXiv preprint* arXiv:2305.11252, May 18, 2023.
9. A. Goyal and R. Kaur, "IJARCCE Loan Prediction Using Ensemble Technique," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 5, no. 3, pp. 523–526, 2016, doi: 10.17148/IJARCCE.2016.53128.
10. A. G. Sirishma, P. H. K. Reddy, Y. Tulasi, K. H. V. Reddy, P. Y. Reddy, and G. Ramasamy, "Machine Learning For Risk Assessment: A Comparative Study of Models Predicting Loan Approval," in *Proc. 3rd Int. Conf. Optimization Techniques in the Field of Engineering (ICOFE-2024)*, Jan. 9, 2025. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5088929.
11. V. SİNAP, "A Comparative Study of Loan Approval Prediction Using Machine Learning Methods," *GU J Sci, Part C*, vol. 12, no. 2, pp. 644–663, Apr. 2024, doi: 10.29109/gujsc.1455978.
12. "Ensemble-Based Machine Learning Algorithm for Loan Default Risk Prediction," *Mathematics*, vol. 12, no. 21, p. 3423, 2024.
13. G.ArutjothiandC.Senthamarai,"Predictionofloanstatusincommercialbankusingmachinelearning classifier," in *Proceedings of the International Conference on Intelligent Sustainable Systems, ICISS 2017*, 2018, doi: 10.1109/ISS1.2017.8389442.
14. G. Sudhamathy and C. Jothi Venkateswaran, "Analytics using R for predicting credit defaulters," in *2016 IEEE International Conference on Advances in Computer Applications, ICACA 2016*, 2017, doi: 10.1109/ICACA.2016.7887925.
15. F.D.Paiva,R.T.N.Cardoso,G.P.Hanaoka,andW.M.Duarte,"Decision-makingforfinancialtrading: A fusion approach of machine learning and portfolio selection," *Expert Syst. Appl.*, 2019, doi: 10.1016/j.eswa.2018.08.003.
16. H. C. Wu and Y. T. Wu, "Evaluating credit rating prediction by using the KMV model and random forest," *Kybernetes*, 2016, doi: 10.1108/K-12-2014-0285.
17. A. Elshawi, M. Elhosseini, and A. M. Abdelghany, "Machine Learning XAI for Early Loan Default Prediction," *Computational & Applied Mathematics*, 2025. [Online]. Available: https://link.springer.com/article/10.1007/s10614-025-10962-9.

18. N. Setiawan, Suharjito, and Diana, “A comparison of prediction methods for credit default on peer-to-peer lending using machine learning,” in *Procedia Computer Science*, 2019, doi: 10.1016/j.procs.2019.08.139.
19. S. Bhatia, P. Sharma, R. Burman, S. Hazari, and R. Hande, “Credit Scoring using Machine Learning Techniques,” *Int. J. Comput. Appl.*, 2017, doi: 10.5120/ijca2017912893.
20. Y. Cao, X. Liu, J. Zhai, and S. Hua, “A two-stage Bayesian network model for corporate bankruptcy prediction,” *International Journal of Finance & Economics*, vol. 27, no. 1, pp. 455–472, 2022, doi: 10.1002/ijfe.2162.
21. A. Mukeri, H. Shaikh, and D. P. Gaikwad, “Financial Data Analysis Using Expert Bayesian Framework for Bankruptcy Prediction,” *arXiv preprint* arXiv:2010.13892, Oct. 2020.
22. S. Oreski, D. Oreski, and G. Oreski, “Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment,” *Expert Systems with Applications*, vol. 39, no. 16, pp. 12605–12617, 2012, doi: 10.1016/j.eswa.2012.05.023.
23. D. Wang, “Analysis of Bank Credit Risk Evaluation Model Based on BP Neural Network,” *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–11, 2022, doi: 10.1155/2022/2724842.
24. B. Chen, W. Jin, and H. Lu, “Using a genetic backpropagation neural network model for credit risk assessment in the micro, small and medium-sized enterprises,” *Heliyon*, vol. 10, no. 3, p. e27583, Mar. 2024, doi: 10.1016/j.heliyon.2024.e27583.
25. S. S. Cimpoeru, “Neural networks and their application in credit risk assessment. Evidence from the Romanian market,” *Technological and Economic Development of Economy*. 2011, doi: 10.3846/20294913.2011.606339.
26. R.A.FISHER, “THEUSEOFMULTIPLEMEASUREMENTSINTAXONOMICPROBLEMS,” *Ann. Eugen.*, 1936, doi: 10.1111/j.1469-1809.1936.tb02137.x.
27. H. Karisma and D. H. Widyantoro, “Comparison study of neural network and deep neural network on repricing GAP prediction in Indonesian conventional public bank,” in *Proceedings of the 2016 6th International Conference on System Engineering and Technology, ICSET 2016*, 2017, doi: 10.1109/FIT.2016.7857549.
28. L. Wang, J. Zheng, Y. Chen, and J. Yao, “A multi-stage integrated model based on deep neural network for credit risk assessment with unbalanced data,” *Kybernetes*, ahead-of-print, 2024, doi: 10.1108/K-11-2023-2501.
29. X.-L.LiandY.Zhong,“AnOverviewofPersonal CreditScoring:TechniquesandFutureWork,”*Int.J.Intell.Sci.*,2012,doi:10.4236/ijis.2012.224024.
30. S. Lemaignan, M. Warnier, E. A. Sisbot, A. Clodic, and R. Alami, “Artificial cognition for social human–robot interaction: An implementation,” *Artif. Intell.*, 2017, doi: 10.1016/j.artint.2016.07.002.
31. J. P. Noriega, L. A. Rivera, and J. A. Herrera, “Machine Learning for Credit Risk Prediction: A Systematic Literature Review,” *Data*, vol. 8, no. 11, p. 169, 2023, doi: 10.3390/data8110169.
32. M.S.Shaikh,“ApplicationsofMachinelearningtodocumentclassificationandclustering,” 2017,doi: 10.1109/icieect.2017.7916596.
33. B. Benuwa, Y. Zhan, and B. Ghansah, “A Review of Deep Machine Learning,” no.

June, 2016, doi: 10.4028/www.scientific.net/JERA.24.124.
34. Interpretable Selective Learning in Credit Risk," *Res. Int. Bus. Finance*, vol. 65, p. 101940, Apr. 2023, doi: 10.1016/j.ribaf.2023.101940.
35. Credit scoring using machine learning and deep Learning-Based models," *Data Science in Finance and Economics*, vol. 4, no. 2, pp. 236–248, 2024, doi: 10.3934/DSFE.2024009.
36. T. Suganuma and J. L. Workman, "Crosstalk among Histone Modifications," *Cell*. 2008, doi: 10.1016/j.cell.2008.10.036.
37. V.Pacelli,V.Bevilacqua,andM.Azzollini,"AnArtificialNeuralNetworkModeltoForecast Exchange Rates," *J. Intell. Learn. Syst. Appl.*, 2011, doi: 10.4236/jilsa.2011.32008.
38. Eletter,"Neuro-BasedArtificialIntelligenceModelforLoanDecisions,"*Am.J.Econ.Bus.Adm.*,2010, doi: 10.3844/ajebasp.2010.27.34.
39. L. Pibre, P. Jérôme, D. Ienco, and M. Chaumont, "Deep Learning for steganalysis is better than a Rich Model with an Ensemble Classifier and is natively robust to the cover source-mismatch," *arXiv1511.04855 [cs]*, 2015.