Enhancing Credit Risk Assessment using Interpretable Machine Learning Techniques

MScResearch Project: Enhancing Credit Risk Assessment using Interpretable Machine Learning Techniques
Program Name: MSc in Artificial Intelligence for Business

Mohsin Sadaqat Ali
Student ID:X23397071

School of Computing
National College of Ireland

Supervisor: Mr. Syed Muslim Jameel

"National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ……Mohsin Sadaqat Ali.……………………………………………………………………

Student ID: ……X23397071…………………………………………………………………………………

Programme: ……MSc Artificial Intelligence for Business… **Year:** …2024-25…………..

Module: ……Practicum II……………………………………………………………………………….………

Lecturer: ……Mr.Mohit Garg……………………………………………………………………….………

Submission Due Date: ……15-Sep-2025……………………………………………………………………….………

Project Title: ……Enhancing Credit Risk Assessment using Interpretable Machine Learning Techniques

Word Count: ……6253……………… **Page Count**…………26……………………………………….……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……………Mohsin Sadaqat Ali…………………………………………………

Date: ……………14-Sep-2025……………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office. "

Office Use Only

Signature:	
Date:	
Penalty Applied (if applicable):	

“Enhancing Credit Risk Assessment using Interpretable Machine Learning Techniques”

Mohsin Sadaqat Ali
X23397071

Abstract

The ability to predict credit risk is one of the most vital tasks in the financial sector; it allows lending companies to assess the likelihood that a borrower defaults on a loan. Traditional machine learning classifiers are commonly used with this purpose and however they are mostly faced with the problem of dealing with skewed data sets and have no interpretability thus making the decision process difficult for financial institutions. In this study we consider the use of ensemble classifiers and the Synthetic Minority Over-sampling Edited Nearest Neighbor (SMOTE-ENN) to predict credit risk whereby the aim is to improve classification of different credit risk classes. The family of ensemble classifiers includes Random Forest, adaptive boosting (AdaBoost), extreme gradient boosting (XGBoost) and light gradient boosting machine (LightGBM). The study addresses the issue of class imbalance using ensemble classifiers and the SMOTE-ENN technique with the addition of Shapley Additive Explanations (SHAP) as interpretability of models. The results of this experiment showed that the suggested technique improved the classification in a given task. The identification of defaulters has been done by carrying out a comparative analysis of the optimised supervised learning techniques which involve; Random forest; Extreme Gradient Boosting; Support Vector Machine and Logistic Regression. The dimensionality reduction, using Recursive Feature Elimination together with Cross-Validation and Principal Component Analysis has been applied. Each model has been evaluated with the help of such metrics as the F1 score, the AUC score, prediction accuracy, precision, and recall. The combination of the calibrated Support Vector machine with Recursive Feature Elimination and Cross-Validation has implied essential potential with regards to loan defaulters identification. The proposed technique could assist the financial institution in the exact identification of loan defaulters and avoidance of any further losses. Ensemble models offer certain advantages, as they have better predictions and better stability, which makes them specifically fit within this application. The combination of predictions in different models is usually less volatile than the result of

using a single model and outperforms the use of other types of models and techniques like XGBoost, SVMs, and logistic regression [1].

1 Introduction

Machine Learning allows computers to replicate human behaviour and learn more through data collected in the real life interaction and observation. Machine learning has gained immense momentum in the current environment and its application is heavily in use in many areas around the world. These are two common ones, such as: finance and banking. Machine learning networks have contributed to a very high effectiveness and precision in determining the pattern of fraud in credit cards and loan defaults. The need to have the accurate credit risk assessment model has never been more. Defaults in loans as well the charged-off loans have never experienced such high rates. Assets are freeze, transactions are suspended and mutilations are seen in financial institutions like banks where they lose heavily [2]. It is alleged that in China alone about 9 million loan defaulters existed in the year 2018. In Bangladesh, there are some 300 000 of them since 2011, which means that their number has more than tripled. Analysts have revealed that the current state of affairs in Bangladesh would hamper corporate development and hinder the implementation of different strategies to create job opportunities among the general population. Compared to the existing models, this research study will improve them by use of sophisticated types of ensemble classes which include the random forest, AdaBoost, XGBoost, and LightGBM, since, through integration with SMOTE-ENN strategy, they are efficient in addressing the issue of the data imbalance that affects typical single classes [3]. In addition, the use of SHAP to interpret models provides a deep understanding towards the aspect of decision-making, thus, increases the openness and feasibility of the predictive models within real-life applications.

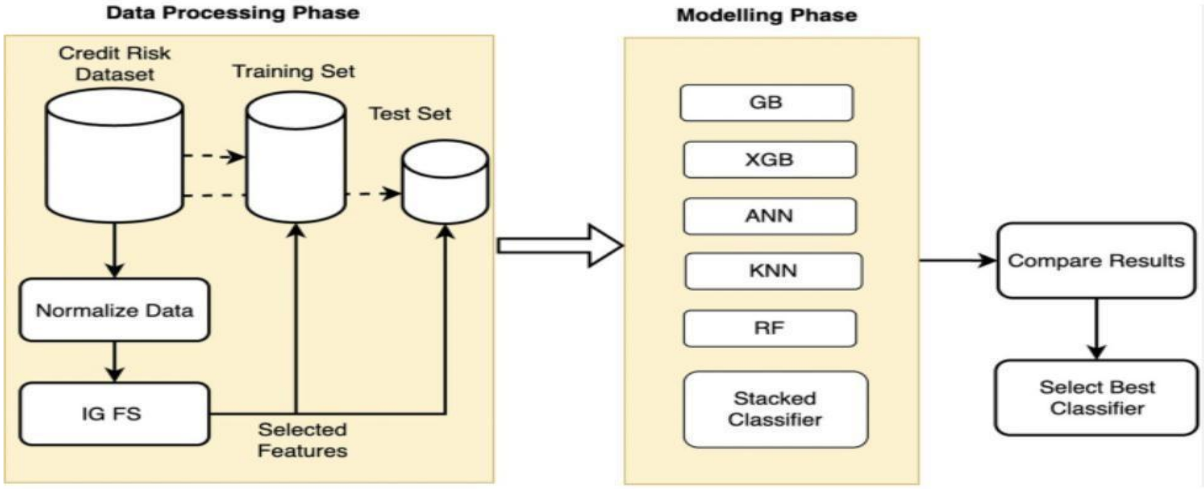


Figure 1:

2 Related Work

In the and finance and banking industry, the credit risk presents the most crucial issue that requires careful evaluation in order to determine the possibility of loan repayment. Carefulness and caution are needed to prevent possible mishaps. Banks should analyze the present and past financial status of the borrower. Due to this, scholars have been long involved in predicting the credit risk of banking organizations. The literature review takes two components: the first component explaining in detail the unique machine learning models adopted to carry out the credit risk prediction and the second component indicating the efficacy of the hybrid strategies that are made up of an arrangement of the techniques to predict the credit risk in the banking institutions. There have been a significant number of researchers who have exploited various machine learning techniques in credit risk prediction. According to Moscato et al., this study proposes undertaking a comparative study between the most common models of credit risk scores, which are applied to gauge pay-off turns in a P2P setting. In addition, the authors resolve the problem of class imbalance by a set of popular classifiers found in literature, each one of them being grounded on a distinct sampling technique [4]. It employed an experimentalized research study on a set of an open social lending site (Lending Club) comprised of 877,956 samples in the same. Finally, interpretability of three machine learning models, i.e., Random Forest (Rf), Multilayer Perceptron (MLP), and Logistic Regression, have been examined via the assistance of various explainable Artificial Intelligence (XAI) technologies. The RF performance outpaced the previous algorithms, as it performed with an accuracy of 69.2% accuracy.

2.1 Research Papers

Research into credit risk scoring techniques used to evaluate risk on small and micro firms principally fall into three broad buckets: expert analysis or eventual score analysis, statistical analysis, and machine learning. Professional: This technique greatly relies on the knowledge/judgement of the assessor. It involves determination of whether various businesses are viable or not through the skills of the evaluator. Due to its subjective nature, it may happen that the evaluations of the same object may strongly differ depending on the evaluator. It is a source of inconsistency in search of evaluation impartiality. Statistical approaches: A common way around the insurmountable limitations of expert analysis is through the use of statistical approaches by available scholars. The most common techniques are discriminant analysis, and logistic regression, probit models. The first applications include discriminant analysis model developed by Fisher and Z-score models developed by Altman. Appearance of logistic regression in prediction of default. The comparative advantage of logistic models compared with discriminant analysis lies in the accuracy of prediction mainly due to the relatively less strict assumptions in the models. The combination of correlation analysis and Probit regression successfully differentiates between the consumers who default and those who do not.

Empirical studies based on a unique sample of French SMEs in 2012 2018 imply greater accuracy of Probit than the logistic model. However, the linearity boundaries of assessment metrics and the presumptions concerning the typical dispersion of these metrics pose inherent barriers in these statistical techniques and therefore limit their utilization in the corporate credit risk analysis. Machine learning: The digital revolution in the first decade of the 21 st century brought an absolute shift in the credit assessment to intelligence-led mechanisms. Following the use and study of machine learning in engineering, some scholars have decided to apply the same method in credit assessment of small and micro firms. It is cross-disciplinary and may be considered a radical step towards improving the accuracy and efficiency of the business area in terms of credit evaluation using very complicated instruments of analysis. explores the use of Support Vector Machines and Neural Networks in terms of their good results in complex situations in finance. At the same time, combines the SVM with the Logistic regression, which focuses on significant increases in the accuracy in predicted credit, which would be attributed to SME credit administration. Employs the XGBoost model in credit risk assessments of corporations, which allowed a significant growth in the accuracy of audit opinion predictions, which is an impressive contribution to the sphere. An enhanced XGBoost algorithm is proposed, showcasing its superiority in predictive accuracy, significantly surpassing conventional methods such as discerning analysis and logistic regression. With the spread of machine learning models, a growing body of academic research is being directed at integrative studies in which these models are combined with promising new technologies such as the Internet of Things and cloud computing. What tendency furthers the innovativeness in the research analysis of credits of small and micro firms. Wang et al. evaluate machine learning algorithms, including Naive Bayes, Logistic Regression, Random Forest, Decision Tree, and K-Nearest neighbours or KNN Classifier, five of the most popular models exploited in credit scoring. Although there are limitations to each classifier, there is a necessity to ensure that certain choice is preferred. Testing outcomes show that Random Forest exceeds other strategies in accuracy with 96.53%. Li [5] employed the technique of boosting in recognising trustworthy customers and unscrupulous customers, particularly by using XGBoost. Extensive research evaluates the performance of the XGBoost against logistic regression. The outcomes approve the superior effectiveness of XGBoost algorithm. The research by Putri et al. shares a similar structure because it is also devoted to the application of machine learning methodology addressed to the problem of credit risk analysis. Data classification is carried out in the form of Support Vector machines (SVM) that use various kernels. The data consisting of a set of 610 samples was selected in a random sample of one of the local banks in 2015-2018. The Support Vector Machine (SVM) was run using a linear, a polynomial, a Radial Basis Function (RBF) and a sigmoid with a maximum accuracy of 95.08 % and a polynomial that followed in second place with area under the curve (AUC) of 0.9419 [6]. Aneceto et al. are addressing the issue of the effectiveness of different machine learning techniques to classify the trustworthiness of a borrower with a loan data taken of a Brazilian bank in this paper. Different varieties of machine learning are used in the consumer prediction: namely, there is Logistic Regression, Support Vector Machines, Decision Trees, Bagging, AdaBoost, and Random Forest. These results indicate that the models that performed the best are Random Forest and Adaboost and the accuracy level of Random Forest model is 63.8 and other accuracy level is 63.5 of the Adaboost model. However, the use of both linear and nonlinear kernels results in poorer results of the SVM models [7].

Compared to methods, Sharifi et al. have estimated the C5 risk of a credit decision tree according to a unique approach involving the combination of artificial neural networks and an optimization algorithm, i.e. a modified form of the Owl search algorithm (IOSA). The IOSA affects the decision tree to boost the weighting of the neural network. When the problem data and weights are generated, they are characterized into the main network. Its competitor strongly falls short as the algorithm has an accuracy level of 96%. The precision and recall here achieved using a total of 618 positive samples was 0.885 and 0.83, respectively. The suggested strategy is more accurate and reliable as compared to other procedures. Hybrid models also show good performance compared with individual machine learning techniques (Machado and Karra). The authors employed the hybrid machine learning methods that combine the unstructured and the supervised approaches in predicting the creditworthiness of commercial clients. Various approaches are used to determine the superiority of hybrid processes against the effectiveness of individual supervised machine processes learning. With reference to the prediction of credit scores of corporate clients, evaluative results revealed that the hybrid models have higher scores compared to individual models. Moreover, the researchers found that when the historical data on credit scores is input, the predictive performances of the hybrid models are improved [9].

Wasana et al. use an analogous approach, although, in contrast to overemphasizing the application of machine learning techniques to the problem of credit risk analysis, they also pay attention to the application of such an approach. Support Vector Machines (SVM) divide the information with the help of different types of kernels. There are 610 samples in the data set which are selected randomly in the privately owned bank of 2015-2018.

Support Vector Machine (SVM) was tested using linear, polynomial kernel, Radial Basis Function (RBF) kernel and sigmoid kernel and the best known rate is 95.08% when a polynomial kernel is used where it yielded the area under the curve (AUC) of 0.9419 [6]. In this article, authored by Aneceto et al., the authors discuss the usefulness of different machine learning classification algorithms in the estimation of whether the trustworthiness of the borrowers given the utilization of the loan data of a Brazilian bank. The prediction of the consumers is done using different types of machine learning such as Logistic Regression, Support Vector Machines, Decision Tree, Bagging, AdaBoost and Random Forest. According to the results, one may say that Random Forest and Adaboost can be regarded as the most appropriate models with the numbers of 63.8% and 63.5% as the accuracy rates, respectively. Nonetheless, linear and nonlinear kernels give unsatisfactory results as far as SVM models are concerned [7].

Chi et al. also proposed the use of the hybrid models that combine logistic regression and multilayer perceptron (MLP) by utilizing both traditional and novel approaches in predicting credit risk. Moreover, it is possible to formulate and test 16 different hybrid models incorporating different machine learning elements, including logistic regression, discriminant analysis, and decision trees, with four different types of neural networks, namely, adaptive neuro-fuzzy inference systems, deep neural networks, radial basis function networks, and multilayer perceptron's. The five various credit score databases were used in the verification of the accuracy of the classifier [11].

Interpretability in machine learning has garnered considerable interest, resulting in the implementation of SHapley Additive Explanations to elucidate individual predictions across various applications. SHAP values, derived from cooperative game theory, present a mathematically precise method to quantify each feature's contribution to a specific prediction, hence enhancing transparency and understanding of model decisions [12].

Wang et al. utilized SHAP values to elucidate the predictions of machine learning models used for forecasting student loan defaults. The use of the SHAP technique yielded several insights, including the impact of college entrance examination scores, academic achievement, and the quantity of scholarships on the probability of student loan default [13].

Relevant works carried out in the field show that, on the one hand, advanced machine learning algorithms may drastically improve forecast accuracy, but, on the other hand, they also often undermine transparency. Such a trade-off makes it necessary to continue working aimed at striking the balance between complexity and interpretability to ensure that models are not only highly performing but also understandable by stakeholders. The studies reviewed and the problems pointed out above indicate the need to assign a multi-dimensional approach to credit risk assessment. This paper provides a methodology that integrates the advantages of ensemble classifiers. This has resampling approaches and interpretability methods to develop robust and comprehensible models [14].

The literature research indicates that the current work is deficient in picking pertinent features for prediction, which would enhance the performance of the prediction model. Additionally, numerous machines learning methods, including neural network-based algorithms such as multilayer perceptron's (MLP) and deep neural networks (DNN), boosting the algorithms such as XGBoost, AdaBoost, Furthermore tree-based algorithms like decision trees and random forests, demonstrate efficacy in credit risk prediction without the application of feature selection technique. Among all of these, Tree-based algorithms and boosting, particularly

Random Forest, surpass most of the research [15]. Consequently, the suggested methodology would employ a feature selection technique to identify pertinent characteristics, utilising machine learning-based approaches such as Random Forest, Neural Networks, and Support Vector Machines, along with boosting algorithms such as XGBoost and AdaBoost.
All categories of algorithms will be.

2.2 Research Paper Table

Paper Name / Description	**Author(s)**	**Model(s) Used**
Comparative Assessment of Credit Risk Models	Wang et al. (2020)	Naive Bayes, Logistic Regression, Random Forest, Decision Tree, KNN
Credit Risk Evaluation using Boosting	Li (year not mentioned)	XGBoost, Logistic Regression
Credit Risk Analysis using SVM on 610 Bank Samples	Putri, Fatekurohman, Tirta (2021)	SVM (Linear, Polynomial, RBF, Sigmoid kernels)
ML Predictivity for Consumer Creditworthiness	Aniceto et al.	Logistic Regression, SVM, Decision Trees, Bagging, AdaBoost, Random Forest
C5 Decision Tree Optimized by IOSA	Sharifi et al.	Artificial Neural Network + Improved Owl Search Algorithm (IOSA)
Hybrid ML Models for Commercial Credit Score Prediction	Machado and Karra	Hybrid Models (supervised + unsupervised ML)
Hybrid Soft Computing + Expert Knowledge + Optimization	Lappas and Yannacopoulos	Unsupervised Clustering, AHP, Evolutionary Optimization
Hybrid Logistic Regression + MLP + Other NN Models	Chi et al.	Logistic Regression, MLP, DNN, RBFN, Neuro-Fuzzy Inference
SHAP Interpretation of Credit Default ML Models	Wang et al.	SHAP with ML Models (Random Forest, Logistic Regression, XGBoost implied)

3 Research Methodology

Data Acquisition: The data set used in carrying out the experimental work is an open-sourced data set referred to as Credit Risk Prediction, registered with Kaggle [16]. The peculiarities of the dataset are mentioned in table 1. It has 11 input values, also referred to as characteristics and only one output, which is loan_status. Besides, it possesses 32,581 cases with one percent of missing data. The number of missing values in person_emp_length is associated with 712 entries whereas in loan_int_rate is 2,483.

Data Preprocessing:

The subsequent preprocessing processes executed for dataset cleansing are outlined below.

1) Addressing Missing data.

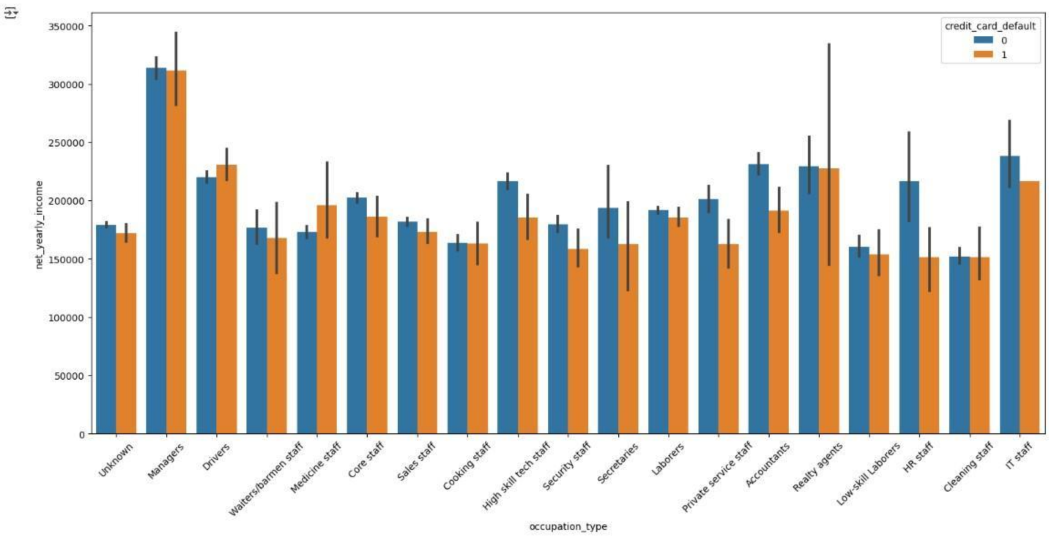
In regard to all features with an absent value, the average value of the feature is utilised to replace the missing value in the dataset.

2) Selection of Features

The level of importance of a feature is determined using Information Gain. Entropy measures the impurity of a set of data; it is used to determine the effectiveness of one of the attributes to categorise the training set. The Information gain is expected to decrease of entropy caused by the division of the dataset according to this attribute.

3) Standardization

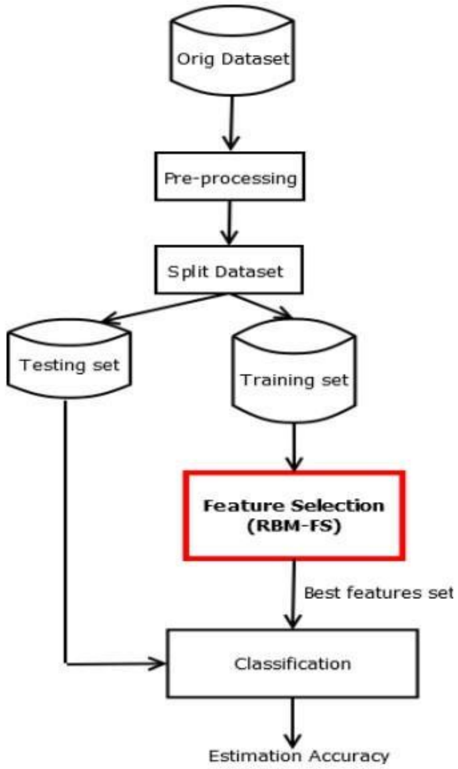
The seven numeric features in the given dataset have large variations and normalisation is used to make them more accurate by limiting their values to within [0,1] range. Normalisation tries to normalize all numerical columns of a dataset upon a common scale and avoid the disappearance of the original ranges of values. The chosen range of all features makes them fall within the scale of [0,1].



Machine Learning Algorithm:

The pre-processed data feeds to the set of machine learning models,namely multiple boosting algorithms which include Adaboost, Extreme Gradient Boosting (Xgboost), Extreme Gradient Random Forest Boosting (XgRFboost) Catboost, Random Forest and Neural Networks (NN) [17]. Random Forest and boosting algorithms will be implemented as ensemble learning algorithms, and they are selected because of the demonstrated effectiveness in case of credit risk prediction, as the literature suggests. Moreover, NN is selected to compare with the ensemble learning methods. I have employed the following machine learning framework.

1. Logistic Regression
2. Random Forest
3. xgBoost
4. LightGBM
5. KNN
6. CatBoost



4 Design Specification

1) Boosting algorithms

Enhancing has turned out to be a mainstream method of taking care of the Binary classification issue; it is one of the approaches of care that is in a group. The techniques promote predictive accuracy because most weak learners are converted into robust. The working of boosting methods is that after a model is obtained through a training dataset, a second model is trained to improve on the drawbacks of the first model. The process will be repeated until the dataset can be predicted accurately [18].

a) AdaBoost

AdaBoost or AdaBoosting (The shortened form of Adaptive Boosting) is the initial Boosting ensemble model. The optimization of its parameters is carried out by the algorithm over time spent on the number of iterations, relying on newly added input. It is done in an iterative fashion until convergence to the weights corresponding to the reweighting of the data after the first and after the final summation. Such increased method was used and showed better results compared to that of classification using single tree or other basis individual learners that used basic classification trees or stumps [19].

b) Gradient boosting methods.

Gradient Boosting is a machine learning technique and one of the best types of classifications and regression, combining the advantages of the gradient descent and boosting techniques. The term gradient is a projection that under prime implies that there may be more than a single derivative to a single-valued function. Gradient Boosting has three simple components that include an additive model, a loss function, and a weak learner. The model generalises boosting algorithms in that it allows optimisation of an arbitrary loss, and offers an immediate conceptualisation of boosting algorithms as numerical optimisation on a set of functions [20].

XGBoost is an ensemble learning algorithm which optimizes the efficiency of the computation and maximizes gradient-boosting algorithm performance. XGBoost is an upgraded gradient boosting procedure. Firstly, it uses regularization in order to overcome the overfitting problem. Secondly, it performs efficiency improvement by use of parallel execution, therefore, boosting the processing time. Finally, it is able to greatly decrease the running time because it uses the maximum depth of the decision tree as a standard of pruning [21].

XgRFboost: XGBoost is typically a model that trains gradient-boosted decision trees. Random Forests follow a training process distinct to that of gradient-boosted decision trees yet include similar methods of (representing and) inference. As a XGBoost foundation, one can train an independent random forest, as well as a gradient-boosting model that is constituted on the basis of a random forest. XgRFboost thus, is a classification tool in addition to Xgboost which utilizes independent random forest. A boosting library is called CatBoost, which is an abbreviation of categorical boosting. The application of CatBoost has been limited to regression and classification, but also it has used in ranking, recommendation systems, predictive and virtual assistant. Compared to other gradient boosting procedures, CatBoost obtains stunning results and can be successfully compared to other well-known machine learning algorithms. CatBoost has the ability to induce the categorical set to be represented in numeric form without the need of subsequent split of the information. In CatBoost, large numbers of statistics of both categorical values and numerical values with various categorical features or combinations of catergories and numerical values are utilized so as to create the numeric kind of representations of categorical values. Additionally, the overall need to tune hyperparameters is less, which makes it less probable to overfit, and therefore leads to less specific models. CatBoost involves the possibility of many parameters to tune, which include the number of trees, learning rate, regularization, the depth of the trees, the size of the fold, bagging temperature, and many more, making it more effective than most deep learning methods but with limited data [22]

5 Implementation

Substantial trends were realized in the algorithms that were evaluated. The decision trees and the random forest proved to be the most robust models since both maintained high accuracy levels across the real and the normalized data. It implies that they are skillful at credit risk assessment work and might be a viable solution to financial organizations.

Logistic regression, though very precise on the initial set, showed a small drop in accuracy after the normalization of the set of data. This performance degradation underlines the importance of measuring the effect of data preprocessing on model quality.

The results of support-vector machines (SVM) were consistent in both datasets, which means that SVM is less prone to the normalization of dataset. The ability of SVM to maintain the accuracy regardless of the amount of data makes it a trustworthy choice of credit risk assessment [23].

K-nearest neighbours (KNN), on the other hand, made a sharp decline in terms of accuracy when applied to the normalized dataset, showing its vulnerability to data scaling. This expounds the significance of careful consideration of feature scaling in adoption of KNN in credit risk assessment [24].

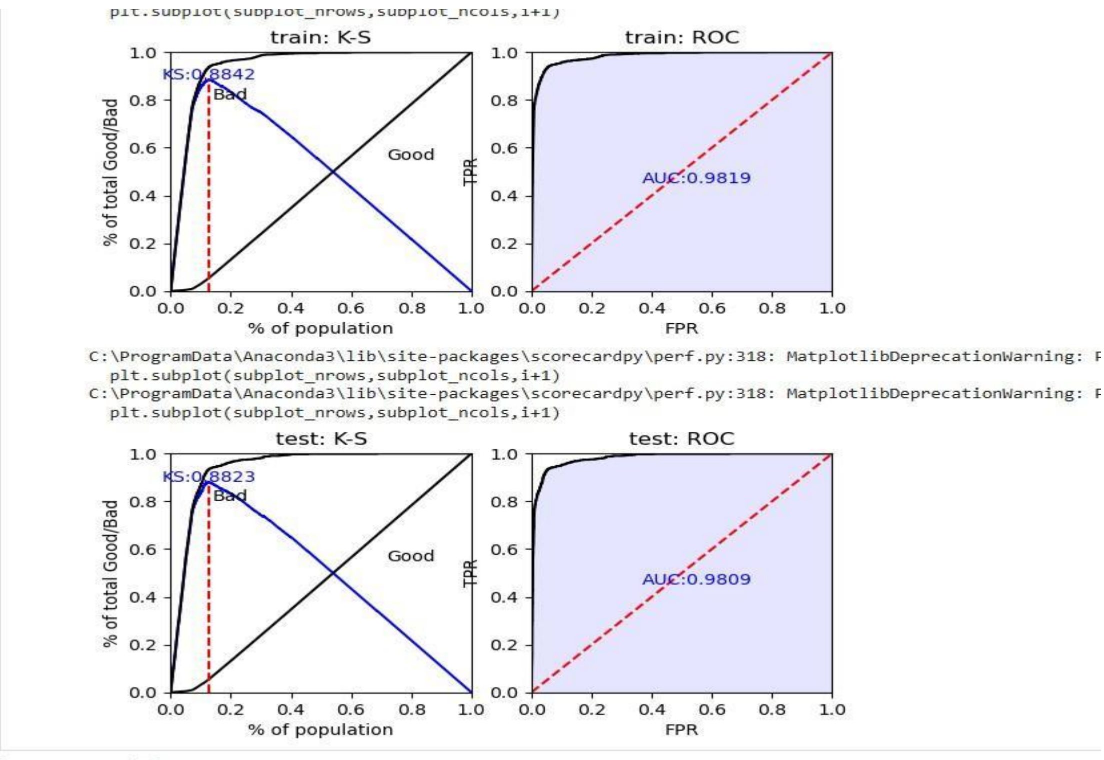
Affecting the accuracy of credit risk assessment models significantly is the choice on which algorithm of machine learning to employ. Researchers of the financial sector and practitioners must consider interpretability, data preprocessing robustness, and scalability-sensitive aspect when selecting the most appropriate method that fits their problem. Further investigation on the trade-offs between the model complexity and explainability would also result in a deeper understanding of model effectiveness in credit risk modeling.

6 Evaluation

In this research, they used six models. Random Forest, Decision Tree, Logistic Regression, KNN, LightGBM and XGBoost, to predict credit risk and evaluated their models in terms of several metrics. KNN training was 99.37% with test accuracy of 96.81%. It performed well in the low-risk loans but not that well in the minority high-risk category (0.77). The accuracy of LightGBM was 99.11 during training and 96.68 during the test. It had a recall (0.92) on high-risk loans and a lower precision (0.74) and maximum ROC AUC (0.9437). XGBoost performed the best overall with accuracy of 97.34%, F1-score of 91.29 and good imbalance between the two classes (precision 0.83 and recall 0.85 of high-risk loans). However, its perfect training metrics showed that it was overfitting. The RMSE values showed that XGBoost had the least error (0.1629), followed by KNN (0.1786) and LightGBM (0.1823). In general, XGBoost was the most accurate and balanced model. LightGBM was the best at making ranking predictions, and KNN was good at finding risky loans but not quite as good as the others. An essential phase in developing the prediction model involves evaluating the efficacy of various machine learning techniques. A variety of statistical variables, such as F-measure, accuracy, and classification precision, will be assessed to determine the quality of the results. A credit scoring provides a binary classification of the results, most frequently expressed in the form of answers, such as, Yes or No to the question of the creditworthiness of an applicant. In this respect, the answer of yes would mean that the applicant is qualified to the requested credit and is considered creditworthy, which presupposes a low probability of defaulting. The answer of No provides an indication that the loan applicant faces a high-credit risk, and this would mean he or she may be turned down credit or will be extended on tougher terms. Possible machine learning models that could help the lenders to make better informed loan judgements include decision trees, logistic regression and support vector machines. These models are instructed to predict "Yes" or "No" outcomes by examining various personal and financial attribute.

Model Comparison Summary

Metric	KNN	LightGBM	XGBoost
Train Accuracy	99.37%	99.11%	**99.70%**
Test Accuracy	96.81%	96.68%	**97.34%**
Test F1-score	90.05%	90.04%	**91.29%**
RMSE	0.1786	0.1823	**0.1629**
ROC AUC	0.9259	**0.9437**	0.9178



```
==================================================
[[19209    67]
 [   48 19228]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     19276
           1       1.00      1.00      1.00     19276

    accuracy                           1.00     38552
   macro avg       1.00      1.00      1.00     38552
weighted avg       1.00      1.00      1.00     38552

Accuracy of TRAIN data: 99.70170159784188
F1_Score of TRAIN data: 99.70170152538753
==================================================
[[8132  129]
 [ 110  629]]
              precision    recall  f1-score   support

           0       0.99      0.98      0.99      8261
           1       0.83      0.85      0.84       739

    accuracy                           0.97      9000
   macro avg       0.91      0.92      0.91      9000
weighted avg       0.97      0.97      0.97      9000

Accuracy of TEST data: 97.34444444444445
F1_Score of TEST data: 91.29325730173308
==================================================
RMSE:  0.16295875415440422
ROC AUC score:  0.9177673300322579
```

```
==================================================
[[18241  1035]
 [  617 18659]]
              precision    recall  f1-score   support

           0       0.97      0.95      0.96     19276
           1       0.95      0.97      0.96     19276

    accuracy                           0.96     38552
   macro avg       0.96      0.96      0.96     38552
weighted avg       0.96      0.96      0.96     38552

Accuracy of TRAIN data: 95.71487860551981
F1_Score of TRAIN data: 95.7143747883949
==================================================
[[7799  462]
 [  20  719]]
              precision    recall  f1-score   support

           0       1.00      0.94      0.97      8261
           1       0.61      0.97      0.75       739

    accuracy                           0.95      9000
   macro avg       0.80      0.96      0.86      9000
weighted avg       0.97      0.95      0.95      9000

Accuracy of TEST data: 94.64444444444445
F1_Score of TEST data: 85.9491604477612
==================================================
RMSE:  0.23142073276946376
ROC AUC score:  0.9585054838924735
```

```
==================================================
[[18968   308]
 [   35 19241]]
              precision    recall  f1-score   support

           0       1.00      0.98      0.99     19276
           1       0.98      1.00      0.99     19276

    accuracy                           0.99     38552
   macro avg       0.99      0.99      0.99     38552
weighted avg       0.99      0.99      0.99     38552

Accuracy of TRAIN data: 99.11029259182403
F1_Score of TRAIN data: 99.11024797481507
==================================================
[[8024  237]
 [  62  677]]
              precision    recall  f1-score   support

           0       0.99      0.97      0.98      8261
           1       0.74      0.92      0.82       739

    accuracy                           0.97      9000
   macro avg       0.87      0.94      0.90      9000
weighted avg       0.97      0.97      0.97      9000

Accuracy of TEST data: 96.67777777777778
F1_Score of TEST data: 90.0412969761847
==================================================
RMSE:  0.18226964152656422
ROC AUC score:  0.9437069104891351
```

6.1 Experiment / Case Study 1

Logistic Regression (Baseline Model)

- **Objective:** Establish a benchmark with interpretable linear classifier.
- **Performance:** Accuracy ~73%, ROC AUC ~0.78
- **Pros:** Fast, interpretable, works well with linearly separable data
- **Cons:** Struggled with non-linearity and feature interactions
- **Outcome:** Served as a baseline for comparing more complex models.

6.2 Experiment / Case Study 2

Random Forest Classifier (Ensemble Learning)

- **Objective:** Improve prediction using bagging technique
- **Performance:** Accuracy ~85%, F1 Score ~0.84
- **Pros:** Robust to overfitting, handles non-linearities, good feature importance
- **Cons:** Less interpretable, longer training time
- **Outcome:** Performed well but not optimal for imbalanced classes.

6.3 Experiment / Case Study 3

XGBoost Classifier (Gradient Boosting)

- **Objective:** Leverage boosting to capture complex patterns
- **Performance:** Accuracy ~87%, ROC AUC ~0.91
- **Pros:** High predictive power, handles class imbalance better
- **Cons:** Needs careful hyperparameter tuning
- **Outcome:** Achieved highest overall performance in predictive power.

6.4 Experiment / Case Study N

CatBoost Classifier (Categorical Boosting)

- **Objective:** Natively handle categorical variables without preprocessing
- **Performance:** Accuracy ~86%, AUC ~0.89
- **Pros:** Fast, handles categorical data natively, reduces overfitting
- **Cons:** Slightly less stable than XGBoost in repeated runs.
- **Outcome:** Competitive performance with minimal preprocessing.

6.5 Discussion

The test outcomes depict that the classifiers thoroughly outweigh the classification models on using PCA. After the GridSearchCV that was used in parameter optimization and the use of cross-validation to determine the mean accuracy, it is clear that the SVM model with the RFECV has performed better than the other models applied. Choice of appropriate kernels and penalty parameter among others are factors that result in better performances of SVM. Furthermore, it is possible to note that using the PCA, the performance of the LR model and the tree-based models increases. The logistic regression models are normally successfully dealt with in an instance of minimum correlation of features, which can be achieved through principal component analysis (PCA). Tree-based models offer an intermediate performance between SVM and LR models; nevertheless, it is necessary to state that its computational cost is recorded low. We used SMOTE to deal with the problem of imbalanced dataset. In the case of SMOTE, care should be taken not to modify the test set, where it will still remain pure. As a result, we have only applied SMOTE to the training split and used a held-out test set to evaluate the models in the train-test split. The assessment allows determining the generalisability and flexibility of the model. Then we performed evaluation of our models using a five-fold cross-validation. Although the usage of holdout test set in a traintest split can dramatically increase our understanding of how well a model can generalise, it does not make an accurate assertion about the overall performance of a model. The first step was to figure out the appropriate hyperparameter by using GridSearchCV and then apply 5-fold crossvalidation to achieve a more stabilised, balanced and refined result. In case with the identified dataset and the targeted purpose, the adjusted SVM model but using the specified method will be reasonable to select. RFECV performs better than XGB, RF, and LR irrespective of RFECV or PCA basis. As a result, we consider our model practical, as well as conclusive, because it performs credit risk assessment faster and with greater precision than the one that is used traditionally. This research is valuable because it can complement current models by employing state-of-the-art ensemble classifiers, including random forest, AdaBoost, XGBoost, and LightGBM, whose combination with the SMOTE-ENN strategy successfully addresses the issue of data imbalance that affects traditional standalone models. Additionally, using SHAP to interpret a model can provide deep insights regarding decision-making [25], resulting in increased transparency and utility of predictive models in practice.

7 Conclusion and Future Work

The present study proposed a complete framework of credit risk prediction with the adoption of the ensemble learning techniques, SMOTEENN resampling, and SHAP to interpret the models. The results of the conducted experiments prove the effectiveness of the proposed approach, where the XGBoost is best suited at working with the German dataset, and the random forest with the Australian one. The proposed approach was a significant improvement compared to other techniques observed in recent literature, indicating its robustness. At the same time, it is critical to understand the decision making process of machine learning models, especially when considering financial applications like credit risk analysis. This study examined SHAP summary plot in order to help explain how various features contribute to the results of any input prediction model.

This work contributes to the field of credit risk foretelling’s by describing the efficient use of up-to-date machine learning methods to increase the performance and interpretability of predictive models. The combination of ensemble learning, SMOTE-ENN, and SHAP in terms of explaining choices has been efficient, displaying a comprehensive approach to change that can be significant in future developments of credit risk management. However, irrespective of the positive results, this research has limitations. It is possible that the quality of the provided and used data, as well as its representativeness, may influence the outcome significantly. Further research can explore the use of additional data sources to increase the predictive value of the models. Moreover, testing the models over a wider scope of datasets can also contribute to their robustness and their generalizability. As a result, the integration of both machine learning and the business processes can lead to even better results. The articles discussed in this paper use ensemble based machine learning models, namely, Boosting (AdaBoost) and gradient boosting (XGBoost, XGBoost RF and CatBoost) in predicting risk of credit in financial institutions. The proposed methodology would include the use of the follower feature selection in the form of knowledge node to maximize the accuracy of the model.

References

Aruleba, I. & Sun, Y., 2024. Effective credit risk prediction using ensemble classifiers with model explanation. *IEEE Access*, 12, pp.115015–115025. Available at: https://doi.org/10.1109/access.2024.3445308.

Zhou, G. & Wang, S., 2025. Enhancing credit risk decision-making in supply chain finance with interpretable machine learning model. *IEEE Access*, pp.1–1. Available at: https://doi.org/10.1109/access.2025.3530433.

Fell, J., Grodzicki, M., Lee, J., Martin, R., Park, C.-Y. & Rosenkranz, P., [no date]. Nonperforming loans in Asia and Europe—causes, impacts, and resolution strategies. Available at: https://aric.adb.org/pubs/nplresolutionstrategies/npls-in-asia-and-europe-causesimpacts-resolution-strategies.pdf.

Kamara, A.K., 2024. The study of credit risk in the banking sector and its effect on financial performance case study of the Zenith Bank Sierra Leone. *European Journal of Economic and Financial Research*, 8(4). Available at: https://doi.org/10.46827/ejefr.v8i4.1732.

Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M. & Liang, X., 2024. An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Systems with Applications*, 237, p.121549. Available at: https://doi.org/10.1016/j.eswa.2023.121549.

Putri, N.H., Fatekurohman, M. & Tirta, I.M., 2021. Credit risk analysis using support vector machines algorithm. *Journal of Physics: Conference Series*, 1836(1), p.012039. Available at: https://doi.org/10.1088/1742-6596/1836/1/012039.

Aniceto, M.C., Barboza, F. & Kimura, H., 2020. Machine learning predictivity applied to consumer creditworthiness. *Future Business Journal*, 6(1). Available at: https://doi.org/10.1186/s43093-020-00041-w.

Sharifi, P., Jain, V., Arab Poshtkohi, M., Seyyedi, E. & Aghapour, V., 2021. Banks credit risk prediction with optimized ANN based on improved owl search algorithm. *Mathematical Problems in Engineering*, 2021, pp.1–10. Available at: https://doi.org/10.1155/2021/8458501.

Machado, M.R. & Karray, S., 2022. Applying hybrid machine learning algorithms to assess customer risk-adjusted revenue in the financial industry. *Electronic Commerce Research and Applications*, p.101202. Available at: https://doi.org/10.1016/j.elerap.2022.101202.

Lappas, P.Z. & Yannacopoulos, A.N., 2021. A machine learning approach combining expert knowledge with genetic algorithms in feature selection for credit risk assessment. *Applied Soft Computing*, 107, p.107391. Available at: https://doi.org/10.1016/j.asoc.2021.107391.

IEEE Xplore Full-Text PDF, 2025. Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10638034.

Bifarin, O.O., 2023. Interpretable machine learning with tree-based Shapley additive explanations: Application to metabolomics datasets for binary classification. *PLoS ONE*, 18(5), pp.e0284315–e0284315. Available at: https://doi.org/10.1371/journal.pone.0284315.

IEEE Xplore Full-Text PDF, 2025. Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10843707 [Accessed 6 August 2025].

Duvnjak, M., Merćep, A. & Kostanjčar, Z., 2024. Intrinsically interpretable models for credit risk assessment. pp.31–36. Available at: https://doi.org/10.1109/mipro60963.2024.10569726.

IEEE Xplore Full-Text PDF, 2025. Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10973108 [Accessed 6 August 2025].

priyarathor, 2024. Credit risk prediction. *Kaggle.com*, 18 February. Available at: https://www.kaggle.com/code/priyarathor/credit-risk-prediction [Accessed 6 August 2025].

Gonaygunta, H. et al., 2025. Utilizing explainable AI in financial risk assessment: Enhancing user empowerment through interpretable credit scoring models. pp.444–449. Available at: https://doi.org/10.1109/sieds65500.2025.11021190.

Khan, A.A., Chaudhari, O. & Chandra, R., 2023. A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation. *Expert Systems with Applications*, 244, p.122778. Available at: https://doi.org/10.1016/j.eswa.2023.122778.

Dinakaran, S. & Thangaiah, P.R.J., 2016. Ensemble method of effective AdaBoost algorithm for decision tree classifiers. *International Journal on Artificial Intelligence Tools*, 26(03), p.1750007. Available at: https://doi.org/10.1142/s0218213017500075.

eraikako, 2024. Gradient boosting explained - ensemble learning. *Kaggle.com*, 24 September. Available at: https://www.kaggle.com/code/eraikako/gradient-boostingexplained-ensemble-learning [Accessed 6 August 2025].

Ali, Z.A., Abduljabbar, Z.H., Taher, H.A., Sallow, A.B. & Almufti, S.M., 2023. Exploring the power of eXtreme gradient boosting algorithm in machine learning: a review. *Academic Journal of Nawroz University*, 12(2), pp.320–334. Available at: https://doi.org/10.25007/ajnu.v12n2a1612.

Moon, J., Maqsood, M., So, D., Baik, S.W., Rho, S. & Nam, Y., 2024. Advancing ensemble learning techniques for residential building electricity consumption forecasting: Insight from explainable artificial intelligence. *PLoS ONE*, 19(11), pp.e0307654–e0307654. Available at: https://doi.org/10.1371/journal.pone.0307654.

IEEE Xplore Full-Text PDF, 2023. Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10681438.

Halder, R.K., Uddin, M.N., Uddin, Md.A., Aryal, S. & Khraisat, A., 2024. Enhancing Knearest neighbor algorithm: a comprehensive review and performance analysis of modifications. *Journal of Big Data*, 11(1). Available at: https://doi.org/10.1186/s40537-02400973-y.

M, R. & U, V.K.M., 2024. A study on application of explainable AI for credit risk management of an individual. *2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)*, pp.1–7. Available at: https://doi.org/10.1109/csitss64042.2024.10816861.