A Systematic Review of Machine Learning and Explainable Artificial Intelligence (XAI) in Credit Risk Modelling

Heng Yi Sheng[1] and Preethi Subramanian[2]

[1, 2] School of Computing, Asia Pacific University, Technology Park Malaysia, 57000 Kuala Lumpur, Kuala Lumpur, Malaysia.

yishenghengapu@gmail.com[1], preethi@apu.edu.my[2]

Abstract. The emergence of machine learning and artificial intelligence has created new opportunities for data-intensive science within the financial industry. The implementation of machine learning algorithms still faces doubt and distrust, mainly in the credit risk domain due to the lack of transparency in terms of decision making. This paper presents a comprehensive review of research dedicated to the application of machine learning in credit risk modelling and how Explainable Artificial Intelligence (XAI) could increase the robustness of a predictive model. In addition to that, some fully developed credit risk software available in the market is also reviewed. It is evident that adopting complex machine learning models produced high performance but had limited interpretability. Thus, the review also studies some XAI techniques that helps to overcome this problem whilst breaking out from the nature of the 'black-box' concept. XAI models mitigate the bias and establish trust and compliance with the regulators to ensure fairness in loan lending in the financial industry.

Keywords: Credit Risk, Explainable Artificial Intelligence (XAI), LIME, Machine Learning, SHAP.

1 Introduction

According to The Malaysian Reserve, the statistics published by Malaysian Department of Insolvency shows that more than 95,000 peoples had their loan defaulted where the defaulted loans were from personal loans (27.76%), hire purchase loans (24.73%), housing loans (14.09%) and credit card (9.91%) between the year 2014 and 2018 [1]. Loan defaults will not only disrupt the individual's credit score but will also introduce monetary losses to banks. This is also witnessed from a related publication released by Bank Negara Malaysia which states that the cumulative amount of impaired loans had reached RM31 billion as of July 2021 [2]. This is a huge loss for the bank sector, and it could lead to significant risk in Malaysia's economy. Thus, financial institutions are invigorated to employ a reliable credit risk model to minimize default risk.

Credit risk is known as the risk of the lender where the lender might not receive the principal and interest from the borrower [3]. Moreover, credit risk assessment plays an important role in financial industries in evaluating the capability of a borrower to repay a loan. Credit scoring has always been a challenge for financial institutions due to the unpredictable certainty of future events. Due to the emergence of machine learning technologies, the focus of credit risk modelling has gained consideration especially in the field of data science. In this paper, research has been carried out using a systematic review of literature such as journals, conference proceedings, academic publications, and books to understand existing investigation and debates relevant to credit risk modelling. The study also narrows down and takes a closer look at the explainability of machine learning models for decision making in the financial industry.

2 Domain Research

2.1 Credit Risk in Financial Industry

There are many types of risks faced by the banking industry, as seen in Fig. 1, which includes credit risk, market risk, liquidity risk, exchange rate risk, interest rate risk and operational risk. Among all the different types of risk aforementioned, credit risk is one of the main risks that most of the bank are facing nowadays [4].

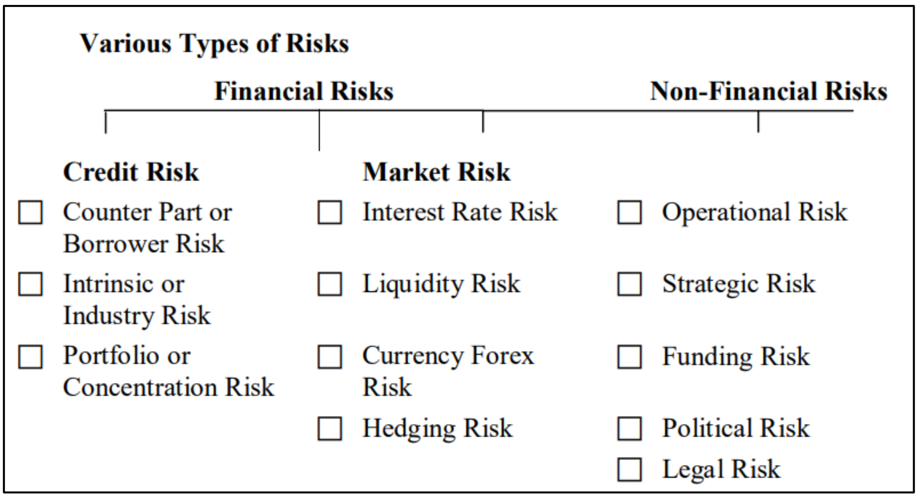

Fig. 1. Various Types of Risks faced by Banks [4]

Credit risk refers to the risk of loss imposed on creditors caused by borrowers due to their inability to meet their obligations [4]–[7]. This type of risk cause ambiguity in terms of the net income and the market value of the shares. Kolapo et al. [8] indicated that the bank is likely to experience financial crisis if the bank is highly vulnerable to credit risk. So, the performance of a bank can be determined using the approach that a bank used to handle credit risk. This is further supported by Chen & Pan [9] who states that credit risk is the most significant risk faced by banks and different banks have different approaches to credit risk management which allows them to adapt to changing environments. In the opinion of Rehman et al. [7], ignorance about credit risk by bank personnel will negatively affect the bank's development and customers' interest. Thus, credit risk can be considered an essential field of study. This is because if some of the borrowers default on the loans issued, it can eventually cause negative impact on banks including the entire banking system whereby banking crisis might occur [10]. This means that banks with high credit risk will face substantial loss mainly because borrowers defaulting on their loan repayment which might potentially lead to bankruptcy and insolvency.

Credit risk can occur due to several factors. For instance, poor management, poor loan underwriting, poor lending procedures, interference by the government bodies, inappropriate credit policies, unstable interest rate, direct lending, low reserves, liquidity level, huge licensing of banks, limited institutional capacity, insufficient supervision by the central bank, lack of strictness in credit assessment and inappropriate laws [11]. Therefore, it is recommended for the bank to consider minimizing the risk such as improvising the lending procedure, maintaining a well-documented information about the borrowers and a stabilized interest rate to potentially reduce the number of loan defaults and non-performing loans.

Effective credit risk management can enhance the compassion of the bank and the confidence of the depositors. Moreover, the financial health of a bank is highly dependent on the possession of good credit risk management. Hence, a good credit risk policy plays an essential role in boosting the banks' performance and its capital adequacy protection [11]. Pradhan & Shah [12] examined the relationship between credit management practices, credit risk mitigation measures and obstacles against loan repayment in Nepal using survey-based primary data and has performed a correlation analysis. The results revealed that credit risk management practices and credit risk mitigation measures have a positive relationship with loan repayment whilst obstacles faced by borrowers have no significant impact on loan repayment. This indicates that credit risk management practices and credit risk mitigation actions taken by the bank can help to reduce credit risk whereby borrowers will repay their loan on time which increases loan repayment behavior.

The Basel Accords was developed with the aim of establishing an international governing framework for controlling market risk and credit risk. This is to make sure banks holds enough capital to protect themselves from the financial crisis. The new Basel Capital Accord (Basel II) stated that banks should implement their internal credit risk model to assess default risk [13]. The effectiveness of credit risk management will not only help to maintain the profitability of the bank's businesses but also helps in sustaining the stability of the economy [14]. Moreover, Basel II relies on the following 3 pillars for its functioning: Minimal capital requirement, Supervisory review process, Market discipline.

According to Basel Committee on Banking Supervision [13], the risk parameters of Basel II are probability of default (PD), exposure at default (EAD) and loss given default (LGD). With these three risk parameters, the expected loss (EL) of the bank can be computed with the formula below:

$$EL = PD * EAD * LGD \quad (1)$$

In general, the banking industry plays an essential role in supporting the financial stability within a country. Thus, it is crucial for financial institutions to fully understand and ensure that data driven decisions are reached by figuring out the Expected Loss as outlined by Basel II in order to avoid the unfortunate impact of credit risk. With data analytics, several machine learning techniques are used in order to predict the credit risk and is reviewed in the following section.

2.2 Machine Learning in Credit Risk Modelling and Scorecard Creation

The advancement of machine learning techniques has provided several alternatives and reliability for loan default classification and prediction instead of manual processing in credit risk assessment [15]. With the rapid growth of big data in the industry, machine learning and deep learning are crucial in credit risk modelling to assist commercial banks in solving financial decision-making problems with the help of financial data [16]. There are many different artificial intelligence and machine learning methods that have been adopted for financial decision making to manage large loan portfolios. Examples of machine learning techniques used for financial decision making are artificial neural networks, decision trees and support vector machines [17]. These models will be able to predict loan applicants as either good credit (accepted) or bad credit (rejected) based on the historical data of the demographic characteristics such as marital condition, age and income [18].

Comparison of Traditional Methods and Machine Learning Models. Credit risk assessment is performed using the traditional methods or machine learning. The traditional methods of credit scoring make decisions based on either subjective scoring or statistical scoring [19]. Vidal & Barbon [20] mentioned that in subjective scoring, the decision is mainly based on qualitative judgement whereby the input from the loan officer and the organization will be used to evaluate the potential borrowers.

Nevertheless, statistical scoring relies on quantified characteristics of the potential borrowers and predict their likelihood of defaulting based on a set of rules and statistical techniques. There are a broad variety of statistical credit scoring models used to predict the probability of default such as Markov chain analysis, decision trees, profit analysis, logistic regression and linear discriminant analysis [21], [22]. After careful review, it has been found that logistic regression is widely used in the banking industry to minimize their credit risk as it is easy to execute and explain. In one of the study conducted by Memic [23], the author has employed traditional statistical methods such as logistic regression and multiple discriminant analysis (MDA) for predicting credit default of companies within the banking market in Bosnia and Herzegovina and its legal entities. The results indicated that both models have produced excellent predictive accuracy where logistic regression was found to have slightly better performance. The models have also identified variables that are significant in predicting credit default. For example, return on assets (ROA) variable is deemed to be statistically significant in logistic regression that has a high influence in predicting credit default as compared to other variables. Obare et al. [24] applied logistic regression to investigate individual loan defaults in Kenya with a sample of 1000 loan applicants. Cross validation was then used to evaluate the prediction results whereby the model achieved an accuracy of 77.27% with the train data and 73.33% with the test data. The authors also disclosed that increasing the sample size will improve the performance of logistic regression model whereby the model performed the best with a sample size of 700. Another paper by Foo et al. [25] discussed about credit scoring model to predict housing loan defaults in Malaysia. The authors have employed logistic regression of different variation using data acquired from the Malaysian Central Credit Reference Information Systems (CCRIS). The variations of logistic regression built involving the use of balance class, unbalanced class, with variable selection and without variable selection. The authors suggested that all four models yield favorable results, but logistic regression based on a balanced dataset with variable selection has obtained a high percentage of correctly classified data and the best sensitivity assuming a 0.5 cut-off value.

However, some of the machine learning techniques are reported to generate better results as compared to statistical techniques. Tsai & Wu [17] has stated that it is much more superior than the traditional statistical models. This can be supported by Bellotti & Crook [22], where the author compared support vector machine (SVM) against traditional methods such as logistic regression and linear discriminant analysis to predict the risk of default. The results indicated that SVM with a linear and Gaussian radial basis function (RBF) kernel produces the best result with an AUC of 0.783 for both algorithms. Nevertheless, the difference in terms of performance between SVM and traditional methods are not significant, but it is proven that SVM can be used as a feature selection to identify important variables in predicting the probability of default. Lee [26] has also implemented support vector machine (SVM) with RBF kernel in corporate credit rating problem and utilized 5-fold cross-validation with grid-search technique to search for the best parameter. Besides, the author compares the SVM's result against multiple discriminant analysis (MDA), case-based reasoning (CBR) and three-layer fully connected back-propagation neural networks (BPN) whereby the results show that SVM transcend other methods without overfitting.

Byanjankar et al. [27] used artificial neural network to predict the default probability of peer-to-peer (P2P) loan applicants. Moreover, comparisons have been conducted between neural network and logistic regression. The result shows that neural network is effective in identifying default borrowers whereas logistic regression is better in identifying non-default borrowers. Even so, neural network's result is deemed promising as it is crucial to forecast default loans in advance to prevent the creditors from investing in bad applicants. In another P2P credit risk study conducted by Bae et al. [28], online P2P lending default prediction models was developed using stepwise logistic regression, classification tree algorithms (CART and C5.0) and multilayer perceptron (MLP) to predict loan default. After evaluating the performance of the models with 5-fold cross-validation, the results reveal that MLP has the highest validation average accuracy, 81.78%, whereas logistic regression has the lowest validation average accuracy, 61.63%.

Moreover, Chandra Blessie & Rekha [29] has proposed a loan default prediction based on Logistic Regression, Decision Tree, Support Vector Machine and Naïve Bayes. The result indicated that Naïve Bayes classifier is tremendously efficient and gave a superior result than other classifiers. Aside from that, data cleaning, feature engineering and exploratory data analysis (EDA) was conducted before training the model. Features that were studied during EDA are application income, co-application income, loan amount, credit history, gender loan status, gender, relation status, education status and property area. Yet another evidence by Mafas developed a predictive model for loan default prediction in peer-to-peer lending communities using Logistic Regression, Random Forest, and Linear SVM with the selected feature set where Random Forest outperformed and achieved an accuracy of 92%. The significant fittest feature subset was obtained using a Genetic Algorithm and was evaluated using a Logistic Regression model [30].

After careful review, it is clear that the machine learning models can easily work with large datasets and generate predictions with high accuracy making it exceptional, but statistical techniques are much simpler and user friendly thereby making it popular for use in the financial industry. Machine learning model fitting also avoids overfitting as it will defeat the purpose of the study. This section discussed the performance of individual statistical and machine learning models. Newer research also experiments the usage of ensemble models also called as stacking approach.

Ensemble Model vs Individual Model. Aside from individual models, some researchers have reported that using ensemble models can yield better accuracy as compared to individual models. Yao [31] experimented with a single Decision Tree and two ensemble learning algorithms such as Adaboost and Bagging (Bootstrap Aggregation) with Decision Tree as a baseline algorithm to predict the creditworthiness of the applicants with the Australian credit dataset. The result indicates that ensemble learning, Adaboost CART with 14 features produced better results than a single Decision Tree without having much complexity. Likewise, another research has also adopted an ensemble model but with a different approach which is an ensemble technique of support vector machine (SVM) for credit risk assessment in Australian and German dataset by Xu et al. [32]. For example, the author experimented with voting ensemble based on single SVM and four SVM based ensemble models of four different kernel functions such as polynomial kernel, linear kernel, RBF kernel and sigmoid kernel against individual SVM models. Besides, Principal Component Analysis (PCA) is implemented before training the model to select credit features and 5-fold cross-validation is utilized for model validation purposes. The results show that the ensemble model of SVM performed better than the individual SVM classifier. Furthermore, the author has also suggested that the use of the ensemble model for credit risk assessment is promising to improve prediction performances.

Madaan et al. [33] proposed using Random Forest and Decision Tree to assess individual loans based on their attributes. The authors had also conducted exploratory data analysis to get acquainted with the dataset and performed data pre-processing. The data are then split into training (70%) and testing (30%) set whereby the selected algorithms will be used to train the model. The results of the classification report show that Random Forest outperforms Decision Tree with an accuracy score of 80% and 73% respectively. Another author, Zhu et al. [34], also proposed Random Forest classification but on a different scenario which is to predict loan default in P2P online lending platform and compare it against other machine learning methods such as

Decision Tree, Support Vector Machine (SVM) and Logistic Regression. The results indicated that Random Forest classification performs significantly better in identifying loan defaults. The authors have overcome the challenge of imbalanced class in the dataset by applying SMOTE (Synthetic Minority Oversampling Technique) method which can generate new samples for the minority class. Furthermore, the authors also suggested using larger datasets and fine-tuning the models can potentially improve the accuracy of the model in future research. Another P2P loan default prediction was conducted by Li et al. [35] based on XGBoost, Logistic Regression and Decision Tree. The result indicated that the predictive accuracy of XGBoost technique (97.705%) outperforms other models under 5-fold cross validation. Other performance comparisons were compared such as AUC value, classification error rate, model robustness and model run time. The result shows that although XGBoost has the best robustness and least error rate, the run time of the XGBoost is the slowest compared to other models. However, the author mentions that XGBoost is drastically better than traditional models in nearly all aspects. Moreover, the author has also visualized the top 10 features that have the most significant influence on loan default rates based on the XGBoost classifier.

Zhao et al. [36] suggested to use ensemble learning classification model such as adaptive boosting (AdaBoost) with decision tree on credit scoring problem. 10-fold cross-validation was performed to assess and compare the performance between AdaBoost-DT, Decision Tree, and Random Forest. The results show that AdaBoost-DT model yields the highest accuracy. Moreover, the author has also recommended to experiment with parameter optimization methods in future research. Udaya Bhanu & Narayana [37] proposed using random forest, logistic regression, decision tree, K-nearest neighbor, and Support Vector Machine for customer loan prediction. The author has also preprocessed the data and apply feature engineering technique to enhance the performance of machine learning algorithms. The comparative study shows that Random Forest shows the best accuracy, 82% in classifying loan candidates with an excellent F1-score.

In addition to the above models, LightGBM is a recently popular machine learning algorithm, which uses histogram algorithm and Leaf-wise strategy with depth limitation. LightGBM model has been used to predict the financing risk profile of 186 enterprises where the researcher conducted comparison experiments using k-nearest-neighbor’s algorithm, decision tree algorithm, and random forest algorithm on the same data set. The experiments show that LightGBM has better prediction results than the other three algorithms for several metrics in corporate financing risk prediction [38].

The reviewed literature has shown that ensemble models perform better compared to individual models. However, there is not much attention given to the voting ensemble model whereby it is a technique of combining the classifiers of different machine learning algorithms which is worth further investigation. A general consensus in the machine learning models either individual or ensemble would be to address data quality issues, handle imbalanced class and tune hyper parameters in order to improve the performance of the model.

Explainable Artificial Intelligence (XAI). The implementation of machine learning algorithms for model building within the credit risk industry faces doubt and distrust mainly due to the lack of transparency in terms of output predictions. According to Dong et al. [39], models such as support vector machine and neural network lacks interpretability and are often portrayed as a 'black box' model. This is primarily due to the output results are not clearly explained to general audiences and the banks will find it hard to provide the reasons for rejecting a loan. This issue is also being stated in recent studies. Hadji Misheva et al. [40] stipulated that complex machine learning has proven to have high predictive accuracy in assessing customer credit risk. Still, these innovative and advanced machine learning algorithms lack transparency that is essential to comprehend the reason behind the rejection and approval of an individual's loan application. The author also added that it is tough to trace back to the steps that an algorithm took to arrive at its decision as these models are developed directly from the data by an algorithm. The lack of credibility, trust and explainability are the major challenges faced by many researchers when introducing machine learning based models to companies in the credit scoring field [41]. Thus, 'black box' models are deemed to be less suitable in financial services due to the lack of interpretation. Even though the machine learning model improves over time and generates excellent predictive results, yet many financial institutions are still reluctant to fully trust the predictive model.

One of the potential solutions would be to incorporate transparent models, statistical models such as linear models or decision trees. Despite having models with high interpretability, it could also result in low predictive accuracy. Conversely, complex machine learning like neural networks gives high predictive accuracy but with limited interpretability [42]. To overcome this problem while also having the freedom to adopt complex machine learning algorithms, explainable artificial intelligence (XAI) should be incorporated to interpret the predictions made by the machine learning model and break out from the nature of the 'black-box' concept. This method not only allows humans to understand the output decision of the model, but it can also allow humans to trust the results of complex machine learning models and eliminate any doubts. Some of the popular XAI techniques commonly used are LIME and SHAP.

The explanation models can be classified into global methods and local methods. Global methods aim to provide a general explanation of a black-box model's behavior by using the overall knowledge of the model, training, and the associated data. For instance, feature importance will determine the top features that contribute the most in predicting the outcome. On the other hand, local methods are responsible for explaining a single outcome or instance of the black-box model. The single prediction performed by the model can be explained by creating local surrogate models that are interpretable and thereby exposing how a black-box model works [42], [43]. Hadji Misheva et al. [40] mentions that LIME is used to obtain local explanations, whereas SHAP can be used to obtain both local and global explanations in the XAI techniques.

LIME which stands for Locally Interpretable Model-Agnostic Explanations is a post-hoc model-agnostic explanation method that seeks to approximate any black-box machine learning model with an interpretable model to explain the single prediction. The author has also mentioned that LIME is a novel approach that explains the prediction of any classifier regardless of the algorithm. LIME will describe the model using a linearly weighted combination of the input features to provide the explanations. Conversely, SHAP, known as Shapley Additive explanations, interprets predictions based on coalitional game theory. It will return Shapley values that indicates how to fairly distribute the ‘payout’ (i.e. The prediction) among the features. Moreover, SHAP can provides a robust and insightful measure of feature importance of a model in a summary plot whereby Shapley value will represents the impact of the features on model output [40], [41]. Some of the recent works have adopted LIME and SHAP in credit risk problems to explain the decision made by the machine learning model.

Provenzano et al. [44] implemented SHAP and LIME techniques to explain the prediction of the high performing Light-GBM classifier that obtains 95% accuracy in default classification. The author stated that adopting SHAP and LIME has helped in understanding the important features in determining an individual result and thereby increasing the confidence in the model. Another study conducted by Visani et al. [45] has compared statistical model, Logistic Regression against machine learning model, Gradient Boosting Trees on credit risk data whereby LIME was tested on machine learning model to check its stability. It is reported that Gradient Boosting Model outperformed Logistic Regression and LIME is a stable and reliable technique when applied to the machine learning model.

Hadji Misheva et al. [40] has also adopted both XAI techniques, LIME and SHAP, in machine learning based credit scoring models on Lending Club dataset. The models that the author train including logistic regression, XGBoost, Random Forest, SVM and Neural Networks. The author has implemented LIME, as shown in Fig. 2, to explain local instances on SVM and tree-based models (XGBoost and Random Forest) whereas SHAP, as shown in Fig. 3, was used to obtain global explanations. The results of the study imply that both LIME and SHAP offer reliable explanation in line with financial reasoning. The author also mentions that SHAP is a powerful and effective techniques in highlighting the feature importance, but it can take a very long time to generate the results. This is supported by Phaure & Robin [46] in their study of model explainability in credit risk management whereby the author indicated that the computational time of SHAP method is proportional to the number of feature, observation and the complexity of the model.

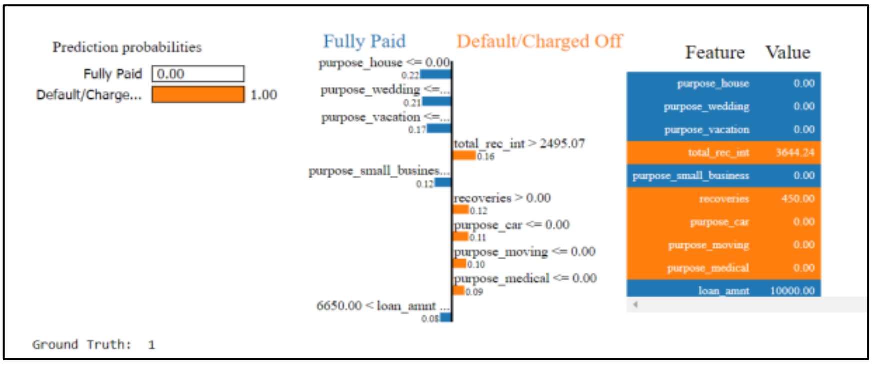

Fig. 2. XGBoost Model with LIME explanation on a customer that classified as a ‘Default’ loan type [40]

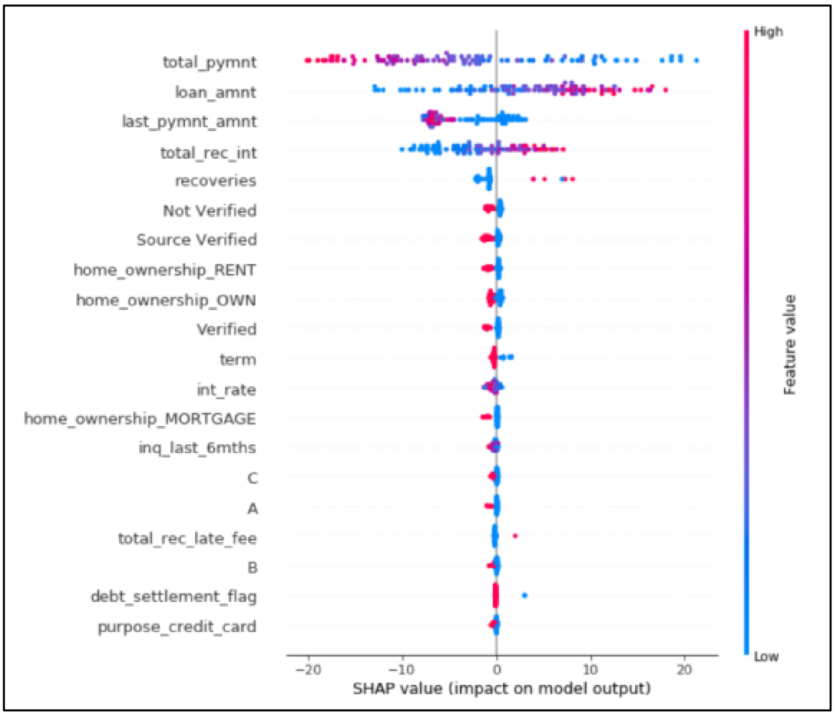

Fig. 3. Summary Plot - XGBoost Model with SHAP Tree Explainer [40]

In short, introducing XAI techniques can help improve the explainability and transparency of the black-box model rather than relying solely on machine learning output for decision making. XAI will not only eliminate bias, but it can also assist in establishing trust and in compliance with the regulators in financial institutions to ensure fairness in loan lending. Therefore, XAI techniques, specifically LIME should be adopted to explain the credit decision of the black-box model.

Credit Scorecards. The banking industry uses credit scorecards as a tool for risk management. Credit scorecards consist of a group of features that are widely used to predict the default probabilities such as classifying good and bad credit risk. There are various techniques used in the development of scorecards such as support vector machine, genetic programming, artificial neural networks, multiple classifier systems, hybrid models, logistic regression, classification tree, linear regression and linear programming [39], [47]. Moreover, Dong et al. [39] stipulated that generating credit scorecards will potentially contribute to effective credit risk management. The author added that the quality of the credit scorecard can be measured such as using Percentage Correctly Classified (PCC) to identify the accuracy of the prediction.

Characteristic Name	Attribute	Scorecard Points
AGE	. -> 23	63
AGE	23 -> 25	76
AGE	25 -> 28	79
AGE	28 -> 34	85
AGE	34 -> 46	94
AGE	46 -> 51	103
AGE	51 -> .	105
CARDS	"AMERICAN EXPRESS," "VISA OTHERS," "VISA MYBANK," "NO CREDIT CARDS"	80
CARDS	"CHEQUE CARD," "MASTERCARD/EUROC," "OTHER CREDIT CARD"	99
EC_CARD	0	86
EC_CARD	1	83
INCOME	. -> 500	93
INCOME	500 -> 1,550	81
INCOME	1,550 -> 1,850	75
INCOME	1,850 -> 2,550	80
INCOME	2,550 -> .	88
STATUS	"E," "T," "U"	79

Fig. 4. Example of Credit Scorecard [48]

Fig. 4 shows an example of a credit scorecard used to evaluate the creditworthiness of a loan applicant. For instance, the features such as age, cards, ec_card, income, and status each will be assigned points based on statistical analysis. The sum of the points accumulated will be the final score of the loan applicant. Therefore, the banks can easily decide which loan should be accepted or rejected. For example, the bank can choose to reject the loan application or charge them a higher interest rate if the applicant scores below a certain range as they possess a greater risk. Hence, a credit scorecard will facilitate a better decision-making process for the financial institution.

3 Related Works

This section will compare and analyze different credit risk models and software that are fully developed and currently available in the market. Most of the credit risk models developed are marketed towards medium and large size companies such as banks and enterprise creditors. Their goal is to assists companies who purchase their system in determining the creditworthiness of potential borrowers and minimizing loan defaults. With a timelier and accurate predictions, lenders can use the result generated to negotiate with the borrowers. As part of the research, comparisons will be conducted between three different commercial systems to understand their structures and functionalities. The three systems selected in this study are GiniMachine, ABLE Scoring and ZAML.

3.1 GiniMachine

Fig. 5. GiniMachine Logo [49]

GiniMachine is an AI-driven credit scoring software that can help lenders make reliable credit decisions within a short amount of time and the logo of GiniMachine can be seen in Fig. 5 [49]. This system will employ machine learning for automated decision-making where it is effective even towards thin-file borrowers. Thus, banks and fintech companies can identify bad loans to avoid unwanted risk without relying on traditional credit scoring or doing manual work that has many shortcomings. For instance, GiniMachine that is based on AI technologies can analyze parameters that traditional method tends to ignore. Furthermore, GiniMachine can easily adapt into changing environment that will fit nicely into specific businesses and risk assessment rules. Let's suppose, if the company has released a new loan product, the system can process the information of the new loan product and adjust accordingly to the needs of the lenders. The system will also generate detailed reports, as shown in Fig. 6, that consist of statistical calculations regarding to the decision made by the model. Moreover, the system is easy to use as it is designed specifically for non-technical individuals to operate the system. Thus, no specific training is required to operate the system.

Fig. 6. GiniMachine's Scoring Details [49]

3.2 ABLE Scoring

Fig. 7. ABLE Scoring's Company Logo [50]

ABLE Scoring is another powerful credit scoring software that will assist in making credit decisions to prevent bad loans and the logo of ABLE Scoring can be seen in Fig. 7. Scorecards along with credit decisions can be easily generated via the scorecard builder, as shown in Fig. 8. Moreover, ABLE Scoring allows lenders to score potential borrowers in batches which will save lots of time. Different machine learning models can be built including the classical logistic regression model. The performance of each of the models can be compared and evaluated in terms of performance and stability. Furthermore, the result of the credit decision will be explained in the scorecards generated, as shown in Fig. 9, which will help the lenders to better understand the output decision made by the machine learning model to eliminate any doubts. It will also check for data formats, consistency, and missing values to ensure the data is in high quality. The software is easy to use without any specific training required. The users will just need to upload an XLS file format to generate a scorecard report. ABLE Scoring promotes fast and smart credit decisions based on AI models and it ensures a stable and high-quality lending process. This software is being trusted by banks and fintech companies such as Eurasian Bank, OTP Bank and Alfa Bank [50].

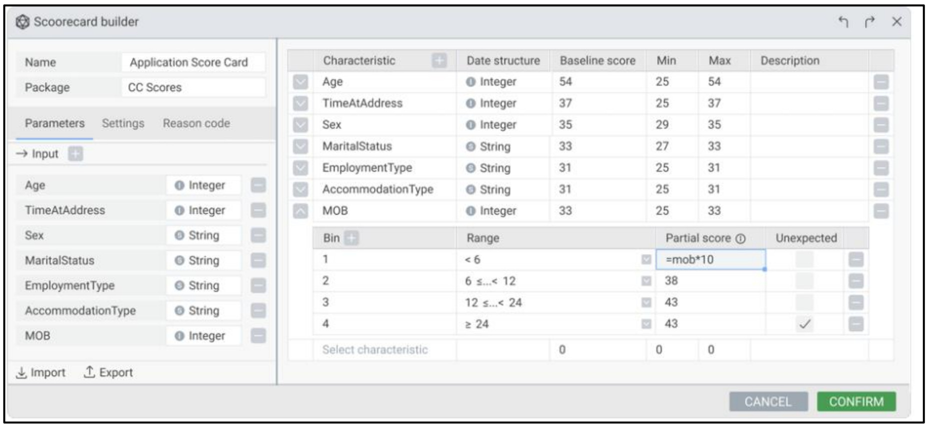

Fig. 8. ABLE Scoring's Scorecard Builder [50]

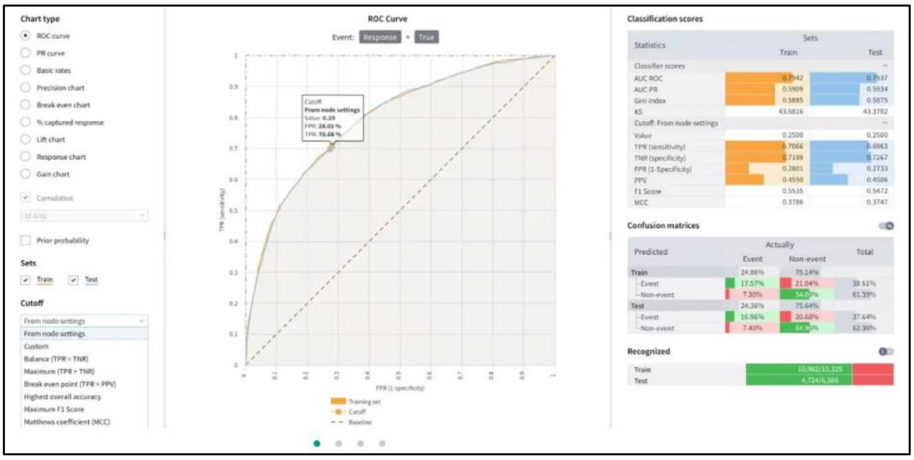

Fig. 9. ABLE Scoring's Scorecard Generation [50]

3.3 Zest AI

Fig. 10. Zest AI Logo [51]

Zest AI is yet another robust machine learning software that assists lenders and underwriters to make better, more timely and transparent credit decisions. The logo of Zest AI is shown in Fig. 10. Zest AI also aims to address the problems of traditional credit scoring tools, such as gaps, errors or structural inequities that lead to the rejection of good applicants [52]. With Zest AI, lenders can easily identify good borrowers and safely increase loan approvals while minimizing the risk and losses. Besides, Zest AI provides a bigger picture of every borrower with full interpretability to comply with the strictest regulators and satisfy doubters [51]. For example, the custom-built logistic regression scorecards in Zest AI will be used to assess the creditworthiness of the borrowers to help lenders in their decision making. Fig. 11 shows a sample of the scorecards generated with Zest AI:

Fig. 11. Zest AI Scorecard Generation [46]

Most importantly, it is a stable software that offers rapid analysis to help lenders make quick business decisions and ensure fairness in lending operations. Thus, this will potentially improve customer experience and make a positive impact on lending businesses. Furthermore, the software owners can also rest assured as Zest AI offers smooth transition and adoption from traditional credit scoring tools with professional support. In addition, the software is also user-friendly whereby it can be operated by non-technical staff without prior machine learning background. Zest AI is also being recognized by one of the largest banks in Turkey, Akbank. Akbank has found Zest AI software extremely effective in identifying good borrowers with minimal risks. Akbank managed to reduce non-performing loans by 45% and less time needed to retrain and build the models with Zest AI that initially took them seven months [53]. Besides, Zest AI can adapt to changing requirements which further increases the confidence of their client. Thus, the adoption of Zest AI can promote sustainable growth among banks and other financial institutions in their lending businesses.

3.4 Evaluation of Related Works

The comparisons between the related work are essential to understand the attributes of the fully developed credit scoring systems. Moreover, new ideas and opportunities can be triggered by analyzing the existing systems, which will benefit future research. Table 1 shows the comparisons between different credit risk systems that are currently available in the market:

Table 1. Comparison of Related work

Attributes	GiniMachine	ABLE Scoring	Zest AI
Features	Automated credit scoring empowered with AI and ML	One button solution to build scorecard for credit decision with AI models and score customers in batch	Employ AI models to make smart lending decisions
Purpose	Avoid bad and non-performing loans	Ensure a continuous and high-quality lending process	Faster loan decisions and ensure fairness in lending
Benefits	Ease of use, save time and adaptable into changing environment	Easy to use, customizable, stable, and transparent result (explainable)	Time and resources saving, easy to operate and comply with the regulators
Target User	Non-technical credit analyst/lenders	Banks and Fintech Companies	Banks and Lending Companies
Cost	Paid	Paid	Paid
Demo	Free demo available	Demo provided upon request	Demo provided upon request

Based on the analysis conducted, all the systems are built to ensure faster, fair, and high-quality loan lending. This is because their target users are mostly banks and other financial institutions whereby the primary goal is to mitigate credit risk and avoid bad loans. The systems are also user-friendly, especially for non-technical staff to operate the system without much training needed. Moreover, it is also important for the output result to be transparent to comply with the regulators. However, it is noted that all three systems solely focus on predicting the output, but it has no dashboard to visualize the trends of loan customers. In that case, it will be an opportunity for the developer to include a dashboard in the web application that will visualize the trends of loan customers.

4 Conclusion and Future Direction

Intensive research has been conducted via Google scholars and APU E-Database to better understand the credit risk field and the machine learning techniques employed to solve the underlying credit risk problem. The findings show that machine learning techniques, especially ensemble models, perform extremely well in identifying loan defaults which can potentially minimize future credit default risk. It is also noticeable that recent research is centered around ensemble learning such as Random Forest, XGBoost and AdaBoost. There are many papers focus on applying machine learning algorithms to solve credit risk problems, such as predicting the likelihood of loan defaults. Some of the papers have compared the performance between statistical methods and artificial intelligence methods. The findings have indicated that artificial intelligence methods produce better classification accuracy as compared to statistical methods. However, in terms of interpretability and simplicity, statistical methods are a better choice as compared to artificial intelligence methods. Furthermore, the research has moved towards a new era of machine learning, explainable artificial intelligence (XAI), that can uncover and explain a black-box machine learning model. Most importantly, implementing XAI in credit risk models will allow humans to better understand the predictions made by the machine learning models whilst establishing trust and compliance with regulatory requirements within the financial institution. The adoption of XAI, such as LIME and SHAP, helps improves the transparency of loan lending while speeding up the loan lending process, which is a more robust approach than the traditional lending procedures.

The future research could consider to study about credit risk in commercial banks and build machine learning models that will be used by credit analysts to identify and predict loan defaults with the intention of assisting them in better decision making and evaluating the profile of potential borrowers whilst minimizing future credit default risk and preventing the recurrence of the global financial crisis. This could also be extended to gathering more diverse and non-conventional data to enhance banks’ approaches to assessing credit risk. Furthermore, future research should also explore different XAI techniques available such as Shapash or Dalex, that are also compatible with many machine learning frameworks in credit risk prediction. More focus on the comparison of XAI models that support both local and global explanations will bring additional value to the credit risk industry.

References

1. A. Hani, Credit cards, personal loans landing Malaysians in debt trap, 2019. https://themalaysianreserve.com/2019/08/08/credit-cards-personal-loans-landing-malaysians-in-debt-trap/
2. Bank Negara Malaysia, Monthly Highlights and Statistics in July 2021, 2021. https://www.bnm.gov.my/-/monthly-highlights-and-statistics-in-july-2021
3. Brock, Credit Risk, 2021. https://www.investopedia.com/terms/c/creditrisk.asp
4. K. A. Goyal and S. Agrawal, "Risk Management in Indian Banks: Some Emerging Issues," *Int. J. Econ. Res.*, vol. 1, no. 1, pp. 102–109, 2010.
5. S. Chenghua and Z. Kui, "Study on commercial bank credit risk based on information asymmetry," *2009 Int. Conf. Bus. Intell. Financ. Eng. BIFE 2009*, pp. 758–761, 2009, doi: 10.1109/BIFE.2009.175.
6. H. Li and S. Pang, "The study of credit risk evaluation based on DEA method," *Proc. - 2010 Int. Conf. Comput. Intell. Secur. CIS 2010*, pp. 81–85, 2010, doi: 10.1109/CIS.2010.25.
7. Z. U. Rehman, N. Muhammad, B. Sarwar, and M. A. Raz, "Impact of risk management strategies on the credit risk faced by commercial banks of Balochistan," *Financ. Innov.*, vol. 5, no. 1, 2019, doi: 10.1186/s40854-019-0159-8.
8. T. F. Kolapo, R. K. Ayeni, and M. O. Oke, "Credit Risk and Commercial Banks' Performance in Nigeria: A Panel Model Approach," *Aust. J. Bus. Manag. Res.*, vol. 2, no. 02, pp. 31–38, 2012.
9. K.-C. Chen and C.-Y. Pan, "An Empirical Study of Credit Risk Efficiency of Banking Industry in Taiwan," *Web J. Chinese Manag. Rev.*, vol. 15(1), no. Febuary, 2012, pp. 1–17, 2012, [Online]. Available: http://cmr.ba.ouhk.edu.hk.
10. W. Waemustafa and S. Sukri, "Bank specific and macroeconomics dynamic determinants of credit risk in islamic banks and conventional banks," *Int. J. Econ. Financ. Issues*, vol. 5, no. 2, pp. 476–481, 2015, doi: 10.6084/m9.figshare.4042992.
11. Y. R. Bhattarai, "The Effect of Credit Risk on Nepalese Commercial Banks," *NRB Econ. Rev.*, vol. 28, no. 1, pp. 41–64, 2016, [Online]. Available: https://nrb.org.np/ecorev/articles/vol28-1_art3.pdf.
12. S. Pradhan and A. K. Shah, "Credit Risk Management of Commercial Banks in Nepal," *J. Bus. Soc. Sci. Res.*, vol. 4, no. 1, pp. 27–37, 2019, doi: 10.3126/jbssr.v4i1.28996.
13. Basel Committee on Banking Supervision, "An Explanatory Note on the Basel II IRB Risk Weight Functions," 2005. [Online]. Available: www.bis.orgb/cbsirbriskweight.pdf.
14. M. Psillaki, I. E. Tsolas, and D. Margaritis, "Evaluation of credit risk based on firm performance," *Eur. J. Oper. Res.*, vol. 201, no. 3, pp. 873–881, 2010, doi: 10.1016/j.ejor.2009.03.032.
15. L. Lai, "Loan Default Prediction with Machine Learning Techniques," *Proc. - 2020 Int. Conf. Comput. Commun. Netw. Secur. CCNS 2020*, pp. 5–9, 2020, doi: 10.1109/CCNS50731.2020.00009.
16. P. M. Addo, D. Guegan, and B. Hassani, "Credit Risk Analysis Using Machine and Deep Learning Models," *SSRN Electron. J.*, no. May 2019, 2018, doi: 10.2139/ssrn.3155047.

17. C. F. Tsai and J. W. Wu, “Using neural network ensembles for bankruptcy prediction and credit scoring,” *Expert Syst. Appl.*, vol. 34, no. 4, pp. 2639–2649, 2008, doi: 10.1016/j.eswa.2007.05.019.
18. M. Chen and S. Huang, “Credit scoring and rejected instances reassigning through evolutionary computation techniques,” *Expert Syst. Appl.*, vol. 24, no. 4, pp. 433–441, 2003, doi: 10.1016/S0957-4174(02)00191-4.
19. M. Schreiner, “Scoring: the next breakthrough in microcredit,” *CGAP*, no. 7, pp. 1–64, 2003.
20. M. F. Vidal and F. Barbon, “Credit Scoring in Financial Inclusion,” *CGAP*, Jul. 2019.
21. Y. L. Eddy and E. M. N. Engku Abu Bakar, “Credit scoring models: Techniques and Issues,” *J. Adv. Res. Bus. Manag. Stud.*, vol. 7, no. 2, pp. 29–41, 2017, [Online]. Available: https://www.akademiabaru.com/submit/index.php/arbms/article/view/1240.
22. T. Bellotti and J. Crook, “Support Vector Machines for Credit Scoring and discovery of significant features,” *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3302–3308, 2009, doi: 10.1016/j.eswa.2008.01.005.
23. D. Memic, “Assessing Credit Default using Logistic Regression and Multiple Discriminant Analysis: Empirical Evidence from Bosnia and Herzegovina,” *Interdiscip. Descr. Complex Syst.*, vol. 13, no. 1, pp. 128–153, 2015, doi: 10.7906/indecs.13.1.13.
24. D. M. Obare, G. G. Njoroge, and M. M. Muraya, “Analysis of Individual Loan Defaults Using Logit under Supervised Machine Learning Approach,” *Asian J. Probab. Stat.*, vol. 3, no. 4, pp. 1–12, 2019, doi: 10.9734/ajpas/2019/v3i430100.
25. L. K. Foo, S. L. Chua, D. Chin, and M. K. Firdaus, “Logistic regression models for Malaysian housing loan default prediction,” 2017. [Online]. Available: https://www.bnm.gov.my/documents/20124/826852/WP11+-+Logistic+Regression.pdf/d22ef5a2-4bdb-4d39-28f3-c19253d2814e?t=1585030599211.
26. Y. C. Lee, “Application of support vector machines to corporate credit rating prediction,” *Expert Syst. Appl.*, vol. 33, no. 1, pp. 67–74, 2007, doi: 10.1016/j.eswa.2006.04.018.
27. A. Byanjankar, M. Heikkila, and J. Mezei, “Predicting credit risk in peer-to-peer lending: A neural network approach,” *Proc. - 2015 IEEE Symp. Ser. Comput. Intell. SSCI 2015*, pp. 719–725, 2015, doi: 10.1109/SSCI.2015.109.
28. J. K. Bae, S. Y. Lee, and H. J. Seo, “Predicting Online Peer-to-Peer (P2P) Lending Default using Data Mining Techniques,” *J. Soc. E-bus. Stud.*, vol. 23, no. 3, pp. 1–6, 2018.
29. E. Chandra Blessie and R. Rekha, “Exploring the machine learning algorithm for prediction the loan sanctioning process,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 1, pp. 2714–2719, 2019, doi: 10.35940/ijitee.A4881.119119.
30. L.Victor, M.Raheem, "Loan Default Prediction Using Genetic Algorithm: A Study Within Peer-To-Peer Lending Communities", International Journal of Innovative Science and Research Technology, Volume 6, Issue 3, March – 2021, ISSN No:-2456-2165.
31. P. Yao, “Credit scoring using ensemble machine learning,” *Proc. - 2009 9th Int. Conf. Hybrid Intell. Syst. HIS 2009*, vol. 3, pp. 244–246, 2009, doi: 10.1109/HIS.2009.264.
32. W. Xu, S. Zhou, D. Duan, and Y. Chen, “A support vector machine based method for credit risk assessment,” *Proc. - IEEE Int. Conf. E-bus. Eng. ICEBE 2010*, pp. 50–55, 2010, doi: 10.1109/ICEBE.2010.44.
33. M. Madaan, A. Kumar, C. Keshri, R. Jain, and P. Nagrath, “Loan default prediction using decision trees and random forest: A comparative study,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1022, no. 1, pp. 1–12, 2021, doi: 10.1088/1757-899X/1022/1/012042.
34. L. Zhu, D. Qiu, D. Ergu, C. Ying, and K. Liu, “A study on predicting loan default based on the random forest algorithm,” *Procedia Comput. Sci.*, vol. 162, no. Itqm 2019, pp. 503–513, 2019, doi: 10.1016/j.procs.2019.12.017.

35. Z. Li, S. Li, Z. Li, Y. Hu, and H. Gao, “Application of XGBoost in P2P Default Prediction,” *J. Phys. Conf. Ser.*, vol. 1871, no. 1, pp. 1–10, 2021, doi: 10.1088/1742-6596/1871/1/012115.
36. J. Zhao, Z. Wu, and B. Wu, “An AdaBoost-DT Model for Credit Scoring,” *WHICEB 2021 Proc.*, vol. 15, 2021.
37. L. Udaya Bhanu and D. S. Narayana, “Customer Loan Prediction Using Supervised Learning Technique,” *Int. J. Sci. Res. Publ.*, vol. 11, no. 6, pp. 403–407, 2021, doi: 10.29322/ijsrp.11.06.2021.p11453.
38. D.N.Wang, L.Li, D.Zhao,"Corporate finance risk prediction based on LightGBM", Information Sciences, Volume 602, 2022, Pages 259-268, ISSN 0020-0255, https://doi.org/10.1016/j.ins.2022.04.058.
39. G. Dong, K. L. Kin, and J. Yen, “Credit scorecard based on logistic regression with random coefficients,” *Procedia Comput. Sci.*, vol. 1, no. 1, pp. 2463–2468, 2010, doi: 10.1016/j.procs.2010.04.278.
40. B. Hadji Misheva, A. Hirsa, J. Osterrieder, O. Kulkarni, and S. Fung Lin, “Explainable AI in Credit Risk Management,” *SSRN Electron. J.*, pp. 1–16, 2021, doi: 10.2139/ssrn.3795322.
41. A. El Qadi, N. Diaz-Rodriguez, M. Trocan, and T. Frossard, “Explaining Credit Risk Scoring through Feature Contribution Alignment with Expert Risk Analysts,” pp. 1–12, 2021, [Online]. Available: http://arxiv.org/abs/2103.08359.
42. M. Wijnands, “Explaining black box decision-making,” University of Twente, 2021.
43. R. Confalonieri, L. Coba, B. Wagner, and T. R. Besold, “A historical perspective of explainable Artificial Intelligence,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 11, no. 1, pp. 1–21, 2021, doi: 10.1002/widm.1391.
44. A. R. Provenzano *et al.*, “Machine Learning approach for Credit Scoring,” 2020. [Online]. Available: http://arxiv.org/abs/2008.01687.
45. G. Visani, E. Bagli, F. Chesani, A. Poluzzi, and D. Capuzzo, “Statistical stability indices for LIME: Obtaining reliable explanations for machine learning models,” *J. Oper. Res. Soc.*, pp. 1–18, 2020, doi: 10.1080/01605682.2020.1865846.
46. H. Phaure and E. Robin, “Explain Artificial Intelligence for Credit Risk Management,” *Deloitte*, Apr. 2020.
47. A. Bequé, K. Coussement, R. Gayler, and S. Lessmann, “Approaches for credit scorecard calibration: An empirical analysis,” *Knowledge-Based Syst.*, vol. 134, pp. 213–227, 2017, doi: 10.1016/j.knosys.2017.07.034.
48. N. Siddiqi, *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. John Wiley & Sons, Inc., 2006.
49. GiniMachine, “Credit Scoring Software,” 2021. https://ginimachine.com/risk-management/credit-scoring/ (accessed Oct. 13, 2021).
50. RND Point, “Credit Scoring Software,” 2021. https://rndpoint.com/solutions/able-scoring/ (accessed Oct. 25, 2021).
51. Zest AI, “Zest AI,” 2021. https://www.zest.ai/ (accessed Oct. 25, 2021).
52. B. Upbin, “ZAML Fair - Our New AI To Reduce Bias In Lending,” 2019. https://www.zest.ai/insights/zaml-fair-our-new-ai-to-reduce-bias-in-lending (accessed Oct. 25, 2021).
53. Zest AI, “Model Management System,” 2021. https://www.zest.ai/product (accessed Oct. 25, 2021).