Journal of Soft Computing Paradigm (ISSN: 2582-2640)
www.irojournals.com/jscp/

Credit Risk Analysis using Explainable Artificial Intelligence

Sowmiya M N.[1], Jaya Sri S.[2], Deepshika S.[3], Hanushya Devi G.[4]

[1]Assistant Professor,[2-4]UG Student, Department of Information Technology, Velammal Engineering College, Tamilnadu, India

E-mail: [1]sowmiya@velammal.edu.in, [2]jayasri@gmail.com, [3]sdeepshika@gmail.com, [4]haunushya@gmail.com

Abstract

The proposed research focuses on enhancing the interpretability of risk evaluation in credit approvals within the banking sector. This work employs LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) to provide explanations for individual predictions: LIME approximates the model locally with an interpretable model, while SHAP offers insights into the contribution of each feature to the prediction through both global and local explanations. The research integrates gradient boosting algorithms (XGBoost, LightGBM) and Random Forest with these Explainable Artificial Intelligence (XAI) techniques to present a more comprehensible framework. The results demonstrate how interpretability methods such as LIME and SHAP enhance the transparency and trustworthiness of machine learning models, which is crucial for applications in credit risk evaluation.

Keywords: LIME, SHAP, Model Prediction, Machine Learning, High Accuracy, Explainable Artificial Intelligence

1. Introduction

Credit risk analysis is a fundamental requirement in making decision, particularly for lending institutions. It involves a thorough evaluation of the potential risks that are associated with lending credit to individuals or business. This analysis aims to evaluate the probability of borrowers missing their payments, thereby providing essential insights that enable lenders to make informed decisions. Credit risk analysis considers various factors, including the borrower's credit history, financial health, and external economic conditions. Evaluating a

DOI: https://doi.org/10.36548/jscp.2024.3.004
Received: 25.06.2024, received in revised form: 28.07.2024, accepted:02.08.2024, published: 13.08.2024

borrower's creditworthiness typically involves analyzing their credit score, which summarizes their past credit behavior, including repayment patterns and outstanding debts. Additionally, financial ratios such as the debt-to-income ratio and liquidity metrics are examined to assess the borrower’s ability to repay debts.

Technological developments, regulatory changes, and shifts in market dynamics have all significantly impacted the evolution of credit risk analysis. Initially, credit risk analysis relied on inaccurate qualitative evaluations and basic statistical models, which lacked consistency and scalability. However, recent advancements in AI and machine learning have brought substantial transformation to the field. Machine learning algorithms can now process data from various sources, including financial records, credit reports, and alternative data, to generate more accurate estimates of credit risk.

The integration of explainable Artificial Intelligence (XAI) in credit risk analysis has introduced greater transparency and interpretability. XAI provides explanations for individual credit decisions, helping to clarify why a particular applicant may have been approved or denied credit. This transparency not only builds trust in the decision-making process but also helps identify potential biases or discriminatory practices within the model.

The proposed work integrates the machine learning models (XGBoost, LightGBM, and Random forest) with XAI to have a more comprehensible frame work that helps in understanding why credit or loan application was rejected or approved.

2. Related Work

The related work presents a review of existing literature on credit risk analysis. Studies show that complex algorithms and methods are employed to predict loan statuses using applicant attributes, income, credit history, and loan details. Pandey et al. [1] surveyed various techniques used in banking to evaluate credit approval risk. Misheva et al. [2] implemented Explainable Artificial Intelligence (XAI) on machine learning models using the Lending Club dataset. The existing research also highlights practical challenges and the potential of XAI in credit risk management on peer review platforms [3,4]. Biecek et al. [5] compared predictive models, finding tree-based models superior to others. Heng et al. [6] reviewed machine learning applications in credit risk modeling, emphasizing XAI's role in improving model predictability and transparency. Hu et al. [7] used XAI to identify causal relationships with restricted datasets, emphasizing cross-validation, regularization, and bootstrapping techniques. De et al. [8] integrated the LightGBM model with SHAP to interpret explanatory variables affecting predictions. Demajo et al. [9] proposed an accurate and interpretable credit scoring model. Burgt et al. [10] cautioned that AI in banking requires regulatory adjustments. Gramespacher et al. [11] discusses the machine learning's adaptability to credit risk evaluation needs. The research collectively underscores the importance of XAI in enhancing trust and compliance in credit risk analysis, focusing on feature relationships and interactions [12-15].

Overall, the continuous development of methodologies and techniques in credit risk analysis highlights the need for ongoing research and innovation in credit risk management within today's dynamic financial landscape.

3. Proposed Work

The proposed methodology enhances existing models by employing sophisticated methods such as LIME and SHAP to make predictions more comprehensible. These techniques contribute to the transparency and reliability of the system by clarifying the decision-making process. Reducing biases and improving user understanding of the decision-making process will increase overall transparency. Figure 1 shows the flowchart of the proposed.

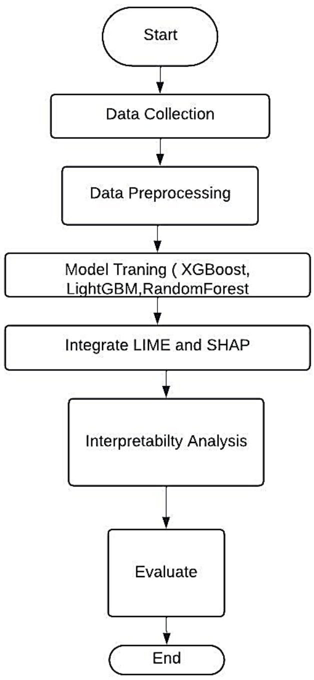


Figure 1. Proposed Flowchart

A. Data Collection

The proposed work uses the Lending Club loan dataset [2,16], which contains data on thousands of loans made through the Lending Club platform. The dataset includes 10,000 observations across 55 variables, such as emp_title, emp_length, loan_intent, loan_grade, and loan_amt. Figure 2 below displays a sample of the dataset used in this study.

	person_age	person_income	person_home_ownership	person_emp_length	loan_intent	loan_grade	loan_amnt	loan_int_rate	loan_status	loan_perc
6061	26	48000	MORTGAGE	10.0	HOMEIMPROVEMENT	A	4000	5.42	0	
3580	25	55200	RENT	0.0	HOMEIMPROVEMENT	C	3600	13.98	0	
6041	24	48000	MORTGAGE	8.0	MEDICAL	A	10000	7.49	0	
27405	35	139000	MORTGAGE	12.0	DEBTCONSOLIDATION	B	35000	12.69	0	
1971	22	24000	RENT	0.0	PERSONAL	B	2100	12.21	0	
18775	31	43200	RENT	NaN	MEDICAL	B	1000	10.74	0	
562	26	81000	RENT	10.0	EDUCATION	E	21000	16.95	1	
16742	21	40800	RENT	3.0	MEDICAL	D	12000	15.21	1	
24605	33	75000	MORTGAGE	6.0	MEDICAL	A	6000	7.51	0	
29949	44	30000	RENT	6.0	VENTURE	C	5000	13.98	0	
2228	21	31000	MORTGAGE	5.0	PERSONAL	A	6000	7.51	0	
28887	35	123000	RENT	2.0	HOMEIMPROVEMENT	E	25000	17.19	1	
22925	30	60000	MORTGAGE	12.0	DEBTCONSOLIDATION	B	2500	12.69	0	
21944	34	100000	RENT	5.0	PERSONAL	B	6000	10.99	0	
29398	44	36000	RENT	0.0	PERSONAL	B	16000	11.49	1	

Figure 2. Sample of Dataset Used

B. Data Preprocessing

The dataset is preprocessed applying data cleaning, handling missing values, scaling, and encoding categorical variables. It ensures that the dataset is properly formatted and ready for modeling. By identifying inconsistencies or missing information. Data visualization techniques is applied to identify and eliminate unwanted or irrelevant data present in credit databases. Visualizations help illustrate the relationships between various features, such as income, credit history, and employment status, while controlling for other variables. This process aids in identifying significant patterns and interactions within the data, which are crucial for accurate credit risk analysis. The Figure 3 depicts the visualization of data.

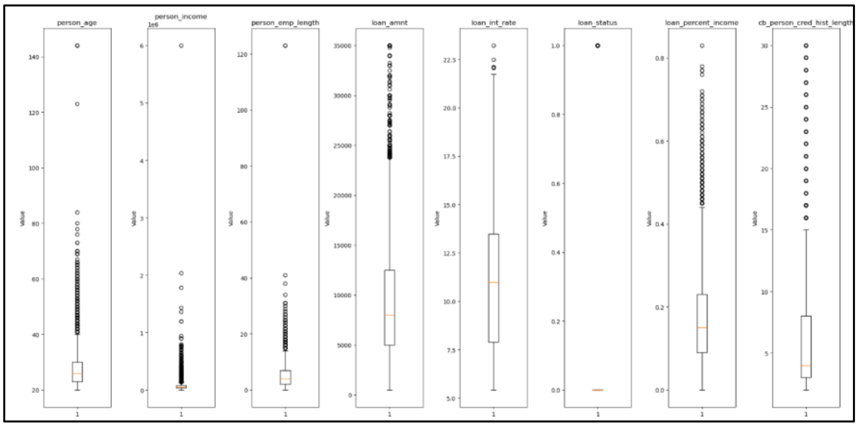


Figure 3. Data Visualization

The Figure.4 Shows the outlier detected for the particular variable in the dataset applying the Elliptical Envelopes and Tukey's method

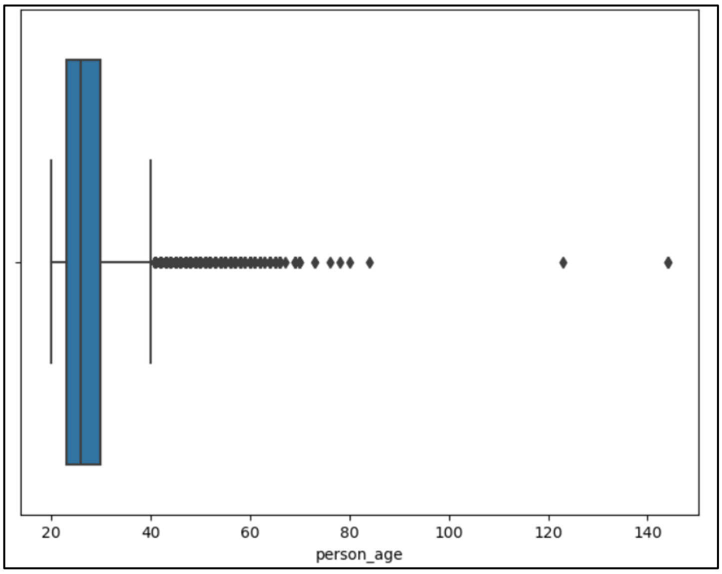


Figure 4. Outlier Detected for Person-Age

C. Machine Learning Algorithms

- **Random Forest Classifier**

Random Forest is a popular machine learning algorithm widely used in credit risk analysis due to its ability to handle complex relationships and provide robust predictions. It constructs multiple decision trees during training and aggregates their predictions to improve accuracy and prevent overfitting. Each tree is trained on a bootstrapped sample of the data and a random subset of features, ensuring diversity among the trees. For Explainable AI (XAI), feature importance in a Random Forest can be analyzed based on the average reduction in impurity (entropy) contributed by each feature across all trees in the forest. Figure 5 shows the tree

- **LightGBM Classifier**

LightGBM (Gradient Boosting Machine) is designed for high efficiency and scalability. It handles large datasets and high-dimensional data faster than many other algorithms, which is beneficial for the extensive and detailed datasets in credit risk analysis. It uses a leaf-wise tree growth strategy, which often results in higher accuracy compared to other tree-based algorithms. Like RF, it provides feature importance metrics, aiding in model interpretability and the identification of key credit risk factors.

- **XGBOOST Classifier**

XGBoost is known for its high predictive performance. It implements gradient boosting with regularization, which improves both the accuracy and robustness of the model. XGBoost can handle missing values and provides various hyperparameters to tune the model, offering flexibility in model building and optimization. XGBoost provides detailed feature importance metrics, which are useful for interpreting the model and understanding the factors driving credit risk predictions. When combined with Explainable Artificial Intelligence (XAI) techniques, XG Boost can provide insights into the model's decisions, making it more interpretable and transparent.

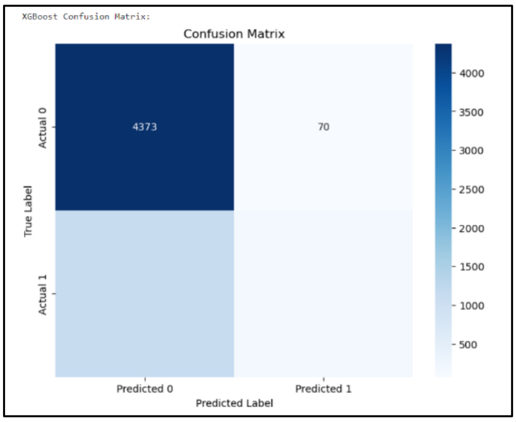
RF, LightGBM, and XGBoost are suitable for credit risk analysis due to their ability to handle complex relationships, efficiency, high predictive performance, and provision of feature importance metrics.

D. Model Integration

LIME and SHAP are integrated to provide interpretability. LIME generates local explanations, offering insights into model decisions at an instance level. SHAP provides a more global perspective by attributing feature importance to predictions, enabling a deeper understanding of the model's behavior.

4. Experimental Result

The proposed method uses various machine learning algorithms to build predictive models. Algorithms such as XGBoost, LightGBM, and Random Forest Classifier are implemented to train models on the prepared datasets. 80% of the dataset is used for training, and 20% is used for testing the models. Each algorithm's suitability and performance score are evaluated. Anaconda Navigator is used to manage environments and packages, ensuring compatibility for machine learning tools, including LIME and SHAP. Jupyter Notebook is utilized for coding, experimenting, and documenting the application of these tools, facilitating interactive development and visualization of model results. Together, they support a comprehensive framework for applying and interpreting machine learning techniques. Figures 5-10 show the confusion matrix and performance scores of the machine learning models.



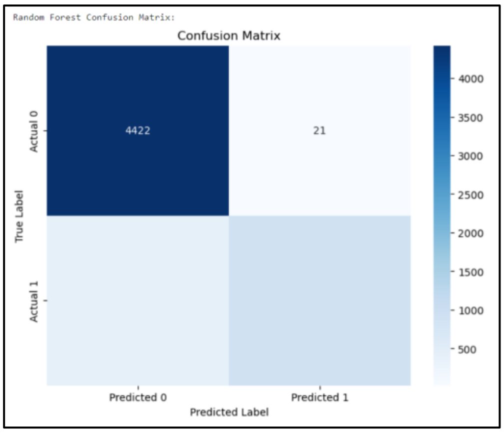

```
XGBoost Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.99      0.97      4443
           1       0.97      0.78      0.86      1285

    accuracy                           0.95      5728
   macro avg       0.96      0.89      0.91      5728
weighted avg       0.95      0.95      0.94      5728

Accuracy: 0.9450069832402235
```

Figure 5. XGBoost Confusion Matrix **Figure 6.** XGBoost Performance Score

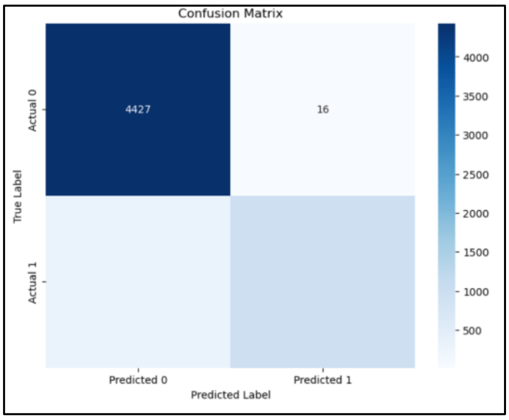



```
Random Forest Classification Report:
              precision    recall  f1-score   support

           0       0.92      1.00      0.95      4443
           1       0.98      0.69      0.81      1285

    accuracy                           0.93      5728
   macro avg       0.95      0.84      0.88      5728
weighted avg       0.93      0.93      0.92      5728
```

Figure 7. Random Forest Confusion Matrix **Figure8.** Random Forest Performance Score




```
LightGBM Classification Report:
              precision    recall  f1-score   support

           0       0.93      1.00      0.96      4443
           1       0.98      0.75      0.85      1285

    accuracy                           0.94      5728
   macro avg       0.96      0.87      0.91      5728
weighted avg       0.94      0.94      0.94      5728

Accuracy: 0.9413407821229051
```

Figure 9. LightGBM Confusion Matrix **Figure10.** LightGBM Performance Score

LIME is a model-agnostic approach used to explain the predictions of machine learning models locally, providing explanations for individual predictions rather than for the model as a whole. SHAP is another model-agnostic method for explaining individual predictions of machine learning models. Figure 11 shows the LIME explanation obtained for the credit risk analysis.

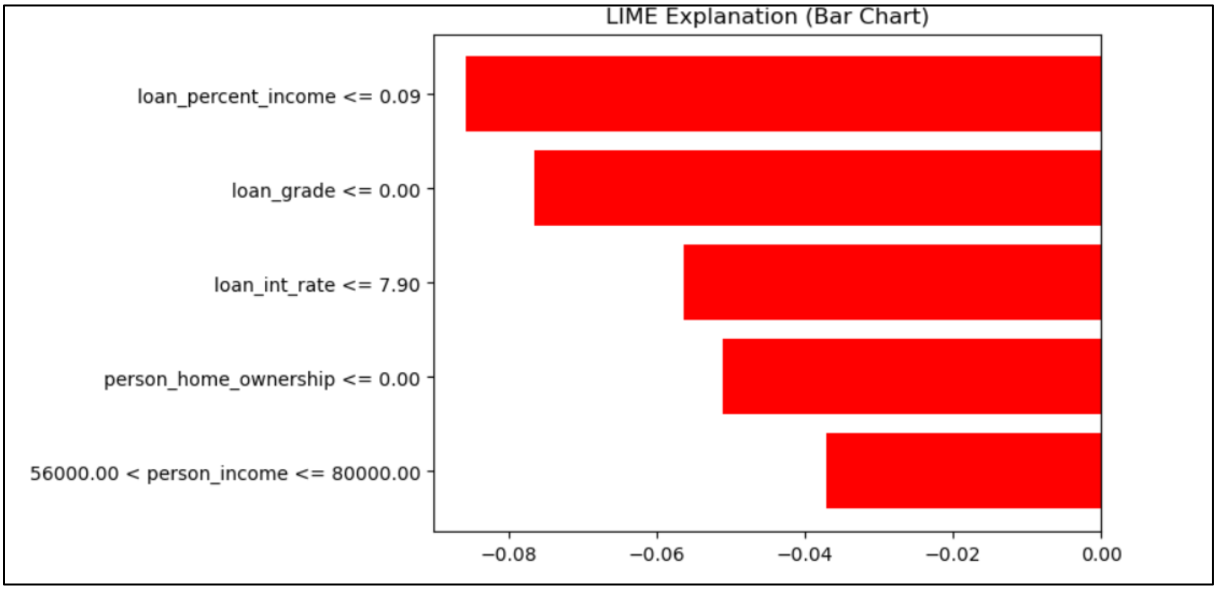


Figure 11. LIME Explanation

By using the Tree Explainer model, we obtain the expected values. To achieve faster runtime, we selected data within a specified range. Warnings are imported, and any warnings that occur will be ignored. The Figure 12 depicts the SHAP inputs.

```
explainer = shap.TreeExplainer(rf_model)
expected_value = explainer.expected_value
if isinstance(expected_value, list):
    expected_value = expected_value[1]
print(f"Explainer Expected Value: {expected_value}")
idx = 100 # row selected for fast runtime
select = range(idx)
features = X_test.iloc[select]
feature_display = X.loc[features.index]
import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)
with warnings.catch_warnings():
    warnings.simplefilter('ignore')
    shap_values = explainer.shap_values(features)[1]
    shap_interaction_values = explainer.shap_interaction_values(features)
if isinstance(shap_interaction_values, list):
    shap_interaction_values = shap_interaction_values[1]
```

Figure 12. SHAP Inputs

The data that are selected in above categories are visualized using SHAP in Figure 13 by this technique we can differentiate the person age and SHAP value for person as shown in Figure 14

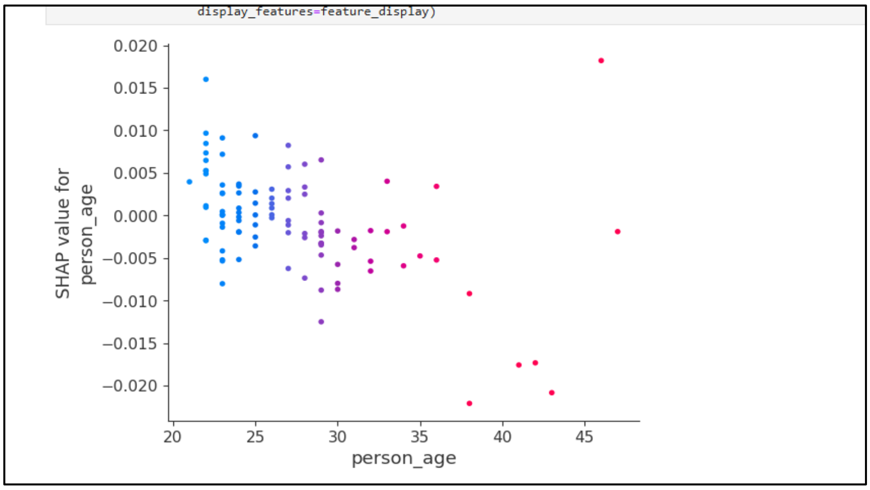


Figure 13. SHAP Outputs

Figure 14 shows the SHAP accuracy calculated for person age, loan percent, person income, person age and loan amount.

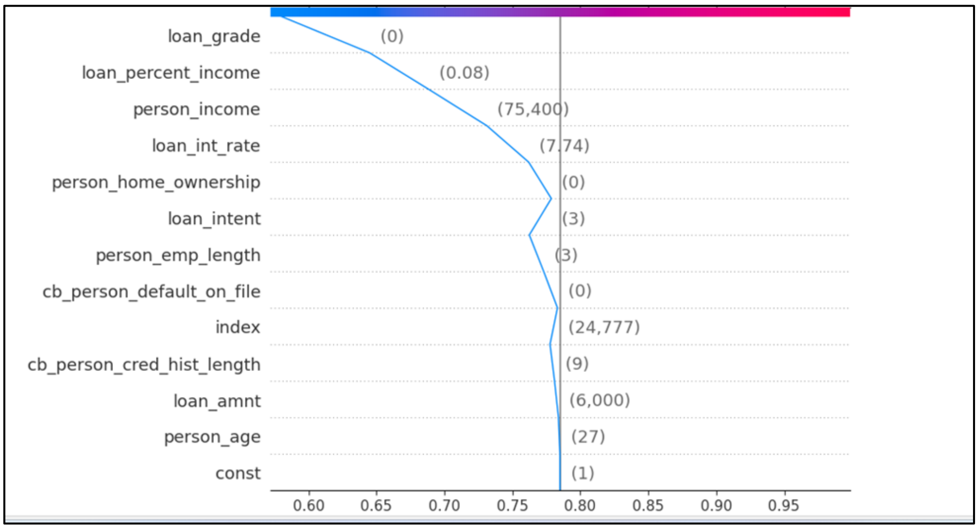


Figure.14. SHAP Accuracy

The Table 1 below shows the accuracy of the machine learning algorithms

Table 1. Accuracy of Machine Learning Algorithms

Machine Learning Algorithms	Accuracy
Random Forest	0.92579

LightGBM	0.94134
XGBoost	0.94500

5. Conclusion

The study presents a comparative analysis of machine learning algorithms—XGBoost, LightGBM, and Random Forest—on the Lending Club loan dataset for credit risk analysis. Addressing outliers using Elliptical Envelopes and Tukey's method significantly improved the models' robustness by mitigating the influence of anomalous data points. Additionally, integrating LIME and SHAP techniques for model interpretability provided valuable insights, enhancing understanding and trust in the model's decision-making.

References

[1] Pandey, Trilok Nath, Alok Kumar Jagadev, Suman Kumar Mohapatra, and Satchidananda Dehuri. "Credit risk analysis using machine learning classifiers." In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India pp. 1850-1854. IEEE, 2017.

[2] Misheva, Branka Hadji, Joerg Osterrieder, Ali Hirsa, Onkar Kulkarni, and Stephen Fung Lin. "Explainable AI in credit risk management." arXiv preprint arXiv:2103.00949 (2021).

[3] Bussmann, Niklas, Paolo Giudici, Dimitri Marinelli, and Jochen Papenbrock. "Explainable machine learning in credit risk management." Computational Economics 57, no. 1 (2021): 203-216.

[4] Bussmann, N., Giudici, P., Marinelli, D., & Papen brock, J. Explainable AI in fintech risk management. Frontiers in Artificial Intelligence, April 2020, vol 3, 26.

[5] Biecek, Przemysław, Marcin Chlebus, Janusz Gajda, Alicja Gosiewska, Anna Kozak, Dominik Ogonowski, Jakub Sztachelski, and Piotr Wojewnik. "Enabling machine learning algorithms for credit scoring--explainable artificial intelligence (XAI) methods for clear understanding complex predictive models." arXiv preprint arXiv:2104.06735 (2021).

[6] Heng, Yi Sheng, and Preethi Subramanian. "A systematic review of machine learning and explainable artificial intelligence (XAI) in credit risk modelling." In Proceedings of the Future Technologies Conference,. Cham: Springer International Publishing, 2022. pp. 596-614

[7] Hu, Bing, and Yi Wu. "Unlocking Causal Relationships in Commercial Banking Risk Management: An Examination of Explainable AI Integration with Multi-Factor Risk Models." Journal of Financial Risk Management 12, no. 3 (2023): 262-274.

[8] De Lange, Petter Eilif, Borger Melsom, Christian Bakke Vennerød, and Sjur Westgaard. "Explainable AI for credit assessment in banks." Journal of Risk and Financial Management 15, no. 12 (2022): 556.

[9] Demajo, Lara Marie, Vince Vella, and Alexiei Dingli. "Explainable ai for interpretable credit scoring." arXiv preprint arXiv:2012.03749 (2020).

[10] Burgt, Joost van der. "Explainable AI in banking." Journal of Digital Banking 4, no. 4 (2020): 344-350.

[11] Gramespacher, Thomas, and Jan-Alexander Posth. "Employing explainable AI to optimize the return target function of a loan portfolio." Frontiers in Artificial Intelligence 4 (2021): 693022.

[12] Fritz-Morgenthal, Sebastian, Bernhard Hein, and Jochen Papenbrock. "Financial risk management and explainable, trustworthy, responsible AI." Frontiers in artificial intelligence 5 (2022): 779799.

[13] Dessain, Jean, Nora Bentaleb, and Fabien Vinas. "Cost of Explainability in AI: An Example with Credit Scoring Models." In World Conference on Explainable Artificial Intelligence, Cham: Springer Nature Switzerland, 2023. pp. 498-516.

[14] Davis, Randall, Andrew W. Lo, Sudhanshu Mishra, Arash Nourian, Manish Singh, Nicholas Wu, and Ruixun Zhang. "Explainable Machine Learning Models of Consumer Credit Risk." Journal of Financial Data Science 5, no. 4 (2023).

[15] Nallakaruppan, M. K., Balamurugan Balusamy, M. Lawanya Shri, V. Malathi, and Siddhartha Bhattacharyya. "An Explainable AI framework for credit evaluation and analysis." Applied Soft Computing 153 (2024): 111307.

[16] https://www.kaggle.com/datasets/utkarshx27/lending-club-loan-dataset?resource=download