Grenze International Journal of Engineering and Technology, June Issue

Explainable AI for Loan Approval Decisions: Integrating Contextual SHAP with Dynamic Temporal Analysis

Muskan Chauhan[1] and Ravinder Kumar[2]
[1-2]Department of Mathematics, University Institutes of Sciences, Chandigarh University, Mohali, Punjab, India-140413
Email: muskanchauhan2207@gmail.com, dr.rkpoonia@gmail.com

***Abstract*— This paper addresses the challenge of achieving both high predictive accuracy and interpretability in AI-based loan approval systems. To solve problems connected with black boxes, a novel approach was introduced C-SHAP-DTA, which includes the elements of Dynamic Temporal Analysis (DTA) and Contextual SHAP (C-SHAP). C-SHAP-DTA understands temporal financial behavior through the means of anomaly detection with autoencoders and provides readable explanations on account of clustering techniques in the field of context. It turns out that the proposed approach is more successful than the alternatives, as it demonstrates 87% accuracy, 89% precision, and the trust score of 4.1 with strong correlation of time. Such results highlight the effectiveness of integrating two important components into a system for evaluating credit risks time dynamics and contextual changes. In practice, the suggested approach can be used not only to make reliable decisions but also improve transparency, personalization, and fairness of operations. Moreover, it helps institutions to build trust, compliance, and increase their overall reliability. As far as performance is concerned, the model demonstrates both accuracy and interpretability by providing time and applicant-specific explanation.**

***Keywords*— Explainable AI, SHAP, Loan Approval, Temporal Analysis, Autoencoder, Anomaly Detection, XGBoost, Credit Scoring.**

I. INTRODUCTION

Systems for loan approval using artificial intelligence have made significant contributions to the prediction of risks associated with loans. The use of artificial neural networks, random forest models, and gradient boosting models allows for capturing the highly non-linear dependencies between features. Nevertheless, the mentioned algorithms are regarded as a black box which does not allow for understanding their decisions, as discussed by Černevičienė and Kabašinskas [1]. The Explainable Artificial Intelligence (XAI) technique serves as a good solution to this challenge. One of the most frequently used approaches in the field is SHAP because of its theoretical background based on cooperative game theory. Lundberg and Lee [2] proposed SHAP and demonstrated its effectiveness in explaining machine learning predictions. Properties such as consistency, local accuracy, and additivity make the technique suitable for application in the financial environment. However, Bussmann et al. [4] highlighted challenges associated with applying explainable AI methods in fintech risk management.

Grenze ID: 01.GIJET.12.2.647_26

Another popular approach is the Local Interpretable Model-Agnostic Explanations (LIME), which is a local interpretability technique applicable to different models. Ribeiro et al. [5] introduced LIME and showed its applicability across different machine learning models. Though it possesses flexibility, the approach is prone to being sensitive to data modifications and may provide different interpretations of similar cases. Attributes like income and credit score vary over time; thus, the significance of temporal modeling in the process of risk assessment has increased. Hochreiter and Schmidhuber [6] introduced Long Short-Term Memory (LSTM) networks, emphasizing the importance of temporal information in sequential data analysis. In the proposed research, temporal behavior is modeled by means of anomalies, which are detected based on features derived from autoencoders using reconstruction errors. On the other hand, anomaly detection techniques have not been combined with an explainability framework yet. Current XAI techniques also fail to capture contextual differences between borrowers. Given the diversity of financial data, borrower profiling may be performed using clustering methods like K-means, originally proposed by MacQueen [8]. The constraints motivate this paper and, therefore, the research is a novel hybrid system for predictive modelling, temporal intelligence, and contextual interpretability called C-SHAP-DTA. The proposed framework introduces three key aspects. First, DTA applies the model to identifying changing risk patterns, rather than fixed snapshots, by using autoencoders to identify temporal anomalies in financial behaviour based on time. Second, C-SHAP builds on the classic SHAP, as it adds feature attributions that can dynamically be altered to suit various groups of applicants, as well as including clustering-based profiling. This is particularly true in financial decision making where the weight of such characteristics as the variability of income or credit history depends on profiles, e.g., salaried and self-employed. Third, Chen and Guestrin [9] introduced XGBoost, which is used as the predictive heart due to its improved performance on structured tabular data and its compatibility with SHAP-based interpretability methods.

II. Literature Review

Recent developments in XAI studies for credit risk prediction concentrate on enhancing both the predictability and interpretability aspects of models. Kumar [10] reported that XGBoost and LightGBM integrated with SHAP achieved better results compared to traditional logistic regression methods. Similarly, Kakkar [11] indicated that gradient boosting methods exhibited high predictive accuracy, while SHAP significantly improved model interpretability.

Hjelkrem and de Lange [12] demonstrated that SHAP provided effective explanations for deep learning-based credit scoring models. Likewise, Li [13] showed that the combination of LightGBM with SHAP and LIME achieved approximately 87% prediction accuracy. Ballegeer et al. [14] noted that although SHAP and LIME enhanced model transparency and predictive performance, they could reduce the stability of explanations. Furthermore, John [15] proposed adaptive SHAP explanations that improved prediction reliability and supported fair evaluation processes.

According to Misheva et al. [16], SHAP provides more accurate and consistent explanations than LIME in credit risk assessment for financial services. In their review of SHAP, LIME, and hybrid explainable AI techniques, Pathi and Pothineni [17] concluded that hybrid explainability models are generally more reliable and easier to interpret.

A recent study by Wang et al. [18] employed a non-parametric oversampling technique combined with SHAP to improve explainability in credit scoring using the GMSC dataset. Moreover, Zhang et al. [19] demonstrated that deep autoencoders can effectively detect financial anomalies through reconstruction errors, making them suitable for identifying risk-related patterns. Finally, Lessmann et al. [20] showed that ensemble and boosting techniques outperform conventional statistical models in credit scoring applications, establishing XGBoost as one of the most effective approaches for explainable AI-based credit risk prediction.

III. Dataset Construction

This paper uses a publicly available dataset from Kaggle under the title “Loan Approval Prediction Dataset,” which contains 4,269 anonymized records of applicants. It is a popular dataset for carrying out studies related to credit risk assessment due to its structured nature of financial attributes, credit attributes, and asset attributes.

This dataset has been chosen for three major reasons. Firstly, it is a structured dataset represented in a tabular format, which is highly compatible with machine learning models like XGBoost. Secondly, it contains some of the most critical financial and credit-related attributes of applicants, such as applicant income, loan amount, and CIBIL score, which are some of the strongest indicators of applicant creditworthiness. Finally, it also contains

asset-related attributes of applicants, which provide a more holistic view of financial conditions of applicants compared to traditional datasets.
While carrying out this paper, emphasis has been placed on attributes such as CIBIL score, loan amount, and applicant income, as these are some of the major determinants of the decision of approving a loan. Furthermore, the attributes are enhanced by incorporating the DTA framework to take into account dynamic financial patterns of applicants. Finally, preprocessing of the dataset is done by handling missing values, encoding categorical variables, and normalizing numerical variables. Temporal sequences are also created for financial attributes to detect anomalies by incorporating the DTA framework.

TABLE I: SUMMARY OF THE DATASET [26]

Feature Name	Feature Description	Data Type
no_of_dependents	Number of dependents supported by the applicant	Numerical
Education	Education level (Graduate / Not Graduate)	Categorical
Self_employed	Employment type (Self-employed or Salaried)	Categorical
income_annum	Annual income of the applicant	Numerical
loan_amount	Total loan amount requested	Numerical
loan_term	Duration of the loan (in months)	Numerical
cibil_score	Credit score indicating repayment reliability	Numerical
residential_assets_value	Value of residential assets owned by applicant	Numerical
commercial_assets_value	Value of commercial assets owned by applicant	Numerical
luxury_assets_value	Value of luxury assets	Numerical
bank_asset_value	Value of bank savings and investments	Numerical
loan_status	Target variable indicating loan approval (Approved / Rejected)	Categorical (Binary)

IV. THEORETICAL FOUNDATIONS

The proposed C-SHAP-DTA framework is grounded in three key theoretical components: Shapley-based feature attribution, clustering-based contextual modelling, and autoencoder-based temporal anomaly detection. These components collectively enable interpretable, context-aware, and temporally sensitive decision-making in loan approval systems.

A. SHAP and Feature Attribution Theory

SHAPley Additive explanations (SHAP) uses cooperative game theory, in which each feature is regarded as a player making its contribution to the final prediction. The feature contribution is measured by the Shapley value, which gives the importance of all features a fair distribution. The feature i Shapley value is given by:

$$\Phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} \; [f(S \cup \{i\}) - f(S)]$$

where $S$is a subset of features and $f(S)$represents the model prediction using features in S. This formulation guarantees properties such as efficiency, symmetry, and consistency, making SHAP a robust and theoretically grounded interpretability method [21].
In this work, feature contributions for every prediction are calculated by applying SHAP to the trained XGBoost model. However, standard SHAP's efficacy in heterogeneous financial datasets is limited by its assumption of uniform feature importance across all instances.

B. Contextual Modelling via Clustering (C-SHAP):

In a bid to overcome the drawback of the static explanations, the framework proposes contextualization using clustering. The clustering of the applicants into homogeneous profiles is based on financial attributes like income and credit history with the help of the K-means clustering method.
The objective of the K-means function is:

$$J = \sum_{k=1}^{k} \sum_{x_i \in C_k} ||x_i - \mu_k||^2$$

After the formation of clusters, the SHAP values are recalculated with the help of cluster-specific weighting and it is possible to create an explanation depending on the profile of an applicant. It is indicative of actual financial

logic: features are not equally important to all groups (e.g., self-employed people are much more concerned with income stability). Contextual adaptation produces greater interpretability and harmonization of explanations with domain specific decision logic [23].

C. DTA using Autoencoders

Financial data are time dependent in nature and it is crucial to record the dynamics of it over time so that the risk assessment is right. The proposed framework uses autoencoders to detect anomalies in financial behaviour in order to model sequence-dependent financial behaviour.

An autoencoder consists of:

Encoder: $h = f_\theta(x)$

Decoder: $\hat{x} = g_\phi(h)$

The reconstruction loss is defined as:

$$L = ||x - \hat{x}||^2$$

where x is the input sequence and $\hat{x}$ is the reconstructed output [24].

The autoencoder is trained on regular financial trends in this research. When reconstruction error is high, this is a sign of anomalies, which are perceived to be temporal irregularities (e.g., sudden drops in income). The model uses these anomaly scores as extra features to allow dynamic risk indicators to be represented.

V. METHODOLOGY

The suggested C-SHAP-DTA framework is designed to be a multi-stage pipeline which consists of a cross of temporal feature extraction, predictive model and context-sensitive interpretability. The methodology will transform the raw financial information into systematic and understandable loan approval decision as follows. The hyperparameters of the proposed framework were optimized using grid search and empirical tuning to achieve improved predictive performance and model stability. The selected configuration provided the best trade-off between classification accuracy and computational efficiency.

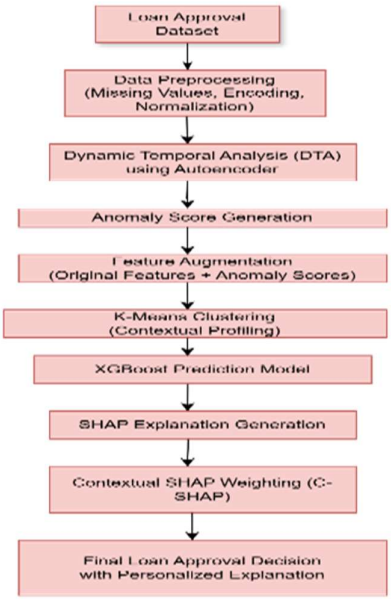

Fig.1: Research Framework Pipeline of the Proposed C-SHAP-DTA Model for Explainable Loan Approval Prediction

A. Data Preprocessing

This data is initially taken through normal preprocessing to maintain consistency and compatibility of the models. Statistical imputation methods are used to fill in missing values, whereas categorical variables are coded through label encoding or one-hot encoding. Numerical values are scaled so that the scales have the same level. This measure will make the data appropriate to the machine learning models and clustering algorithms.

The dataset is split into 80% training data and 20% test data. The performance of the model is validated against the test data to check for generalization and 5-fold cross-validation is also done to check for robustness.

B. DTA Module: Temporal Feature Engineering

To capture various financial behaviors, pseudo-temporal representations are developed using some critical financial attributes like income of the applicant. The abnormal behaviors are then detected by processing these representations with an autoencoder that has been trained on normal financial behaviors.
The reconstruction error obtained from the autoencoder is used to calculate an anomaly score that reflects the variations in financial behaviour. This anomaly score is added as another feature to the model, such as IncomeAnomalyScore, allowing it to identify possible risk patterns related to abnormal financial characteristics.
The autoencoder has been trained with a dense model and the mean squared error loss function for 50 epochs to detect normal patterns of financial behaviors.

C. Predictive Modelling

The classifier XGBoost model uses augmented training data that includes both original features and time series-based anomaly features. The reason behind using XGBoost algorithm is because of its excellent efficiency while working on structured data, ability to handle non-linear feature interaction, and use in SHAP-based interpretability.
For parameter optimization of the model, grid search approach will be adopted to optimize parameters like learning rate, tree depth, and number of estimators. After optimal value selection for each hyperparameter, the model will be implemented with a learning rate of 0.1, tree depth of 6, and 100 estimators for loan approval prediction.

D. Clustering based Contextual Profiling

K means clustering helps put applicants into groups that are similar based on financial details. Income and loan amounts along with credit history are the main things considered. The applicants end up in three clusters for low medium and high risk profiles. This kind of segmentation makes explanations more aware of the context. Feature importance can be different for each group. It seems like this helps with interpretability. Local trends become easier to analyze too. I am not totally sure how much the groups affect everything though.

E. C-SHAP Explanation

SHAP values are calculated for each prediction to determine the feature importance, and they are further fine-tuned using cluster-specific weights to obtain profiled explanations. For instance, one set of applicants might be concerned about income stability, whereas another set might depend on credit history more, leading to more tailored and realistic explanation methods.

VI. Experiments and Results

All the experiments were carried out in a Python-based platform. Jupyter Notebook and PyCharm have been used as development platforms. The implementation of the proposed solution has been carried out through the Anaconda platform by making use of the standard libraries such as Scikit-learn for machine learning, XGBoost for predictive modelling, and TensorFlow/Keras for auto-encoder implementation.
Experimental Results:
The proposed C-SHAP-DTA framework was compared with the baseline explainability methods, which are LIME + Gradient Boosting (GB) and SHAP + Gradient Boosting (GB). The test dataset was evaluated based on the predictive and interpretability measures.

Table II: Results

Method	Accuracy	Precision	Recall	F1-Score	Trust Score	Temporal Corr.
LIME+GB	0.82	0.84	0.81	0.83	3.1	0.65
SHAP+GB	0.83	0.85	0.82	0.84	3.3	0.68
C-SHAP_DTA	0.87	0.89	0.86	0.88	4.1	0.82

The baseline models, namely, LIME + GB and SHAP + GB, are chosen because of their popularity in explainable credit risk modelling.
From the SHAP analysis, it is evident that the most impacting factors in the decision to approve the loan include the CIBIL score, the income of the applicant, and the asset-based features. A high CIBIL score and steady income levels have a positive impact on the loan approval decision, while high loan amount and anomaly-based

features have a negative impact. This is in accordance with the real-world scenario and validates the proposed framework.

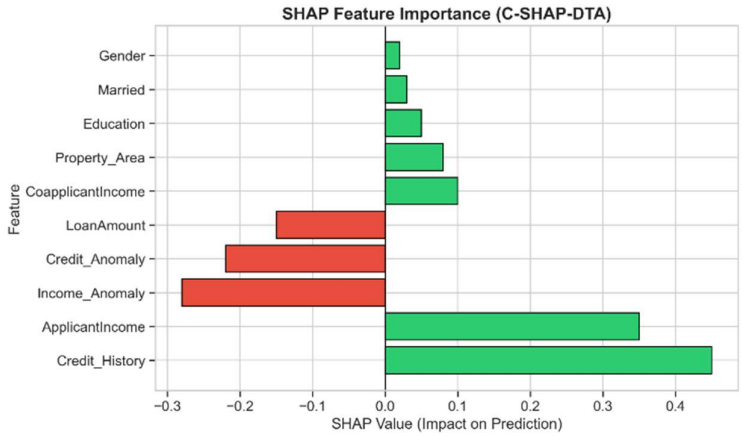

Fig.2: SHAP-Based Feature Importance Analysis Illustrating the Impact of Financial and Anomaly-Based Features on Loan Approval Decisions

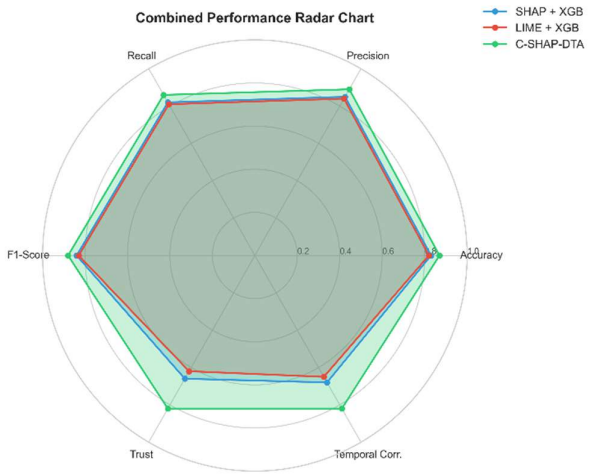

Fig.3: Comparative Performance Analysis of LIME+GB, SHAP+GB, and Proposed C-SHAP-DTA Framework Using Radar Chart Representation

A. Technological Comparative Analysis

LIME with Gradient Boosting

The model offers local explanations; however, the model has poor performance due to its low trust score (3.1), implying that the explanation is unreliable and cannot account for temporal financial behavior.

SHAP with Gradient Boosting

Compared to LIME, SHAP offers more reliable and consistent explanations; hence, the model performs slightly better with a higher trust score (3.3); however, SHAP assumes uniform feature importance and fails to consider temporal changes in the finances.

Proposed C-SHAP-DTA Framework

The proposed model produced the best overall performance after incorporating both temporal and contextual factors into explanation generation. Temporal features using anomaly detection increased risk prediction accuracy (0.87), and contextual SHAP generated consistent explanations with higher reliability based on the higher trust score (4.1). The high temporal correlation (0.82) further highlights the effectiveness of the DTA module.

VII. Conclusion and Future work

This paper is devoted to the problem of immediate high predictive accuracy and interpretability of AI-based loan approval systems. In order to address the shortcomings of black-box models, a new framework, C-SHAP-DTA, was put forward, which combines both DTA and C-SHAP. The framework believes the temporal financial behaviour with the help of autoencoders with anomaly detection and improves readability with the help of cluster-based contextual explanations. The findings show that C-SHAP-DTA is a better method compared to the baseline strategies, with an 87% accuracy, 89% precision, and a trust score of 4.1, and good time correlation. These profits are indicative of the effectiveness of utilizing both time-related and contextual change in assessing credit risks.

Practically, the framework enables making decisions in a transparent, personal, and reliable way and assists financial institutions to improve the trust, fairness, and compliance with the regulations. The model is both accurate and interpretable through being able to provide an explanation which is time and applicant profile adaptable. Altogether, the presented research proves that the use of temporal intelligence in combination with explainable AI is a good line to follow in fintech. This approach can be improved by future work by using real-world longitudinal datasets and sophisticated temporal models to increase robustness and applicability.

However, there are some limitations to the proposed framework. Firstly, the framework is based on the evaluation of a single dataset. This might restrict the generalization of the proposed framework for various financial environments. In the proposed framework, the temporal behaviour is not directly represented by the feature construction method based on anomalies. In the future, the proposed framework should be extended to include real temporal datasets. In the proposed framework, the trust score is based on the subjective evaluation criteria. The model is evaluated on a single dataset, which may limit generalizability. In the future, the framework will be extended to real-life financial data with actual temporal patterns to make accurate predictions of borrower behavior. The integration of sophisticated sequential models like Long Short-Term Memory (LSTM) networks might also improve the temporal modelling of the framework. Furthermore, fairness-based learning methods might be used to mitigate any possible biases in the machine-based loan approval process. Moreover, extensive user studies and domain expert tests will be conducted to test the efficacy of the explanations generated by the framework.

References

[1] Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence in finance: A systematic review. Artificial Intelligence Review, 57(11), Article 297. https://doi.org/10.1007/s10462-024-10854-8

[2] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774. https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[3] Molnar, C. (2022). Interpretable machine learning (2nd ed.). Christoph Molnar. https://christophm.github.io/interpretable-ml-book/

[4] Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2020). Explainable AI in fintech risk management. Frontiers in Artificial Intelligence, 3, Article 26. https://doi.org/10.3389/frai.2020.00026

[5] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144). ACM. https://doi.org/10.1145/2939672.2939778

[6] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[7] Hawkins, D. M. (1980). Identification of outliers. Chapman and Hall. https://doi.org/10.1007/978-94-015-3994-4

[8] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1, pp. 281–297). University of California Press. https://projecteuclid.org/euclid.bsmsp/1200512992

[9] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

[10] Kumar, D. (2025). Explainable machine learning models for credit risk prediction in retail lending. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5341125

[11] Kakkar, S. (2025). Explainable AI models for credit risk scoring in banking. International Journal of Financial Data Science, 3(2), 1–15. https://doi.org/10.34218/IJFDS_03_02_001

[12] Hjelkrem, L. O., & de Lange, P. E. (2023). Explaining deep learning models for credit scoring with SHAP. Journal of Risk and Financial Management, 16(4), 221. https://doi.org/10.3390/jrfm16040221

[13] Li, X. (2025). Explainable AI-based LightGBM model for borrower default prediction. Intelligent Systems with Applications, 27, Article 200514. https://doi.org/10.1016/j.iswa.2025.200514
[14] Ballegeer, M., Bogaert, M., & Benoit, D. (2025). Evaluating the stability of model explanations in credit scoring. arXiv. https://arxiv.org/abs/2509.01409
[15] John, S. (2025). Fair and explainable credit scoring under concept drift. arXiv. https://arxiv.org/abs/2511.03807
[16] Misheva, B. H., Osterrieder, J., Hirsa, A., Kulkarni, O., & Lin, S. F. (2021). Explainable AI in credit risk management. Applied Sciences, 11(16), 7434. https://doi.org/10.3390/app11167434
[17] Pathi, S. P., & Pothineni, J. S. (2025). Interpretable AI in credit scoring: A comparative survey. The American Journal of Engineering and Technology, 7(11), 1–12. https://doi.org/10.37547/tajet/v7i11-304
[18] Wang, Y., Zhang, H., Liu, X., & Chen, J. (2024). A non-parametric oversampling framework for explainable credit scoring. Scientific Reports, 14, Article 78055. https://doi.org/10.1038/s41598-024-78055-5
[19] Zhang, Y., Chen, X., & Li, J. (2022). Anomaly detection in financial time series using deep autoencoders. Expert Systems with Applications, 202, 117128. https://doi.org/10.1016/j.eswa.2022.117128
[20] Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, 247(1), 124–136. https://doi.org/10.1016/j.ejor.2015.05.030
[21] Shapley, L. S. (1953). A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), Contributions to the Theory of Games (Vol. 2, pp. 307–317). Princeton University Press. https://doi.org/10.1515/9781400881970-018
[22] Lloyd, S. P. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489
[23] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052
[24] Sakurada, M., & Yairi, T. (2014). Anomaly detection using autoencoders with nonlinear dimensionality reduction. In Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis (pp. 4–11). ACM. https://doi.org/10.1145/2689746.2689747
[25] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451
[26] https://www.kaggle.com/datasets/architsharma01/loan-approval-prediction-dataset