UDC: 004.8 https://doi.org/10.15407/jai2025.01.132

M. Fostyak[1], **L. Demkiv**[2]
[1,2]Ivan Franko National University of Lviv, Ukraine
50 Drahomanova St., Lviv, 79005
[2]lidia.demkiv@gmail.com
[1]https://orcid.org/0000-0003-4670-5834
[2]https://orcid.org/0009-0002-0185-6364

A DATA-CENTRIC APPROACH TO BUILDING AI MODELS FOR DETERMINING THE CREDIT RATING OF FINTECH COMPANY CLIENTS BASED ON OPEN BANKING

Abstract. This paper implements a data-centric approach to training and testing models that assess a client's credit rating based on their transactions that are obtained using open banking. Well-defined logic rules were used to create synthetic data, and additional synthetic datasets were generated using GANs. The training and testing of a machine learning model (Random Forest) and neural network models implemented in TensorFlow and PyTorch with identical architectures were performed. The process of adapting the models to data updates is examined. Data clustering was used to identify patterns of behaviour among clients who did not repay their loans. The quality of the generated data was evaluated. Classification results are presented depending on the type of model, the amount of data, and the consideration of client behaviour patterns. The results showed that Random Forest models provide the highest accuracy regardless of the type of dataset. The accuracy of training neural networks depends significantly on the structure of the datasets.

Keywords: machine learning, neural networks, synthetic data, generative adversarial networks, open banking, credit rating.

Introduction

In today's world, new financial technologies (fintech) have become an integral part of the investment sector. These technologies contribute to economic growth by increasing the efficiency of financial services. Of particular note is the development of the open banking (Open Banking) concept, which provides access to financial data through standardized APIs. This concept enables a system in which clients, financial institutions, and third-party developers can interact in real-time using a shared financial data repository. The client grants permission to a third-party organization to access their banking data. The bank opens access to the necessary information via a secure API. The third-party uses the data to create personalised services, for instance, offering quick loans. The National Bank of Ukraine has approved the Open Banking Concept, and the Law of Ukraine "On Payment Services" stipulates that open banking will be implemented in Ukraine by August 2025.

Artificial intelligence (AI) plays a key role in the development of the fintech industry. Thanks to machine learning algorithms, neural networks, and generative models, financial companies can automate processes, determine credit ratings, forecast market trends, and make optimal investment decisions. However, the introduction of artificial intelligence in Open Banking faces limited access to real data for training AI models. Furthermore, some important factor combinations may be missing from real datasets due to limited samples or the specific nature of the data. In this context, the need arises for creating synthetic data and artificially generated datasets that retain the statistical properties of real data but do not contain any sensitive information. Synthetic data makes it possible to generate combinations of characteristics not yet observed in real data, to create balanced samples where all possible classes or groups are represented, and to implement specific patterns of customer behaviour. A particularly promising tool for generating these synthetic datasets is generative adversarial networks (GANs), which allow the creation of realistic synthetic data. These data are then used to train models, test their accuracy, and adapt them to new conditions.

This paper discusses a data-centered approach that includes creating synthetic datasets based on client transaction information obtained from Open Banking, for training and testing models that determine a client's credit rating. Implementing a data-centred approach in AI development significantly changes the paradigm of how AI models are handled. Whereas traditionally the focus was on optimizing algorithms, today more attention is being paid to the quality and structure of the data used for training. The relevance of such research is driven by the growth of fintech investment companies at the intersection of technological innovations, where Open Banking, artificial intelligence, and synthetic data lay the groundwork for new opportunities.

Problem Statement

At this stage of development, data is a key asset for decision-making, yet access to high-quality and sufficient volumes of data remains a significant challenge for researchers. This is particularly true in areas such as finance and healthcare, where confidentiality requirements greatly limit data collection and processing capabilities. Building an effective dataset for a new investment fintech company may encounter challenges such as the absence or limited number of real clients. Therefore, it is necessary to explore methods for creating synthetic data that accurately represent real financial data.

One approach is the use of Generative Adversarial Networks (GANs) to generate synthetic data. Additionally, a model needs to be developed for assessing the creditworthiness of clients of a financial investment company that operates under the concept of open banking, which is set to launch in Ukraine in 2025. Real customer data within open banking is not yet available due to the novelty of this concept.

Constructing synthetic data based on a clear logic and generating artificial datasets using modern technologies such as GANs allows for the creation of a representative dataset for modeling, testing, and training decision-making systems. Moreover, a systematic approach for continuously updating data – considering client behaviours such as loan defaults – and adapting credit scoring models must be developed. This involves studying changes in credit rating distributions to minimize financial risks, solving optimization problems, and ensuring a more reliable decision-making process in lending systems.

Literature Review

The rapid growth of information volume is accompanied by the emergence of data that may be incomplete in terms of reflecting key characteristics, interconnections, or contextual features necessary for accurate analysis and decision-making. To address decision-making problems caused by incomplete input data, the study in [1] explores the Belief Rule-Based (BRB) approach. The synthesis of the BRB system integrates rules that use a probabilistic approach to model complex dependencies between parameters. The BRB learning process involves a selection strategy that searches for and optimizes rules, eliminates noise and redundant rules, and combines expert knowledge with data to create rules that account for both individual features and their interactions. However, BRB has limitations related to the complexity of rule updates. To address these issues, an improved version called Dynamic Belief Rule Base (DBRB) was proposed. DBRB dynamically adapts the rule base based on new data and modifies system parameters in response to changing conditions [2]. By integrating optimization algorithms and automated structural updates, DBRB enhances accuracy and stability in complex systems handling large datasets, such as in the financial sector or risk modeling.

In [3], a method combining fuzzy rule-based models with gradient boosting techniques is presented, aiming to balance model interpretability and performance.

Generative Adversarial Networks (GANs) are among the most advanced AI technologies due to their ability to generate high-quality synthetic data that mimic complex real-world data distributions. These synthetic datasets are widely used to build effective machine learning models. The advantages and limitations of generative methods for creating realistic data are thoroughly analyzed in [4].

The study in [5] explores the theory behind GAN and WGAN, comparing classification performance on five datasets using WGAN and the Synthetic Minority Over-sampling Technique (SMOTE). It evaluates classification parameters concerning data feature distributions, generation algorithms, and classification metrics. In three out of five datasets, WGAN outperformed SMOTE.

Reference [6] presents a methodology for assessing the performance of models trained on synthetic data when applied to real-world data. A real dataset comprising one million individuals was divided into three categories: financial data, social network connections, and social attributes. Synthetic data were generated using CTGAN and TVAE. A training framework was developed to analyze models trained on synthetic data and observe their effectiveness on real-world data. Additionally, model performance was monitored a year after training to examine data sensitivity over time. Results indicate that synthetic data quality deteriorates as the number of attributes increases. Credit scoring models trained on synthetic data showed a 3% drop in AUC and a 6% drop in KS compared to those trained on real data.

At the same time, the financial sector requires high-quality synthetic data that align with real customer behaviour patterns while ensuring confidentiality and data protection. The development of specialized models, such as FinGAN, is focused on adapting traditional GAN architectures for financial applications. In [7], FinGAN is introduced as an enhanced Generative Adversarial Network model designed for financial synthetic data generation. This model captures complex patterns in the input data by fine-tuning hyperparameters, including layer depth, neuron configurations, early stopping criteria, learning rate, and activation functions. FinGAN's performance and data generation quality are positively evaluated in comparison with the baseline GAN model. Reference [8] examines the generation of synthetic data using Python libraries such as SynthCity, SDV, DataSynthesizer, Faker, and Metadata-to-Data for various data modalities, including tabular, time series, event sequences, and unstructured financial data from both market and retail applications. The study also investigates the relationship between AUC-ROC scores across different machine learning algorithms and synthetic data generators.

Open banking is evolving in many countries. Reference [9] discusses the impact of alternative data from fintech companies on credit scoring model training. Results from two countries show that these new data sources are particularly useful for predicting financial behaviour among low-income and young individuals, who are more likely to engage with alternative lenders. Data modeling strategies for such companies are proposed in [10,11]. Modern AI and machine learning systems emphasize explainability. The development of explainability metrics (Explainability Metrics) has become increasingly relevant. Tools like SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) help interpret the impact of individual features on model outcomes [12,13].

Reference [14] explores credit card default prediction based on transaction data using a combination of deep learning and Explainable AI (XAI) techniques. The model achieves a credit risk assessment accuracy of 0.8350. The most important features identified by the model include late payments and outstanding bill amounts.

Credit scoring models can be built not only on numerical data but also on textual descriptions of transactions obtained from open banking APIs. In [15], it is shown that a deep learning model trained from scratch outperforms the BERT transformer in accuracy. This is explained by the fact that transaction data consists of text strings without natural sentences, making word context less relevant. Although BERT excels in contextual understanding, this advantage does not significantly improve performance in transaction text analysis.

Research Objective

The objective of this study is to develop and analyze classification models for assessing the credit rating of fintech company clients based on open banking transaction data. A data-centered approach with a well-defined logic will be employed to create datasets that ensure high model accuracy.

In generating synthetic data, statistical indicators of income and expenditure distribution, as well as the percentage-based composition of expenses across different categories in the country, will be considered.

Generative Adversarial Networks (GANs) will be used to generate datasets for training credit rating models. Quantitative metrics, such as the Kolmogorov-Smirnov (KS) test for statistical distribution similarity and the minimal difference between dataset correlation structures, will be applied to evaluate the quality of generated data.

Machine learning models (Random Forest) and neural network models (TensorFlow, PyTorch) will be trained and compared. The study will analyze how classification accuracy depends on the volume of generated data and model parameter settings. Additionally, the impact of real financial transaction behaviour patterns on model accuracy will be assessed.

Considering that approximately 3–5% of clients typically default on loans, the dataset will be updated to reflect this fact. Cluster analysis will be conducted to identify groups of clients with similar financial behavioural patterns. The k-nearest neighbours method will be applied to analyze the characteristics of "problematic" clients. The impact of clustering on credit rating prediction accuracy will be evaluated, and classification accuracy within different client clusters will be analyzed.

Furthermore, the study will investigate how data modifications influence the distribution of credit ratings and how structural changes in datasets affect model prediction accuracy. The relevance of updating datasets with information on clients who received but did not repay loans is driven by the need to minimize financial risks and enhance the effectiveness of creditworthiness prediction models. Incorporating data on such clients allows models to better adapt to real-world conditions, refine credit rating parameters, and identify potential risks at the application assessment stage.

Data synthesis and generation

The data structure is built on the premise that, with the client's consent, financial transaction data is retrieved from an open banking API in JSON format. This file contains detailed information about the client's daily transactions, including the date, amount, transaction category, and additional technical details. To create a dataset in CSV format, the client's daily income and expenses must be anonymized and categorized. As a result, the dataset contains columns belonging to specific categories: regular income (salaries), banking transactions (transfers to and from accounts, servicing other loans, monitoring debt obligations), regular mandatory expenses (rent, utilities, groceries, transportation), high-risk expenses (gambling transactions), low-risk expenses (medical costs, clothing, repairs), non-periodic expenses (travel, luxury goods, electronics), and others. Averaging transaction data over three to six months enables the formation of data for assessing a client's creditworthiness or risk level, which financial institutions use as primary or supplementary information for loan approval decisions. The Fail column represents a binary classification of clients.

The input data is formed based on rule-based logic, incorporating a hidden parameter — the difference between income and expenses (diff), which is not explicitly included as a column in the dataset. A loan can be granted if the difference exceeds diff = 5000. The rules define credit eligibility conditions in percentage terms. For example, high-risk expenses cannot exceed 15% of income, and servicing other loans must not include outstanding debt obligations. The dataset considers the statistical distribution of average salaries and national expenditure statistics, along with customer behaviour patterns. According to statistical data, 60% of income is spent on mandatory payments for individuals earning an average salary. Clients with significantly higher incomes typically engage with financial institutions regarding deposits or investments, so such income levels were excluded. Filters were applied to the dataset to establish dependencies between variables, ensuring realistic behavioural patterns and defining two test behavioural templates for clients, considering that banking transactions generate a significant volume of data with complex functional dependencies.

A well-known statistical fact is that 3–5% of loan recipients default on their payments. Incorporating these clients into the dataset and analyzing their impact on credit rating model accuracy requires additional behavioural pattern analysis. The behaviour of five such clients within a dataset of 100 clients was examined using k-nearest neighbours (KNN) and DBSCAN (Density-Based Spatial Clustering of Applications with Noise). The results showed that four clients were clustered with 21 others who exhibited high non-periodic spending, and one client belonged to a group of eight characterized by two zero-value columns in mandatory payments and high spending in the low-risk expense category.

Thus, two datasets—with and without defaulting clients—can be represented as three groups of clients. The notation introduced includes: Y_0 and Y_{01} for the first behavioural pattern without and with defaulting clients, respectively; Z_0 and Z_{01} for the second behavioural pattern without and with defaulting clients, respectively; and X_{01} as the input data. Two datasets A (with defaulters) and B (without defaulters) can be represented as: $A=X_{01}+Y_{01}+Z_{01}$ and $B=X_{01}+Y_0+Z_0$.

Correlation matrix analysis of the datasets revealed that the Fail column has the highest correlation (0.4) with overdue payments, while regular income shows a maximum correlation of 0.4 with mandatory expenses. The highest correlation coefficient (0.8) was found in columns related to high-risk and low-risk expense behavioural patterns.

To obtain a larger numerical synthetic dataset (excluding categorical, identification, and ordinal data), data generation was implemented using Generative Adversarial Networks (GANs). The YData Synthetic tool, which specializes in tabular data, was used for this purpose. YData Synthetic is optimized for generating data that accurately reproduces the distributions, structure, and correlations of the original dataset. It also effectively generates data for tasks with imbalanced classes while preserving feature relationships through various models, including GAN, CGAN, WGAN, and WGAN-GP. The ydata-sdk version features an advanced API and automated selection of the best generative model for the user's data.

The quality of the generated synthetic data was evaluated using the Kolmogorov-Smirnov (KS) test, correlation matrix differences, and functional metrics that measure the extent to which synthetic data improves classification model performance for credit rating assessment. The KS test results indicate that for all variables (except Fail), the differences between real and synthetic data do not reach statistical significance. However, for the Fail variable, real and synthetic distributions differ significantly (KS statistic = 0.235, P-value = 0.003) due to class balancing during generation. The average absolute difference between correlation matrices of real and synthetic data is 0.11, indicating high accuracy in data generation. However, for a small number of variable pairs related to high-risk expenses and specific customer behavioural patterns, the maximum correlation difference reached 0.22.

These findings confirm that GAN-generated synthetic data effectively replicate the structural dependencies of the input data used for model testing.

Classification models

As mentioned earlier, statistical data indicate that 3–5% of clients who receive a loan fail to repay it. Therefore, we analyze datasets where defaulting clients are absent (B) and those that include such clients (A). Table 1 presents the training accuracy results of the Random Forest model and neural networks implemented using TensorFlow and PyTorch on GAN-generated datasets of varying sizes.

The training accuracy on dataset A (with 4000 rows) for neural networks implemented in TensorFlow and PyTorch is 5–6% lower than the accuracy for dataset B. The neural network models in TensorFlow and PyTorch have identical architectures, consisting of four fully connected layers. The first three layers contain 128, 64, and 32 neurons, respectively, with ReLU activation functions. The output layer consists of a single neuron with a sigmoid activation function, which represents the neural network's output as a probability of the credit rating required for binary classification.

Among machine learning algorithms such as XGBoost, AdaBoost, and Random Forest, Random Forest with max depth=20 demonstrated the highest training accuracy for both datasets A and B, even with the smallest amount of generated data. Furthermore, Random Forest exhibited the smallest difference in training accuracy between datasets of different sizes (A and B). This difference can be attributed to the structured nature and limited volume of the data. As dataset size increases, accuracy improves and reaches a maximum value of 0.99.

Table 1. Dependencies of model training accuracy implemented in Random Forest, TensorFlow, and PyTorch based on dataset size

	RandomForest		**TensorFlow**		**PyTorch**	
	Data sets					
Amount of Data	**A**	**B**	**A**	**B**	**A**	B
250	0.91 ± 0.03	0.95 ± 0.01	0.79 ± 0.05	0.91 ± 0.03	0.64 ± 0.04	0.94 ± 0.01
1000	0.98 ± 0.01	0.99 ± 0.01	0.88 ± 0.02	0.95 ± 0.02	0.87 ± 0.02	0.95 ± 0.02
2000	0.98 ± 0.01	0.99 ± 0.01	0.89 ± 0.02	0.96 ± 0.01	0.87 ± 0.01	0.95 ± 0.01
4000	0.99 ± 0.01	0.99 ± 0.01	0.92 ± 0.02	0.98 ± 0.01	0.92 ± 0.02	0.97 ± 0.01

The increase in training accuracy of the TensorFlow neural network for both datasets A (curve 1) and B (curve 2) with the growth of GAN-generated data (number of clients N) is illustrated in Figure 1.

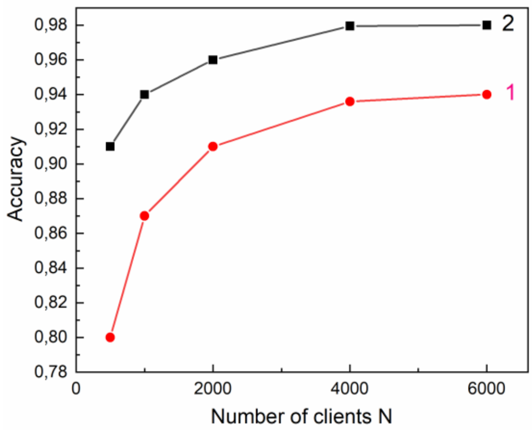

Fig. 1. Dependence of TensorFlow neural network training accuracy for bank client credit rating assessment on the amount of data N generated using GAN

Increasing the dataset size provides the model with more information about data patterns and improves accuracy. The model trained on dataset B (curve 2) shows higher accuracy compared to the model working with the full dataset A (curve 1). For dataset B (curve 2), accuracy stabilizes at smaller dataset sizes (~4000 samples) compared to dataset A (curve 1), reaching a value of 0.98. Accuracy initially increases faster on dataset A, but stabilizes at a lower level (~0.92). These patterns are due to the fact that dataset A (curve 1) has a more complex structure compared to dataset B (curve 2), which is more homogeneous and less variable. Because of this, models for dataset B achieve higher accuracy more easily and outperform all other classification metrics compared to dataset A (Table 2). Table 2 presents classification metrics after testing models built on datasets B, A-Z_{01}, and A. From Table 2, it is clear that the first model demonstrates the best results across all parameters. It is the most efficient in terms of precision, recall (ability to detect true positives), and also excellently distinguishes between classes (AUC-ROC = 0.97). The third model has the lowest values for all metrics. Key drawbacks include a recall lower than precision and a gap between test and validation accuracy. This means the model misclassifies some positive examples, which is a significant downside, especially in the financial sector. The second model outperforms the third on all parameters, especially recall. The significant increase in recall is related to the removal of the Z_{01} client cluster from the dataset.

Figure 2 illustrates the dependence of training accuracy (Accuracy) and loss (Loss) of the TensorFlow-based model on the number of epochs (N_e) for datasets A (curve 3), B (curve 1), and A-Z01 (curve 2), each with N = 2000 samples. A common trend across all datasets is the steady increase in accuracy and the gradual decrease in loss with each training epoch. This dynamic indicates an efficient learning process, where the model progressively adapts to the training data while maintaining its ability to generalize to the validation set.

Table 2. Classification Model Performance Metrics

Model	Data	precision	recall	f1 score	AUC-ROC	accuracy test	accuracy validation
1	B	0.91	0.98	0.94	0.97	0.96	0.96
2	A-Z_{01}	0.88	0.83	0.86	0.90	0.88	0.92
3	A	0.87	0.76	0.82	0.87	0.83	0.90

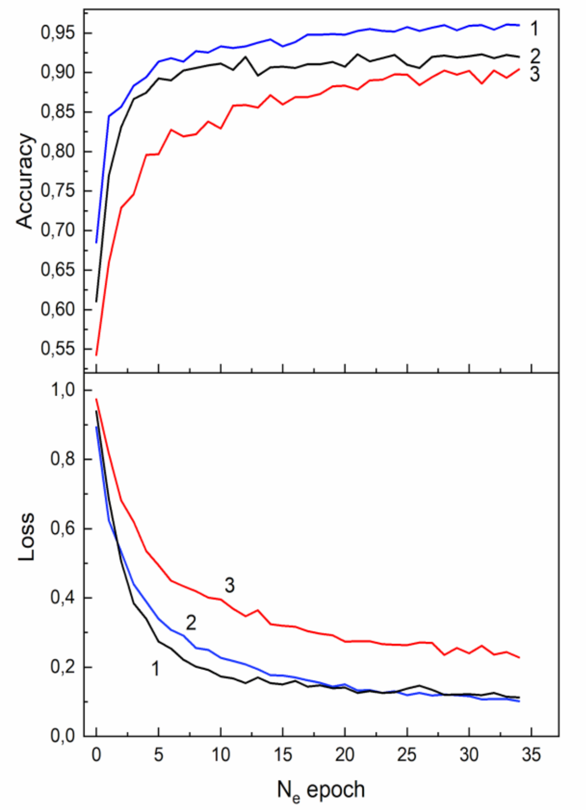

Fig. 2. Dependence of TensorFlow model training accuracy (Accuracy) and loss (Loss) on the number of epochs N_e.

The model trained on dataset A requires a greater number of epochs compared to B and A-Z_{01} before accuracy reaches saturation. Similarly, after a certain number of epochs, the loss values stabilize, indicating that the minimum training error has been reached. There are no sections where loss increases with additional epochs, suggesting that overfitting does not occur. Analysis shows that the number of epochs significantly influences optimal model tuning. At N_e = 25, model accuracy remains consistently high, and loss remains low across all datasets.

Figure 3 presents histograms of credit rating distributions for the test dataset. The predicted numerical values of credit ratings for the test set of 100 clients were obtained using the Random Forest machine learning model, trained on datasets A and B, respectively. The credit rating distribution histogram allows for an analysis of the frequency of different rating values among clients and can be used to construct an optimization model for credit rating assessment.

From Figure 3, it is evident that training the model on dataset B (curve 1) results in a more localized distribution of credit ratings around maximum and minimum values compared to the distribution obtained from training on dataset A (curve 2). The more uniform distribution of credit ratings (curve 2) is caused by the lower accuracy of the model due to the inclusion of defaulting clients in the dataset, who exhibit the same behavioural patterns as clients who successfully repaid their loans.

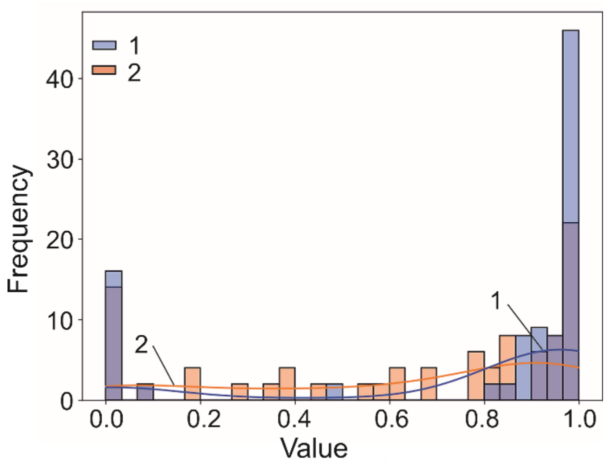

Fig. 3. Histogram of credit rating distributions for test data

For datasets A and B, generative adversarial networks and machine learning models demonstrate the ability to generalize information. A small number of test data points received a credit rating value of 0.5 (curve 1), despite the high accuracy of the model trained on dataset B. These credit rating values are associated with data where the difference (diff) is closest to the loan rejection threshold diff < 5000. The diff column and the loan rejection condition were not explicitly included in the datasets. However, the analysis of results indicates that the model successfully identified this condition and accounted for it when constructing data models based on generalized characteristics across income and expense categories.

Conclusions

The data-centered approach involves creating high-quality datasets that capture the maximum number of behavioural patterns observed in real data. This study proposed a method for generating datasets for credit rating models based on client transactions obtained from open banking. A combination of rule-based logic and GANs was used to generate synthetic data, which was then employed to train artificial intelligence models for credit rating assessment.

A comparative analysis of classification models applied to generated datasets with different behavioural patterns was conducted. The Random Forest machine learning method demonstrated the best accuracy and generalization ability, particularly on dataset B, which had a simpler structure and greater homogeneity. TensorFlow and PyTorch neural networks also performed well, but their accuracy was highly dependent on dataset size and structural complexity. Models trained on dataset A exhibited lower accuracy due to its higher complexity. The credit rating distribution histogram confirmed that models trained on dataset A produced a more uniform distribution of client credit ratings.

The study results emphasize the importance of generating high-quality synthetic data and fine-tuning models to improve classification accuracy. The identified patterns can be valuable for enhancing risk assessment credit systems in the banking sector by leveraging open banking data processing models.

References

1. Fu Y-G., Fang C., Liu Y. (2023) Disjunctive belief rule-based reasoning for decision making with incomplete information, Information Sciences, V. 625, P. 49 – 64. https://doi.org/10.1016/j.ins.2023.01.010
2. Gao F., Zhang A., Bi W., Ma J. (2021) A greedy belief rule base generation and learning method for classification problem, Applied Soft Computing, V. 98, 106856. https://doi.org/10.1016/j.asoc.2020.106856
3. Li J., Liu P, Chen L., Pedrycz W., Ding W. (2024). An Integrated Fusion Framework for Ensemble Learning Leveraging Gradient-Boosting and Fuzzy Rule-Based Models. IEEE Transactions on Artificial Intelligence. V:5, I:11. DOI: 10.1109/TAI.2024.3424427
4. E. Strelcenia, S. Prakoonwit , "A Survey on GAN Techniques for Data Augmentation to Address the Imbalanced Data Issues in Credit Card Fraud Detection," Mach. Learn. Knowl. Extr. 2023, 5(1), 304-329, [Online]. Available: https://doi.org/10.3390/make5010019
5. Ngwenduna K.S., Mbuvha R. (2021) Alleviating Class Imbalance in Actuarial Applications Using Generative Adversarial Networks. Risks 9(3). 49, https://doi.org/10.3390/risks9030049
6. Muñoz-Cancino R., Bravo C., Ríos S., Graña M. (2022). Assessment of creditworthiness models privacy-preserving training with synthetic data. Quantitative Finance. Risk Management. arXiv:2301.01212. https://doi.org/10.48550/arXiv.2301.01212
7. Ramzan F., Sartori C., Consoli S., Recupero D.R. (2024) Generative Adversarial Networks for Synthetic Data Generation in Finance: Evaluating Statistical Similarities and Quality Assessment. AI 2024, 5(2), 667-685. https://doi.org/10.3390/ai5020035
8. Potluru V.K., Borrajo D. (2024) Synthetic Data Applications in Finance. Computer Science. Machine Learning. arXiv:2401.00081 https://doi.org/10.48550/arXiv.2401.00081
9. Roa L., Correa-Bahnsen A., Suarez G., Cortés-Tejada F. (2021) Super-app behavioral patterns in credit risk models: Financial, statistical and regulatory implications. Expert Systems with Applications. V. 169, 114486. https://doi.org/10.1016/j.eswa.2020.114486
10. Fostyak M., Demkiv L. (2024) Development of data mesh data platform with ml domain of data analysism, Electronics and information technologies. 2024. 27. doi: https://doi.org/10.30970/eli
11. Fostyak M., (2024) Development of an AI domain in a data mesh network for customer credit classification using transaction data, IEEE 19th International Conference on Computer Science and Information Technologies (CSIT) (in press).
12. Aas K., Jullum M., Loland A. (2021) Explaining individual predictions when features are dependent: More accurate approximations to Shapley values, Artificial Intelligence. V. 298, 103502. https://doi.org/10.1016/j.artint.2021.103502
13. Branka Hadji Misheva, Joerg Osterrieder, Ali Hirsa, Onkar Kulkarni, Stephen Fung Lin (2021) Explainable AI in Credit Risk Management. arXiv:2103.00949. https://doi.org/10.48550/arXiv.2103.00949
14. Talaat, F.M., Aljadani, A., Badawy, M. et al. Toward interpretable credit scoring: integrating explainable artificial intelligence with deep learning for credit card default prediction. Neural Comput & Applic 36, 4847–4865 (2024).

https://doi.org/10.1007/s00521-023-09232-2

15. L. O. Hjelkrem, P. E. Lange, (2023) Explaining Deep Learning Models for Credit Scoring with SHAP: A Case Study Using Open Banking Data," J. Risk Financial Manag. 16(4), 221; https://doi.org/10.3390/jrfm16040221

The article has been sent to the editors 30.01.25.
After processing 28.02.25.
Submitted for printing 30.03.25.