КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 2 (34), 2026
ISSN 2663 – 4023
DOI 10.28925/2663-4023.2026.34.1241
UDC 004.85:336.717

George Gaina
Candidate of Technical Sciences, Prof.
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
ORCID:0000-0003-0260-0950
ggaina@knu.ua

Dmytro Masiuk
PhD student
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
ORCID: 0009-0002-0840-5558
Masiukdmitry96@gmail.com

GAN FOR DATA SYNTHESIS WHILE MINIMIZING PREDICTIVE POWER LOSS

Abstract. In the modern financial environment, the ability to leverage data for developing intelligent risk detection methods is a huge ethical challenge. However, this does not make it less of a necessity. That is something I need to solve for my research as well. The quality and quantity of the data I would be able to employ for my research hinges on how good would be my anonymization proposal. Anonymization is the process of transforming datasets to prevent identification of individuals or organizations. It has become of paramount importance in order to ensure privacy and compliance in experiments involving sensitive financial information. As regulatory frameworks such as the General Data Protection Regulation (GDPR) in the European Union and similar local laws worldwide are on the rise, ethical approaches to anonymization are no longer optional but essential to achieve legal compliance and to conduct responsible research. Advances in machine learning and big data analytics have increased the likelihood of re-identification, especially when anonymized data is combined with external datasets [1]. This raises critical questions about whether traditional anonymization techniques remain sufficient in high-risk financial contexts and can be still used to full extent. However, it is known that over-anonymization will degrade data quality, sometimes severely, limiting the ability of machine learning algorithms to detect necessary patterns in order to locate fraudulent transactions efficiently. Ethical practice requires balancing privacy protections with the need for accurate and efficient models. Finding this balance is proving to be even more challenging as the demand for the data gets bigger. Removing key values will unintentionally bias models, particularly in fraud detection, potentially worsening their performance as well as their interpretability.

Keywords: GAN, anonymization, financial risks, machine learning, confidentiality, data quality, fraud.

INTRODUCTION

The rapid digitalization of financial ecosystems has led to an unprecedented volume of online transactions and a corresponding surge in fraudulent activities. If before in order to conduct financial fraud perpetrator would need to visit a physical brick-and-mortar establishment, now they can do so in a matter of a few clicks. Modern fraud schemes are increasingly adaptive, leveraging automation, synthetic identities, and behavioral camouflage to bypass conventional rule-based and machine-learning detection systems. Traditional fraud detection approaches, such as supervised classifiers trained on historical labeled data suffer from two fundamental limitations: the extreme scarcity and concept drift of both fraudulent and normal data examples, and the lack of proactive mechanisms to anticipate new attack strategies. As a result, the arms race between fraudsters and antifraud systems demands methods capable not only of detection, but also of strategic simulation and anticipation of adversarial behavior. There is also a big question of which data can be used to train and test models. Financial sector is heavily regulated and most of the time transaction logs in their raw form cannot be used. For example, in Europe GDPR law makes sure that there need to be a lawful explainable basis to use data collected from customers, and PCI DSS regulation makes it so that you are unable to just store any data needed for subsequent usage. Thus, anonymization of the data becomes a mandatory step in any fraud prevention pipeline.

A novel technique named SEAL [2], which stands for Self-Refining Small Language Model Anonymizers, trains smaller language models (SLMs) to anonymize text effectively, without relying on expensive or external large models like ChatGPT, version 4. These SLMs are capable of not only anonymizing, but also critiquing their own outputs, that way iteratively improving privacy. What is even more impressive, 8-billion parameter models are able to achieve privacy–utility trade-offs comparable to GPT-4 and can even surpass it in some scenarios. For example, on the synthetic dataset SynthPAI which consists of personal profiles with text comments, an 8B SLM trained using SEAL achieves a privacy-utility balance on par with GPT-4 anonymizer. With further self-refinement, it even surpasses GPT-4 in privacy – while retaining high readability [2]. This method employs a technique called "adversarial distillation". Firstly, LLM anonymizer processes original text. Secondly, another model tries to guess private parameters that were previously anonymized. Thirdly, yet another model measures readability, semantic preservation and privacy based on previous results. Basically, these models select a "trajectory" of sorts, which aims to increase privacy, as well as maintain readability and preservation. Afterwards those models are being fine-tuned controllably in order to prefer anonymization with stronger privacy and better utility, refining its outputs iteratively. What is also important, this method is capable of context-aware anonymization. That means it effectively hides not just names or numbers, but hints and patterns that can be exploited by sophisticated models. In my opinion this quality is critical for keeping transactional data safe.

There is also an approach of using GAN, which were originally used to create pictures. However, they have long been adapted to work with tabular data as well and as good. Which bring an idea, that such models can be used to synthesize data from real one, thus both "pseudo-anonymizing" and creating new data to train predictive models on. Generative Adversarial Networks (GANs) offer a powerful approach for modeling complex, high-dimensional data through the method of an adversarial learning. In the context of fraud detection, GANs have traditionally been used for synthetic data augmentation – generating realistic but anonymized fraudulent transactions to mitigate class imbalance and improve model generalization. Recent advances in technology, however, suggest that the adversarial nature of GANs can be extended beyond data generation toward strategy generation: simulating the evolving tactics of fraudsters and the corresponding counter-strategies of antifraud systems. In this emerging framework, the generator can be viewed as an intelligent agent, a "white hacker" of sorts – producing fraudulent transaction patterns or behavioral strategies designed to evade detection, while the discriminator acts as a proxy for the antifraud system, learning to distinguish genuine behavior from the fraudulent one.

Through iterative adversarial training, both agents evolve simultaneously – mirroring the real-world dynamics of attack and defense in financial ecosystems. This interaction enables the creation of synthetic but plausible fraud scenarios that expose vulnerabilities in existing rules and models, guiding the design of more robust and most importantly adaptive antifraud mechanisms. The concept of GAN based fraud strategy generation introduces several research challenges. These includes maintaining semantic validity and compliance with business constraints in generated samples, preventing overfitting and information leakage from sensitive datasets, and defining quantitative metrics for "strategic realism." Furthermore, integrating such generative agents into operational risk management pipelines raises ethical and regulatory considerations related to privacy, explainability, and responsible AI. The potential of GAN architectures for fraud strategy generation, outlining their conceptual foundations, experimental design, and implications for the next generation of proactive antifraud systems is going to be explored in future works. GANs are not typically used as the final fraud classifier. However, they can be used for the following tasks.

Generating realistic synthetic fraudulent transactions to augment scarce fraud labels

Producing adversarial fraud examples to harden detectors

Simulating attacker strategies, when combined with r conditional generation, to discover weak spots in rules/ML models

Modeling legitimate/fraud distributions to detect anomalies (e.g., GAN anomaly detectors).

In this paper the point of interest lies in synthesizing data, as this is a crucial problem given the scarcity of financial data and severity of financial regulations. GAN is proposed to be used not as a full on "strategy generator" yet, but as a training data synthesizer. This is a first step, which could solidify the promise of such methods and open a new avenue in fraud detection.

GAN consist of a generator G with maps latent noise z to samples x, and a discriminator D estimating the probability if the sample is real.

$$G_0: \mathrm{Z} \rightarrow \mathrm{X} \tag{1}$$

$$z \sim p(z) \tag{2}$$

$$\hat{x} = G_0(z) \quad (3)$$

They basically play a minimax game, with a canonical minimax objective, as stated in a paper [5]:

$$\min_{\theta} \max_{\varphi} E_{x\sim p_{data}}[\log D_{\varphi}(x)] + E_{z\sim p(z)}[\log(1 - D_{\varphi}(G_0(z)))]. \quad (4)$$

However, for the specific case that is being reviewed in this paper, there a much better suited variety of this model class, CTGAN or tabular GAN. The papers on CTGAN[6, 7, 8, 9] identifies tabular-specific challenges: mixed data types, non-Gaussian and multimodal continuous distributions, sparse one-hot vectors, and highly imbalanced categorical columns that exacerbate mode collapse. These issues directly map to fraud settings – rare class and heterogeneous patterns, even when features are largely numeric. Thus, such CTGAN is proposed as a best solution for the task.

METHODOLOGY

1. Data Preparation. The study utilizes a labeled transactional dataset containing both legitimate and fraudulent operations. For this experiment we are going to use an open source Kaggle dataset with credit card fraud. The dataset is divided into training and testing subsets, following the standard data preprocessing procedure, where the test set consists exclusively of real-world observations and is held out throughout the experiment to ensure unbiased evaluation. Standard preprocessing procedures are applied, including feature scaling and formatting of categorical and numerical variables to ensure compatibility with both machine learning models and generative algorithms.

2. Baseline Model. As a reference point, a supervised classification model is trained on the real training dataset. The model aims to distinguish between fraudulent and legitimate transactions. Performance is evaluated on the real test set using metrics appropriate for imbalanced classification tasks, including the Area Under the Receiver Operating Characteristic Curve (ROC-AUC) and the Area Under the Precision-Recall Curve (PR-AUC). This configuration serves as the baseline for subsequent comparisons.

3. Synthetic Data Generation. To assess the utility of synthetic data, a generative model based on CTGAN is employed. The model is trained on the real training dataset to learn the joint distribution of features. Special attention is given to correctly specifying metadata, including the distinction between categorical and continuous variables, to ensure the quality of generated samples. Following training, the generative model is used to produce a synthetic dataset that replicates the structure of the original data. The generated data maintains the same schema and feature space as the real dataset, allowing it to be used interchangeably in downstream tasks.

4. Experimental Design. To evaluate the effectiveness of synthetic data, three experimental setups are considered:

a) Real-only scenario. The classification model is trained on real training data and evaluated on the real test set.

b) Synthetic-only scenario. The model is trained exclusively on synthetic data generated by the generative model and evaluated on the real test set. This setup assesses whether synthetic data alone can capture sufficient information for fraud detection. This is a scenario which we are the most interested in.

c) Hybrid scenario. The model is trained on a combined dataset consisting of both real and synthetic observations. Evaluation is again performed on the real test set. This setup examines whether synthetic data can enhance model performance when used alongside real data.

All experiments are conducted using the same simple model architecture and evaluation protocol to ensure comparability of results. The model itself is inconsequential to the experiment, so we are going to use something simple.

5. Evaluation Metrics. Given the highly imbalanced nature of fraud detection tasks, model performance is assessed using not one, but multiple metrics. ROC-AUC is used to measure the overall ranking capability of the classifier, while PR-AUC is emphasized as a more informative metric for minority class detection. These metrics provide complementary perspectives on model effectiveness, particularly in scenarios of fraud detection in e-commerce, where accurate identification of rare fraudulent events is critical.

RESULTS

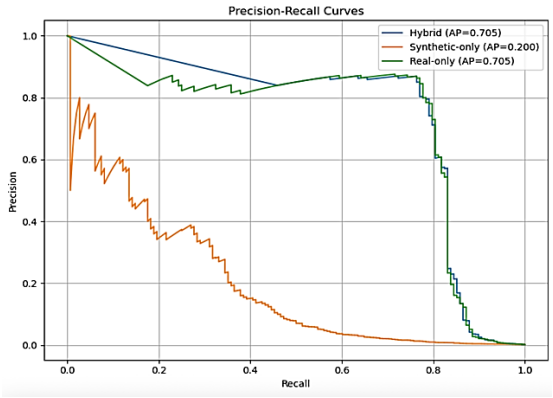

Figure 1. Precision-Recall curves

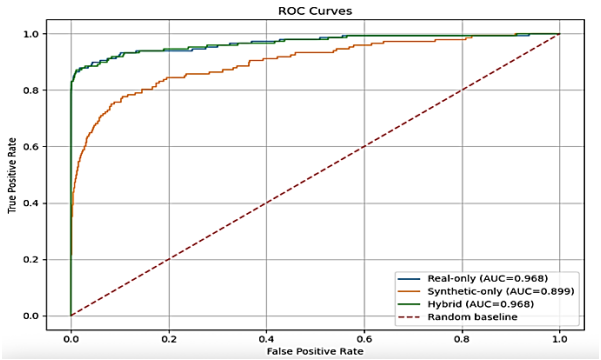

Figure 2. ROC-AUC curves

Table 1

Precision-recall for models trained on different data

Recall	Hybrid Precision	Real-only precision	Synthetic-only precision
0.00	1.00	1.00	1.00
0.05	0.98	0.98	0.75
0.10	0.96	0.92	0.58
0.20	0.92	0.84	0.38
0.30	0.88	0.83	0.32
0.40	0.85	0.82	0.15
0.50	0.86	0.86	0.08

It can be seen that models trained on synthetic-only data significantly loses on both metrics, which unfortunately requires more tuning in order to be used as a full-on data synthesizer.

However, the model trained on real data augmented with synthetic one shows better results in terms of precision, albeit marginally and similar results in terms of ROC-AUC scores.

CONCLUSION

The analysis of the precision recall curves demonstrates a clear difference in performance depending on which dataset the model was trained – real, synthetic, or hybrid. The model trained exclusively on synthetic data shows significantly worse performance, with precision rapidly declining as recall increases. This shows that the synthetic data created by the GAN fails to adequately and wholly capture the underlying distribution and complexity of real-world transaction patterns, leading to a bad generalization. In contrast, both the real-only and hybrid approaches show nearly identical performance across the majority of the recall range, maintaining high precision values (approximately 0.85-0.88) up to a recall level of around 0.75-0.80. This suggests that the inclusion of synthetic data in the hybrid approach does not degrade model performance, and also provides an improvement over training on real data alone, albeit on low recall values. From a practical perspective the results

indicate that synthetic data, in its current form, is not good enough to replace real data in fraud detection tasks just yet. However, it can still be useful as an additional resource, particularly in scenarios where real data is limited or subject to strict privacy constraints. The hybrid approach, demonstrates stability and preserves predictive quality, while also improving on the baseline, making it a viable option for augmenting datasets without introducing risk. The tuning avenues that are going to be explored in further papers:

- Increasing number of epochs above 1000 on a more potent machines
- Increasing discriminator step parameter, the number of times discriminator updates per generator cycle
- Using conditional sampling on fraud class, to counter severe class imbalance.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review, 57*(6), 1701-1777.
2. Kim, K. (2025). *Self-refining language model anonymizers via adversarial distillation*.
3. Rocha, Á. (n.d.). *Generative AI in FinTech: Revolutionizing finance through intelligent algorithms*.
4. Kaplan, J. (2025). *Generative artificial intelligence: What everyone needs to know*. Oxford University Press.
5. Goodfellow, I. J., et al. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems, 27*.
6. Xu, L., et al. (2019). Modeling tabular data using conditional GAN. *Advances in Neural Information Processing Systems*.
7. Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning* (pp. 214-223).
8. Alqulaity, M., & Yang, P. (2024). Enhanced conditional GAN for high-quality synthetic tabular data generation in mobile-based cardiovascular healthcare. *Sensors, 24*(23), 7673.
9. Engelmann, J., & Lessmann, S. (2020). *Conditional Wasserstein GAN-based oversampling of tabular data for imbalanced learning*.
10. Radford, A., Metz, L., & Chintala, S. (2016). *Unsupervised representation learning with deep convolutional generative adversarial networks*.

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 2 (34), 2026
ISSN 2663 – 4023

Гайна Георгій Анатолійович
Кандидат технічних наук, професор
Київський національний університет імені Тараса Шевченка, Київ, Україна
ORCID:0000-0003-0260-0950
ggaina@knu.ua

Масюк Дмитро Володимирович
Аспірант
Київський національний університет імені Тараса Шевченка, Київ, Україна
ORCID:0009-0002-0840-5558
Masiukdmitry96@gmail.com

GAN ДЛЯ СИНТЕЗУ ДАНИХ, З МІНІМІЗАЦІЄЮ ВТРАТ ПРЕДИКТИВНОЇ СИЛИ

Анотація. У сучасному фінансовому середовищі здатність ефективно використовувати дані для розробки інтелектуальних методів виявлення ризиків становить значний етичний виклик. Водночас це не зменшує їхньої необхідності. Це є однією з ключових проблем, яку необхідно вирішити в межах даного дослідження. Якість і обсяг даних, які можуть бути використані, значною мірою залежать від ефективності запропонованого підходу до анонімізації. Анонімізація це процес трансформації наборів даних з метою унеможливлення ідентифікації окремих осіб або організацій. Вона набуває критичного значення для забезпечення конфіденційності та відповідності нормативним вимогам у дослідженнях, що працюють із чутливою фінансовою інформацією. Із посиленням регуляторних рамок, таких як GDPR у Європейському Союзі та подібних законодавчих ініціатив у світі, етичні підходи до анонімізації стають не просто бажаними, а обов'язковими для забезпечення правової відповідності та відповідального проведення досліджень. Розвиток методів машинного навчання та аналітики великих даних підвищив ризик повторної ідентифікації, особливо у випадках, коли анонімізовані дані комбінуються із зовнішніми джерелами [1]. Це ставить під сумнів ефективність традиційних методів анонімізації в умовах високоризикових фінансових застосувань і обмежує можливість їх повноцінного використання. Водночас надмірна анонімізація призводить до деградації якості даних, іноді суттєвої, що обмежує здатність алгоритмів виявляти необхідні закономірності для ефективного визначення шахрайських операцій. Етична практика вимагає досягнення балансу між захистом конфіденційності та забезпеченням точності й практичної цінності моделей. Пошук цього балансу ускладнюється зі зростанням попиту на анонімні дані. Видалення ключових демографічних або поведінкових характеристик може ненавмисно призводити до зміщення моделей, особливо у задачах виявлення шахрайства або скорингу, що негативно впливає як на їхню ефективність, так і на інтерпретованість, яка сьогодні є критично важливою.

Ключові слова: GAN, анонімізація даних, фінансові ризики, машинне навчання, конфіденційність, якість даних, шахрайство.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review, 57*(6), 1701-1777.
2. Kim, K. (2025). *Self-refining language model anonymizers via adversarial distillation*.
3. Rocha, Á. (n.d.). *Generative AI in FinTech: Revolutionizing finance through intelligent algorithms*.
4. Kaplan, J. (2025). *Generative artificial intelligence: What everyone needs to know*. Oxford University Press.
5. Goodfellow, I. J., et al. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems, 27*.
6. Xu, L., et al. (2019). Modeling tabular data using conditional GAN. *Advances in Neural Information Processing Systems*.
7. Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning* (pp. 214-223).
8. Alqulaity, M., & Yang, P. (2024). Enhanced conditional GAN for high-quality synthetic tabular data generation in mobile-based cardiovascular healthcare. *Sensors, 24*(23), 7673.

9. Engelmann, J., & Lessmann, S. (2020). *Conditional Wasserstein GAN-based oversampling of tabular data for imbalanced learning*.
10. Radford, A., Metz, L., & Chintala, S. (2016). *Unsupervised representation learning with deep convolutional generative adversarial networks*.

Отримано редакцією журналу / Received: 05.02.26
Прорецензовано / Revised: 18.02.26
Схвалено до друку / Accepted: 24.09.26