TOWARDS ROBUST ACCOUNT TAKEOVER DETECTION IN ONLINE MARKETPLACES: DATA-CENTRIC RESEARCH AGENDA AND BENCHMARKING FRAMEWORK

DOI: https://doi.org/10.37458/nstf.27.1.7
Original scientific paper
Received: January 19, 2026
Accepted: March 19, 2026

Marko Jurišić [*], **Igor Tomičić** [**]

Abstract: Account Takeover (ATO) fraud is an escalating threat in online marketplaces, but research progress remains limited due to the absence of domain-specific, publicly available

[*] Marko Jurišić, Faculty of Organization and Informatics, (marko.jurisic@foi.unizg.hr) is a senior software engineer with extensive industry experience in large-scale systems and event-driven architectures. His PhD research explores machine learning methods for user behavior analysis and anomaly detection, with an emphasis on applying these techniques to real-world systems.

[**] Igor Tomičić is an information security professional whose career bridges academia and practice. He is an Associate Professor at the University of Zagreb Faculty of Organization and Informatics, where he works at the Department of Computing and Technology and develops and teaches security courses. In parallel, he runs a security consulting and auditing firm, applying his expertise to real-world challenges. His interests are broad and include offensive security, applied cryptography, data security, social engineering and GRC, with a healthy dose of IoT and AI research mixed in. He also holds ISO 27001 and ISO 22301 Lead Auditor certifications.

datasets. This gap hinders benchmarking and reproducibility, slows methodological innovation, and prevents systematic comparisons between classical probabilistic models and modern deep learning approaches.

This paper proposes MAATO, a new marketplace ATO dataset and a conceptual roadmap toward robust ATO detection by outlining the design principles for new synthetic datasets, intended to emulate realistic behavioral patterns and fraud scenarios. In parallel, we introduce a benchmarking framework grounded in the CRISP-DM methodology to support reproducible evaluation across model families, feature-engineering strategies, and anomaly-scoring paradigms. Instead of reporting empirical findings, this paper articulates hypotheses regarding (1) the relative importance of engineered behavioral features, (2) the comparative performance of classical vs. deep learning architectures, and (3) the scalability of per-user versus global detection models. The work aims to guide future empirical studies and establish a foundation for shared community standards.

Keywords: account takeover, anomaly detection, fraud analytics, behavioral modeling, synthetic datasets

Introduction

The rise of e-commerce has dramatically expanded the attack surface for credential-based fraud, which increased even more during the COVID pandemic (Kemp et al., 2021), with Account Takeover (ATO) emerging as one of the most disruptive threats (Kawase et al., 2019). ATO attacks are difficult to detect because compromised accounts may behave normally for

extended periods or blend malicious actions with legitimate user behavior.

The research community faces several barriers, many of which have been repeatedly documented (Jurisic et al., 2023):

1. Data scarcity: No public datasets exist that reflect the particularities of marketplace ecosystems.

2. Severe class imbalance: Real ATO events occur extremely rarely (Makki et al., 2019).

3. Methodological incomparability: Studies differ widely in data sources and metrics.

4. Privacy constraints: GDPR and confidentiality concerns limit sharing real logs.

This paper introduces a research agenda aimed at addressing these challenges through data-centric standardization using the CRISP-DM model (Shearer, 2000) and conceptual benchmarking tools.

Problem space: why a new benchmark is needed

The rapid evolution of Account Takeover (ATO) attacks in online commerce, and, more specifically, online marketplaces has outpaced the development of public datasets used to train and validate detection models. Current benchmarks frequently rely on generated datasets depicting legacy corporate environments, that fail to capture the behavioral nuances of modern, mobile-first consumer platforms. This paper identifies the critical gaps in the existing research landscape and introduces the conceptual framework for MAATO (MArketplace ATO), a synthetic benchmarking

environment designed to simulate realistic marketplace dynamics and adversarial patterns.

To systematically address these barriers (specifically data scarcity, class imbalance, and methodological fragmentation) this framework is proposed to serve as a testbed for specific theoretical questions. The MAATO framework allows us to test hypotheses regarding the necessity of deep learning and the trade-offs of privacy-preserving feature sets, which we articulate formally in a later section, by isolating variables such as attack velocity and feature granularity.

Limitations of Current Datasets

Existing public datasets such as CERT (Glasser and Lindauer, 2013) simulate insider threat scenarios, not external ATO attacks, and while BankSim (Lopez-Rojas and Axelsson, 2014) and PaySim (Lopez-Rojas et al., 2016) exist as standards for synthetic mobile money fraud, they lack the session and text layer that MAATO provides. The structural assumptions of corporate environments and workstation logs in CERT do not reflect the behavioral signatures observed in marketplaces (Jurisic et al., 2023), such as irregular mobile access patterns, mixed device ecosystems, and conversational fraud strategies, e.g., trying to move the conversation away from the platform and subsequently sending phishing links. Therefore, we propose a design blueprint for a new dataset MAATO (MArketplace ATO) - a fully synthetic environment intended to emulate behavioral patterns and attack scenarios characteristic of online marketplaces.

Methodological Approach

The dataset generation process is proposed as a hybrid of three mechanisms:

- *Agent-based simulation*, where users and adversaries are modeled as agents with behavioral rules governing login, messaging, and transactions.
- *Stochastic temporal modelling*, where inter-event times are drawn from empirically motivated distributions (e.g., circadian rhythms) to simulate irregular access.
- *Location modelling*, where benign users exhibit realistic mobility patterns (e.g., vacations, commute routines), while fraudulent sequences may originate from disjoint high-risk geolocations (Bruce et al., 2024), or proxy exits.

Fraudulent sequences are injected by modifying agent behavior under predefined ATO scenarios producing anomalies that are statistically distinct (e.g., latent drifts, as described in Dal Pozzolo et al. (2017) and Lo et al. (2017)) rather than just rule-violations.

The proposed hybrid architecture is thus designed to move beyond the static, rule-based generation found in traditional benchmarks, which often fail to capture the "interlocking" nature of marketplace fraud. MAATO would ensure that an attack is not merely a single suspicious event, but a coherent sequence of anomalies, and would achieve this by integrating agent-based intent with stochastic and geospatial layers.

Another possibility would be to use CTGAN (Xu et al., 2019) as a deep learning generation approach, but we

decided on Agent-based simulation for better coherence and interpretable causality.

For example, a fraudulent login (Location) only becomes a detectable threat when correlated with a deviation in messaging cadence (Temporal) and a breach of typical user preferences (Agent-based). This multi-dimensional approach allows for the creation of "grey-area" scenarios, where malicious activity mimics benign behavior-challenging detection models to identify subtle behavioral drift rather than obvious, high-volume spikes.

Statistical Validation

To validate the synthetic dataset, we propose comparing the statistical distributions (e.g., spacing between events, session lengths) against anonymized aggregate metrics from real-world traffic.

Furthermore, we propose an adversarial validation step, where a discriminator model is trained to distinguish real anonymized data from synthetic logs; inability to distinguish the two will serve as the acceptance criterion for the generator.

More specifically, in order to establish the utility of a fully synthetic environment, we propose a two-tiered validation framework designed to ensure that the generated data is not only plausible but also mathematically robust for model training.

The first layer of the proposed validation involves benchmarking the synthetic output against anonymized, aggregate telemetry from real-world marketplace traffic (macro-level distributional alignment). By comparing distributions, such as the heavy-tailed nature of session

durations and the 1/f noise characteristic of human messaging rhythms, we aim to ensure the "global" physics of the dataset are preserved. We suggest using metrics like Kullback–Leibler (KL) divergence (Kullback and Leibler, 1951) as the primary measure of success; a low divergence would indicate that the synthetic generator has successfully captured the macro-scale behavioral regularities of the target environment. Figure 1 shows the MAATO hybrid generation pipeline.

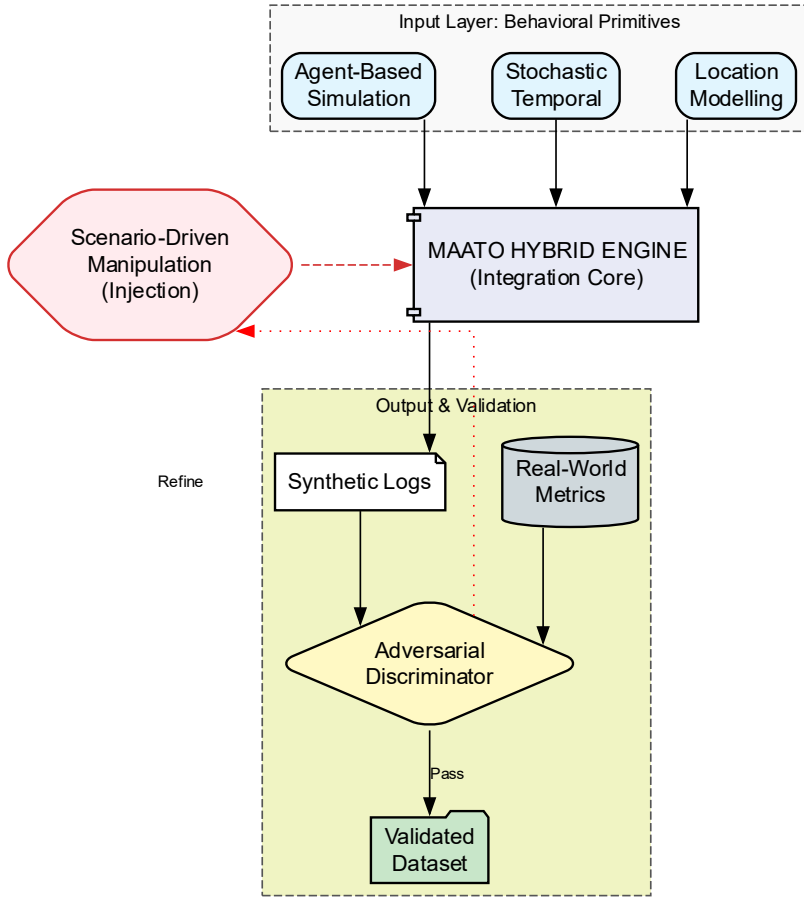

Figure 1: MAATO hybrid generation pipeline (source: author's work)

Beyond simple statistics, we propose an adversarial validation step to test the structural integrity of the sequences (adversarial "Turing test" for logs). In this phase, a discriminator model (e.g., a Random Forest (Breiman, 2001) or LSTM-based classifier) would be tasked with distinguishing between real-world aggregate logs and the synthetic MAATO sequences. The theoretical acceptance criterion for the generator is the achievement of "adversarial indistinguishability," where the discriminator's performance approaches a random-guess baseline (AUC is approx. 0.5). This step is critical to our proposal, as it ensures that the high-dimensional correlations, such as the subtle link between a specific device fingerprint and a subsequent transaction velocity, are contextually coherent and free of "synthetic artifacts."

Proposed Event Taxonomy

To support cross-platform research and ensure interoperability with existing detection frameworks, we propose a unified event schema for MAATO. In contrast to existing datasets that focus primarily on system logs this taxonomy is designed to capture the full lifecycle of a marketplace interaction. The schema categorizes events into four functional domains, reflecting the multi-stage nature of modern account takeovers:

1. *Authentication and Identity*, which captures the entry point of the session, including login attempts, multi-factor authentication (MFA) challenges, and device fingerprinting (e.g., User-Agent strings, screen resolution). This allows for the study of credential stuffing and session hijacking.

2. *Messaging and Social Dynamics,* that defines the communication layer, including message frequency, response latencies, and high-level content abstractions (e.g., text embeddings or URL-to-text ratios). This is critical for identifying "off-platforming" attempts where attackers lure victims to third-party phishing sites.
3. *Browsing/Intent,* which records telemetry related to navigation, such as page views and search queries. This domain is intended to help researchers distinguish between the targeted "search-and-buy" behavior of a benign user and the "account harvesting" or "silent reconnaissance" patterns of an adversary.
4. *Transaction and Fulfillment,* that maps the final stage of the attack, covering changes to shipping addresses, payment methods, and checkout attempts.

The aim is to move the benchmark beyond simple binary classification (fraud vs. non-fraud) and toward early-stage detection. This taxonomy ensures that the dataset provides sufficient "hooks" for researchers to develop models that can flag suspicious behavior during the messaging or browsing phase, long before a financial transaction occurs.

Design Principles & Objectives

The development of the MAATO framework is guided by four foundational principles. These objectives are designed to ensure that the resulting synthetic data is not only a viable proxy for real-world traffic but also a superior tool for stress-testing modern anomaly detection systems.

Behavioral Realism and Human-Centricity - rather than generating uniform activity, the framework is designed to replicate the "messiness" of human behavior. This includes modeling diurnal rhythms (peaks in evening usage) and preference drift (gradual changes in user categories). We argue that by proposing a realistic device split (e.g., 80% mobile, 20% desktop) there would be insurance that models are trained on the same data imbalances they will encounter in production environments.

Adversarial Scenario Diversity - a primary objective is to move beyond "obvious" fraud. We propose the inclusion of heterogeneous attacker profiles, ranging from high-volume automated spamming to "low-and-slow" silent reconnaissance. This diversity allows researchers to measure a model's sensitivity across a spectrum of attack velocities, preventing the development of systems that only catch the most aggressive (and therefore easiest) threats.

High-Resolution, Fine-Grained Logging - to support the next generation of Deep Learning and Graph Neural Network models, MAATO prioritizes resolution over abstraction. Instead of providing summarized session counts, the framework envisions capturing high-fidelity activity streams, including the precise millisecond-level spacing between a "page view" and a "message sent", which often contains the most vital signals for distinguishing bots from humans.

Privacy-by-Design and Regulatory Compliance - MAATO sidesteps the legal and ethical barriers (such as GDPR (“Regulation (EU) 2016/679, 2016) or CCPA (Bonta, 2022)) that might prevent the sharing of high-

resolution marketplace data by committing to a 100% synthetic generation approach. This makes the dataset "open-science ready," and allows for true benchmarking and reproducibility across the global research community.

Limitations and Risks of Synthetic Data

While fully synthetic data offers a pragmatic and privacy-aware foundation for reproducible research in Account Takeover (ATO) detection, it cannot serve as a wholesale replacement for real-world behavioral distributions. Although MAATO's agent-based architecture minimizes the risk of memorizing specific user records, we have to explicitly acknowledge the limitations and risks associated with synthetic-only data regimes (Jordon et al., 2022). Synthetic data should not automatically be interpreted as inherently privacy-safe or used as a direct proxy for real-world behavioral distributions.

Prior research has demonstrated that generative models trained on sensitive datasets may inadvertently retain and reproduce rare or identifying patterns (Fung et al., 2010), enabling membership inference attacks (Yeom et al., 2018; Van Breugel, Sun, et al., 2023) or attribute disclosure (Hittmeir et al., 2020) attacks, even in seemingly completely synthetic outputs. Consequently, any dissemination or reuse of MAATO-like datasets should be accompanied by explicit privacy risk assessments rather than relying solely on the assumption that synthetic generation guarantees anonymity. The proposed MAATO agentic approach, combined with manually injected attacks alleviates most of these concerns, but they still must be evaluated and

documented, rather than being assumed anonymous by default.

Synthetic data generation also introduces non-trivial risks related to model validity and generalization (Rankin et al., 2020; Van Breugel et al., 2023). Generators necessarily encode assumptions about user behavior, attacker strategies, and temporal dynamics; these can oversimplify, smooth, or bias phenomena that are inherently noisy and adversarial in real environments. As a result, detection models trained or benchmarked exclusively on synthetic data may overfit to generator-specific artifacts rather than learning robust indicators of genuine compromise. This risk is particularly pronounced under distribution shift, where models that appear stable and effective in controlled synthetic benchmarks may exhibit degraded performance when exposed to unanticipated behaviors, leading to elevated false-positive rates or missed "low-and-slow" attack scenarios in production systems. Another consideration is the notion of concept drift, the constant changing of attacker behavior in effort to circumvent defensive measures, which requires constant monitoring and tuning or retraining the models.

Crucially, from a security perspective, the synthetic data pipeline itself becomes a part of the attack surface (Lapid and Dubin, 2025). If the statistical assumptions, seeds, or rule definitions underlying a generator are compromised or maliciously influenced, the resulting synthetic dataset may systematically encode blind spots or backdoor patterns that propagate downstream into trained detection models. These hybrid threats – where weaknesses are introduced indirectly through data generation rather than model code can result in long-

lived and difficult-to-diagnose failures, including the systematic under-detection of specific attack classes or the disproportionate targeting of particular user populations, further emphasizing the need for constant monitoring and eventual tuning of models and processes.

MAATO is therefore positioned explicitly as a controlled methodological testbed, not a direct proxy for deployment readiness. Results obtained should be interpreted as comparative signals about modeling strategies, feature sensitivities, and scenario robustness, not as evidence of deployment readiness. Future work building on this framework should incorporate complementary validation strategies, including privacy attack simulations, robustness testing, and, where possible, validation on real data, using TSTR (Train on Synthetic, Test on Real) paradigm (Esteban et al., 2017), and transparently report generator assumptions to separate between benchmarking insights from operational claims.

A Benchmarking Framework for ATO Detection

To support systematic comparison across methodological paradigms, we focus on four representative families of machine learning models. Each family embodies distinct modeling assumptions regarding structure (probabilistic, temporal, or neural network) allowing us to isolate which properties interact best with marketplace behavioral patterns.

Naive Bayes (NB): The Probabilistic Baseline

While its independence assumptions are rarely satisfied, NB serves as a critical lower bound for performance and many authors use it either as a benchmark method

(Bertrand et al., 2023), or as a primary method (Kawase et al., 2019), while some authors use more advanced Bayesian Network approaches (Roberts et al., 2016). In the ATO context, NB is particularly informative for detecting population-wide anomalies (e.g., impossible travel, device switching) where feature distributions shift drastically, offering high interpretability and training efficiency.

Hidden Markov Models (HMM): Structured Temporal Dependence

HMMs explicitly model latent behavioral states, making them well-suited for capturing transitions between routine and anomalous sessions, modeling user activity as a series of events (Rashid et al., 2016) or (Saaudi et al., 2019). Unlike deep learning approaches, HMMs remain computationally tractable and interpretable, providing the best balance between the training time and performance (Le and Zincir-Heywood, 2018). We hypothesize HMMs will effectively identify structural disruptions in user rhythms (e.g., login followed by immediate mass messaging) that stateless models like NB miss.

Long Short-Term Memory (LSTM): Long-Range Dependencies

LSTMs represent the upper bound for sequence modeling capacity and are often used on the CERT dataset, e.g. in (Tuor et al., 2017a; Saaudi et al., 2018; Matterer and Lejeune, 2018; Tuor et al., 2017b). They are theoretically ideal for ATO scenarios where fraud unfolds gradually (e.g., social engineering) or relies on context spanning days. However, their inclusion in this benchmark is primarily to determine if the high

computational cost and data requirements yield statistically significant gains over simpler Markovian models.

Graph-Based Methods

A key development in the field, as shown also on the CERT dataset in (Al-Shehari et al., 2023) and (Peccatiello et al., 2023) involves moving above the analysis of isolated user timelines to a graph-based modeling of interrelationships between entities - such as users, devices, and listings. This approach effectively identifies account linkage and organized fraud rings that traditional sequential models often overlook by using Graph Neural Networks (GNNs) (Zheng et al., 2022) and Graph Convolutional Networks (GCNs) (Jiang et al., 2019). We propose constructing a heterogeneous graph where nodes represent Users, Devices, and IP addresses. Edges represent interactions (e.g., 'UserAloggedinfromIPB'). GNNs can then propagate 'guilt' through these shared attributes, identifying disjoint accounts that share a suspicious device fingerprint.

Global vs. Per-User Modeling

An open question in User Behavior Analysis (UBA) is whether detection should rely on Per-user models (learning personal baselines) or Global models (learning shared population norms). Per-user models can better capture individual user behavior patterns but have problems with cold-start (new users) and scaling (might be unfeasible to train millions of models) while global models generalize across population and might not catch nuanced changes in behavior. We hypothesize that the optimal choice depends on the density of meaningful

features and user heterogeneity, which this framework aims to test.

Alignment with the CRISP-DM Methodology

To ensure that MAATO provides a reproducible workflow for researchers, we propose a benchmarking framework structured around the CRISP-DM methodology. This systematic approach ensures that the transition from synthetic data generation to model evaluation is consistent, transparent, and aligned with the operational realities of marketplace security.

The protocol begins with a formal mapping of the ATO Threat Model. This involves defining the specific adversarial goals (e.g., account draining vs. reputation hijacking) and examining the intrinsic challenges of the synthetic logs, such as data sparsity and behavioral heterogeneity. This stage ensures that any model being benchmarked is evaluated against the specific "pain points" of online marketplaces rather than generic anomaly detection tasks.

Next, we propose a dual-track feature engineering phase. Researchers can evaluate models using "Simple" feature sets (e.g., count-based aggregations, analogue to (Rashid et al., 2016) or "Exploded" high-resolution feature sets (e.g., raw sequence embeddings like in Bertrand et al. (2023)). This allows the benchmark to measure the trade-off between computational efficiency and detection accuracy, which is a critical consideration for real-time marketplace deployments.

The framework also provides a standardized environment for training and comparing distinct model families (NB, HMM, LSTMs, Graph Neural Networks).

We aim to eliminate "optimization bias," and ensure that performance differences reflect the models' structural capabilities rather than variations in tuning effort, by prescribing a consistent protocol for hyperparameter tuning.

As for the operational evaluation metrics, traditional accuracy metrics are often insufficient for imbalanced fraud datasets. We therefore propose an evaluation suite focused on Operational Impact; Precision-Recall (PR) Curves to assess performance under extreme class imbalance; recall at Low False Positive Rates (FPR $\leq$ 1%), like in (Tuor et al., 2017a; Le et al., 2021) to prioritize the user experience - in a marketplace, "false positives" result in blocked legitimate customers, which is commercially unsustainable and Detection Delay: Measuring the "time-to-detection" to evaluate how early in the attack sequence a model can trigger an alert as described in (Le and Nur Zincir-Heywood, 2019).

Hypotheses For Future Empirical Study

We articulate the following testable hypotheses:

1. **H1 (Feature Engineering and Privacy):** We hypothesize that a small, privacy-preserving subset of behavioral features will provide discriminatory power rivaling complex, invasive feature sets. If validated, this refutes the assumption that deep content inspection is required for robust ATO detection, directly addressing the privacy constraints that currently limit cross-organizational data sharing.

Current research cannot definitively say if invasive, high-dimensional logging is necessary for ATO detection. Thus, by testing H1 on MAATO, we aim to

prove whether privacy-preserving, low-dimensional feature sets are sufficient, addressing the privacy barriers that limit real-log sharing.

2. **H2 (Model Selection and Specificity**): We hypothesize that no single architecture will dominate; rather, performance will be strictly scenario-dependent (e.g., LSTMs for 'low-and-slow' social engineering vs. Naive Bayes for abrupt velocity attacks). This hypothesis seeks to replace generic 'accuracy' metrics (which are statistically misleading due to the severe class imbalance of ATO events) with scenario-specific benchmarks. We thus aim to resolve the methodological incomparability that currently plagues studies using different data sources - by standardizing evaluation across distinct attack types.

3. **H3 (Hybrid Methods)**: Hybrid architectures that combine sequential state tracking (LSTM) with either HMM or structural relational data (GNNs) will offer the best trade-off between detection accuracy and computational overhead.

Marketplace fraud is not a single event but a sequence of anomalies across devices and accounts. We propose H3 to formally test if integrating the 'social' graph of a marketplace (users, devices, IPs) provides a detection uplift over pure sequence modeling. Testing this validates the necessity of relational structures, directly addressing the scarcity of datasets that reflect the particularities of marketplace ecosystems compared to the limited flat-file datasets (like CERT) currently available.

Conclusion:

This paper has outlined a data-centric research agenda for advancing Account Takeover (ATO) detection in online marketplaces. The central contribution is the conceptual foundation of the MAATO synthetic dataset, and a unified evaluation framework grounded in CRISP-DM. This roadmap positions high-quality behavioral data and methodological transparency as the cornerstones of future research, moving beyond isolated experimentation.

The paper has introduced a set of hypotheses concerning the role of feature engineering and the expressive capacity of different model families. These hypotheses are intended to guide empirical inquiry and encourage systematic exploration rather than presuppose outcomes. Future work must focus on the implementation of the MAATO generation engine and the rigorous validation of these models against the proposed scenarios.

Literature:

1. Al-Shehari, T., Al-Razgan, M., Alfakih, T., Alsowail, R. A., & Pandiaraj, S. (2023). Insider Threat Detection Model Using Anomaly-Based Isolation Forest Algorithm. IEEE Access, 11, 118170–118185. https://doi.org/10.1109/ACCESS.2023.3326750
2. Bertrand, S., Desharnais, J., & Tawbi, N. (2023). Unsupervised User-Based Insider Threat Detection Using Bayesian Gaussian Mixture Models. 2023 20th Annual International Conference on Privacy, Security and Trust, PST 2023. https://doi.org/10.1109/PST58708.2023.10320169

3. Bonta, R. (2022). California consumer privacy act (CCPA). Retrieved from State of California Department of Justice: Https://Oag. ca. Gov/Privacy/Ccpa.

4. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

5. Bruce, M., Lusthaus, J., Kashyap, R., Phair, N., & Varese, F. (2024). Mapping the global geography of cybercrime with the World Cybercrime Index. Plos One, 19(4), e0297312.

6. Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2017). Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE Transactions on Neural Networks and Learning Systems, 29(8), 3784–3797.

7. Esteban, C., Hyland, S. L., & Rätsch, G. (2017). Real-valued (medical) time series generation with recurrent conditional gans. arXiv Preprint arXiv:1706.02633.

8. Fung, B. C., Wang, K., Chen, R., & Yu, P. S. (2010). Privacy-preserving data publishing: A survey of recent developments. ACM Computing Surveys (Csur), 42(4), 1–53.

9. Glasser, J., & Lindauer, B. (2013). Bridging the Gap: A Pragmatic Approach to Generating Insider Threat Data. 2013 IEEE Security and Privacy Workshops, 98–104. https://doi.org/10.1109/SPW.2013.37

10. Hittmeir, M., Mayer, R., & Ekelhart, A. (2020). A baseline for attribute disclosure risk in synthetic data. Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy, 133–143.

11. Jiang, J., Chen, J., Gu, T., Choo, K.-K. R., Liu, C., Yu, M., Huang, W., & Mohapatra, P. (2019). Anomaly Detection with Graph Convolutional Networks for Insider Threat and Fraud Detection. Proceedings - IEEE Military Communications Conference MILCOM, 2019-November. https://doi.org/10.1109/MILCOM47813.2019.9020760

12. Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C., Cohen, S. N., & Weller, A. (2022). Synthetic Data–what, why and how? arXiv Preprint arXiv:2205.03257.

13. Jurisic, M., Tomicic, I., & Grd, P. (2023). User Behavior Analysis for Detecting Compromised User Accounts: A Review Paper. CYBERNETICS AND INFORMATION TECHNOLOGIES, 23(3), 102–113. https://doi.org/10.2478/cait-2023-0027

14. Kawase, R., Diana, F., Czeladka, M., Schüler, M., & Faust, M. (2019). Internet Fraud: The Case of Account Takeover in Online Marketplace. Proceedings of the 30th ACM Conference on Hypertext and Social Media, 181–190.

15. Kemp, S., Buil-Gil, D., Moneva, A., Miró-Llinares, F., & Díaz-Castaño, N. (2021). Empty streets, busy internet: A time-series analysis of cybercrime and fraud trends during COVID-19. Journal of Contemporary Criminal Justice, 37(4), 480–501.

16. Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. The Annals of Mathematical Statistics, 22(1), 79–86. JSTOR.

17. Lapid, R., & Dubin, A. (2025). Backdoors in Conditional Diffusion: Threats to Responsible Synthetic Data Pipelines. arXiv Preprint arXiv:2507.04726.

18. Le, D. C., & Nur Zincir-Heywood, A. (2019). Machine learning based insider threat modelling and detection. 2019 IFIP/IEEE Symposium on Integrated Network and Service Management, IM 2019, 1–6.

19. Le, D. C., & Zincir-Heywood, A. N. (2018). Evaluating insider threat detection workflow using supervised and unsupervised learning. Proceedings - 2018 IEEE Symposium on Security and Privacy Workshops, SPW 2018, 270–275. https://doi.org/10.1109/SPW.2018.00043

20. Le, D. C., Zincir-Heywood, N., & Heywood, M. (2021). Training regime influences to semi-supervised learning for insider threat detection. 2021 IEEE SYMPOSIUM ON

SECURITY AND PRIVACY WORKSHOPS (SPW 2021), 13–18. https://doi.org/10.1109/SPW53761.2021.00010

21. Lo, Y.-Y., Liao, W., Chang, C.-S., & Lee, Y.-C. (2017). Temporal matrix factorization for tracking concept drift in individual user preferences. IEEE Transactions on Computational Social Systems, 5(1), 156–168.

22. Lopez-Rojas, E. A., & Axelsson, S. (2014, September). BankSim: A Bank Payment Simulation for Fraud Detection Research. 26th European Modeling and Simulation Symposium, EMSS 2014.

23. Lopez-Rojas, E., Elmir, A., & Axelsson, S. (2016). PaySim: A financial mobile money simulator for fraud detection. 28th European Modeling and Simulation Symposium, EMSS, Larnaca, 249–255.

24. Makki, S., Assaghir, Z., Taher, Y., Haque, R., Hacid, M.-S., & Zeineddine, H. (2019). An experimental study with imbalanced classification approaches for credit card fraud detection. IEEE Access, 7, 93010–93022.

25. Matterer, J., & Lejeune, D. (2018). Peer group metadata-informed LSTM ensembles for insider threat detection. Proceedings of the 31st International Florida Artificial Intelligence Research Society Conference, FLAIRS 2018, 62–67.

26. Peccatiello, R. B., Gondim, J. J. C., & Garcia, L. P. F. (2023). Applying One-Class Algorithms for Data Stream-Based Insider Threat Detection. IEEE Access, 11, 70560–70573. https://doi.org/10.1109/ACCESS.2023.3293825

27. Rankin, D., Black, M., Bond, R., Wallace, J., Mulvenna, M., & Epelde, G. (2020). Reliability of supervised machine learning using synthetic data in health care: Model to preserve privacy for data sharing. JMIR Medical Informatics, 8(7), e18910.

28. Rashid, T., Agrafiotis, I., & Nurse, J. R. (2016). A new take on detecting insider threats: Exploring the use of hidden markov models. Proceedings of the 8th ACM CCS International Workshop on Managing Insider Security Threats, 47–56.

29. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). (2016). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

30. Roberts, S. C., Holodnak, J. T., Nguyen, T., Yuditskaya, S., Milosavljevic, M., & Streilein, W. W. (2016). A Model-Based Approach to Predicting the Performance of Insider Threat Detection Systems. Proceedings - 2016 IEEE Symposium on Security and Privacy Workshops, SPW 2016, 314–323. https://doi.org/10.1109/SPW.2016.14

31. Saaudi, A., Al-Ibadi, Z., Tong, Y., & Farkas, C. (2018). Insider threats detection using CNN-LSTM model. Proceedings - 2018 International Conference on Computational Science and Computational Intelligence, CSCI 2018, 94–99. https://doi.org/10.1109/CSCI46756.2018.00025

32. Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. Journal of Data Warehousing, 5(4), 13–22.

33. Tuor, A., Kaplan, S., Hutchinson, B., Nichols, N., & Robinson, S. (2017a). Deep learning for unsupervised insider threat detection in structured cybersecurity data streams. AAAI Workshop - Technical Report, WS-17-01-WS-17-15, 224–234.

34. Tuor, A., Kaplan, S., Hutchinson, B., Nichols, N., & Robinson, S. (2017b). Predicting user roles from computer logs using recurrent neural networks. 31st AAAI Conference on Artificial Intelligence, AAAI 2017, 4993–4994.

35. Van Breugel, B., Qian, Z., & Van Der Schaar, M. (2023). Synthetic data, real errors: How (not) to publish and use synthetic data. International Conference on Machine Learning, 34793–34808.

36. Van Breugel, B., Sun, H., Qian, Z., & van der Schaar, M. (2023). Membership inference attacks against synthetic

data through overfitting detection. arXiv Preprint arXiv:2302.12580.

37. Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Modeling tabular data using conditional gan. Advances in Neural Information Processing Systems, 32.

38. Yeom, S., Giacomelli, I., Fredrikson, M., & Jha, S. (2018). Privacy risk in machine learning: Analyzing the connection to overfitting. 2018 IEEE 31st Computer Security Foundations Symposium (CSF), 268–282.

39. Zheng, C., Hu, W., Li, T., Liu, X., Zhang, J., & Wang, L. (2022). An Insider Threat Detection Method Based on Heterogeneous Graph Embedding. Proceedings - 2022 IEEE 8th International Conference on Big Data Security on Cloud, IEEE International Conference on High Performance and Smart Computing, and IEEE International Conference on Intelligent Data and Security, BigDataSecurity/HPSC/IDS 2022, 11–16. https://doi.org/10.1109/BigDataSecurityHPSCIDS54978.2022.00013