Balancing Utility and Privacy: Evaluating the Privacy-Preserving Capabilities of Synthetic Data from Deep Generative Models

Julia L. Wang
University of Toronto
Supervisor: Prof. Yuri Lawryshyn

B.A.Sc. Thesis

Division of Engineering Science
UNIVERSITY OF TORONTO

Abstract

This thesis presents an in-depth investigation into the privacy-preserving properties of synthetic financial datasets generated by advanced generative AI models. In the age of big data, where financial institutions increasingly rely on machine learning for decision-making, the protection of personal information has never been more critical. This work critically assesses four prominent generative models: CTGAN, TVAE, DoppelGANger, and Banksformer. The research methodology includes comprehensive privacy evaluation metrics, such as Jensen-Shannon divergence, Wasserstein distance, and entropy differences, to measure the statistical similarities between synthetic and real datasets. The re-identification risk, attribute disclosure risk, and susceptibility to membership inference attacks are examined, revealing the strengths and limitations of each model in terms of privacy protection. The results demonstrate that while some models maintain higher fidelity to the original data distributions, others offer stronger privacy guarantees by deviating more significantly from the real datasets. Notably, Banksformer and DoppelGANger reveal higher risks of privacy breaches due to their propensity for high classification accuracies in membership inference attacks. Concluding that the current generative models exhibit varying levels of privacy preservation, the thesis advocates for future research to establish a quantifiable balance between data utility and privacy.

Acknowledgements

I would like to express my gratitude to everyone at the CMTE group for the valuable conversations we had during our weekly meetings. I would also like to express special thanks to Professor Yuri Lawryshyn and Lucy L. for all your mentorship and support throughout this research.

Contents

1 Introduction **7**

1.1 Background . . . 7

1.2 Objective . . . 7

1.3 Deliverables . . . 8

1.4 Outline . . . 8

2 Literature Review **9**

2.1 Privacy Attacks . . . 9

2.2 Privacy-protection Methods . . . 11

2.2.1 Privacy in Centralized Machine Learning . . . 11

2.2.2 Privacy in Distributed Machine Learning . . . 12

2.2.3 Privacy in Federated Learning . . . 14

2.2.4 Privacy Protection in Machine Learning Summary . . . 15

2.3 Generative Models for Synthetic Data . . . 17

2.3.1 CTGAN . . . 18

2.3.2 TVAE . . . 19

2.3.3 TimeGAN . . . 21

2.3.4 DoppelGANger . . . 22

2.3.5 Banksformer . . . 23

2.3.6 VAE-GAN Combination Models . . . 25

2.3.7 Generative Models Summary . . . 26

2.4 Chapter Summary . . . 27

3 Methodology **29**

3.1 Dataset . . . 29

3.2 Synthetic Data Generation . . . 30

3.3 Synthetic Data Evaluation Metrics . . . 31

3.4 Privacy Evaluation Metrics . . . 32

3.4.1 Data Analysis Metrics . . . 32

3.4.2 Re-identification Risk . . . 33

3.4.3 Attribute Disclosure Risk . . . 35

3.4.4 Membership Inference Attack Success Rate . . . 36

3.4.5 Model Inversion Attack Success Rate . . . 37

3.5 Chapter Summary . . . 38

4 Results and Discussion **39**
4.1 Statistical Analysis . . . 39
4.2 Data Analysis Metrics Results . . . 41
4.3 Re-identification Risk Results . . . 42
4.4 Attribute Disclosure Risk Results . . . 44
4.5 Membership Inference Attack Results . . . 46
4.6 Mutual Information Investigation Results . . . 49
4.7 Chapter Summary . . . 53

5 Conclusions and Future Work **54**

6 Appendices **56**
6.1 Generative Models Architecture Overviews . . . 56
6.1.1 CTGAN . . . 56
6.1.2 TVAE . . . 56
6.1.3 TimeGAN . . . 57
6.1.4 DoppelGANger . . . 57
6.1.5 Banksformer . . . 58
6.1.6 VAE-GAN Model . . . 59
6.2 Binary Classifier Parameters . . . 62
6.3 Individual Distributions . . . 63
6.4 Membership Inference All Results . . . 64
6.5 Membership Inference Confusion Matrices . . . 65
6.6 Mutual Information Results . . . 67

List of Figures

1 Privacy Attacks: identification, inference, and linkage (Liu et al. 2021) . . . 10
2 Univariate Distributions of Categorical Features 40
3 Univariate Distribution of Continuous Feature Amount 40
4 Model Inference Attack Accuracy and F1 Score over 5 Folds 47
5 Mutual Information of DoppelGANger and Banksformer Features 50
6 DoppelGANger Scaled Dates Membership Inference Results 52
7 CTGAN Overall Architecture (Xu et al. 2019) 56
8 TVAE Overall Architecture (Ishfaq, Hoogi, and Rubin 2023) 57
9 TimeGAN Overall Architecture (Yoon, Jarrett, and Schaar 2019) 58
10 DoppelGANger Overall Architecture (Lin et al. 2019) 58
11 Banksformer Overall Architecture (Nickerson et al. 2023) 59
12 VRNN proposed by Chung et al. (2015) 60
13 Individual Amount Distributions 63
14 Model Inference Attack Accuracy, Precision, Recall, f1 over 5 Folds 64
15 CTGAN Membership Inference Matrices 65
16 TVAE Membership Inference Matrices 66
17 DoppelGANger Membership Inference Matrices 66
18 Banksformer Membership Inference Matrices 67
19 Mutual Information of CTGAN and TVAE Features 68

List of Tables

1 Overview of Privacy Protection Methods in Machine Learning 17
2 Overview of Generative AI Models for Synthetic Data Generation 27
3 Summary of the Czech Dataset 29
4 Czech Dataset Features 29
5 Synthetic Data Generation Models Parameters 31
6 Column-wise Comparisons of Distance Metrics 39
7 Comparison of Privacy Metrics across Datasets 41
8 Re-identification Risk Assessment for Synthetic Data Models 43
9 Attribute Disclosure Risk Assessment 44
10 Range of Amounts and Dates in Real and Synthetic Data 51
11 Binary Classifier Models Parameters 62

1 Introduction

1.1 Background

In an age marked by unprecedented advancements in deep learning and machine learning (ML) applications, the imperative for robust privacy-preserving solutions has become increasingly critical. The ubiquity of data-driven decision-making across industries necessitates a paradigm shift towards safeguarding sensitive personal information (PI) in a manner that aligns with ethical data practices. This need is especially pronounced in the financial sector, where privacy and confidentiality are paramount considerations for customers entrusting institutions with their data. This thesis aims to contribute towards solving this pressing challenge, emphasizing the potential role of generative artificial intelligence models in synthesizing and safeguarding financial data.

As organizations delve into the transformative power of data analytics, there is a conspicuous gap in the current landscape of privacy-preserving ML solutions, particularly in the financial domain. Financial institutions stand to gain immensely from harnessing the potential of machine learning, but this potential can only be realized if robust measures are in place to protect the sensitive data underpinning these models. The significance of this gap is underscored by the immense opportunities for innovation and efficiency that lie untapped in the application of ML. The proposed research not only contributes to the broader field of privacy-preserving ML but also unlocks the potential for financial institutions to leverage cutting-edge machine learning techniques securely.

1.2 Objective

The objective of this thesis is to evaluate how generative AI models manage and mask sensitive financial information within synthetic datasets. This research seeks to dissect the effectiveness of various deep learning models in striking a critical balance between

maintaining data utility for ML applications and ensuring the anonymity and security of underlying PI. By achieving this dual aim, the thesis contributes to filling the prevalent research void and equipping financial institutions with the ability to adopt ML innovations responsibly and securely.

1.3 Deliverables

The various key deliverables for this thesis are listed below. All code will be written in Python including an overall Jupyter Notebook containing results and code which financial institutions may run.

- An implementation of all generative models described in the thesis.
- An implementation of privacy metrics used to measure the similarities between the synthetically generated data and original data.
- An analysis of how the proposed models perform concerning privacy preservation compared to baseline models.

1.4 Outline

Chapter 2 Literature Review: Surveys existing knowledge on privacy attacks, privacy-protection methods, and the role of generative models in creating synthetic datasets.
Chapter 3 Methodology: Describes the privacy metrics and methods applied to assess the privacy aspects of synthetic data generated by various models.
Chapter 4 Results and Discussion: Presents the findings of the experiments and analyses the implications for privacy preservation in synthetic data.
Chapter 5 Conclusions and Future Work: Summarizes the main conclusions regarding the privacy-preserving properties of the evaluated models and proposes directions for future research.

2 Literature Review

The evolving landscape of data privacy and security, alongside rapid advancements in ML, has precipitated an urgent need for robust privacy-preserving methods within these domains. This chapter outlines the progression of privacy attacks and delineates current privacy-preserving ML methods, setting the stage for an in-depth exploration of generative models for synthetic data generation. It underscores the dual challenge of fostering data privacy while ensuring the utility and accessibility of generated datasets for analytical purposes. Particularly in financial sectors, where the sensitivity of data is paramount, the development and evaluation of generative models present promising avenues for secure data augmentation and analysis. This discourse extends into the realm of synthetic data generation, focusing on the imperative to balance statistical fidelity with stringent privacy considerations, thereby mitigating the risks of personal information exposure. The following chapter delves into various generative models implemented in this thesis, examining their efficacy in crafting synthetic datasets that uphold privacy without compromising data integrity.

2.1 Privacy Attacks

Privacy attacks can take various forms, and among them are identification attacks, inference attacks, and linkage attacks (Liu et al. 2021). An identification attack is characterized by its focus on discerning a user's name or identity by exploiting information available in a public dataset. In contrast, an inference attack involves the unauthorized analysis of data to gain knowledge about a specific subject. This type of attack seeks to extract valuable insights from the data without proper authorization. Linkage attacks, on the other hand, have a distinct approach. In a linkage attack, the adversary endeavours to obtain information about a target by strategically correlating data from multiple sources. The objective is to establish connections between different datasets, allowing the attacker to uncover sensitive

information about the target through the amalgamation of diverse data points. These privacy attacks highlight the diverse strategies employed by malicious actors to compromise the confidentiality and security of individuals' information in various contexts.

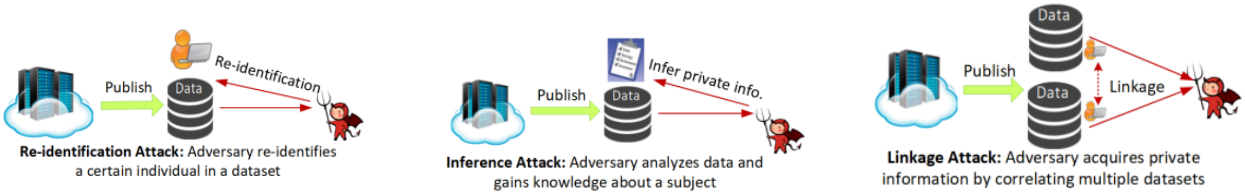


Figure 1: Privacy Attacks: identification, inference, and linkage (Liu et al. 2021)

Shedding light on machine learning attacks, Rodríguez, Otero, and Canal (2023) delves into the realm of membership inference attacks, which directly target ML models. This category also encompasses 3 distinct types of attacks. (1) Model inversion attacks involve extracting training data through ML model predictions. (2) Model stealing attacks focus on recovering internal training parameters, particularly in deep neural networks where information about the training data may be memorized. (3) Property inference attacks aim to discern sensitive patterns within ML models. The significance of protecting financial data from such privacy attacks cannot be overstated, as the consequences of breaches in this domain are particularly severe. Financial datasets contain sensitive information that, if compromised through identification, inference, or linkage attacks, could lead to significant financial loss, identity theft, and erosion of trust in financial institutions. Given the intricacies of financial data, which often include transaction histories, account balances, and personal identification information, the strategies employed by malicious actors to compromise data privacy pose unique challenges. Ensuring the security of such data against these sophisticated privacy attacks is paramount, underscoring the need for advanced privacy-preserving methods tailored to the financial sector's distinct requirements. This necessity highlights the critical role of developing robust mechanisms that not only thwart attempts at unauthorized data access but also safeguard the integrity and confidentiality of financial information in an ever-evolving digital landscape.

2.2 Privacy-protection Methods

In response to the proliferation of data privacy attacks, current research in ML-based privacy solutions utilizes combinations of models with protection techniques to maintain data privacy. These models output data which can then be utilized for ML and data analysis purposes without exposing any personal information. Such advancements in data privacy can therefore be applied to financial applications in recent services such as protecting voice recognition and masking data before uploading it to the cloud. Common learning architectures focus on centralized and distributed learning, combined with protection techniques such as encryption and differential privacy (Rodríguez, Otero, and Canal 2023). The following section analyzes current research on these learning architectures, which is summarized in Table 1.

2.2.1 Privacy in Centralized Machine Learning

Centralized machine learning (CML)-based privacy solutions involve centralizing training data in a machine or data center, leading to enhanced efficiency and accuracy owing to the easy accessibility of data. Various encryption techniques, including homomorphic encryption, multi-party computation, attribute access control, and lightweight cryptographic techniques, are employed for privacy protection. Rahulamathavan et al. (2014) and Zhu et al. (2017) implemented privacy-preserving solutions in eHealth scenarios using Support Vector Machine (SVM) models, while W. Wang et al. (2017) utilized multi-layer extreme learning machine (ML-ELM) for image classification with encryption. Homomorphic encryption, in particular, has gained attention for its strong privacy advantages, leading to the development of lighter algorithms like Somewhat Homomorphic Encryption (SHE) and Partially Homomorphic Encryption (PHE). These techniques, while effective in preserving user privacy for ML, come with high computational complexity and resource requirements. Rodríguez, Otero, and Canal (2023) illustrate how the addition of homomorphic encryption techniques has the disadvantage of a less efficient network, especially when used for training deep neural

networks. As such, ongoing research into encryption currently explores enhancing efficiency and reducing training and execution time.

2.2.2 Privacy in Distributed Machine Learning

Distributed Learning-based solutions have also become a significant focus for privacy preservation. Distributed Machine Learning (DML) is where learning model generation takes place on participant devices, and a coordinating server generates a global model shared among the distributed participants. Shokri and Shmatikov (2015) and Servia-Rodriguez et al. (2017) have conducted pioneering work in collaborative deep learning systems that preserve user privacy by allowing multiple users to learn Neural Network models based on their inputs without sharing them. They used SGD and Multi-Layer Perceptron (MLP) models, where the solutions are tested on datasets like MNIST (LeCun, Cortes, and Burges 2010), SVHN (Netzer et al. 2011), WISDM Human Activity Recognition (Kwapisz, Weiss, and Moore 2010), NIPS, and Wikipedia (Foundation, n.d.), achieving high accuracy levels. However, such distributed models are susceptible to privacy inference attacks, such as membership inference or model memorization, which focus on retrieving sensitive data from trained ML models. To address this, common privacy protection techniques such as encryption and differential privacy are integrated into distributed learning.

In terms of encryption, homomorphic encryption is also widely used in DML to enhance privacy in IoT systems. Researchers like Phong et al. (2017) and González-Serrano, Navia-Vázquez, and Amor-Martín (2017) have proposed systems based on MLP and CNN models, utilizing homomorphic encryption for secure computations. Almaiah et al. (2022) similarly used VAE techniques for privacy, which introduced a lightweight encryption-decryption combination based on homomorphic encryption. However, all these encryption solutions have limitations such as the sharing of a common homomorphic encryption key among participants, which mitigates its privacy effects.

In addition to encryption, differential privacy is a technique that enhances data privacy

by adding perturbations to the original data including swapping, additive perturbation, condensation, randomized response, micro-aggregation, random rotation and projection, and geometric and hybrid perturbation (Rodríguez, Otero, and Canal 2023). Early works by Abadi et al. (2016) and Hitaj, Ateniese, and Perez-Cruz (2017) laid the foundation for differential privacy in algorithmic learning techniques, achieving computational efficiency in distributed collaborative learning systems based on CNNs. J. Wang et al. (2018) and Zhang, He, and Lee (2018) then integrated differential privacy into DML, proposing frameworks like Arden and obfuscating function-based approaches. These solutions preserve data privacy through arbitrary data nullification, random noise addition, and obfuscation functions, demonstrating their efficacy in maintaining privacy while improving inference performance. To address the computational overhead associated with encryption in high-volume data systems, further obfuscation techniques are considered. For instance, Alguliyev, Aliguliyev, and Abdullayeva (2019) demonstrated the effectiveness of using an autoencoder-based DNN approach to obtain high accuracy in transforming sensitive medical data into non-sensitive data. They used a sparse denoising autoencoder to transform data and then a Deep CNN to classify the transformed data perturbation. Similarly, Du et al. (2018) focused on big data privacy in Multi-Access Edge Computing, employing differential privacy while achieving high accuracy. They employed 2 methods to maximize query accuracy while minimizing privacy leakage: adding Laplacian noise to the output value and adding noise to the objective function, then using SGD and GANs. However, it is essential to note that these obfuscation-based techniques exhibit vulnerabilities to minimality attacks and foreground knowledge attacks (Rodríguez, Otero, and Canal 2023). Comparing differential privacy to homomorphic encryption, the encryption-based solutions achieve higher privacy guarantees although these systems are less efficient. Ongoing research in this area aims to overcome these limitations, emphasizing the continuous evolution of privacy-preserving techniques in distributed learning environments.

2.2.3 Privacy in Federated Learning

Federated Learning (FL) is an approach that focuses on learning global models based on decentralized data, providing increased privacy to distributed systems. This methodology ensures that the training of machine learning models occurs locally on client devices, allowing distributed nodes to communicate parameters with each other for the parallel development of models (Rieke et al. 2020). Thereby, this decentralized approach eliminates the need for exchanging data between client devices and global servers.

Researchers have explored various encryption techniques to increase user privacy in FL. Zhang et al. (2020) introduced 2 privacy-preserving deep learning models DeepPAR and DeepDPA for applications in industrial systems. DeepPAR utilizes proxy re-encryption preserving data privacy for each person called dynamic update secrecy and DeepDPA implements group key management providing backward secrecy in lightweight dynamic-preserving DL. Building off current encryption techniques, recent research by Ma et al. (2022) details improvements to homomorphic encryption schemes with federated learning by encrypting model updates via an aggregated public key. This addition outperforms other communication and computational cost schemes while preserving model accuracy. Fu et al. (2020) also introduced a verifiable FL framework, utilizing encryption techniques to securely aggregate gradients.

Similarly, differential privacy has been demonstrated to improve privacy preservation in FL systems. Kumar, Gupta, and Tripathi (2021) proposed an approach that employs perturbation techniques instead of encryption in smart cities, providing a trustworthy and privacy-preserving secured framework. Lu et al. (2020b) employed gradient perturbation for a privacy-preserving edge network by proposing a solution based on asynchronous FL and self-adaptive threshold gradient compression. This addition demonstrated success in learning multi-party data without sharing sensitive information. Kairouz et al. (2021) also introduces a differentially private (DP) variant of the Follow-The-Regularized-Leader (FTRL) algorithm, as an alternative to differentially private SGD. DP-SGD by Abadi

et al. (2016) is the SOTA method for training private learning models, utilizing noisy minibatch gradients with privacy amplification by sampling or shuffling. However, the requirement for perfectly random subsets in each mini-batch poses challenges, particularly in practical scenarios like federated learning, where data access patterns are less controllable. DP-FTRL is designed as an online learning-based DP algorithm that competes favourably with DP-SGD in terms of privacy, utility, and computation trade-offs, without relying on privacy amplification. Notably, DP-FTRL outperforms un-amplified DP-SGD across various privacy levels and exhibits superior performance, especially in the higher-accuracy and lower-privacy regimes. The algorithm leverages the properties of FTRL, including linearizing the loss function, optimizing models in previous timesteps for better performance in subsequent steps, and utilizing tree aggregation protocols to privately track the summed loss over all timesteps. The key distinction lies in DP-FTRL's use of tree aggregation protocols for adding noise, offering more flexibility in data access patterns compared to the sampling and shuffling requirements of DP-SGD (Kairouz et al. 2021). Blockchain technologies have additionally been incorporated into differentially private FL, enhancing customer learning on IoT devices while preserving privacy. Zhao et al. (2020) introduced a blockchain-based crowd-sourcing FL system which combined blockchain-distributed storage and mobile edge computing (MEC) server technologies. This system enforces customer privacy and achieves improved accuracy in FL models through a normalization technique that outperforms batch normalization when features are under differential privacy protection.

2.2.4 Privacy Protection in Machine Learning Summary

Rodríguez, Otero, and Canal (2023) emphasizes the primary hurdle faced by privacy-preserving machine learning (ML) techniques is the delicate balance between accuracy and complexity. Overall, training neural networks, particularly Deep Neural Networks, over encrypted data presents significant challenges revolving around computational complexity. Even when trained on plain text data, networks already face

slowness. The introduction of homomorphic encryption to the process exacerbates this issue, making it at least an order of magnitude slower. The computational complexity of training neural networks is notably affected by the additional layer of homomorphic encryption. The process of computing polynomials is a fundamental aspect, where the level of the computed polynomial is directly proportional to the number of back-propagation steps undertaken during training Rodríguez, Otero, and Canal (2023). As a result, the deceleration introduced by homomorphic encryption is more likely to worsen, posing a substantial hurdle in achieving efficient and practical training of neural networks over encrypted data. Similarly, differential privacy also comes with notable drawbacks around accuracy. Striking the right balance between privacy and utility can be challenging, as the introduction of noise or perturbation to data may compromise the accuracy. Implementing differential privacy can be computationally expensive, impacting efficiency, and tuning parameters require careful consideration to avoid either overly conservative privacy measures or inadequate protection. The framework may have limitations in dynamic datasets and collaborative scenarios, and it may not fully guard against all side-channel attacks (Wright and Rumsey 2018). Ongoing research aims to address these challenges and enhance the practicality of differential privacy.

Table 1: Overview of Privacy Protection Methods in Machine Learning

Paper	**Privacy Method**	**ML Methods**
Rahulamathavan et al. (2014), Zhu et al. (2017)	Encryption (CML)	SVM
W. Wang et al. (2017)	Encryption (CML)	ML-ELM
Phong et al. (2017), González-Serrano, Navia-Vázquez, and Amor-Martín (2017)	Encryption (DML)	MLP, CNN
Almaiah et al. (2022)	Encryption (DML)	VAE
Alguliyev, Aliguliyev, and Abdullayeva (2019)	Differential Privacy (DML)	AE, Deep CNN
Du et al. (2018)	Differential Privacy (Big Data)	SGD, GANs
Zhang et al. (2020)	Encryption (FL)	DeepPAR, DeepDPA
Ma et al. (2022)	Encryption (FL)	Federated Averaging
Fu et al. (2020)	Encryption (FL)	Verifiable FL Framework
Kumar, Gupta, and Tripathi (2021)	Blockchain & Differential Privacy (FL)	PCA, XGBoost
Lu et al. (2020b)	Differential Privacy (Asynchronous FL)	Gradient Perturbation
Kairouz et al. (2021)	Differential Privacy (FL)	DP-FTRL Algorithm
Zhao et al. (2020)	Blockchain & Differential Privacy (FL)	CNN, MEC

2.3 Generative Models for Synthetic Data

With the current advancement of generative AI models, there is a research opportunity to investigate how privacy-preserving such methods are and how they can further enhance data protection. Generative Adversarial Networks (GANs) consist of a generator which creates synthetic data and a discriminator which tries to distinguish the real data, where they work together adversarially to generate synthetic data. Variational Autoencoders (VAEs), on the other hand, are based on the principle of encoding inputs into a latent space representation and subsequently reconstructing the output from this space. The generation of synthetic data through these models presents a potent tool for expanding datasets, especially in domains where data sensitivity or scarcity is a concern. However, a fundamental challenge lies in assessing how privacy-preserving the synthetic data produced by these models is. This is particularly critical in financial contexts where preserving the statistical fidelity of data

without exposing any personal identifiers is paramount. The following sections describe various state-of-the-art models for generating synthetic datasets which are implemented in this thesis.

2.3.1 CTGAN

The Conditional Tabular GAN model (CTGAN) introduced by Xu et al. (2019) represents an advancement in the generation of synthetic tabular data, addressing the unique challenges that arise from the mixed data types, non-Gaussian distributions, multimodal distributions, learning from sparse one-hot-encoded vectors, and highly imbalanced categorical columns typically found in tabular datasets. Unlike generative models designed for images or text, CTGAN is specifically tailored for tabular data, which often consists of a mix of discrete and continuous columns, with continuous columns possibly exhibiting multiple modes and discrete columns sometimes being imbalanced. CTGAN introduces several methodologies to overcome such challenges, including mode-specific normalization for handling non-Gaussian and multimodal distributions, and a conditional generator combined with training-by-sampling to address the issue of data imbalance in categorical columns. The model employs a variational Gaussian mixture model to estimate the number of modes in each continuous column and normalize the values within each mode. This mode-specific normalization enables the representation of continuous values with arbitrary distributions in a form that is more amenable to neural network processing. The architecture of CTGAN is built around fully connected networks for both the generator and discriminator, incorporating techniques like batch normalization, ReLU activation, Gumbel softmax for generating discrete outputs, and the use of Wasserstein GAN with gradient penalty for training stability. The conditional generator aspect of CTGAN allows for the generation of synthetic data under specific conditions, making it possible to generate samples that match given criteria in discrete columns, thus enabling a more balanced representation of all categories within the synthetic dataset. Note that an illustration of this architecture can be

found in Appendix 6.1.1.

While CTGAN demonstrates commendable performance in modelling complex tabular datasets, surpassing conventional models and Bayesian methods in benchmark comparisons, its application in generating synthetic data raises important considerations for privacy preservation. Specifically, its potential in financial contexts underscores the need for further investigation into how effectively it can safeguard personal information, ensuring that synthetic outputs do not compromise privacy. Additionally, CTGAN's focus on tabular data, rather than time-series information, presents an area for future exploration, particularly in evaluating its capability to extend privacy-preserving advantages to the generation of synthetic time-series financial data.

2.3.2 TVAE

The TVAE (Triplet-based Variational Autoencoder) model by Ishfaq, Hoogi, and Rubin (2023) introduces a novel approach by integrating deep metric learning with the VAE framework to generate high-quality synthetic data while preserving semantic similarities. This integration is designed to overcome limitations in traditional VAEs and representation learning techniques, which often fail to capture fine-grained semantic structures and relations within data. TVAE introduces a new triplet loss alongside the standard evidence lower bound (ELBO) of VAE, optimizing the latent space embeddings to be more informative and semantically structured. The main innovation of TVAE's architecture is the modification of the VAE encoder to produce latent vectors that are then subjected to triplet loss minimization. This process involves selecting triplets of data points (anchor, positive, and negative samples) such that the anchor is more similar to the positive sample than to the negative sample. By minimizing the triplet loss, which encourages the anchor and positive samples to be closer in the latent space than the anchor and negative samples, TVAE ensures that the learned embeddings reflect the semantic relationships inherent in the data. Note that an illustration of this architecture can be found in Appendix 6.1.2.

The TVAE model was extensively tested on MNIST (LeCun, Cortes, and Burges 2010), demonstrating its superiority over traditional VAEs in capturing semantic similarities. The experiments showcased a significant improvement in triplet accuracy, with TVAE achieving a 95.60% accuracy compared to the 75.08% of traditional VAEs. This improvement indicates that TVAE is more adept at preserving the relative distances among different classes in the latent space, making it highly effective for tasks that rely on semantic understanding and similarity.

The TVAE model's innovative approach to learning semantically rich latent embeddings offers potential for applications in generating synthetic financial data. Financial datasets often contain complex, multidimensional relationships crucial for fraud detection, risk assessment, and customer segmentation. The ability of TVAE to preserve semantic similarities and detailed relationships in the latent space makes it particularly suitable for synthesizing financial data that retains the intricate patterns and structures of real datasets. This capability is vital for creating synthetic datasets that can be used for model training and testing without exposing sensitive financial information, thus mitigating privacy and confidentiality concerns. Moreover, the enhanced semantic clustering observed with TVAE's embeddings suggests that the synthetic financial data generated would not only be diverse but also representative of the original data's underlying distributions. This aspect is crucial for developing robust financial models that are trained on comprehensive and varied datasets, ensuring their effectiveness in real-world scenarios. However, the application of TVAE in generating synthetic financial data underscores the need for a thorough investigation into its privacy-preserving capabilities. While the model's effectiveness in embedding semantic similarity is clear, it is crucial to ensure that the synthetic data does not allow for the reconstruction of, or inference about, real individual data points. This concern becomes particularly significant in the context of financial information, where the unauthorized disclosure of data could have serious consequences.

2.3.3 TimeGAN

Time-series Generative Adversarial Networks (TimeGAN) model by Yoon, Jarrett, and Schaar (2019) presents an innovative approach for generating realistic time-series data by capturing temporal dynamics. TimeGAN's novelty lies in its dual emphasis on maintaining the unsupervised flexibility of GANs while incorporating supervised learning to ensure the generated data adhere closely to the temporal dynamics observed in the training data. This amalgamation of techniques addresses a crucial gap in existing methods, where either the temporal correlations unique to time-series data are not fully captured or the models are inherently deterministic and lack generative flexibility.

TimeGAN is comprised of four key components. (1) an embedding network transforms the original high-dimensional time-series data into a lower-dimensional latent space, effectively capturing the essential temporal dynamics in a more manageable form. The embedding process is designed to be reversible, allowing for the recovery of data that closely matches the original time series in structure and content. (2) Operating in tandem with the embedding network, a recovery network maps the latent low-dimensional representations back to the original data space. This network ensures that the process of compression and decompression (embedding and recovery) retains the critical information necessary for realistic time-series generation. (3) A generator then uses the latent representations as inputs to produce synthetic time-series data and (4) a discriminator evaluates both real and synthetic data sequences to challenge the generator by distinguishing between them. This process thereby pushes the Generator to produce increasingly realistic data. These components work together to learn a low-dimensional representation of time-series data, generate new data points within this space, and ensure that these points follow the temporal patterns of the original dataset. Note that an illustration of this model can be found in Appendix 6.1.3.

Given its sophisticated handling of time-series data, TimeGAN presents an opportunity for application within financial scenarios, where time-series data is abundant and complex.

Financial transaction data is characterized by intricate temporal dependencies, making the generation of realistic synthetic sequences both a technical challenge and a privacy concern. As this thesis aims to evaluate the privacy-preserving capabilities of generative models, TimeGAN's ability to produce convincing financial time-series data—while potentially maintaining individual privacy—merits further exploration. Specifically, investigating how TimeGAN could be tailored or applied to financial datasets could shed light on its utility in generating synthetic data that preserves the statistical characteristics of original financial transactions without compromising personal and sensitive information.

2.3.4 DoppelGANger

Lin et al. 2019 introduced the current SOTA method utilizing GANs for the generation of synthetic time series datasets with their DoppelGANger (DG) framework. DG was specifically designed to overcome the complexities inherent in time series data such as preserving long-term dependencies, multidimensional inter-relationships, and preventing mode collapses. Overall, DG comprises two generators and two discriminators, working in concert to produce high-fidelity synthetic data. The first generator, an MLP, is dedicated to generating metadata, capturing the intricate relationships between different metadata attributes. Simultaneously, an RNN generator, specifically tailored with batched generation techniques, focuses on producing time series data, ensuring the capture of temporal dependencies and dynamics over extended sequences. To address possible mode collapse, DG integrates an auto-normalization technique within its generation process, dynamically adjusting the range of data during training. The architecture is complemented by two discriminators: the primary discriminator assesses the authenticity of the combined metadata and time series output, while an auxiliary discriminator specifically evaluates the metadata, enhancing the model's ability to generate realistic and diverse metadata. This dual-discriminator approach not only improves the overall data fidelity but also facilitates a nuanced balance between the generation of metadata and time series data, ensuring that

the intricate interdependencies within the data are preserved. Note that the architecture for DoppelGANger is visualized in Appendix 6.1.4.

The evaluation of DG across several real-world datasets showcases its ability to replicate the statistical properties of the original data, markedly outperforming existing techniques in capturing crucial data characteristics. This facilitates more accurate and reliable data sharing, addressing both the fidelity and privacy concerns associated with synthetic data generation. However, the paper also critically examines the privacy implications inherent in GAN-based synthetic data generation, highlighting the challenges of preventing data memorization and ensuring a meaningful balance between data utility and privacy. While Lin et al. (2019) introduces methods to partially address these concerns, such as modifiable metadata distributions, it underscores the necessity for ongoing research to refine and evaluate GAN models further, particularly in enhancing privacy protections without significantly compromising data utility.

2.3.5 Banksformer

Banksformer (Nickerson et al. 2023) introduces a transformative approach to generating synthetic multivariate sequences of banking transactions, addressing the challenge of irregular transaction intervals and their dependence on timestamp-based features. This novel Deep Generative Model (DGM), distinguished by its transformer-based architecture, excels in emulating date-based patterns in transactional data more effectively than contemporary models. Central to Banksformer's architecture are two main components: the Transformer-Decoder (TD) for sequence generation and an innovative mechanism for date feature handling. The model is structured around a unique preprocessing step that translates timestamps into multiple features, enhancing the model's capability to learn and replicate date-based transaction patterns. This preprocessing includes encoding the day of the month, days to the month's end, day of the week, and the month itself using a sinusoidal function, aiding in the representation of temporal information. Banksformer

utilizes a sequential generation process, where each transaction attribute (transaction code, date, and amount) is generated in order, conditioned on the history of previous transactions. This approach allows Banksformer to capture the joint distribution of transaction attributes accurately. The architecture comprises an input layer that maps multivariate transactions to a higher-dimensional space, a decoder stack that processes the encoded sequence to generate a context sequence, and an output layer that transforms the context into transaction predictions. The output layer includes a conditional generation mechanism for generating transaction attributes sequentially, enhancing the model's fidelity to real-world data distributions. Note that the architecture for Banksformer is illustrated in Appendix 6.1.5. Banksformer's performance is rigorously evaluated against two datasets of banking transactions, demonstrating its superior ability to generate high-quality synthetic sequences that closely mirror the statistical properties of real transactional data. Through a comprehensive evaluation, including univariate distribution comparisons, N-gram frequency analysis, and joint distribution assessments, Banksformer exhibits a clear advantage in capturing the intricate relationships between transaction types and dates, as well as the overall statistical patterns found in transactional banking data.

Nickerson et al. (2023) briefly addresses the importance of privacy considerations in the use of deep generative models for creating synthetic data. The document highlights the motivation behind developing synthetic banking transaction data as a means to minimize reliance on real, private data, thereby addressing privacy concerns. However, it also underscores a significant future direction for the work: examining the privacy implications of models like Banksformer more closely. Before these models can be utilized to generate synthetic datasets that could replace real datasets with genuine privacy concerns, it is crucial to understand any potential information that could be leaked through the synthetic datasets. This suggests that while Banksformer presents a novel and effective approach to generating synthetic financial data, a comprehensive analysis of its privacy-preserving capabilities remains an area for future investigation to ensure that the synthetic data does

not inadvertently compromise sensitive information.

2.3.6 VAE-GAN Combination Models

Razghandi et al. (2023) introduces a novel VAE-GAN approach for generating synthetic time-series data related to energy consumption in smart homes, specifically targeting the challenges associated with the limited availability of publicly accessible datasets in this domain. The VAE-GAN hybrid model addresses the issue of mode collapse in GANs by collapsing the VAE decoder and GAN generator into one entity, sharing parameters and training them jointly. The architecture thus generates synthetic data samples based on a regulated latent space, deviating from the traditional GAN approach of producing samples from noise input. To evaluate the performance of the VAE-GAN-generated dataset, the study employs metrics such as KL divergence, maximum mean discrepancy (MMD), and Wasserstein distance, comparing the probability distributions of the real and synthetic data. The experiments demonstrate that the synthetic data closely matches the distribution of real data, highlighting the effectiveness of the VAE-GAN approach while avoiding mode collapse. Furthermore, the study explores integrating a Q-learning-based Home Energy Management System (HEMS). The online performance of the Q-learning-based HEMS, tested with real-world smart home data, demonstrated how the generated synthetic data proves valuable for training a higher-performance Q-learning-based HEMS compared to baseline approaches. Given its success in the application of smart homes, this research could be further extended to evaluate the approach to other domain-specific data generation problems, such as generating financial datasets.

Another paper by Lee (2022) proposed a similar novel approach for synthetic time-series generation, introducing a Recurrent VAE-GAN model. The model employs a variational recurrent neural network (VRNN) for the generator and recurrent neural network (RNN) architectures as the discriminator, with the addition of time-dependent stochastic variables

in the hidden state of the RNN. This augmentation aims to enhance the model's ability to capture temporal dependencies inherent in sequential data. Lee (2022) highlights the success of RNN-based GAN frameworks that use standard non-stochastic architectures. In this context, the incorporation of time-dependent stochastic variables is explored to improve the model's performance in capturing the complexities of sequential data. The cost function involves minimizing the Kullback-Leibler (KL) divergence as a regularizer, optimized using gradient descent. Comparing the performance of the proposed VRNN-based GAN framework against the existing non-stochastic RNN-based GAN methods, it was concluded that the recurrent VAE-GAN model outperforms traditional RNN-based GAN approaches. More detail regarding this VAE-GAN architecture can be found in Appendix subsubsection 6.1.6.

These combinations of generative models warrant further investigation, particularly in financial scenarios where privacy concerns are paramount. While both studies demonstrate the effectiveness of their respective models in generating high-fidelity time-series synthetic data, the aspect of privacy preservation, especially in the context of sensitive financial information, remains to be explored. Integrating privacy metrics into the evaluation framework would provide valuable insights into the model's ability to safeguard sensitive data. Additionally, transitioning from generic time-series data to financial time-series data introduces new challenges related to the unique characteristics and dependencies inherent in financial datasets. Investigating the adaptability of the proposed VAE-GAN frameworks to the intricacies of financial time-series data presents an opportunity to assess their ability to maintain data utility while ensuring no compromise on privacy.

2.3.7 Generative Models Summary

Table 2 summarizes the generative models discussed in subsection 2.3 above.

Table 2: Overview of Generative AI Models for Synthetic Data Generation

Model	Data Type	Description
CTGAN (Xu et al. 2019)	Tabular	Addresses the challenge of imbalanced data and mixed types through conditional generation and mode-specific normalization.
TVAE (Ishfaq, Hoogi, and Rubin 2023)	Tabular	Enhances VAE with deep metric learning, using triplet loss to capture finer-grained information in latent space for generating tabular data.
TimeGAN (Yoon, Jarrett, and Schaar 2019)	Time-series	Combines GAN with supervised learning to generate realistic time-series data, emphasizing the preservation of temporal dynamics through a learned embedding space and stepwise supervised loss.
DoppelGANger (Lin et al. 2019)	Time-series, Tabular	Focuses on generating networked time series data, leveraging GANs to produce realistic sequences with metadata, addressing mode collapse through batched RNN generation and auto-normalization techniques.
Banksformer (Nickerson et al. 2023)	Time-series, Tabular	Utilizes transformer models within a GAN framework to generate synthetic financial time-series data, aiming to capture complex temporal relationships and market dynamics effectively.
VAE-GAN (Razghandi et al. 2023)	Time-series, Sequential	Integrates Variational Autoencoders (VAE) with GANs to generate time-series data related to energy consumption in smart homes, mitigating mode collapse and leveraging a shared latent space for robust synthetic data generation.
Recurrent VAE-GAN (Lee 2022)	Time-series, Sequential	Employs a variational recurrent neural network (VRNN) for the generator and RNN architecture for the discriminator, incorporating time-dependent stochastic variables to enhance temporal dependency capture in sequential data generation.

2.4 Chapter Summary

This chapter provides background context into ongoing research in privacy and generative models, emphasizing their widespread adoption in data analysis and machine learning applications across industries. The literature review underscores the existing gap in research, particularly within the financial domain, highlighting the susceptibility to privacy breaches and attacks. With the increasing exploration of machine learning applications, the imperative for domain-specific privacy measures becomes evident. This synthesis illustrates current research into generative AI models to synthesize and protect data, thereby propelling the

discourse beyond general privacy-preserving methods and advocating for tailored solutions in financial data management.

3 Methodology

The following sections delve into the intricacies of the methodologies employed in this study, covering the dataset characteristics, the process of generating synthetic data, evaluating synthetic data through various metrics, and an extensive exploration of privacy evaluation metrics. These components collectively aim to scrutinize the privacy preservation capabilities of synthetic datasets generated by advanced models.

3.1 Dataset

The dataset utilized for this analysis is a real anonymized transaction dataset from Czech banks, made publicly available for the Principles and Practice of Knowledge Discovery (PKKD) conference in 1999. It encompasses over one million transactions spanning from January 1st, 1993, to December 31st, 1998. To offer a structured overview of the dataset, Table 3 and Table 4 are provided. Table 3 summarizes the dataset in terms of accounts and transactions, while Table 4 details the specific columns utilized within the dataset.

Table 3: Summary of the Czech Dataset

Accounts	**Transactions**	**Min Sequence Length**	**Max Sequence Length**	**Avg. Seq. Length**
4500	1,056,320	9	675	235

Table 4: Czech Dataset Features

Column Name	**Description**	**Property**
trans_id	Transaction ID	ID
account_id	Account associated with transaction	ID
date	Date of transaction	Temporal
type	Credit or withdrawal	Categorical
operation	Mode of transaction	Categorical
amount	Transaction amount	Numerical
balance	Account balance after transaction	Numerical
k_symbol	Characterization of the transaction	Categorical
bank	Partnering bank of the transaction	Categorical
account	Partnering account of the transaction	Categorical

For this thesis, the features utilized were `account_id`, `date`, `type`, `operation`, `amount`, and `k_symbol`. These features were selected based on their relevance to the analysis objectives and their potential to provide insightful results. The dataset can be accessed through the following link: https://data.world/lpetrocelli/czech-financial-dataset-real-anonymized-transactions.

3.2 Synthetic Data Generation

The goal of generating synthetic data in this thesis is to assess the privacy preservation capabilities of different generative models. Specifically, four distinct models for generating synthetic data are utilized: CTGAN and TVAE, which are primarily designed for tabular data, and DoppelGANger and Banksformer, which are adept at handling time-series tabular data. This diverse selection facilitates a comprehensive comparison across different data types and synthesis methodologies.

Due to the constraints of the experimental setup, particularly in relation to training duration and the feasibility of deploying models in production settings, TimeGAN and VAE-GAN architectures were not explored. These were excluded from experimentation as their training durations extended to approximately 2 days, which contrasts sharply with the roughly 6-hour training time for both CTGAN and TVAE synthesizers. The extended training time for TimeGAN and VAE-GAN was attributed to their architectures being inherently less efficient at handling tabular time-series data, incurring overhead.

All training processes were conducted on Google CoLab, leveraging a subscription to CoLab Pro+ to ensure access to enhanced computational resources and thereby optimize the training efficiency of the employed models. To provide clarity on the configuration of each generative model utilized in this study, Table 5 outlines the key parameters set for each model.

Table 5: Synthetic Data Generation Models Parameters

Banksformer	**DoppelGANger**	**CTGAN & TVAE**
epochs: 80	max_len: 700	epochs: 500
training_max_seq_len: 80	batch_size: 200	enforce_min_max_values: True
training_min_seq_len: 20	sample_len: 10	enforce_rounding: True
early_stop: 2		
dropout_rate: 0.1		
opt_name: adam		
d_model: 128		
num_layers_dec: 4		
num_heads: 4		
bs: 64		

3.3 Synthetic Data Evaluation Metrics

Following previous research by Liu (2023) who evaluated the quality of synthetic data generated by Banksformer and DoppelGANger, similar column-wise statistical similarity metrics were used for this thesis through a privacy lens.

As discussed in Chapter 2, Differential Privacy (DP) offers a rigorous mathematical framework for evaluating the privacy guarantees of data analysis and synthesis processes. It quantifies the privacy loss associated with the release of statistical information, ensuring that the inclusion or exclusion of a single individual's data in a dataset does not significantly affect the outcome of any analysis. In the context of this thesis, DP serves as an evaluation metric for assessing the privacy-preserving efficacy of synthetic financial datasets. This thesis approximates the application of DP by contrasting distance metrics between authentic datasets and their synthetic counterparts, concentrating on columns harbouring confidential financial data. For this evaluation, three principal distance metrics between distributions are utilized.

(1) The Jensen-Shannon divergence which is defined as $JS(P||Q) = \frac{1}{2}D(P||M) + \frac{1}{2}D(Q||M)$, where D is the Kullback-Leibler divergence, $M = \frac{1}{2}(P + Q)$, and P, Q are the probability distributions being compared. This is used for the analysis of distributions for categorical columns.

(2) Similarly, differences in entropy are calculated for categorical columns via equation $H(P) = -\sum_{x \in X} p(x) \log p(x)$, where H is the entropy and $p(x)$ is the probability mass function of P).

(3) For numerical columns, the distance between distributions are evaluated using the Wasserstein distance for continuous distributions, which is defined as $W(P, Q) = \inf_{\gamma \in \Gamma(P,Q)} \int_{\mathbb{R} \times \mathbb{R}} |x - y| d\gamma(x, y)$, where $\Gamma(P, Q)$ represents all couplings of P and Q.

These metrics provide insights into the statistical similarity between the distributions of the real and synthetic datasets, with a particular emphasis on ensuring that the synthetic data does not reveal sensitive individual-specific information. Further, column-wise comparisons of the direct univariate distributions were also conducted.

3.4 Privacy Evaluation Metrics

To evaluate the privacy preservation capabilities of the generative models mentioned above in Chapter 3.2, various privacy evaluation metrics are employed. The selected metrics are data analysis privacy metrics, re-identification risk, attribute disclosure risk, and membership inference attack success rate. The following sections describe how these evaluation metrics are assessed.

3.4.1 Data Analysis Metrics

Three metrics for evaluating privacy preservation are K-anonymity, L-Diversity, and T-closeness. Each of these metrics offers a distinct perspective on how well a synthetic dataset can protect individual identities and sensitive attributes from potential privacy attacks, thus providing a comprehensive view of the dataset's privacy characteristics.

K-Anonymity measures the extent to which individual records in a dataset can be de-identified (Sweeney 2002). A dataset is said to satisfy K-Anonymity if each record is

indistinguishable from at least $k-1$ other records concerning certain identifying attributes, known as quasi-identifiers. In the context of financial data, quasi-identifiers might include transaction types, operations, or symbols associated with transactions. The higher the value of k, the greater the level of anonymity, as it becomes increasingly difficult for an attacker to single out an individual based on the quasi-identifiers.

L-Diversity extends K-Anonymity by not only considering the identifiability of records but also the diversity of sensitive attributes within groups of records that are indistinguishable from each other. An L-Diverse dataset ensures that for every group of records sharing the same quasi-identifiers, there are at least l "well-represented" values for the sensitive attributes (Machanavajjhala et al. 2007). This metric is particularly relevant for financial datasets where sensitive attributes might include transaction amounts or account balances. L-Diversity protects against attribute disclosure by ensuring that an attacker cannot accurately infer the value of a sensitive attribute from the quasi-identifiers.

T-Closeness further refines the concept of privacy preservation by addressing the limitation of L-Diversity regarding the distribution of sensitive attributes. A dataset achieves T-Closeness if the distribution of a sensitive attribute for any group of records with the same quasi-identifiers is no further than a threshold t from the distribution of the attribute in the full dataset (Li, Li, and Venkatasubramanian 2007). This metric is crucial for financial datasets as it ensures that the presence of an individual in the dataset does not significantly alter the overall distribution of a sensitive attribute, such as the number of transactions, thereby preventing inference attacks based on statistical analysis.

Each of the generated synthetic datasets was evaluated against these 3 privacy metrics for this thesis.

3.4.2 Re-identification Risk

Re-identification risk refers to the probability that an individual's record in a dataset, even after being anonymized or synthesized, can be traced back to the individual, compromising

their privacy (Benitez and Malin 2010). This risk is particularly significant in contexts where sensitive information is involved, such as financial or medical data. An essential aspect of synthetic data generation is thus to ensure that while the synthetic data retains the utility of the real data for analytical purposes, it does not allow for the re-identification of individuals whose data contributed to the synthetic dataset's creation.

To assess the re-identification risk associated with synthetic data, (1) the frequency distributions of the categorical variables between the real and synthetic data are compared and (2) the congruity of the numerical variables' distributions is gauged through a distance metric. For categorical variables such as `operation`, `type`, and `k_symbol`, the total difference in distribution is evaluated by summing the absolute value differences of category frequencies in both datasets, normalized over two to account for distribution in either dataset. This summation serves as an indicator of the degree to which both datasets resemble each other in terms of category representation. For numeric variables, the metric selected is the Kolmogorov-Smirnov (KS) Test (Dimitrova, Kaishev, and Tan 2020), a non-parametric test used to compare the distributions of two datasets. Specifically, the KS statistic quantifies the distance between the empirical distribution functions of the synthetic and real datasets, reflecting the maximum discrepancy observed between two cumulative distributions. This statistic is defined by the equation

$$D_{n,n'} = \sup_x |F_{1,n}(x) - F_{2,n'}(x)|,$$

where $D_{n,n'}$ is the KS statistic, $F_{1,n}$ and $F_{2,n'}$ are the empirical distribution functions of the real and synthetic datasets respectively, and sup denotes the supremum function capturing the greatest discrepancy across all values of x. A closer alignment of the synthetic data's distribution to the actual data's distribution raises the re-identification risk. Conversely, a substantive disparity suggests an obfuscation of individual records, diminishing the threat of re-identification.

3.4.3 Attribute Disclosure Risk

Attribute disclosure risk quantifies the likelihood that sensitive information about individuals can be inferred from a dataset (Hittmeir, Mayer, and Ekelhart 2020). In the context of synthetic data, this risk assessment helps evaluate the degree to which synthetic data generation processes preserve individual privacy by preventing the accurate prediction of sensitive attributes based on other available data attributes.

To compute attribute disclosure risk, a systematic approach involving predictive modelling is employed. Specifically, Random Forest (Breiman 2001) constructs a multitude of decision trees during training and outputs the class that is the mode of the classes of the individual trees, offering a practical way to evaluate how well a sensitive attribute can be predicted from other attributes in the dataset. The assessment process follows a 70-30 train-test split methodology and predicts sensitive attributes based on other attributes within the dataset. Attribute disclosure risk can be evaluated by comparing the prediction accuracy between models trained on real versus synthetic data. High accuracy indicates a potential risk of attribute disclosure, as it suggests that the synthetic data closely mirrors the real data in terms of the relationships between attributes, allowing sensitive attributes to be predicted accurately. Conversely, lower accuracy on synthetic data, as compared to real data, implies that the synthetic data generation process introduces sufficient ambiguity to reduce the risk of attribute disclosure, enhancing privacy protection.

This methodology provides a quantifiable measure of attribute disclosure risk, informing the development of synthetic data generation processes that balance data utility and privacy protection. By meticulously evaluating this risk, it becomes possible to identify and mitigate potential privacy vulnerabilities in synthetic datasets, ensuring that they can be used safely for research and analysis without compromising individual privacy.

3.4.4 Membership Inference Attack Success Rate

Membership Inference Attacks (MIAs) pose a significant privacy risk in machine learning, aiming to determine whether a particular data record was used in the training set of a model (Liu et al. 2021). In the context of synthetic data generation, MIAs seek to identify if the synthetic data retains identifiable traces of the real training data, thereby compromising individual privacy. Such attacks exploit differences in the model's behaviour on seen (training) versus unseen data, leveraging these discrepancies to infer membership.

The methodology used for assessing the vulnerability of generative models to MIAs involves an experiment that mirrors the attack process. The experiment uses both real data, which the generative model was trained on, and synthetic data, which the model generated. The core idea is to see if an external classifier can distinguish between real and synthetic data, effectively performing a membership inference in a 4 step process.

(1) The real training data is combined with the generated synthetic data and assigned binary labels respectively. This labelling creates a dataset that reflects potential knowledge an attacker might possess. (2) Binary classifiers are trained as attack models to learn to differentiate between real and synthetic records. (3) Performance metrics of the attack models are evaluated. High performance in distinguishing real from synthetic data implies a higher risk of re-identification and privacy breaches. Conversely, if the attack model cannot reliably differentiate between the two, it suggests that the synthetic data generation process effectively obfuscates the characteristics of individual records, enhancing privacy. Note that the success rate of such attacks should ideally be close to random guessing (50%) for the synthetic data to be considered privacy-preserving. (4) Steps 1-3 are done multiple times while varying n, the number of accounts to which an attacker has access to labels.

The binary classifiers used in this thesis are logistic regression, Naive Bayes, Random Forest, XGBOOST, GBM, and a feedforward neural network. The hyperparameters and training details for these models can be found in Appendix 6.2 and the performance metrics for evaluation are balanced accuracy, precision, recall, and f1 scores using 5-fold

cross-validation. As such, all experiments were run five times with different random splits of the training and test data, and the results were averaged. The n values investigated were $1, 10, 25, 50, 100, 250, 500, 750$, and 1000. Using Google Colab Pro+, fitting classifiers for each generated synthetic dataset took approximately 4 hours each.

3.4.5 Model Inversion Attack Success Rate

As described in Chapter 2, model inversion attacks represent a class of privacy threats where an adversary aims to reconstruct input data or extract sensitive information about the dataset from the model's output or behavior. In the context of synthetic data generation, model inversion attacks could theoretically be used to evaluate the privacy of generated synthetic datasets by attempting to reverse-engineer the generative model to reveal or approximate the real data that was used for training. The core idea is to exploit the model to infer sensitive attributes of the individuals in the dataset, thereby compromising their privacy. However, applying model inversion attacks to assess the privacy of synthetic financial datasets generated in this thesis is not explored due to the inherent challenge of establishing a one-to-one mapping between synthetic and real accounts. Synthetic data generation processes are designed to ensure that the synthetic data, while statistically similar to the real data, does not directly correspond to or reveal information about specific real-world entities or individuals. Therefore, the lack of a direct mapping between synthetic and real accounts makes it impractical to use model inversion attacks as a privacy evaluation mechanism in this context. A potential method to circumvent this issue could involve clustering the data to identify which synthetic accounts are most similar to real accounts and then using the real amounts as targets for inversion attacks. However, this approach is not feasible with the dataset used in this thesis, as the data cannot be clustered (Pita 2024). This limitation underscores the challenge of directly assessing the privacy-preserving properties of synthetic datasets using model inversion attacks and highlights the need for alternative privacy evaluation methods that are better suited to the characteristics of synthetic financial

data.

3.5 Chapter Summary

This chapter has detailed the methodologies adopted in this thesis, encompassing the examination of a real anonymized transaction dataset, the generation of synthetic data via state-of-the-art models, and the evaluation of this data against rigorous privacy and data analysis metrics. The thorough exploration of generative models such as CTGAN, TVAE, DoppelGANger, and Banksformer, alongside the assessment of their outputs through differential privacy simulation and privacy evaluation metrics like K-Anonymity, L-Diversity, and T-Closeness, underscores the study's commitment to investigating the privacy-preserving capabilities of synthetic financial data. This comprehensive methodological framework not only evaluates the fidelity and utility of synthetic datasets but also ensures their alignment with privacy standards, thereby contributing to the field of privacy-preserving synthetic data generation.

4 Results and Discussion

This chapter presents and discusses the results from methods outlined in Chapter 3.

4.1 Statistical Analysis

Table 6 presents a comprehensive evaluation of the differential privacy aspects of synthetic datasets generated by CTGAN, TVAE, DoppelGANger, and Banksformer, employing metrics such as Entropy Difference, Wasserstein Distance, and Jensen-Shannon (J-S) Divergence across different attributes.

Table 6: Column-wise Comparisons of Distance Metrics

Model	**Type Entropy Difference**	**Operation Entropy Difference**	**K_Symbol Entropy Difference**	**Amount Wasserstein Distance**
CTGAN	0.0029	0.0474	0.0529	667.3731
TVAE	0.0001	0.0300	0.0480	920.6838
DoppelGANger	8.4480	6.5106	4.0596	10904.6761
Banksformer	8.1683	5.2654	3.3487	4378.2261
Model	**Type J-S Divergence**	**Operation J-S Divergence**	**K_Symbol J-S Divergence**	
CTGAN	0.0269	0.1131	0.1228	
TVAE	0.0047	0.0847	0.1014	
DoppelGANger	0.7069	0.7293	0.4470	
Banksformer	0.6398	0.5144	0.4457	

Notably, CTGAN and TVAE exhibit minimal entropy differences and J-S divergence values across `type`, `operation`, and `k_symbol` categories, indicating a close statistical similarity to the real dataset while maintaining a lower risk of privacy breaches. In contrast, Banksformer and DoppelGANger show significantly higher values in these metrics, suggesting a greater departure from the real data distribution, which could imply a higher degree of privacy but at the potential cost of utility. This evaluation is also evident upon visual inspection of the univariate distributions for the columns, as displayed in Figure 2.

The Wasserstein distance, particularly for the `amount` attribute, further differentiates

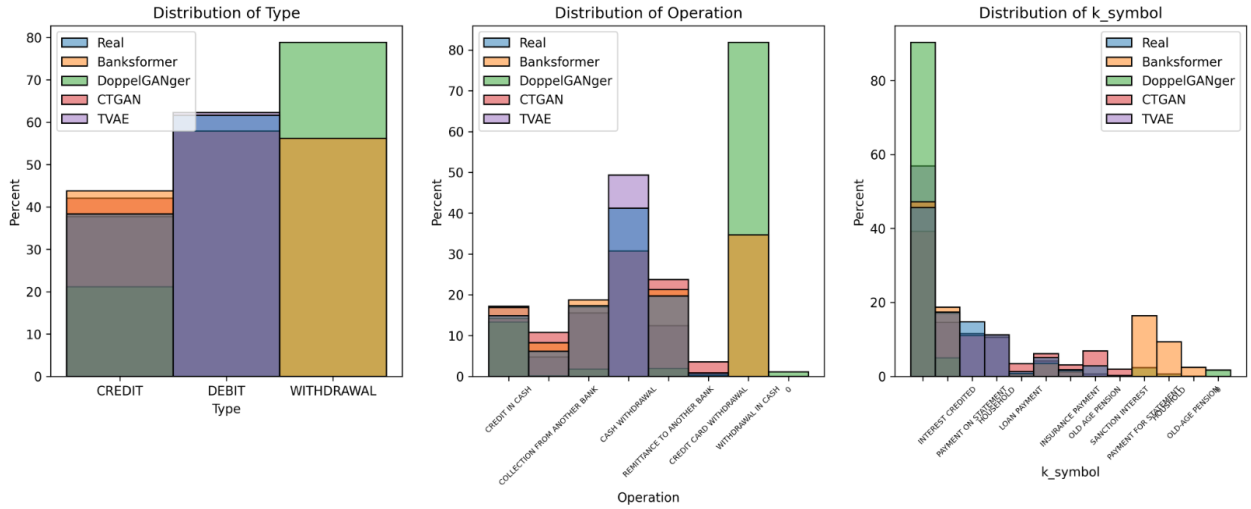


Figure 2: Univariate Distributions of Categorical Features

the models. CTGAN and TVAE show lower distances, again indicating a closer match with the real data, which corroborates their lower entropy differences and J-S divergence values. However, the high Wasserstein distance observed with DoppelGANger underscores a substantial discrepancy in the distribution of transaction amounts, potentially indicating either a higher level of privacy preservation or a loss in data utility. This result is supported by Figure 3 which visualizes the real and synthetic distributions for the `amount` column. Individual distributions compared to the real data can be viewed in Appendix 6.3.

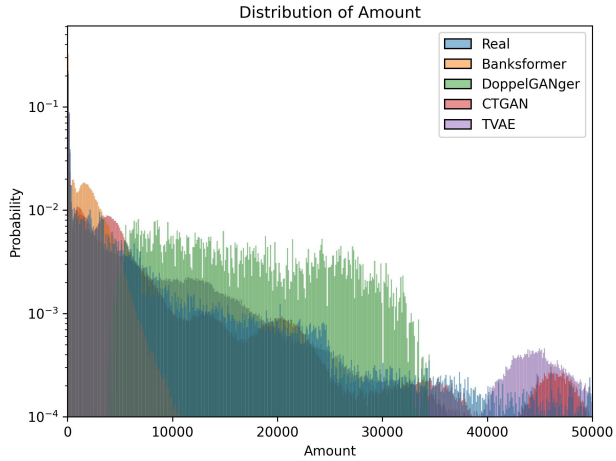


Figure 3: Univariate Distribution of Continuous Feature Amount

Overall, these results underscore the nuanced balance between privacy preservation

and data utility inherent in the generation of synthetic datasets. CTGAN and TVAE demonstrate an ability to closely mimic the real data's distribution with minimal privacy risk indicators, making them potentially more suitable for applications requiring high data fidelity. Conversely, the larger discrepancies observed with Banksformer and DoppelGANger may render these models more appropriate for scenarios where maximizing privacy is the primary concern, albeit with a possible compromise on the synthetic data's utility for certain analyses.

4.2 Data Analysis Metrics Results

Table 7 displays the results from each of the three privacy evaluation metrics for the generated synthetic data. The results presented in Table 7 highlight significant variations in privacy

Table 7: Comparison of Privacy Metrics across Datasets

Model	**K-Anonymity**	**L-Diversity**	**T-Closeness**
Real	23	3	-
CTGAN	1	1	0.4787
TVAE	1	1	0.5060
DoppelGANger	1	1	0.0781
Banksformer	14	14	0.3827

preservation capabilities across different generative models when compared with the real dataset. The real dataset exhibits a K-Anonymity of 23 and an L-Diversity of 3, indicating a relatively high level of privacy protection, with each record being indistinguishable from at least 22 others based on quasi-identifiers, and each group of indistinguishable records containing at least 3 well-represented sensitive attributes.

In contrast, synthetic datasets generated by CTGAN, TVAE, and DoppelGANger all exhibit a K-Anonymity of 1 and an L-Diversity of 1, significantly lower than the real dataset, suggesting a diminished ability to protect against re-identification and attribute disclosure attacks. This is particularly concerning as it implies that each record in these synthetic datasets is unique concerning the quasi-identifiers, and there is no diversity in the sensitive

attributes within these groups, making individuals potentially identifiable.

T-Closeness, which measures the similarity in the distribution of a sensitive attribute between the synthetic and real datasets, further elucidates the privacy implications. DoppelGANger exhibits a remarkably low T-Closeness (0.0781), indicating a significant departure from the real data distribution and potentially better privacy preservation. However, this could also indicate a loss of utility or representativeness of the synthetic data. CTGAN and TVAE have higher T-Closeness values (0.4787 and 0.5060, respectively), suggesting a closer match to the real data distribution but potentially at the cost of privacy. Banksformer stands out with a K-Anonymity of 14 and an L-Diversity of 14, alongside a moderate T-Closeness of 0.3827, indicating a more balanced approach to privacy preservation without severely compromising data utility.

These results suggest that while some synthetic data generators like Banksformer are capable of producing data that is both useful and relatively privacy-preserving, others may require further refinement to enhance their privacy protection capabilities. The trade-off between data utility and privacy preservation remains a critical consideration in the design and evaluation of generative models for synthetic data generation.

4.3 Re-identification Risk Results

The re-identification risk assessment for the four synthetic data generation models reveals significant differences in their ability to mimic the original data distribution while potentially preserving privacy. Table 8 displays the categorical distribution similarities and the Kolmogorov- Smirnov (KS) statistic for the `amount` column, providing insights into each model's performance. Note that here the P-values for all models are 0 for the `amount` column, indicating statistically significant differences from the original data's distribution. However, the practical implications of these differences for privacy and utility need further evaluation within specific use cases.

CTGAN shows relatively minor differences in the `operation`, `type`, and `k_symbol`

Table 8: Re-identification Risk Assessment for Synthetic Data Models

Model	**Operation Distribution Difference**	**Type Distribution Difference**	**K_Symbol Distribution Difference**	**Amount KS Statistic (P)**
CTGAN	0.1139	0.0372	0.1040	0.1183 (0)
TVAE	0.2291	0.0064	0.4717	0.0775 (0)
DoppelGANger	0.4148	0.3944	0.4705	0.6687 (0)
Banksformer	0.2432	0.3355	0.1707	0.2726 (0)

distribution differences (0.1139, 0.0372, and 0.1040 respectively), indicating a closer resemblance to the real data distributions for these categories. However, the KS statistic for the `amount` column is 0.1183 with a p-value of 0, suggesting that while the synthetic data may look similar in categorical terms, the numerical distribution is discernibly different, potentially mitigating re-identification risks. This result suggests that CTGAN is most effective at generating synthetic data that closely matches the original dataset's distributions while potentially offering a balance between privacy preservation and data utility. TVAE, while maintaining a close similarity in the `type` category (0.0064), shows a marked departure from the original data in the `k_symbol` category (0.4717), which is the highest among all models. The KS statistic for `amount` is lower than CTGAN at 0.0775, which suggests that TVAE may better obscure the individual record characteristics in the numerical data, thereby reducing re-identification risk.

DoppelGANger presents the highest distribution differences across all categorical variables and the largest KS statistic (0.6687) for the `amount` column. This significant divergence from the real data implies a stronger obfuscation of the original dataset characteristics, potentially lowering the re-identification risk but also potentially impacting the utility of the synthetic data for analytical purposes. Banksformer displays moderate distribution differences in `operation` and `type` (0.2432 and 0.3355 respectively), and a lower difference for `k_symbol` (0.1707) compared to DoppelGANger. However, the KS statistic for `amount` (0.2726) with a p-value of 0 indicates a substantial difference from the real data distribution, suggesting an effective reduction in re-identification risk, albeit possibly at the

expense of data utility.

While all models exhibit varying degrees of success in mimicking the real data distributions and potentially reducing re-identification risks, trade-offs between data utility and privacy preservation are evident. DoppelGANger's significant distribution differences may offer the strongest privacy protection but raise concerns about data utility. CTGAN and TVAE provide a more balanced approach, with close categorical similarities to the real data yet differences in the numerical distributions, suggesting a nuanced balance between preserving privacy and maintaining data utility. CTGAN demonstrates the best overall performance in terms of minimizing the total difference in distribution for categorical variables while maintaining a low KS statistic. Banksformer's performance sits between these extremes, offering moderate privacy protection without overly compromising data utility. In light of these findings, future work could involve investigating the optimal distribution differences that strike a balance between privacy and utility, particularly in the context of re-identification risk analysis, to enhance the applicability of synthetic datasets in sensitive domains.

4.4 Attribute Disclosure Risk Results

Table 9 outlines the accuracy of predicting categorical attributes (operation, type, and k_symbol) on synthetic data generated by different models, in comparison with real data accuracy. This assessment provides insights into each model's ability to obfuscate sensitive attributes, hence mitigating the risk of attribute disclosure.

Table 9: Attribute Disclosure Risk Assessment

Model	**Operation Accuracy**	**Type Accuracy**	**K_Symbol Accuracy**
Real Data	0.9856	1.0000	0.9859
CTGAN	0.9175	0.9964	0.8335
TVAE	0.9017	0.9933	0.8544
DoppelGANger	0.8335	0.8180	0.9188
Banksformer	0.9190	1.0000	0.9278

The benchmark accuracies obtained from the real data highlight the inherent predictability of attributes such as `operation`, `type`, and `k_symbol`, with accuracies ranging from approximately 98.56 to 100%. These high accuracies underscore the potential vulnerability to attribute disclosure risks if such patterns are not sufficiently obfuscated in the synthetic data.

Comparatively, the synthetic data generated by all models exhibit lower accuracies in predicting these attributes. This reduction in accuracy represents an improvement in terms of privacy preservation, as it indicates a decrease in the predictability of sensitive attributes. Notably, none of the synthetic data models achieve higher accuracy than the real data, suggesting that all models introduce a level of randomness or complexity that effectively reduces the attribute disclosure risk. This outcome signals progress in the development of privacy-preserving synthetic data generation methods.

Among the models assessed, CTGAN and TVAE stand out for their substantial reduction in prediction accuracy for certain attributes, particularly `operation` and `k_symbol`, compared to the real data benchmark. However, the `type` attribute remains relatively predictable across most models, with CTGAN and Banksformer showing high accuracies that closely mirror the real data. This persistence of predictability for the `type` attribute points to an area where further refinement of synthetic data generation techniques is needed to enhance privacy protection.

The results collectively indicate a move towards more privacy-preserving synthetic data models, as evidenced by the generally lower prediction accuracies for synthetic data compared to real data. However, the continued presence of high accuracy in predicting certain attributes in synthetic datasets calls for ongoing efforts to improve the mechanisms by which these models obfuscate sensitive information. Future advancements should aim to reduce the predictability of all attributes further, ensuring that synthetic data can be used for analysis and research without compromising individual privacy.

4.5 Membership Inference Attack Results

Figure 4 displays the accuracy and f1 results from simulating model inference attacks using binary classifiers logistic regression, XGBoost, random forest, Naive Bayes, GBM, and a simple neural network. These results reveal several insights into the privacy preservation capabilities of models CTGAN, TVAE, DoppelGANger, and Banksfomer. The full results including precision and recall graphs can be found in Appendix 6.4 and example confusion matrices for 1 fold of each model can be found in Appendix 6.5.

It is observed that logistic regression tends to overfit and predict a single class, as indicated by a consistent 50% balanced accuracy across all synthetic datasets. This behaviour indicates that logistic regression may not be a reliable indicator of privacy risk since its performance does not truly reflect the classifier's ability to differentiate between real and synthetic data. This overfitting can also be visualized in Appendix 6.5 where the confusion matrices for logistic regression illustrate only 1 class being predicted.

Overall, the ideal 50% accuracy rate, indicative of a perfect balance between privacy and data utility where the attacker is essentially guessing, is not consistently achieved by any model. However, the general trend before convergence indicates that with an increased number of data points available to the attacker (n), classifiers become better at distinguishing between real and synthetic data. This convergence suggests that beyond this threshold, an attacker gains no additional advantage with more information, highlighting a potential limit in the usefulness of additional data for MIAs.

The results for Banksformer reveal a troubling trend with recall, precision, and F1 score metrics starting high and remaining high across most classifiers as n increases. The model's sudden convergence towards 90% accuracy before $n = 100$ suggests that it may be retaining too many characteristics from the training data, making it easier for an attacker to identify the synthetic data as being derived from real data. This level of accuracy far exceeds the 50% threshold that would indicate a good balance between data utility and privacy,

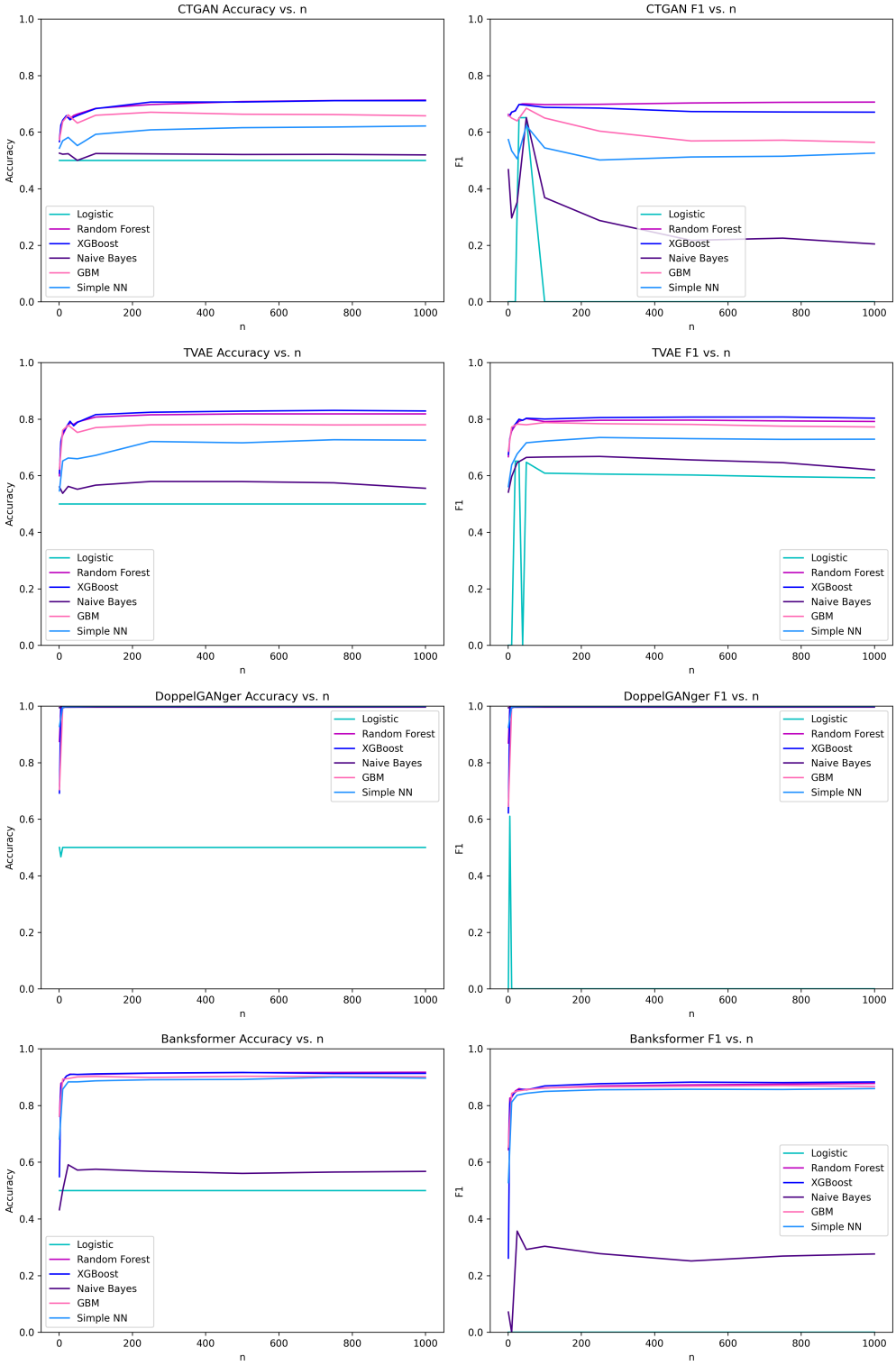


Figure 4: Model Inference Attack Accuracy and F1 Score over 5 Folds

implying a potential risk of re-identification in the synthetic data generated by Banksformer. Similarly for DoppelGANger, while the performance metrics initially vary, they eventually converge to near 100% accuracy, indicating a very high risk of privacy breaches. This swift convergence to high accuracy indicates that an attacker can easily distinguish between real and synthetic data, suggesting that the synthetic data generated by DoppelGANger is potentially too similar to the original training data, thus failing to effectively obfuscate the sensitive information.

CTGAN shows more favourable results for privacy preservation, with metrics indicating less confidence in distinguishing real from synthetic data. The accuracy, precision, recall, and F1 scores for CTGAN do not approach the high levels seen with Banksformer and DoppelGANger, suggesting that CTGAN may offer a higher degree of privacy protection. This can be seen in Figure 14 where the balanced accuracies for CTGAN converge within the 50% to 75% range. This aligns with the expectation that synthetic data should not be easily classifiable with respect to its source, thereby better safeguarding against MIAs. TVAE's metrics are similarly lower than Banksformer and DoppelGANger, converging to an accuracy range around 55% to 84% indicating a stronger performance in terms of privacy preservation. TVAE's synthetic data appears to introduce enough uncertainty to prevent an attacker from accurately identifying the data's origin, which is a positive sign for its use in privacy-sensitive applications. However, the metrics for TVAE, especially regarding precision and recall, do show an increase before plateauing, which suggests a potential for privacy risk, albeit lower than Banksformer and DoppelGANger.

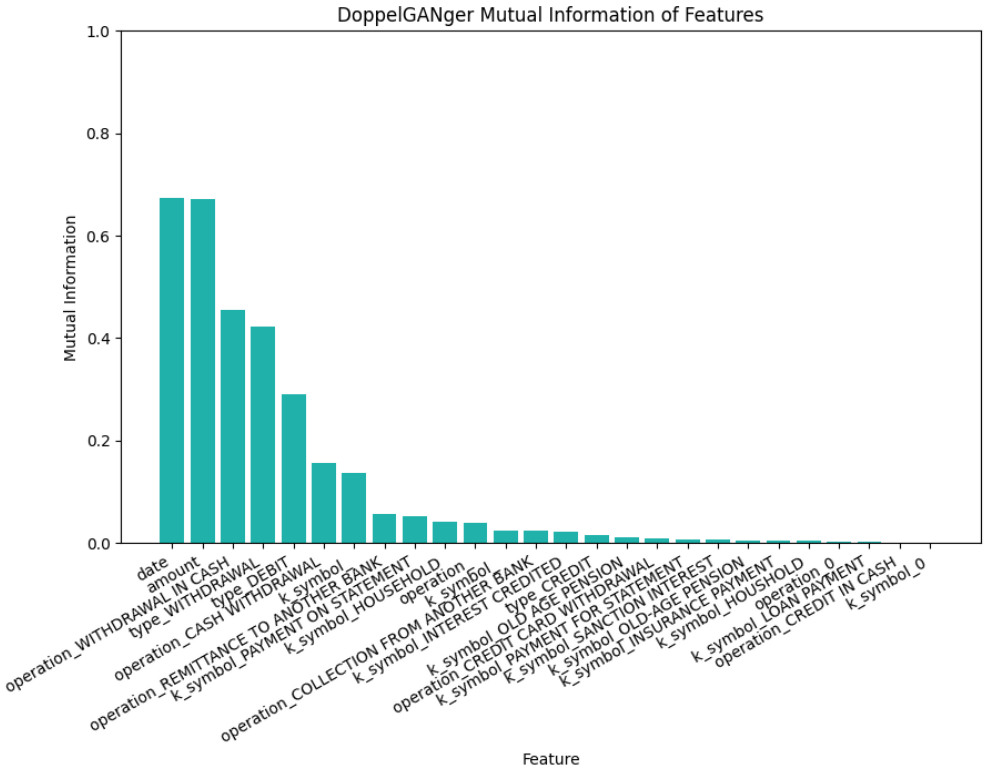
It is also observed that most of the binary classifiers converge almost immediately to around the same performance with Banksformer and DoppelGANger. The exception to this is Naive Bayes which converges to a lower accuracy of around 60% for Banksformer. With CTGAN and TVAE, there is noticeably more variation between the performance of different binary classifiers compared to Banksformer and DoppelGANger, resulting in larger accuracy ranges. Across the binary classifiers, XGBoost and Random Forest exhibit similar trends,

which is expected given that they are both tree-based models. These classifiers generally achieve higher accuracies compared to GBM, the simple neural network, and particularly Naive Bayes, which tends to perform the worst. The higher performance of tree-based models may be attributed to their robustness and ability to model non-linear interactions in the data, potentially making them more adept at detecting patterns that differentiate real from synthetic data.

The privacy implications based on these results are significant. High classification accuracy in Banksformer and DoppelGANger means that synthetic data generated by these models could potentially be reverse-engineered to reveal sensitive information from the original data. In contrast, the lower classification accuracies in CTGAN and TVAE suggest that these models are more effective at generating synthetic data that protects individual privacy, making them more suitable for situations where privacy is a paramount concern. However, it's essential to balance these findings with the practical utility of the synthetic data; while privacy is critical, the generated data must still be representative enough to be useful for its intended analytical purposes. The ideal synthetic data generation model would, therefore, be one that not only provides utility and representativeness but also ensures that individual privacy is not compromised in the process.

4.6 Mutual Information Investigation Results

The prompt convergence to high accuracies observed in membership inference attacks on the Banksformer and DoppelGANger models in the previous Section 4.5 necessitated further examination. Consequently, for a selected value of $n = 100$, the mutual information for each feature in relation to the classification label (real or synthetic) was computed to understand the depth of information shared between the variables and the potential for privacy breaches. The mutual information results for DoppelGANger and Banksformer are illustrated in Figure 5 below, where the same results for CTGAN and TVAE can be found in Appendix 6.6.



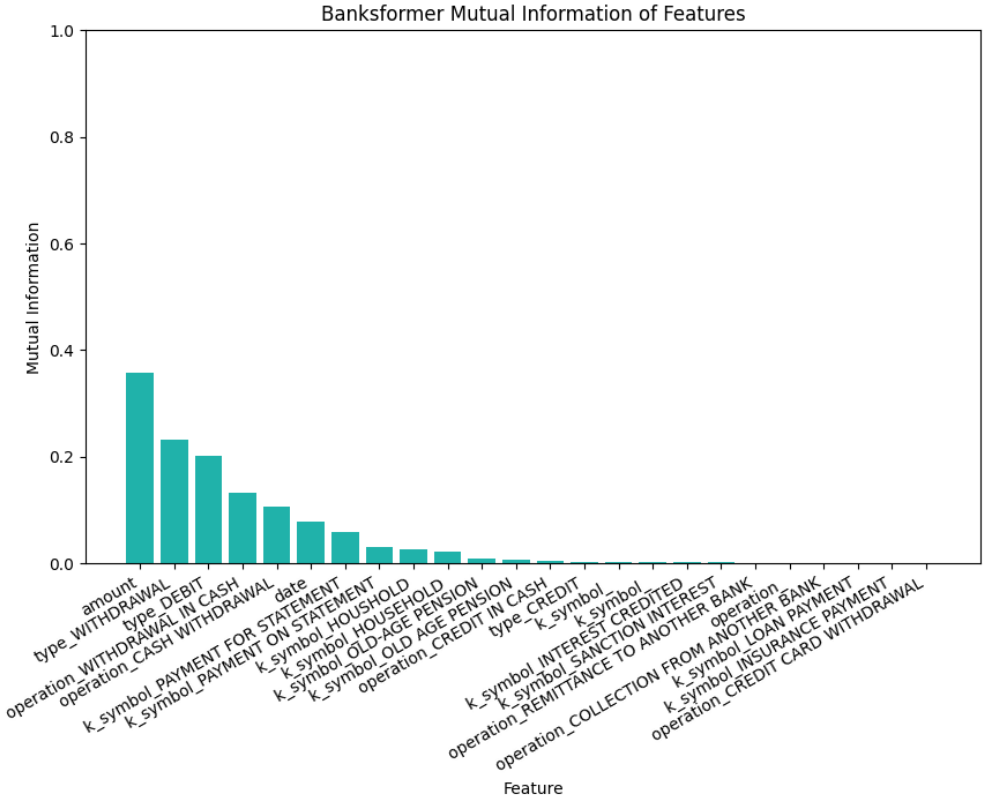


Figure 5: Mutual Information of DoppelGANger and Banksformer Features

As depicted in Figure 5, the DoppelGANger model exhibits significantly high mutual information for the features `date` and `amount`, with values around 0.7. This indicates a strong dependency between these features and the classification label, suggesting that these attributes could be pivotal in distinguishing real from synthetic data. Conversely, Banksformer presents a less pronounced but still substantial mutual information for the `amount` feature, which does not extend to the `date` feature. The distinction between the mutual information values for `date` in both models prompts a further investigation into the range of values for these features, to elucidate their impact on the re-identification risk and understand their contribution to the overall privacy preservation of the synthetic datasets. This investigation is illustrated by Table 10.

Table 10: Range of Amounts and Dates in Real and Synthetic Data

Data	**Amount Min**	**Amount Max**	**Date Min**	**Date Max**
Real	0.00	87,400.00	1993-01-01	1998-12-31
DoppelGANger	3,687.45	35,496.73	0651-10-06	2228-02-14
Banksformer	0.00	24,066.96	1993-01-01	2002-07-31

For DoppelGANger, the deviation in dates is particularly noteworthy, with the synthetic range extending from the year 0651 to 2228, well beyond the real data's span from 1993 to 1998. Binary classifiers could easily leverage such an anomalous temporal range in the synthetic data to identify synthetic instances, as any date falling outside of the real dataset's time frame would be labelled as fake. This is a clear vector for the high accuracy rates seen in the membership inference attacks, as it presents a simple, deterministic rule for the classifier to follow, thus compromising the synthetic data's privacy. The amount fields show that the real data's range is substantially broader than both synthetic datasets, with the real maximum amount being nearly double the maximum for Banksformer and more than double for DoppelGANger. DoppelGANger's synthetic data does not contain any of the smallest transactions present in the real data, with its minimum amount being higher than any amount found in the real dataset. This discrepancy in the financial figures could be another factor contributing to the binary classifier's high performance, as transactions

falling below a certain threshold would automatically be identified as real.

The significant distortions in the key features of the synthetic datasets, particularly for DoppelGANger, highlight a privacy risk through potential re-identification. While these distortions may offer some protection against simpler re-identification strategies, they also render the data vulnerable to classifiers trained to detect such anomalies. To counter this, future work in synthetic data generation should focus on ensuring that key features such as dates and amounts maintain a distribution that is close to the real data, without introducing easy patterns for classifiers to exploit, thereby achieving a better balance between data utility and privacy. An experiment scaling the DoppelGANger dates to match the real data's date distribution was run over 1 fold, where the results are displayed in Figure 6.

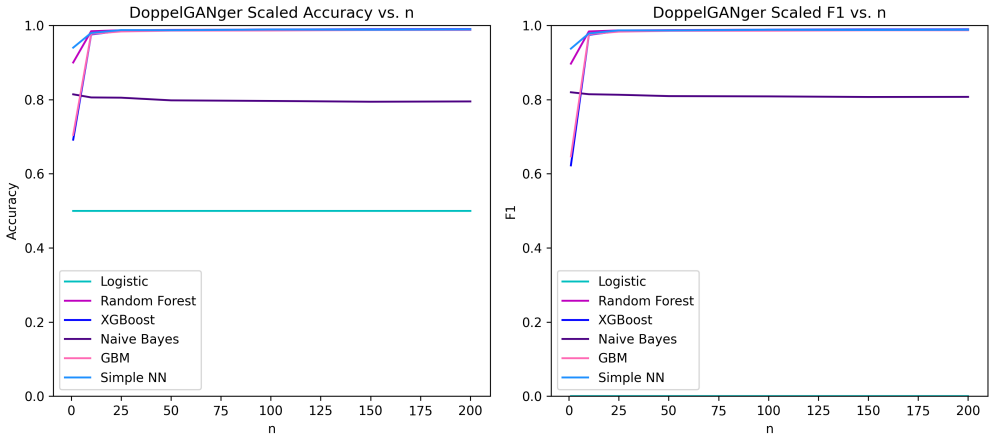


Figure 6: DoppelGANger Scaled Dates Membership Inference Results

The scaled data yields results that align closely with those observed for Banksformer, with Naive Bayes underperforming in comparison to other classifiers. Notwithstanding this adjustment, the privacy preservation outcomes are still profoundly concerning; classification accuracy immediately converges to 100% when trying to distinguish between real and fake data. This indicates that attributes other than dates allow for the synthetic data to be easily classified, underscoring the existence of additional features that contribute to the compromised privacy of the synthetic dataset. The persistent high accuracy across classifiers, apart from Naive Bayes, emphasizes the need for further scrutiny of the feature set to

identify and mitigate the factors contributing to such elevated re-identification risks in the DoppelGANger synthetic data.

4.7 Chapter Summary

This chapter presents a critical analysis of the privacy-preserving capabilities of various generative models used to synthesize financial datasets. It explores the differential privacy measures of CTGAN, TVAE, DoppelGANger, and Banksformer, utilizing entropy difference, Wasserstein distance, and Jensen-Shannon divergence to compare synthetic data against real data across several attributes. The findings reveal a spectrum of efficacy; CTGAN and TVAE exhibit closer statistical alignments with real data, while Banksformer and DoppelGANger display greater distribution discrepancies, which could impact data utility.

The chapter further discusses re-identification and attribute disclosure risks, showing that all synthetic models demonstrate lower predictability of sensitive attributes compared to real data, signaling progress in obfuscating potentially identifying information. Nevertheless, the analysis underscores a trade-off between data utility and privacy preservation, with certain models achieving balance more adeptly than others. The examination of membership inference attacks highlights particular vulnerabilities in the Banksformer and DoppelGANger models, where high classification accuracies raise concerns over privacy protection. In contrast, CTGAN and TVAE suggest stronger privacy safeguards due to lower classification accuracies. The comprehensive evaluation across these sections provides a nuanced understanding of the complex interplay between maintaining data utility and ensuring privacy in the realm of synthetic data generation.

5 Conclusions and Future Work

This thesis has critically evaluated the privacy-preserving characteristics of synthetic data generated by four distinct models: CTGAN, TVAE, DoppelGANger, and Banksformer. Through comprehensive analyses employing differential privacy metrics, re-identification risk assessments, attribute disclosure risk evaluations, and membership inference attack simulations, this research has illuminated the varying degrees of privacy preservation these models afford. CTGAN and TVAE generally demonstrate a higher fidelity to the original data's distribution while suggesting a lower risk to privacy breaches. On the other hand, Banksformer and DoppelGANger, although ensuring a greater level of privacy through larger distribution differences, exhibit higher re-identification and attribute disclosure risks. The membership inference attacks in particular have shed light on the susceptibility of Banksformer and DoppelGANger to privacy compromises, as evidenced by their tendency towards high classification accuracies. The findings underscore a critical need in the field of synthetic data generation: to delineate a balance between the utility of synthetic data and the privacy of individuals represented in the datasets. Establishing a threshold that quantitatively defines this balance is a key avenue for future research. Such a threshold would serve as a benchmark for model evaluation, guiding the development of generative models toward outputs that are both analytically useful and privacy-compliant.

Further research could explore the integration of differential privacy directly into the generative process of these models. By incorporating differential privacy mechanisms into the data generation phase, it may be possible to create synthetic datasets with inherent privacy guarantees. This could enhance the models' capability to generate data that is representative for analysis while simultaneously protecting against the re-identification of individuals. Additionally, future work may investigate advanced machine learning techniques and architectural improvements in generative models to better control the trade-offs between data utility and privacy. These might include more sophisticated feature engineering, the

application of newer and more complex generative neural network architectures, or the utilization of privacy-preserving techniques like federated learning in the data synthesis process. Lastly, due to the evolving nature of both data privacy regulations and technological advancements, ongoing monitoring of legal and ethical standards is imperative. Ensuring that synthetic data generation aligns with these standards is essential for its acceptance and utility in real-world applications.

In conclusion, while this thesis has contributed valuable insights into the current state of privacy-preserving synthetic data generation, there is a pressing need for continued research and development to refine these generative models, ensuring they effectively balance the dual imperatives of data utility and privacy.

6 Appendices

6.1 Generative Models Architecture Overviews

6.1.1 CTGAN

CTGAN, the Conditional Tabular Generative Adversarial Network, displayed in Figure 7 is specially designed for generating synthetic tabular data, addressing the challenges posed by mixed data types, non-Gaussian distributions, and imbalanced categorical columns. It utilizes mode-specific normalization and a conditional generator to produce realistic synthetic data, with an architecture that includes fully connected networks for both the generator and discriminator. CTGAN's unique approach ensures the generation of diverse and statistically representative data while maintaining privacy and data utility.

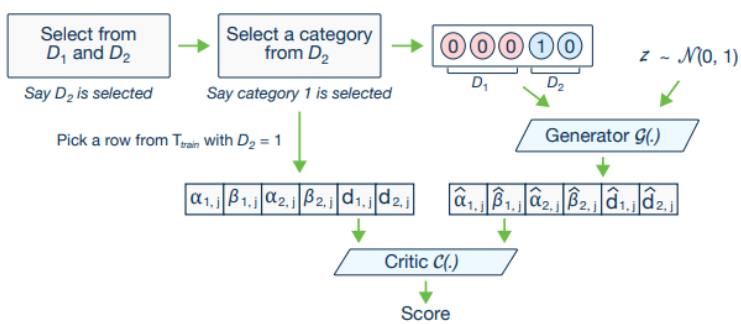


Figure 7: CTGAN Overall Architecture (Xu et al. 2019)

6.1.2 TVAE

TVAE, or Triplet-based Variational Autoencoder, enhances the traditional VAE framework by incorporating deep metric learning through triplet loss to generate high-quality synthetic data. This integration optimizes the latent space embeddings to capture semantic similarities more effectively, making TVAE particularly adept at handling tabular data. A diagram of the models architecture is illustrated in Figure 8.

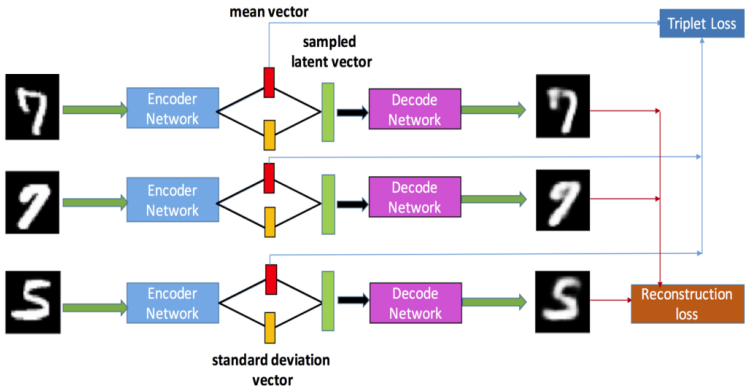


Figure 8: TVAE Overall Architecture (Ishfaq, Hoogi, and Rubin 2023)

6.1.3 TimeGAN

TimeGAN, or Time-series Generative Adversarial Network, introduces a novel approach for generating synthetic time-series data by integrating unsupervised and supervised learning to capture temporal dynamics accurately. It consists of four key components: an embedding network to transform time-series data into a manageable latent space, a recovery network to map back to the original data space, a generator to produce synthetic sequences, and a discriminator to ensure the realism of generated data. This architecture can be viewed in Figure 9. TimeGAN's ability to maintain the essential characteristics and temporal correlations of original datasets makes it particularly valuable for applications requiring high-fidelity time-series data synthesis, including financial data modelling.

6.1.4 DoppelGANger

DoppelGANger is a state-of-the-art framework using GANs to generate synthetic time-series datasets, overcoming challenges like long-term dependencies and mode collapse. It features two generators and two discriminators, with an MLP for generating metadata and an RNN generator for time series data, complemented by an auto-normalization technique (seen in Figure 10). DoppelGANger's architecture fosters the production of realistic and statistically accurate sequences, making it suitable for applications requiring detailed temporal data analysis.

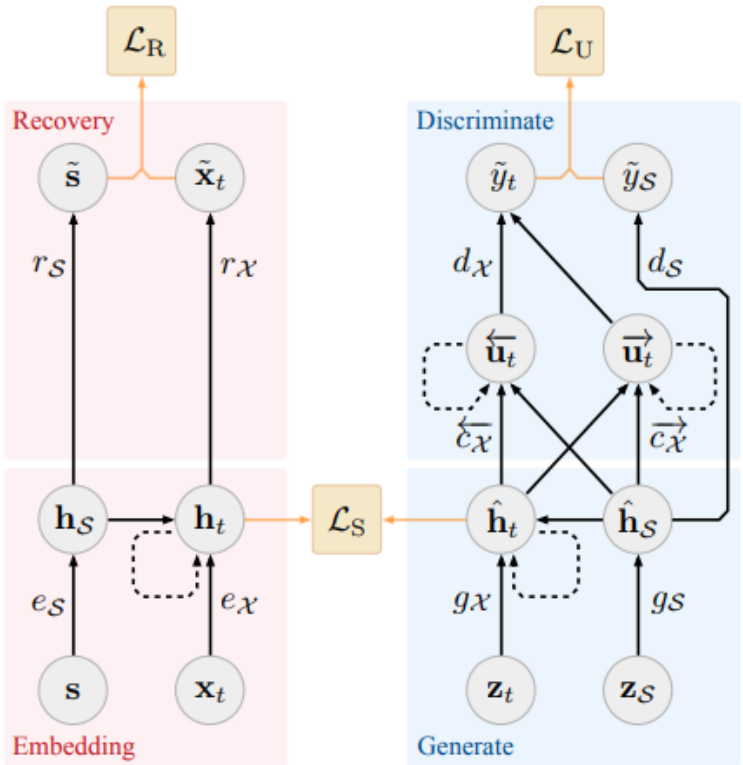


Figure 9: TimeGAN Overall Architecture (Yoon, Jarrett, and Schaar 2019)

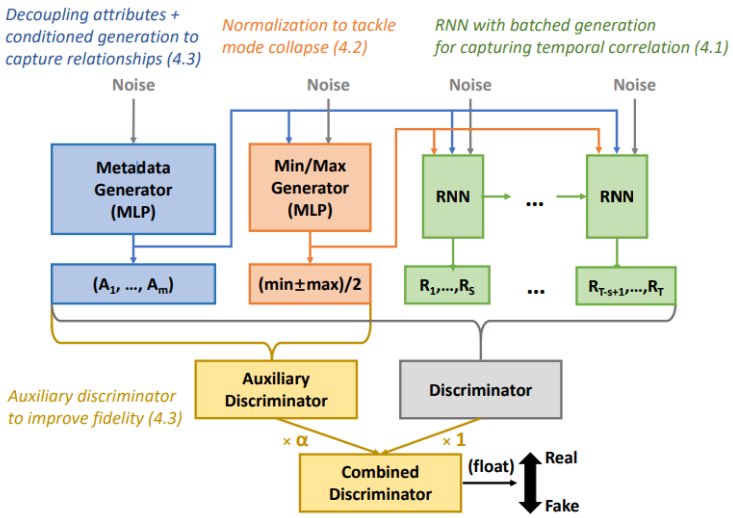


Figure 10: DoppelGANger Overall Architecture (Lin et al. 2019)

6.1.5 Banksformer

Banksformer leverages a transformer-based architecture to generate synthetic sequences of banking transactions, addressing the challenge of irregular transaction intervals and their

dependence on timestamp-based features. Its architecture includes a Transformer-Decoder for sequence generation and an innovative mechanism for handling date features, ensuring the generation of high-fidelity synthetic data. Figure 11 illustrates this architecture where Banksformer stands out for its effectiveness in emulating complex temporal relationships and market dynamics in financial datasets.

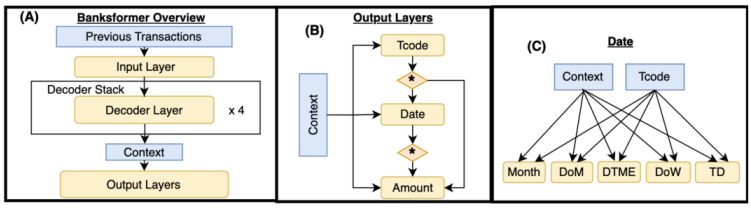


Figure 11: Banksformer Overall Architecture (Nickerson et al. 2023)

6.1.6 VAE-GAN Model

Drawing from the success of Chung et al. (2015), who introduced the VRNN model that outperformed standard RNNs in modelling sequential data, a time-step VAE loss function will be incorporated into the generator. Moreover, inspired by the Continuous RNN-GAN (CRNNGAN) approach introduced by Mogren (2016), which utilizes RNNs in both the generator and discriminator, this thesis aims to evaluate and enhance the model's ability to capture temporal dependencies in the financial data. This amalgamation of techniques seeks to create a robust and privacy-aware model for generating synthetic financial time-series data.

Chung et al. (2015) proposed a Variational Recurrent Neural Network (VRNN) that exhibits better performance in modelling sequential data compared to standard RNNs. The key innovation involves incorporating a VAE at each timestep of the RNN. This is achieved through a time-step VAE loss, encoding sequential information into time-dependent stochastic variables. Unlike a standard Gaussian prior, the prior distribution in VRNN (Figure 12a) is not centred on 0 and is represented as a diagonal matrix, thereby having no covariance. The VRNN model follows a series of steps in each timestep: first, a generating

function conditions on the hidden state from the previous timestep, h_{t-1} and the encoder output z_t to generate the next step (Figure 12b). The hidden state is then updated using a deterministic non-linear transition function (Figure 12c). Subsequently, the inference of the approximate posterior is a function of the hidden state from the previous timestep, which is employed to update the encoder (Figure 12d). This intricate design enhances the model's ability to capture complex temporal dependencies in sequential data.

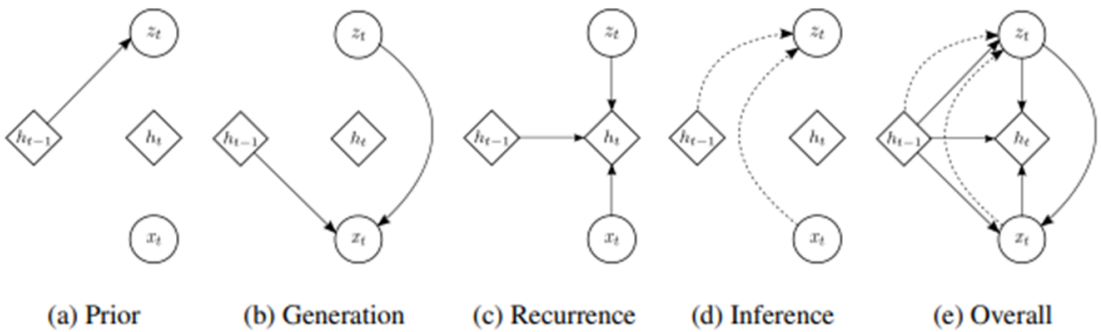


Figure 12: VRNN proposed by Chung et al. (2015)

Variational inference (Figure 12d) is utilized to learn a time-varying approximate posterior distribution for the latent variables $z_1, \ldots, z_T$. The approximate posterior distribution is denoted as q_θ. The joint conditional probability factorization is given by

$$q_\theta(z_{\leq T}|x_{\leq T}) = \prod_{t=1}^{T} q_\theta(z_t|x_{\leq t}, z_{<t}),$$

where

$$z_t|x_t \sim \mathcal{N}(\mu_{z,t}, \mathrm{diag}((\sigma_{z,t})^2)),$$

and

$$[\mu_{z,t}, \sigma_{z,t}] = \phi_\theta^{\mathrm{enc}}(\phi_x^\phi(x_t), h_{t-1}).$$

Here, $\phi_{\mathrm{enc}\theta}$ and ϕ_x^ϕ are flexible functions, typically implemented as feed-forward networks. Notably, the latent parameter estimates in the inference model, unlike the generation model, depend on both the observations x_t and the previous hidden state h_{t-1}.

Similarly, interactions between the hidden states h_{t-1}, h_t, latent variable z_t, and output x_t in the generation model are described by the following joint probability factorization:

$$p(x_{\leq T}, z_{\leq T}) = \prod_{i=1}^{T} p(x_t|x_{<t}, z_{\leq t})p(z_t|x_{<t}, z_{<t})$$

The relationships between the variables are specified as

$$z_t \sim \mathcal{N}(\mu_{0,t}, \mathrm{diag}(\sigma_{0,t}^2)),$$

$$[\mu_{0,t}, \sigma_{0,t}] = \phi_\tau^{\mathrm{prior}}(h_{t-1}),$$

and

$$x_t|z_t \sim \mathcal{N}(\mu_{x,t}, \mathrm{diag}(\sigma_{x,t}^2)),$$

$$[\mu_{x,t}, \sigma_{x,t}] = \phi_\theta^{\mathrm{dec}}(\phi_z^\phi(z), h_{t-1}).$$

Here, $\phi_\tau^{\mathrm{prior}}$, $\phi_\theta^{\mathrm{dec}}$, and ϕ_z^ϕ are flexible functions, typically implemented as feed-forward networks in this paper. h_{t-1} refers to the hidden state calculated by the RNN model.

The discriminator for the model will be a bidirectional RNN. A bidirectional RNN is a type of neural network architecture that processes input data in both forward and backward directions. Unlike traditional RNNs that only consider the sequential information in one direction from the past to the future or vice versa, bidirectional RNNs simultaneously analyze sequences in both directions. This dual-directional approach allows the network to capture dependencies and patterns that may exist in both the past and future context of each data point. In the context of sequence modelling or time-series analysis, bidirectional RNNs are particularly useful for tasks where understanding the entire context, both preceding and following a given point in the sequence, is crucial for making accurate predictions or classifications. The bidirectional nature of the RNN helps to enhance the network's ability to capture long-range dependencies and context in sequential data.

6.2 Binary Classifier Parameters

Table 11: Binary Classifier Models Parameters

Logistic Regression	**Random Forest**	**XGBoost**
max_iter: 200 random_state: 42	n_estimators: 200 random_state: 42	random_state: 42
Naive Bayes	**GBM**	**Simple Neural Network**
Gaussian NB	n_estimators: 100 random_state: 42	neurons: 64 epochs: 100 Adam optimizer BCE Loss ReLU and sigmoid activations

6.3 Individual Distributions

Figure 13 illustrates each of the individual generated synthetic univariate distributions of the `amount` column compared to the real distribution.

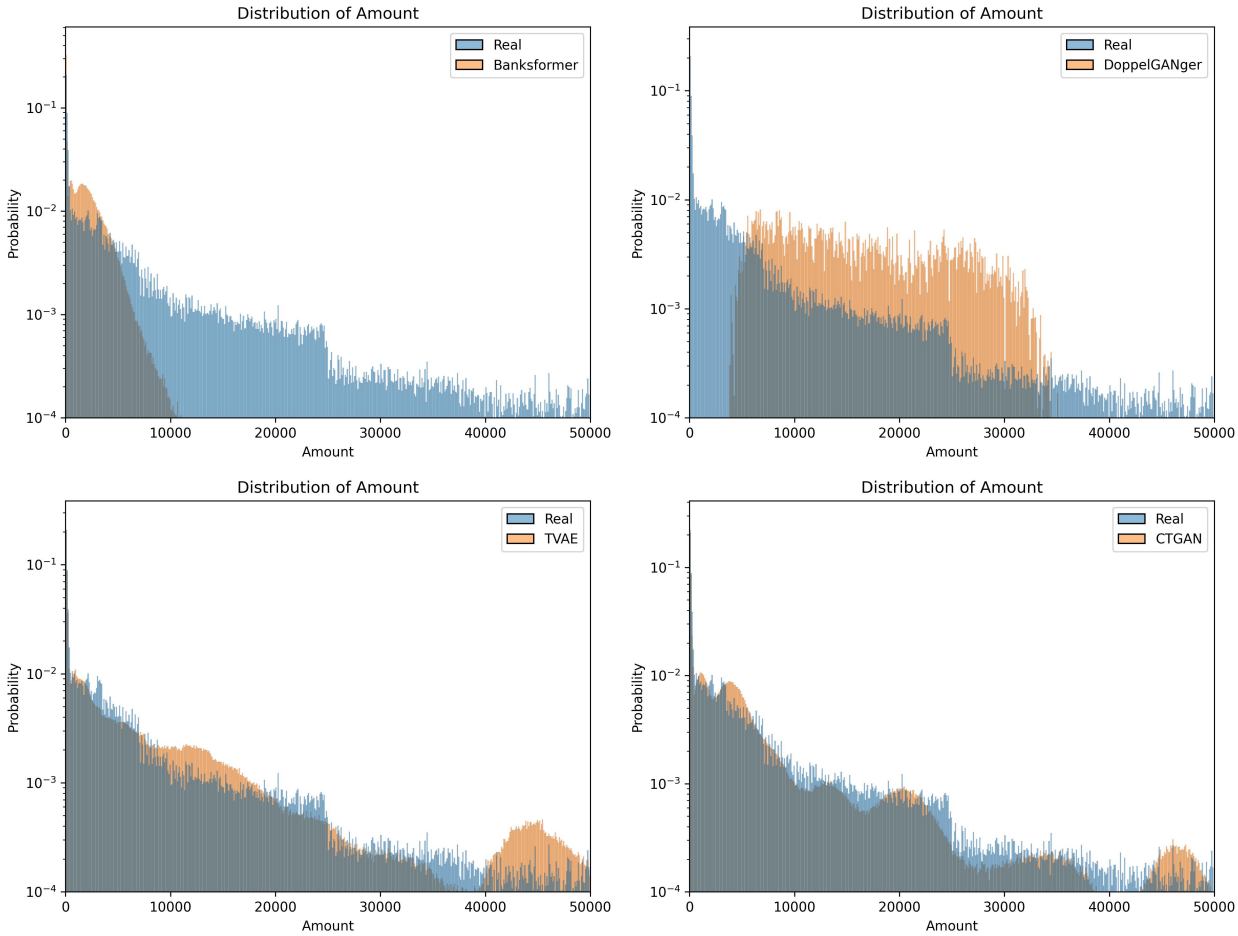


Figure 13: Individual Amount Distributions

6.4 Membership Inference All Results

Figure 14 illustrates the full results from membership inference attacks averaged over 5 folds for balanced accuracy, precision, recall, and f1 scores. In order, the figure shows results on the synthetic data generated from CTGAN, TVAE, DoppelGANger, and Banksformer from top to bottom.

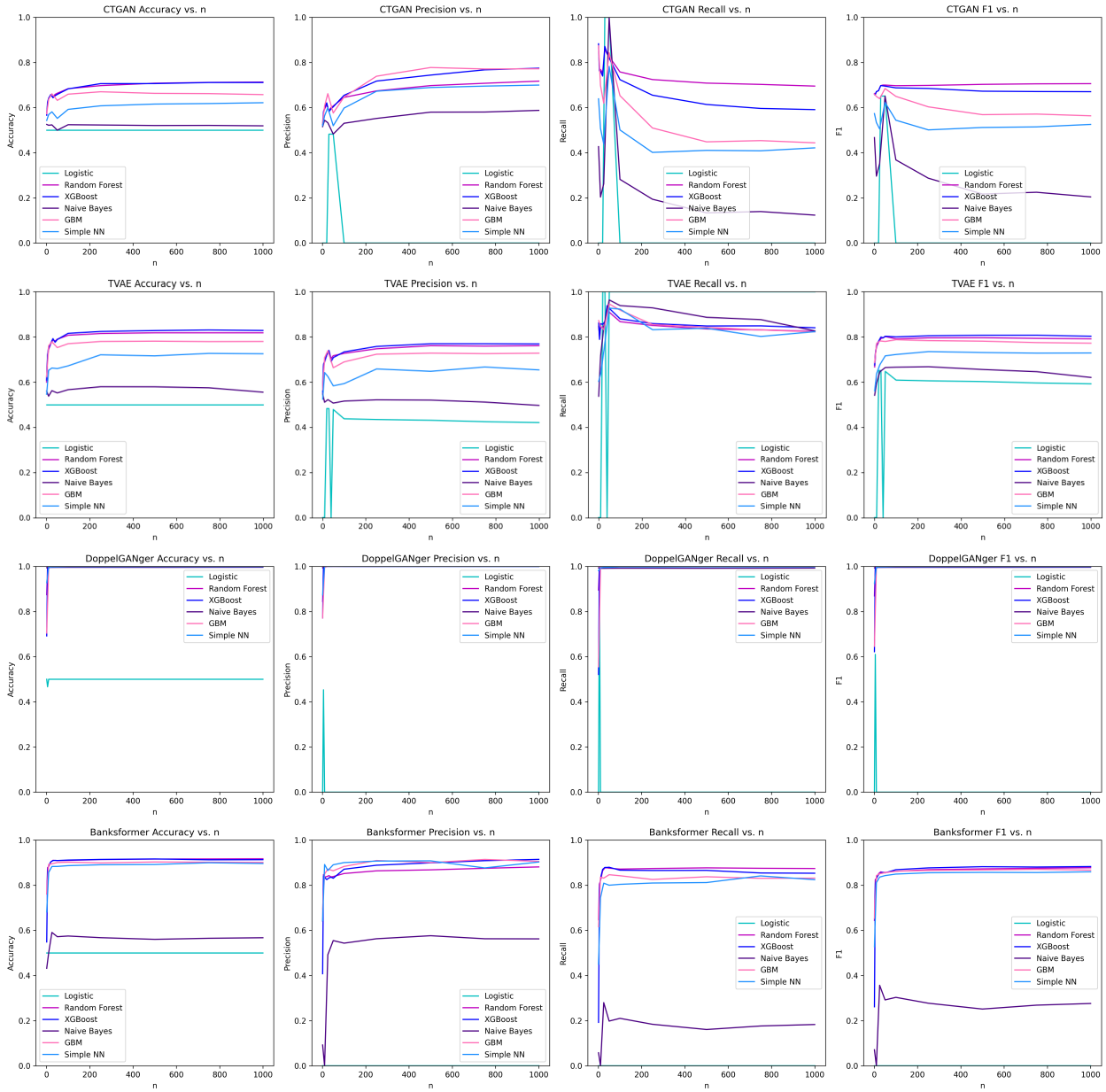


Figure 14: Model Inference Attack Accuracy, Precision, Recall, f1 over 5 Folds

6.5 Membership Inference Confusion Matrices

The figures below illustrate examples of confusion matrix outputs from the membership inference results on a single split of the data where $n = 1000$. From left to right, top to bottom, we see a result from logistic regression, XGBoost, random forest, Naive Bayes, GBM, and a feedforward neural network. Note that logistic regression (top left) predicts a single label for everything for each generated synthetic dataset.

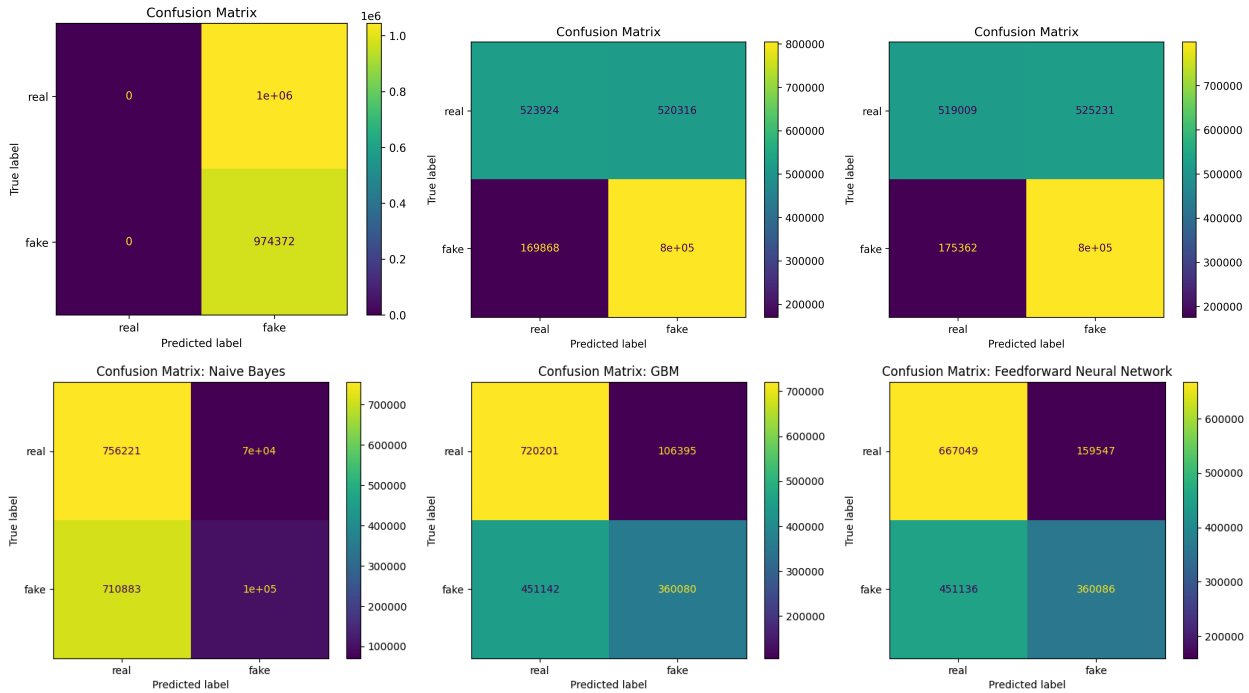


Figure 15: CTGAN Membership Inference Matrices

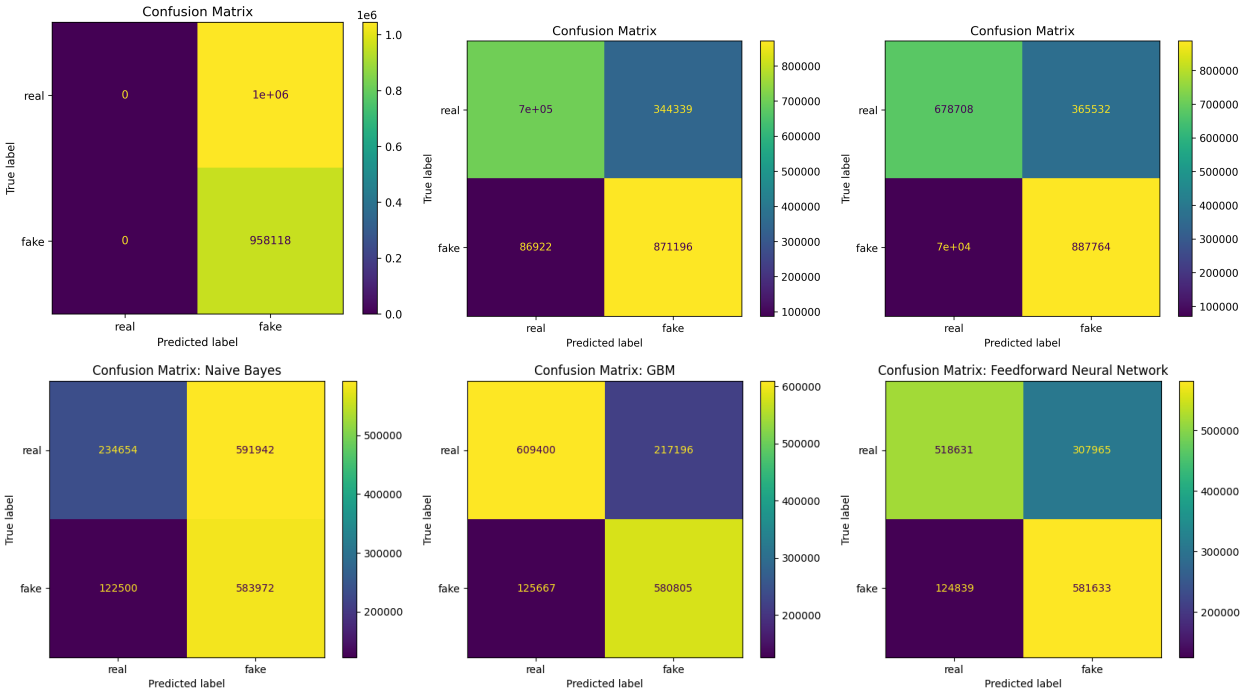


Figure 16: TVAE Membership Inference Matrices

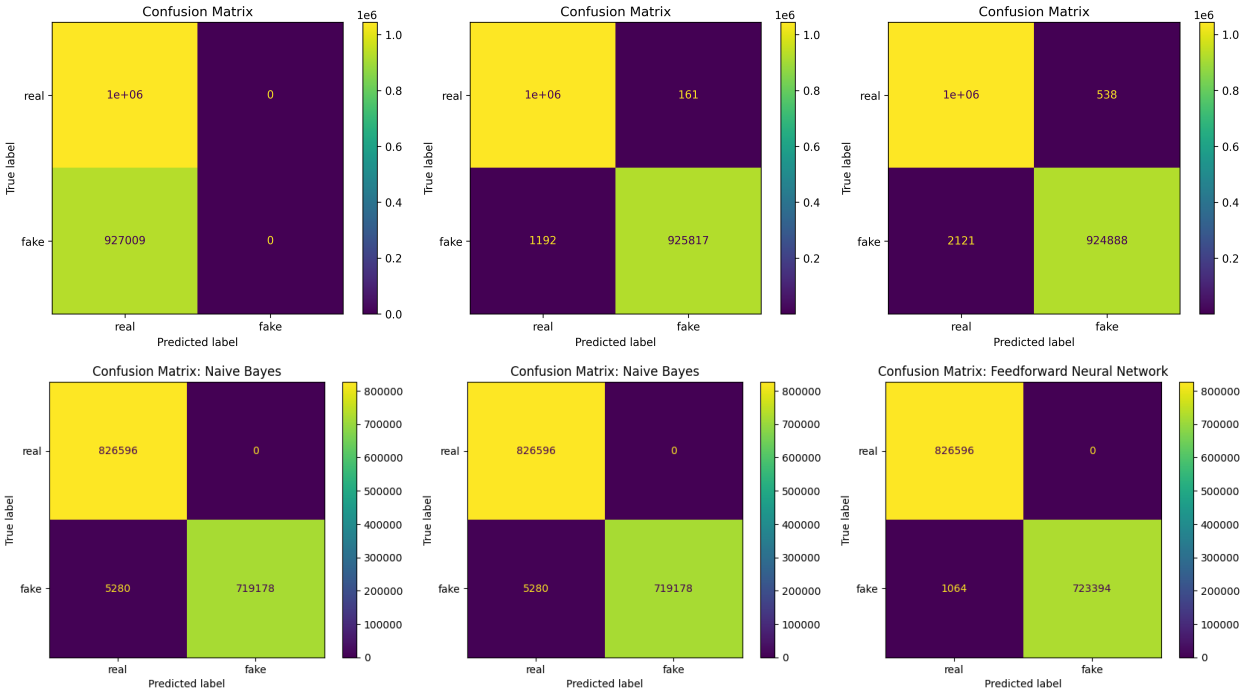


Figure 17: DoppelGANger Membership Inference Matrices

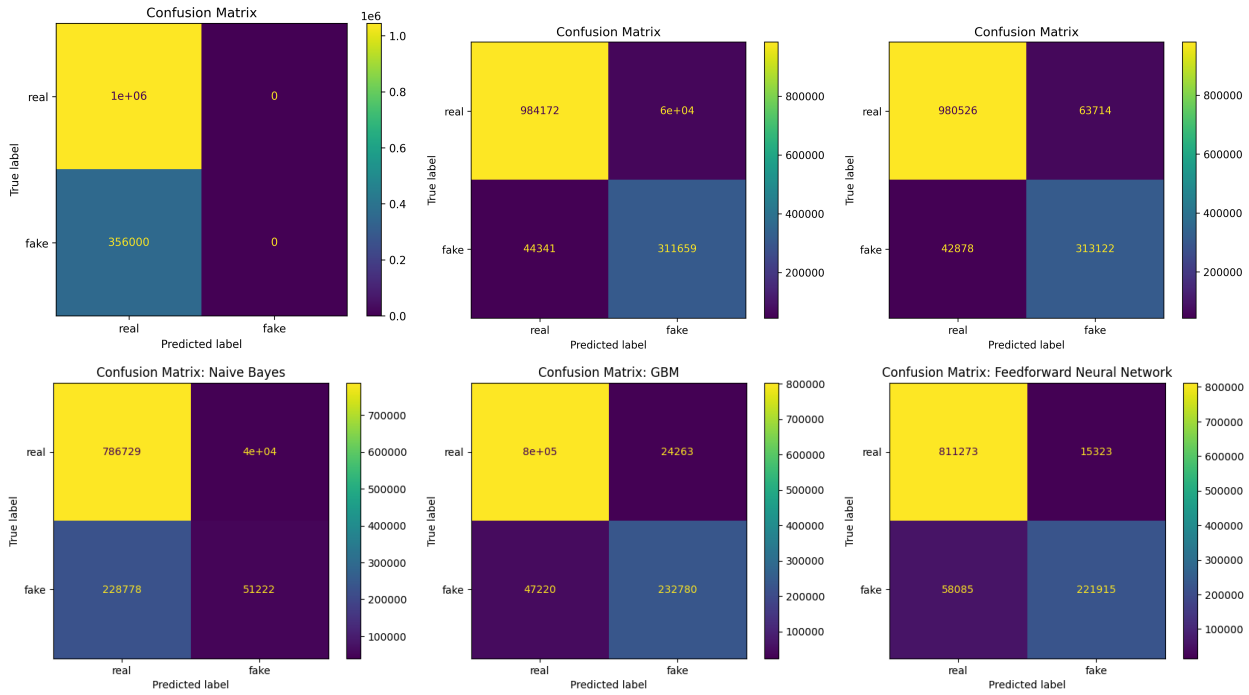


Figure 18: Banksformer Membership Inference Matrices

6.6 Mutual Information Results

Figure 19 below illustrates the mutual information of features for the generated synthetic data from CTGAN and TVAE. Note that mutual information here evaluates how much information specific features give you about the label (real versus fake). Compared with the results in the paper from DoppelGANger and Banksformer, these datasets do not result in any features with high mutual information values and were excluded from the comparison in Chapter 4.6.

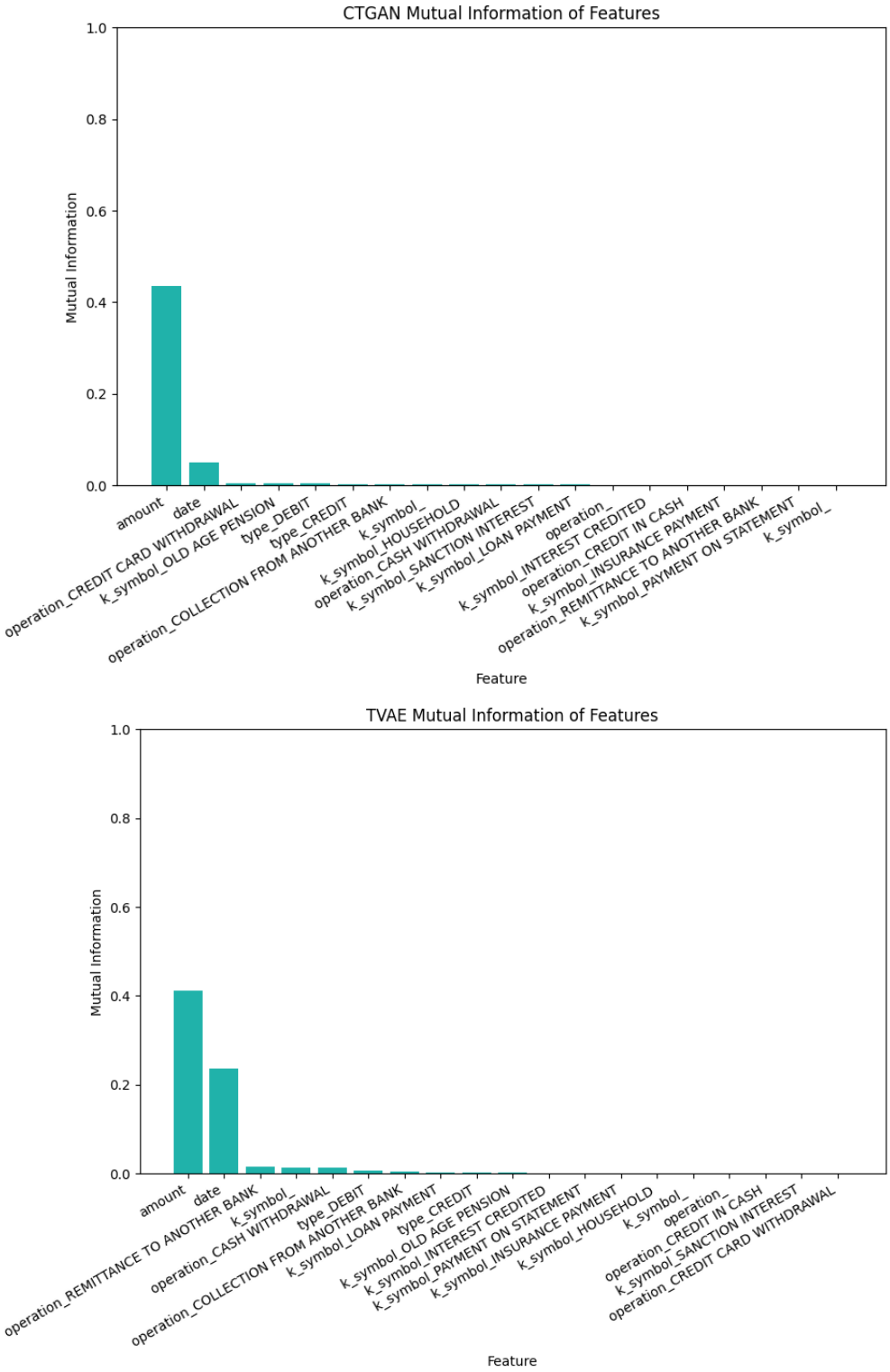


Figure 19: Mutual Information of CTGAN and TVAE Features

References

Abadi, Martin, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. 2016. "Deep learning with differential privacy." In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security,* 308–318.

Alguliyev, Rasim M., Ramiz M. Aliguliyev, and Fargana J. Abdullayeva. 2019. "Privacy-preserving deep learning algorithm for big personal data analysis." *Journal of Industrial Information Integration* 15:1–14. ISSN: 2452-414X. https://doi.org/https://doi.org/10.1016/j.jii.2019.07.002. https://www.sciencedirect.com/science/article/pii/S2452414X18301456.

Almaiah, Mohammed Amin, Aitizaz Ali, Fahima Hajjej, Muhammad Fermi Pasha, and Manal Abdullah Alohali. 2022. "A lightweight hybrid deep learning privacy preserving model for FC-based industrial internet of medical things." *Sensors* 22 (6): 2112.

Benitez, Kathleen, and Bradley Malin. 2010. "Evaluating re-identification risks with respect to the HIPAA privacy rule." *Journal of the American Medical Informatics Association* 17, no. 2 (March): 169–177. ISSN: 1067-5027. https://doi.org/10.1136/jamia.2009.000026. eprint: https://academic.oup.com/jamia/article-pdf/17/2/169/17373689/17-2-169.pdf. https://doi.org/10.1136/jamia.2009.000026.

Breiman, Leo. 2001. "Random Forests." *Mach. Learn.* (USA) 45, no. 1 (October): 5–32. ISSN: 0885-6125. https://doi.org/10.1023/A:1010933404324. https://doi.org/10.1023/A:1010933404324.

Chung, Junyoung, Kyle Kastner, Laurent Dinh, Kratarth Goel, Aaron C. Courville, and Yoshua Bengio. 2015. "A Recurrent Latent Variable Model for Sequential Data." *CoRR* abs/1506.02216. arXiv: 1506.02216. http://arxiv.org/abs/1506.02216.

Dimitrova, Dimitrina S., Vladimir K. Kaishev, and Senren Tan. 2020. "Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed, or Continuous." *Journal of Statistical Software* 95 (10): 1–42. https://doi.org/10.18637/jss.v095.i10. https://www.jstatsoft.org/index.php/jss/article/view/v095i10.

Du, Miao, Kun Wang, Yuanfang Chen, Xiaoyan Wang, and Yanfei Sun. 2018. "Big Data Privacy Preserving in Multi-Access Edge Computing for Heterogeneous Internet of Things." *IEEE Communications Magazine* 56 (8): 62–67. https://doi.org/10.1109/MCOM.2018.1701148.

Foundation, Wikimedia. n.d. "Wikimedia Downloads." https://dumps.wikimedia.org.

Fu, Anmin, Xianglong Zhang, Naixue Xiong, Yansong Gao, Huaqun Wang, and Jing Zhang. 2020. "VFL: A verifiable federated learning with privacy-preserving for big data in industrial IoT." *IEEE Transactions on Industrial Informatics* 18 (5): 3316–3326.

González-Serrano, Francisco-Javier, Ángel Navia-Vázquez, and Adrián Amor-Martín. 2017. "Training support vector machines with privacy-protected data." *Pattern Recognition* 72:93–107.

Gulrajani, Ishaan, Faruk Ahmed, Martín Arjovsky, Vincent Dumoulin, and Aaron C. Courville. 2017. "Improved Training of Wasserstein GANs." *CoRR* abs/1704.00028. arXiv: 1704.00028. http://arxiv.org/abs/1704.00028.

Hitaj, Briland, Giuseppe Ateniese, and Fernando Perez-Cruz. 2017. "Deep models under the GAN: information leakage from collaborative deep learning." In *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security,* 603–618.

Hittmeir, Markus, Rudolf Mayer, and Andreas Ekelhart. 2020. "A Baseline for Attribute Disclosure Risk in Synthetic Data." In *Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy,* 133–143. CODASPY '20. New Orleans, LA, USA: Association for Computing Machinery. ISBN: 9781450371070. https://doi.org/10.1145/3374664.3375722. https://doi.org/10.1145/3374664.3375722.

Hugging Face. 2023. https://huggingface.co/datasets/imodels/compas-recidivism.

Ishfaq, Haque, Assaf Hoogi, and Daniel Rubin. 2023. *TVAE: Triplet-Based Variational Autoencoder using Metric Learning.* arXiv: 1802.04403 [stat.ML].

Kaggle. 2023. https://www.kaggle.com/datasets/ealaxi/paysim1/data.

Kairouz, Peter, Brendan McMahan, Shuang Song, Om Thakkar, Abhradeep Thakurta, and Zheng Xu. 2021. "Practical and Private (Deep) Learning without Sampling or Shuffling," arXiv: 2103.00039 [cs.CR].

Kumar, Prabhat, Govind P Gupta, and Rakesh Tripathi. 2021. "TP2SF: A Trustworthy Privacy-Preserving Secured Framework for sustainable smart cities by leveraging blockchain and machine learning." *Journal of Systems Architecture* 115:101954.

Kwapisz, Jennifer R., Gary M. Weiss, and Samuel A. Moore. 2010. "Activity Recognition using Cell Phone Accelerometers." *Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data (at KDD-10)* (Washington DC).

LeCun, Yann, Corinna Cortes, and CJ Burges. 2010. "MNIST handwritten digit database." *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist* 2.

Lee, Justin. 2022. "VRNNGAN: A Recurrent VAE-GAN Framework for Synthetic Time-Series Generation" [in English]. *ProQuest,* 41. http://myaccess.library.utoronto.ca/login?qurl=https%3A%2F%2Fwww.proquest.com%2Fdissertations-theses%2Fvrnngan-recurrent-vae-gan-framework-synthetic%2Fdocview%2F2738613407%2Fse-2%3Faccountid%3D14771.

Li, Ninghui, Tiancheng Li, and Suresh Venkatasubramanian. 2007. "t-closeness: Privacy beyond k-anonymity and l-diversity." In *2007 IEEE 23rd international conference on data engineering,* 106–115. IEEE.

Lin, Zinan, Alankar Jain, Chen Wang, Giulia Fanti, and Vyas Sekar. 2019. "Generating High-fidelity, Synthetic Time Series Datasets with DoppelGANger." *CoRR* abs/1909.13403. arXiv: 1909.13403. http://arxiv.org/abs/1909.13403.

Liu, Bo, Ming Ding, Sina Shaham, Wenny Rahayu, Farhad Farokhi, and Zihuai Lin. 2021. "When Machine Learning Meets Privacy: A Survey and Outlook." Doi: 10.1145/3436755, *ACM Computing Surveys* 5, no. 2 (March).

Lu, Xiaofeng, Yuying Liao, Pietro Lio, and Pan Hui. 2020a. "Privacy-Preserving Asynchronous Federated Learning Mechanism for Edge Network Computing." *IEEE Access* 8:48970–48981. https://doi.org/10.1109/ACCESS.2020.2978082.

———. 2020b. "Privacy-preserving asynchronous federated learning mechanism for edge network computing." *IEEE Access* 8:48970–48981.

Ma, Jing, Si-Ahmed Naas, Stephan Sigg, and Xixiang Lyu. 2022. "Privacy-preserving federated learning based on multi-key homomorphic encryption." *International Journal of Intelligent Systems* 37 (9): 5880–5901.

Machanavajjhala, Ashwin, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. 2007. "l-diversity: Privacy beyond k-anonymity." *Acm transactions on knowledge discovery from data (tkdd)* 1 (1): 3–es.

Mogren, Olof. 2016. "C-RNN-GAN: Continuous recurrent neural networks with adversarial training." *CoRR* abs/1611.09904. arXiv: 1611.09904. http://arxiv.org/abs/1611.09904.

Netzer, Yuval, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. 2011. "Reading Digits in Natural Images with Unsupervised Feature Learning."

Nickerson, Kyle, Terrence Tricco, Antonina Kolokolova, Farzaneh Shoeleh, Charles Robertson, John Hawkin, and Ting Hu. 2023. "Banksformer: A Deep Generative Model fornbsp;Synthetic Transaction Sequences." In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2022, Grenoble, France, September 19–23, 2022, Proceedings, Part VI,* 121–136. Grenoble, France: Springer-Verlag. ISBN: 978-3-031-26421-4. https://doi.org/10.1007/978-3-031-26422-1_8. https://doi.org/10.1007/978-3-031-26422-1_8.

Phong, Le Trieu, Yoshinori Aono, Takuya Hayashi, Lihua Wang, and Shiho Moriai. 2017. "Privacy-preserving deep learning: Revisited and enhanced." In *Applications and Techniques in Information Security: 8th International Conference, ATIS 2017, Auckland, New Zealand, July 6–7, 2017, Proceedings,* 100–110. Springer.

Pita, Guilherme Leal Cardoso. 2024. "Improving Synthetic Transaction Data Generation with Network Graph Clustering," University of Toronto.

Rahulamathavan, Yogachandran, Raphael C.-W. Phan, Suresh Veluru, Kanapathippillai Cumanan, and Muttukrishnan Rajarajan. 2014. "Privacy-Preserving Multi-Class Support Vector Machine for Outsourcing the Data Classification in Cloud." *IEEE Transactions on Dependable and Secure Computing* 11 (5): 467–479. https://doi.org/10.1109/TDSC.2013.51.

Razghandi, Mina, Hao Zhou, Melike Erol-Kantarci, and Damla Turgut. 2023. "Smart Home Energy Management: VAE-GAN synthetic dataset generator and Q-learning," arXiv: 2305.08885 [cs.LG].

Rieke, Nicola, Jonny Hancox, Wenqi Li, Fausto Milletari, Holger R Roth, Shadi Albarqouni, Spyridon Bakas, Mathieu N Galtier, Bennett A Landman, Klaus Maier-Hein, et al. 2020. "The future of digital health with federated learning." *NPJ digital medicine* 3 (1): 119.

Rodríguez, E., B. Otero, and R. Canal. 2023. "A Survey of Machine and Deep Learning Methods for Privacy Protection in the Internet of Things." Doi: 10.3390/s23031252, *Sensors* 23, no. 3 (January): 1252.

Servia-Rodriguez, Sandra, Liang Wang, Jianxin R Zhao, Richard Mortier, and Hamed Haddadi. 2017. "Personal model training under privacy constraints." *training* 40 (33): 24–38.

Shokri, Reza, and Vitaly Shmatikov. 2015. "Privacy-Preserving Deep Learning." In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security,* 1310–1321. CCS '15. Denver, Colorado, USA: Association for Computing Machinery. ISBN: 9781450338325. https://doi.org/10.1145/2810103.2813687. https://doi.org/10.1145/2810103.2813687.

Sweeney, Latanya. 2002. "k-anonymity: A model for protecting privacy." *International journal of uncertainty, fuzziness and knowledge-based systems* 10 (05): 557–570.

TensorFlow. 2023. https://www.tensorflow.org/datasets/catalog/speech_commands#:~:text=An%20audio%20dataset%20of%20spoken,background%20noise%20or%20unrelated%20speech..

Triastcyn, Aleksei, and Boi Faltings. 2018. "Generating Differentially Private Datasets Using GANs." *CoRR* abs/1803.03148. arXiv: 1803.03148. http://arxiv.org/abs/1803.03148.

Wang, Ji, Jianguo Zhang, Weidong Bao, Xiaomin Zhu, Bokai Cao, and Philip S Yu. 2018. "Not just privacy: Improving performance of private deep learning in mobile cloud." In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining,* 2407–2416.

Wang, W., CM. Vong, and Y. Yang. 2017. "Encrypted image classification based on multilayer extreme learning machine." *Multidim Syst Sign Process* 28:851–865. https://doi.org/10.1007/s11045-016-0408-1.

Wang, Weiru, Chi-Man Vong, Yilong Yang, and Pak-Kin Wong. 2017. "Encrypted image classification based on multilayer extreme learning machine." *Multidimensional Systems and Signal Processing* 28:851–865. https://doi.org/10.1007/s11045-016-0419-y. https://link.springer.com/article/10.1007/s11045-016-0419-y.

Wright, Catherine, and Kellin Rumsey. 2018. "The Strengths, Weaknesses and Promise of Differential Privacy as a Privacy-Protection Framework." *University of New Mexico Dept. of Computer Science.*

Xu, Lei, Maria Skoularidou, Alfredo Cuesta-Infante, and Kalyan Veeramachaneni. 2019. "Modeling Tabular data using Conditional GAN." In *Advances in Neural Information Processing Systems.*

Yoon, Jinsung, Daniel Jarrett, and Mihaela van der Schaar. 2019. "Time-series Generative Adversarial Networks." In *Advances in Neural Information Processing Systems,* edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, vol. 32. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2019/file/c9efe5f26cd17ba6216bbe2a7d26d490-Paper.pdf.

Zhang, Tianwei, Zecheng He, and Ruby B Lee. 2018. "Privacy-preserving machine learning through data obfuscation." *arXiv preprint arXiv:1807.01860.*

Zhang, Xiaoyu, Xiaofeng Chen, Joseph K. Liu, and Yang Xiang. 2020. "DeepPAR and DeepDPA: Privacy Preserving and Asynchronous Deep Learning for Industrial IoT." *IEEE Transactions on Industrial Informatics* 16 (3): 2081–2090. https://doi.org/10.1109/TII.2019.2941244.

Zhao, Yang, Jun Zhao, Linshan Jiang, Rui Tan, Dusit Niyato, Zengxiang Li, Lingjuan Lyu, and Yingbo Liu. 2020. "Privacy-preserving blockchain-based federated learning for IoT devices." *IEEE Internet of Things Journal* 8 (3): 1817–1829.

Zhu, Hui, Xiaoxia Liu, Rongxing Lu, and Hui Li. 2017. "Efficient and Privacy-Preserving Online Medical Prediagnosis Framework Using Nonlinear SVM." *IEEE Journal of Biomedical and Health Informatics* 21 (3): 838–850. https://doi.org/10.1109/JBHI.2016.2548248.