

Article

Not peer-reviewed version

---

# Approximate Unlearning in Finance

---

[Chang Qi](#) \*

Posted Date: 24 September 2024

doi: [10.20944/preprints202409.1840.v1](https://doi.org/10.20944/preprints202409.1840.v1)

Keywords: Approximate Unlearning; Machine Learning; Deep Learning; Digital Asset; Cryptocurrency



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Approximate Unlearning in Finance

Chang Qi

Independent Researcher, Seattle, WA; changchannieqi1@gmail.com

**Abstract:** Approximate unlearning is an emerging field in machine learning that addresses the challenge of efficiently removing specific data points from models without the need for full retraining. In the financial sector, where models are built on sensitive customer data and market information, the ability to unlearn data is crucial for meeting privacy regulations like GDPR and ensuring model adaptability in real-time environments. This paper explores the categorization of approximate unlearning into two main approaches: data-driven approximation and model-driven approximation. Data-driven approximation focuses on selectively retraining parts of the dataset to simulate the removal of data, while model-driven approximation adjusts the model's internal parameters to nullify the influence of unwanted data points. Both methods offer computational and memory-efficient ways to balance the need for privacy compliance with the performance demands of financial models. The paper discusses practical applications in algorithmic trading, fraud detection, and risk assessment, and highlights the challenges associated with each approach. Through case studies and literature references, we demonstrate how approximate unlearning can be applied to maintain system efficiency, data privacy, and regulatory compliance in the financial domain. This research provides a roadmap for future developments in approximate unlearning, particularly in the context of real-time financial systems and federated learning environments.

**Keywords:** approximate unlearning; machine learning; deep learning; digital asset; cryptocurrency

---

## Introduction

Approximate unlearning is an emerging concept in machine learning that focuses on the efficient removal of data or information from models, while ensuring that the system behaves as if the data were never incorporated in the first place. This technique is particularly relevant in the financial sector, where the need for data privacy, adaptability, and compliance with regulatory frameworks is paramount. Financial institutions often handle vast amounts of sensitive data, including customer profiles, transaction histories, and trading strategies. Given this, there is an increasing demand to remove specific data points efficiently while ensuring that the performance of the model is preserved.

In contrast to exact unlearning, which aims to completely remove all traces of the data, approximate unlearning focuses on providing a close approximation by selectively retraining or recalibrating models. The goal is to balance computational efficiency with legal and regulatory compliance. In finance, this balance becomes critical due to the high frequency of transactions and the need for real-time analysis. Approximate unlearning provides a pathway to maintain system efficiency while respecting privacy and compliance needs.

## Literature Review

[1] and [2] categorize approximate unlearning into two classes: data-driven approximation and model-driven approximation. Both approaches aim to remove the influence of specific data points from machine learning models, but they differ in their methodologies and the specific components of the learning system they target.

Strategies that focus on manipulating the data are categorized as data driven approximation. [3], [4] and [5] use data isolation strategy for data-driven approximate unlearning. Although data-driven approximate unlearning is applicable to most model types, the major drawbacks is that unlearning is not complete. [6], [7], [8], [9] and [10] use data modification approach to accomplish the

unlearning process. The unlearning process is easy to achieve, but it impacts the model's utility and consumes storage resources. This introduces some resource constraints to the overall model.

Model-driven approximation, on the other hand, focuses on the internal components of the model itself rather than the data. This approach modifies the learned parameters of the model or its architecture to approximate the removal of data influence. Model-driven techniques aim to directly alter the weight configurations or the memory of the model without retraining the entire system, ensuring computational efficiency while simulating the model as if the unwanted data was never present. Graph neural network introduced in [11] can be used to assist in the model-driven approximation. [12], [13], [14], [15] and [16] uses the influence function to evaluate the trained model. These are called influenced based model driven approximation. At the same time, we have Fisher-based model driven approximation from [17], [18] and [19]. All of the approaches above can be applied to treasury [20] or crypto trading [21] applications. The detailed comparison between data-drive and model driven approach can be found in **Error! Reference source not found..**

**Table 1.** Comparison Between Data-driven and Model-driven Approximation.

Aspect	Data-Driven Approximation	Model-Driven Approximation
Focus	Training data adjustments	Direct changes to the model's parameters
Data-Removal Approach	Remove or modify data points in the dataset	Modify model weights and architectures
Efficiency	Efficient for small data removals, avoids full retraining	Highly efficient for real-time model updates
Memory/Resource Usage	Typically requires storing multiple subsets of data	Optimized for memory-efficient updates
Residual Influence	May leave residual traces of unlearned data	Minimizes residual influence with parameter recalibration
Use Cases	Customer data deletion, fraud detection	High-frequency trading, risk models

### Approximate Unlearning Techniques

Several techniques have been proposed for approximate unlearning, each with different advantages and limitations, particularly when applied to the financial sector. Some methods involve efficiently updating model parameters to remove the influence of the specific data point or set of points that need to be erased. Other approaches involve leveraging gradient information or memory-efficient representations to reduce the footprint of the data without retraining the entire model. Below are some prominent techniques in approximate unlearning:

**Gradient-based Approximate Unlearning:** This technique calculates how specific data points influence the model and adjusts the model's parameters accordingly. In financial systems where models continuously learn from real-time data, this approach is beneficial as it allows for the efficient removal of certain data points without degrading the overall model performance. Studies such as those by Ginart et al. (2019) have demonstrated the feasibility of gradient-based unlearning for sensitive financial data.

**Subspace Projection:** This approach seeks to project the model into a subspace that nullifies the contribution of specific data. In a financial context, this could be particularly useful for time-sensitive applications such as algorithmic trading, where the removal of older, irrelevant data must be fast and efficient.

**Memory-efficient Recalibration:** This method focuses on recalibrating certain model components to ensure that the model behaves as if it never learned from specific data. It is especially applicable in risk management, where sensitivity to outliers or erroneous data can have far-reaching consequences. Approximate unlearning helps in adjusting for these without requiring a full retrain.

In high-frequency trading (HFT), for example, where models need to adapt quickly to new information, traditional unlearning would require retraining large and complex models each time some data is removed or updated. With approximate unlearning, traders can remove irrelevant or faulty data from the model without compromising the model's decision-making capabilities. Ginart et al. (2019) showed how approximate unlearning could be applied to decision tree models, which are commonly used in financial decision-making, allowing for efficient updates without the need for full retraining.

Another practical example can be found in the domain of credit risk modeling, where financial institutions use large datasets of historical credit information to assess an individual's risk. If a customer requests their data be erased, approximate unlearning can remove that specific data while still maintaining the integrity of the overall model.

### **Applications in the Financial Arena**

The application of approximate unlearning in the financial sector spans various areas, including but not limited to algorithmic trading, fraud detection, and customer profiling. In each of these domains, models are built using large volumes of data that often need to be updated or removed due to changing market conditions or regulatory requirements.

One of the most sensitive and high-stakes applications of machine learning in finance is algorithmic trading, where the speed and accuracy of models can lead to significant profits or losses. In this domain, models are constantly learning from vast streams of financial data. Approximate unlearning allows these models to stay current without retraining from scratch, helping traders remain competitive in rapidly changing markets.

Fraud detection systems rely heavily on machine learning to identify anomalies in transaction patterns. These models must remain accurate and adaptive, given that financial crimes evolve over time. Approximate unlearning becomes essential when fraud-related data needs to be erased, such as when a false positive is identified, or in cases where customer information must be deleted due to privacy concerns. Literature on privacy-preserving learning has explored approximate unlearning in fraud detection systems, helping banks ensure that their models comply with regulations like GDPR.

Customer profiling is at the core of personalized financial services, from credit scoring to loan approval and investment advice. These models are built on detailed historical data, including transaction histories, credit scores, and personal information. However, customers now have the right to request their data be deleted. Approximate unlearning allows financial institutions to maintain accurate models without having to retrain them from scratch, ensuring that customer preferences and personalized services are still effective even after some data has been removed.

### **Challenges and Future Directions**

Although approximate unlearning holds great promise, several challenges must be addressed for it to be widely adopted in the financial industry. One significant issue is that approximations, by their nature, may not provide a perfect removal of data, potentially leaving traces that could be exploited in adversarial attacks. In finance, where the stakes are high, even small inaccuracies could result in substantial financial loss or legal ramifications.

Federated learning, a distributed machine learning approach that trains models across decentralized data sources, is becoming increasingly important in the financial sector due to privacy and data governance concerns. Approximate unlearning can be particularly useful in federated learning frameworks, where individual nodes may need to remove local data without compromising the global model.

### **Conclusion**

Approximate unlearning represents a promising advancement for the financial industry, providing a means to efficiently remove sensitive data while maintaining the performance of machine learning models. As financial systems continue to evolve and become more reliant on real-time data

processing, the need for quick, efficient unlearning methods will become increasingly vital. Approximate unlearning balances the need for privacy and regulatory compliance with the practical demands of financial modeling, offering a way forward in an increasingly data-driven world. However, future research is required to address the potential limitations and integrate these methods more seamlessly into the broader financial ecosystem.

## References

1. T. Shaik, X. Tao, H. Xie, L. Li, X. Zhu and Q. Li, "Exploring the landscape of machine unlearning: A comprehensive survey and taxonomy," *arXiv preprint arXiv:2305.06360*, 2023.
2. N. Li, C. Zhou, Y. Gao, H. Chen, A. Fu, Z. Zhang and Y. Shui, "Machine Unlearning: Taxonomy, Metrics, Applications, Challenges, and Prospects," *arXiv preprint arXiv:2403.08254*, 2024.
3. S. Neel, A. Roth and S. Sharifi-Malvajerdi, "Descent-to-delete: Gradient-based methods for machine unlearning," in *Algorithmic Learning Theory*, 2021.
4. V. Gupta, C. Jung, S. Neel, A. Roth, S. Sharifi-Malvajerdi and C. Waites, "Adaptive machine unlearning," *Advances in Neural Information Processing Systems*, vol. 34, p. 16319–16330, 2021.
5. Y. He, G. Meng, K. Chen, J. He and X. Hu, "Deepoblivate: a powerful charm for erasing data residual memory in deep neural networks," *arXiv preprint arXiv:2105.06209*, 2021.
6. L. Graves, V. Nagisetty and V. Ganesh, "Amnesiac machine learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.
7. V. S. Chundawat, A. K. Tarun, M. Mandal and M. Kankanhalli, "Zero-shot machine unlearning," *IEEE Transactions on Information Forensics and Security*, vol. 18, p. 2345–2354, 2023.
8. A. K. Tarun, V. S. Chundawat, M. Mandal and M. Kankanhalli, "Fast yet effective machine unlearning," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.
9. M. Chen, W. Gao, G. Liu, K. Peng and C. Wang, "Boundary unlearning: Rapid forgetting of deep networks via shifting the decision boundary," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.
10. D. L. Felps, A. D. Schwickerath, J. D. Williams, T. N. Vuong, A. Briggs, M. Hunt, E. Sakmar, D. D. Saranchak and T. Shumaker, "Class clown: Data redaction in machine unlearning at enterprise scale," *arXiv preprint arXiv:2012.04699*, 2020.
11. Z. Wang, Y. Zhu, Z. Li, Z. Wang, H. Qin and X. Liu, "Graph neural network recommendation system for football formation," *Applied Science and Biotechnology Journal for Advanced Research*, vol. 3, no. 3, p. 33–39, 2024. doi: 10.5281/zenodo.12198843
12. C. Guo, T. Goldstein, A. Hannun and L. Van Der Maaten, "Certified Data Removal from Machine Learning Models," in *International Conference on Machine Learning*, 2020.
13. Z. Izzo, M. A. Smart, K. Chaudhuri and J. Zou, "Approximate data deletion from machine learning models," in *International Conference on Artificial Intelligence and Statistics*, 2021.
14. A. Warnecke, L. Pirch, C. Wressnegger and K. Rieck, "Machine unlearning of features and labels," *arXiv preprint arXiv:2108.11577*, 2021.
15. J. Wu, Y. Yang, Y. Qian, Y. Sui, X. Wang and X. He, "Gif: A general graph unlearning strategy via influence function," in *Proceedings of the ACM Web Conference 2023*, 2023.
16. K. Wu, J. Shen, Y. Ning, T. Wang and W. H. Wang, "Certified edge unlearning for graph neural networks," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023.
17. A. Golatkar, A. Achille and S. Soatto, "Eternal sunshine of the spotless net: Selective forgetting in deep networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.
18. A. Golatkar, A. Achille and S. Soatto, "Forgetting outside the box: Scrubbing deep networks of information accessible from input-output observations," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16*, 2020.
19. A. Golatkar, A. Achille, A. Ravichandran, M. Polito and S. Soatto, "Mixed-privacy forgetting in deep networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021.
20. Z. Li, B. Wang and Y. Chen, "Incorporating economic indicators and market sentiment effect into US Treasury bond yield prediction with machine learning," *Journal of Infrastructure, Policy and Development*, vol. 8, p. 7671, 2024.
21. Z. Li, B. Wang and Y. Chen, "A Contrastive Deep Learning Approach to Cryptocurrency Portfolio with US Treasuries," *Journal of Computer Technology and Applied Mathematics*, vol. 1, pp. 1-10, 2024

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.