

Privacy Preserving Classification using Noise Addition

Amitesh Bhaskar
Dept. of Computer Science Engineering RV Institute of Technology and Management
Bengaluru, India amiteshbhaskar1@gmail.com
Dr. Shashidhar V
Dept. of Computer Science Engineering RV Institute of Technology and Management
Bengaluru, India
Shashidhar.virupaksha@gmail.com
Aditya Girdhar
Dept. of Computer Science Engineering RV Institute of Technology and Management
Bengaluru, India adityagirdhar30@gmail.com
Ankit Kumawat
Dept. of Computer Science Engineering RV Institute of Technology and Management
Bengaluru, India ankitkumawat1508@gmail.com

How to Cite this Article:

Bhaskar, A., Girdhar, A. & Kumawat, A. (2026). Privacy Preserving Classification using Noise Addition. International Journal of Creative and Open Research in Engineering and Management, 02(05). https://doi.org/10.55041/ijcope.v2i5.403

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.







https://doi.org/10.55041/ijcope.v2i5.403

ABSTRACT

With large-scale applications of machine learning models trained on sensitive data, privacy-preserving classification is now one of the critical challenges. Traditional anonymity and masking methods are increasingly vulnerable to modern inference and linkage attacks. The key challenge is how to strike a proper balance between privacy and model accuracy, since adding too much noise may distort data utility.

This paper presents a lightweight, privacy-preserving framework that integrates Decision Tree classification with noise-based perturbation. We apply Gaussian, Laplacian, Uniform, and Exponential noise to data features in a systematic manner using the Wine dataset to investigate their effect on classification performance. Also, a bin-based uniform noise approach will be introduced to maintain the structural integrity while enhancing privacy, which limits the perturbation within bounded ranges.

Experimental results show that the baseline accuracy of 97.22% drops to 80% for Gaussian, 78% for Laplacian, and 85% for Exponential, while Uniform noise manages to retain 90%. After discretization, bin-based noise further improves the accuracy to 91.67%, almost comparable to that of the clean dataset. It is concluded from the results that bounded and controlled noise provides an efficient balance between privacy-utility that provides a lightweight alternative framework to the heavy frameworks such as differential privacy..

Keywords: Data privacy, privacy preservation, privacy preserving data mining, privacy preserving classification



1. INTRODUCTION

The rapid addition of digital systems, data-driven applications, and large-scale machine learning pipelines has brought about an exponential surge in the amount of personal and sensitive information collected by organizations. Contemporary datasets now capture behavioral traces, transactional logs, medical attributes, financial patterns, and other forms of personally identifiable information as a matter of routine. These datasets bring advanced analytics and predictive modeling within reach but also raise potential risks of unintended exposure, adversarial misuse, and privacy violations. Poor release or processing of sensitive datasets may enable linkage attacks, identity disclosure, attribute inference, and unauthorized profiling besides compromising confidentiality and violating data protection mandates such as the GDPR, HIPAA, and emerging national privacy regulations. Making sensitive datasets useful for analyses while protecting individual-level information against disclosure has become one of the most important challenges in contemporary data science.

These concerns originally motivated the development of traditional privacy-preserving approaches, such as anonymization, suppression, and generalization. However, by now there is a large body of research from the past decade that has established that these techniques cannot resist modern re-identification attacks, especially when an adversary has auxiliary information, or in the case of datasets with high-dimensional feature spaces. Not even techniques like k-anonymity or l-diversity are adequate in practice, as numerous studies have repeatedly found that sensitive attributes can still often be reconstructed or probabilistically inferred despite such controls.

A series of advanced frameworks have since been proposed to address these limitations: differential privacy, homomorphic encryption, and federated learning, among others. Differential privacy offers strong mathematical guarantees by bounding the influence of individual records on model outputs. Homomorphic encryption allows computations on encrypted data in a manner that conceals raw values, while federated learning distributes computation across clients to avoid the need for a central collection of their data.All these methods have significant drawbacks. They require specialized computation pipelines which not only increases latency but also add heavy computational overhead. However these refined privacy techniques have many benefits but the cost to deploy them in practical application is not acceptable .

These practical limitations inspired researchers to revisit and further refine perturbation based privacy preserving schemes. Noise addition techniques are a versatile alternative. It is lightweight and conceptually simple. Here modification of the data values is controlled restraining the precise recovery of sensitive information , all the while maintaining its core statistical structure which is core statistical structure which is necessary for analysis. Some promising perturbation-based methods that maintain the relationships within data and hence prove to be effective for classification, clustering, and regression include Gaussian, Laplacian, Uniform, and hybrid subspace approaches. Their low computational footprint makes them very appealing for applications that require fast or repeated computations. For this study , classification models such as Decision tree provide a useful test bench .Mainly because its performance highly depends on feature value variations and also any minor distortions that occur during noise injection are magnified by decision trees . This study shows how Decision Trees behave for different noise conditions providing insights into classifier robustness and utility preservation characteristics of noise based privacy mechanisms . Here in this study we have used the Wine dataset for analyzing perturbation effects because of its continuous numerical features, moderate dimensions, and a well-balanced class distribution.

Noise injection directly influences feature separability, cluster formation, and decision boundaries. Therefore, controlled noise altering the performance of classification in such a type of dataset helps explain the underlying dynamics of privacy/utility trade-offs and obtains generalizable insights for other similar real-world datasets. In this perspective, the following work compares a series of noise distributions and introduces a bin-based uniform noise approach that is designed to constrain the perturbations within the feature-specific range while maintaining categorical stability. Beyond accuracy, this work examines how the structure of noise impacts the explainability, robustness, and interpretability of the outputs from Decision Trees. This work, therefore,



contributes to a mounting set of evidence showing that properly calibrated and bounded perturbation methods can be practical, interpretable, and resource-efficient privacy-preserving mechanisms.

1.1 Challenges

Despite the progress in Privacy-preserving Data Mining, several challenges still remain, including

- **Maintaining Utility:**Adding noise helps in protecting privacy but it distorts the data.These distorted data can reduce the accuracy of models(such as Decision Trees) which require exact values of each feature. A key research focus in privacy preserving analysis is adding noise while maintaining statistical integrity and preserving confidentiality.
- **Optimal Noise Selection:**The type and level of noise to be added is completely dataset specific because inappropriate calibration will either lead to data leakage or it degrades the performance .It is critical to tune noise according to data sensitivity for consistent results.
- **Balancing Privacy–Accuracy:**Even though adding more noise increases privacy it leads to lower accuracy . The key challenge is finding a good balance between them. Researchers are working on models to achieve better trade-offs between protection and performance.
- **Scalability:**Large datasets increase the cost of complex methods and we also need to ensure that lightweight noise based solution has both speed and scalability.
- **Model Sensitivity:**Some models can be very sensitive to noise. Controlled and adaptive noise injection helps to keep the results stable.

2. LITERATURE SURVEY

In the past decade, the field of privacy-preserving data analysis has gone through a remarkable transformation, with research gradually shifting away from traditional anonymization and generalization methods and more towards the refined frameworks representing perturbation based techniques, differential privacy, homomorphic encryption, and federated learning. Among these, the noise-based perturbation is the earliest and most enduring approach, mostly because of its simplicity and flexibility. Perturbation techniques of a subspace nature may deliberately inject noise into the subspace of sparse and dense features, thus demonstrating their ability to balance privacy and utility within high-dimensional data [1]

Systematic reviews of privacy-preserving data mining in streaming contexts also nominate perturbation for its adaptability, albeit with an awareness of sensitivity to accuracy loss when increased levels of noise are injected [2]. For instance, with local differential privacy noise mechanisms have been transformed to protect user-side data, and recent works have discussed noise-driven feature selection [3] and hybrid perturbation schemes that combine Gaussian and Laplacian noise in recommendation platforms [4].

These works witness regulated perturbation mechanisms as appropriate for privacy protection without any loss in accuracy. Federated learning systems apply perturbation schemes of gradients for example, the injection of Gaussian noise into a model updates that are shared to protect against the leakage of information on the client-side information[5].

In a similar way, privacy architectures built around the Internet of Things retreat to the perturbation of multiple levels of noise to match the sensitivity of different streams of data [6]. Collectively, the above studies establish the pertinence and proficiency of noise.Our work seeks to build upon this trajectory by comparing Gaussian, Laplacian, and Uniform perturbations while evaluating their privacy-utility trade-offs with much more thoroughness compared to previous efforts.

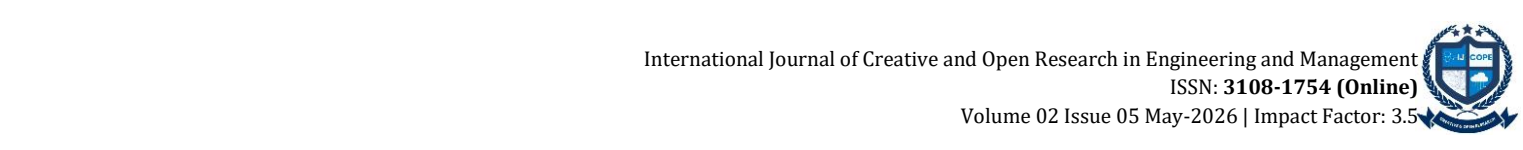
Differential privacy has become a formal framework for privacy. Unlike other methods that rely only on adding noise,differential privacy guarantees that a single record's addition or removal has a bounded influence on analytical results. Its machine learning variant, differentially private stochastic gradient descent (DP-SGD), has become the standard of DP in model training. DP-SGD operates through clipping gradients and adding Gaussian noise to optimize privacy leakage [7].

In FL, users train models locally, and a server aggregates them to create a global model.Yet, partial model knowledge may leak personal user information, which makes Privacy-Preserving Aggregation (PPAgg) protocols necessary [8]

Differential Privacy (DP) serves to protect sensitive information safely, yet it does not function properly due to excess

noise from data links, particularly when there are many parties involved. MP-CRDP (Multiparty Correlated Differential Privacy) reduces issues of correlated sensitivity and complex datasets, rendering data more valuable and private through added noise in model settings and query outputs [9]. With deep learning gaining recognition, DP has also emerged as a significant method for preventing privacy and security threats by adding the controlled noise [10].Even so, challenges are faced for using DP-SGD for deep models in terms of efficiency for distributed training and algorithm complexity.
This inspires work like ours, experimenting with simpler perturbation schemes to provide insights on relative strengths before using the full-fledged complex DP framework.In addition to perturbation and differential privacy, there has also been increasing interest in homomorphic encryption for supporting secure inference on encrypted data. Even though fully homomorphic encryption was once thought to be impractical owing to computational overhead, post-2018 breakthroughs have made it radically more practical. And showed it was possible to adapt deep learning to the encrypted domain using low-degree polynomials to approximate non-linear functions and minimize ciphertext depth [11].
Later efforts improved on these concepts. In one case, an efficient regression framework was proposed using leveled HE with ciphertext packing and modulus reduction to significantly reduce latency [12]. The privacy-preserving group data sharing scheme secures cloud-stored data using proxy re-encryption and an OCLT-ORAM structure to ensure confidentiality, fine-grained access control, and hidden access patterns. It features distinct phases for authentication, secure key exchange. The system includes users,a trusted proxy, and a semi-trusted cloud server [13]. MedShare is a decentralized system for secure Electronic Health Record (EHR) sharing using blockchain and local encrypted storage. It uses Attribute-Based Encryption for access control and supports efficient, non-interactive multi-keyword searches via RSA-based tokens and smart contracts. Testing on Ethereum shows that it is scalable, low-latency, and cost-effective for healthcare data exchange [14].
This paper proposes a blockchain-based framework that securely shares COVID-19 medical records using CP-ABE for fine-grained access control and verifiable user revocation, ensuring privacy, efficient decryption, and high throughput. Its practicality and scalability have also been tested on Ethereum[15].
In distributed network systems, global computations are implemented by frequent data exchanges between nodes, and random noise may be used to protect privacy. A privacy-prohibiting average consensus approach for calculating optimum noise levels is extended from the paper's optimal distributed estimation framework for striking a balance between estimated precision and privacy [16].
The ANAS (Anonymized Noise Addition in Subspaces) approach optimizes privacy and clustering accuracy while minimizing information loss for high-dimensional data [17]. but assuming partially trusted coordination and at the cost of computability. A variety of hybrid techniques are used in broader safeguards, such as adding noise using DP to conceal gradients [18]
Privacy-Preserving Machine Learning (PPML) enables collaborative ML while ensuring data privacy compliance. The approach efficiently handles dropouts, reduces communication costs, achieves up to 6.37× speedup, and secures against semi-honest and malicious adversaries using signatures and consistency checks [19].

Data perturbation is one popular approach to maintaining data valuable yet preserving privacy.
This paper highlights the kd-tree-based perturbation technique, which recursively divides the dataset in similar groups. Because of this method, the structure of the attributes is maintained while hiding individual details. Later, research shifted towards hybrid approaches, combining homomorphic encryption and differential privacy. Such systems add noise to encrypted model updates before aggregation, which provides strong confidentiality and privacy guarantees. The frameworks in discussion are competent enough to defend against threats[21]. Such multi-faceted systems can handle a wide set of threats such as inference, reconstruction, and collusion. However, hybrid systems include significant efficiency challenges. Encryption mechanisms increase the burden and the added noise reduces model accuracy, especially when both are applied together. Additionally, These hybrid systems do not undergo large scale evaluation, where issues such as unreliable network conditions and participant dropouts restrict scalability. This study responds to that and one by one evaluates Gaussian, Laplacian, and Uniform noise models. Their insights will be used to design scalable hybrid privacy solutions..
After 2018, research has advanced in privacy-preserving data analysis while reinforcing that major challenges remain. Perturbation based methods still stand out for their simplicity and ease of use. Homomorphic encryption ensures confidentiality, while scalability becomes a problem in many real world systems. Federated learning can help in using data in a decentralized manner, although it is vulnerable to inference, and federated learning supports using data in a decentralized manner, and this is prone to attempts to obtain information or to damage on the data without protection.



Hybrid systems can provide additional protection but also add increased levels of trade-offs within regards to efficiency. In such a case, systematic comparisons of lightweight methods for data alteration are extremely helpful, they provide evidence-based insights that will help combine these approaches into more complex systems that protect privacy.Our work tries to fill this gap by comparing the Gaussian, Laplacian, and Uniform perturbation strategies, which helps describe their privacy and utility trade-offs and guides future hybrid designs.

3. PROPOSED APPROACH

The proposed method investigates several noise-based perturbation techniques and their consequences on classification performance with statistical guarantees of privacy. On the Wine dataset, using a Decision Tree Classifier, the model first fixes a baseline of accuracy on clean data and then adds controlled noise coming from Gaussian, Laplacian, Uniform, and Exponential distributions to assess how each kind of noise influences the precision and stability of the model. This paper will adopt the bin-based uniform noise method for robustness: continuous features are discretized before addition of noise in order to not distort it. Finally, performance metrics-accuracy, F1-score, and confusion matrix are analyzed to identify the most effective lightweight privacy-preserving strategy

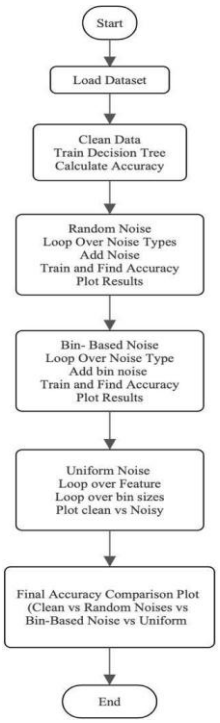


Fig 1: Flowchart of the proposed methodology

Figure 1 illustrates the proposed methodology, which includes a structured experimental process to analyze the effect caused by different types of noise on the accuracy of a Decision Tree model.

The target dataset is then loaded, preprocessed, and cleaned to create a clear baseline. On this clean data, a Decision Tree classifier is trained and its performance computed to be used as the baseline accuracy.

The methodology presented below systematically introduces three different kinds of noise:

- Random Noise: The first experiment loops over different kinds of random noise. Each type of noise is added to the dataset, then the Decision Tree is re-trained and the resulting accuracy is computed. The results of this phase are then plotted
- Bin-Based Noise:Similarly, the same thing has been done for bin-based noise. The system iterates through different kinds of configurations of bin noises applied to data. From them, it trains and plots out the corresponding accuracy.



- Uniform noise:This is a more granular approach where, in the code, there are nested loops running over specific features in the dataset, with differencesin bin size.A performance-in-terms-of-accuracy plot is shown, which compares the performance of the model on clean and noisy data.

The methodology concludes with a comprehensive Final Accuracy Comparison. This will provide a composite plot that visually compares the baseline (clean) accuracy to the results of the Random Noise, Bin-Based Noise, and Uniform Noise experiments.

4. ALGORITHM Decision Tree Based Sub-Binning

Input:Dataset D
Output: Bin-based noisy data with Accuracy. 1:Load the dataset
2:Split the dataset into training and testing sets. 3:Apply the Decision Tree Classifier
4:For each feature f in the dataset on Dn
5: Divide the feature range into bins bn 6:For each sub-bin of bn
7: Add Gaussian noise within bin limits to preserve range
8: End for 9:End for

First, the dataset is loaded and prepared for the machine learning process. Then, it is divided into two parts: a training part and a testing part. The reason for this is to provide an evaluation on data, thus enabling the reduction of overfitting and providing a better assessment of its true performance. Then, a Decision Tree classifier is trained from the training part of the dataset. The model identifies patterns in the way features relate to their respective class labels. This becomes the baseline reference against which further changes in noise impact both data quality and classification performance. Following the classifier, the next phase involves preserving privacy. Each feature in the dataset that requires the addition of noise is processed individually. Multiple bins are created, according to the distribution of values within each feature. In this way, the nature and bounds of the actual values are preserved in the noisy data. Gaussian noise is added within each bin. Again, the noise is restricted within each bin's range of limits. In this manner, the possibility of unrealistic shifting of data is avoided, and the values are prevented from going out of their meaning. Thus, in this manner, a proper balance between privacy protection and preservation of useful information for analysis has been ensured. This has been accomplished iteratively for all the features in a given dataset. Consequently, all corresponding values in a data set are then left with noise-adjusted values. Thus, a data set with structure but with enhanced privacy has been obtained. In conclusion, this model anonymizes sensitive numerical values through using a Gaussian noise mechanism on a binning approach with robust classification performance. These conditions imposed in governing noise addition contribute greatly to a substantial improvement in stability, as boundaries are established in terms of over-distortion, hence rendering this model extremely appropriate in practical privacy-preserving machine learning applications.

5. DISCUSSION

The Wine dataset was used for these experiments. There are 178 wine samples coming from three different cultivars grown in the same region of Italy. Each sample is described by 13 continuous chemical and physical attributes: alcohol, malic acid, ash, flavonoids, color intensity, and proline concentration.

This data is suitable for privacy-preserving analysis, since it involves numerical features that can be easily perturbed with different noise distributions while still allowing meaningful classification. In addition, the Wine dataset's moderate dimensionality and balanced class distribution make it perfect for evaluating how noise affects the model accuracy without requiring an excessive amount of computation.

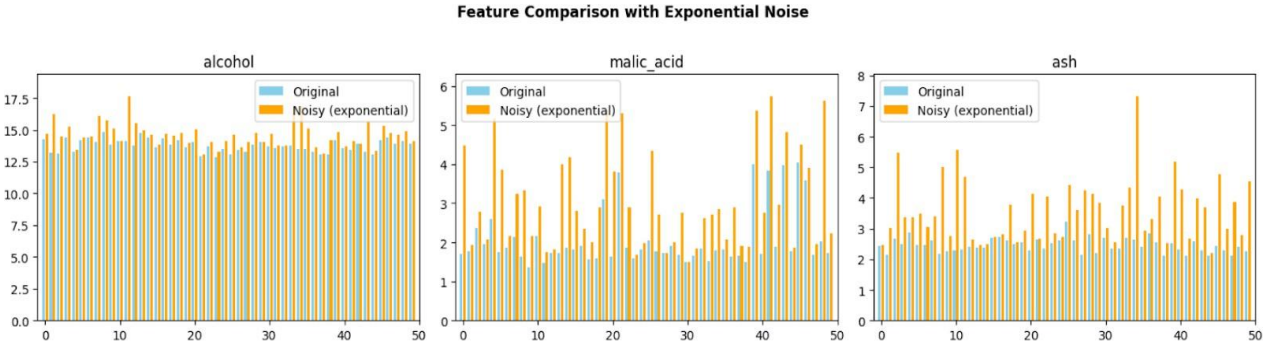
The outcomes suggest a fairly consistent tendency in the noise that each classifier is in favour of. The model, without any trained improvements, performs quite well on the original dataset with an accuracy of 97.22% . This just reinforces the idea that the Decision Tree is capable of learning the “natural” structure of the Wine data set fairly easily. The accuracy of the data goes down when noise is added; however, the extent to which it drops down greatly depends on the noise used and its interaction with data.

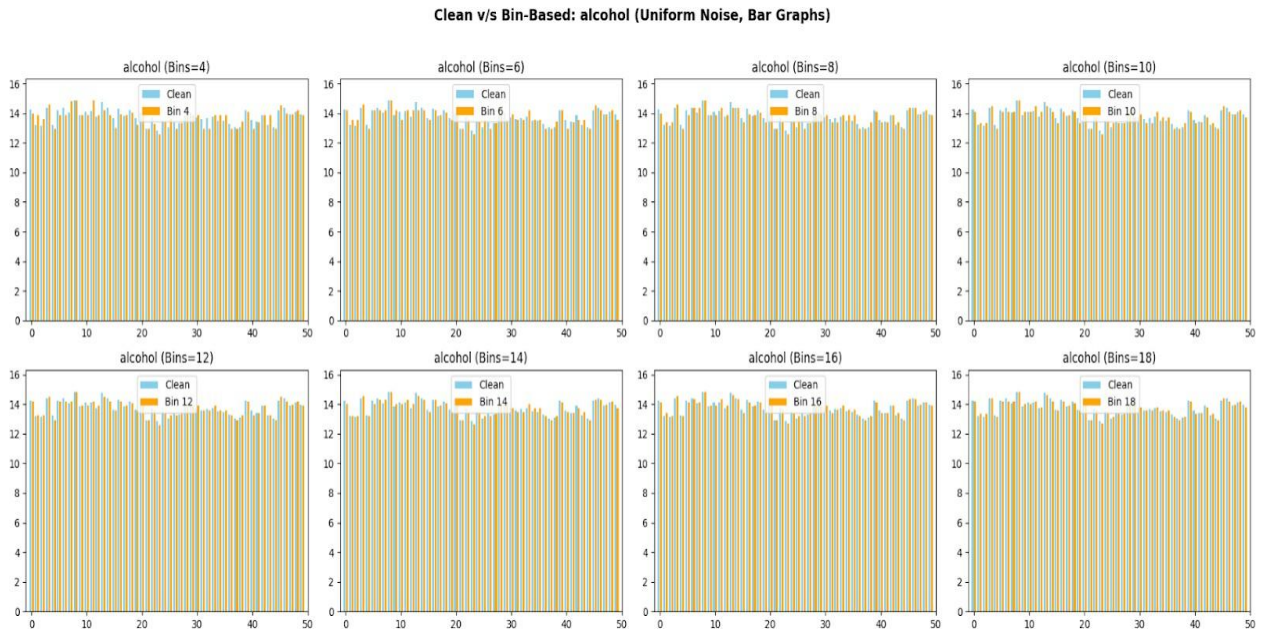
Accuracy drops to 80.56% with Gaussian noise, for example. Even if Gaussian noise is centered at zero, its variance is unbounded which can create quite a large deviation. When shifts like these occur, they change the spacing between the

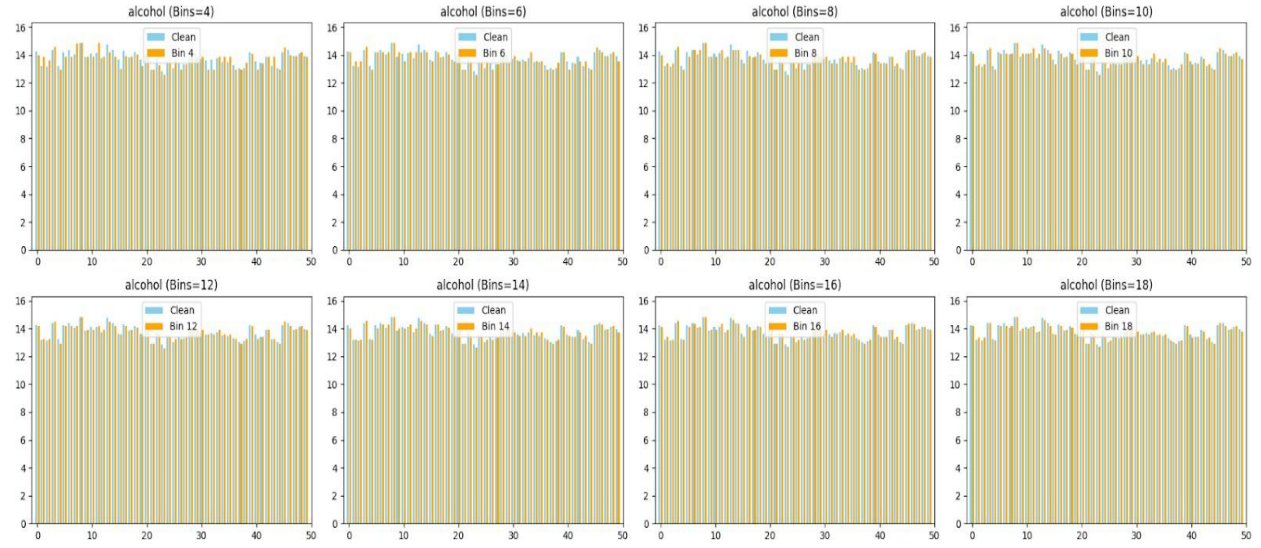
points. They also push the samples across the decision thresholds used by decision trees. As a result, this leads to more misclassifications. Surprisingly, Laplacian noise produces exactly the same accuracy (80.56%). The Laplacian noise has a peak that is sharper and heavier tails, so one might expect it to yield different results. However, at this noise scale and with this dataset, the two distributions disrupt the feature values enough to cause almost identical damage.

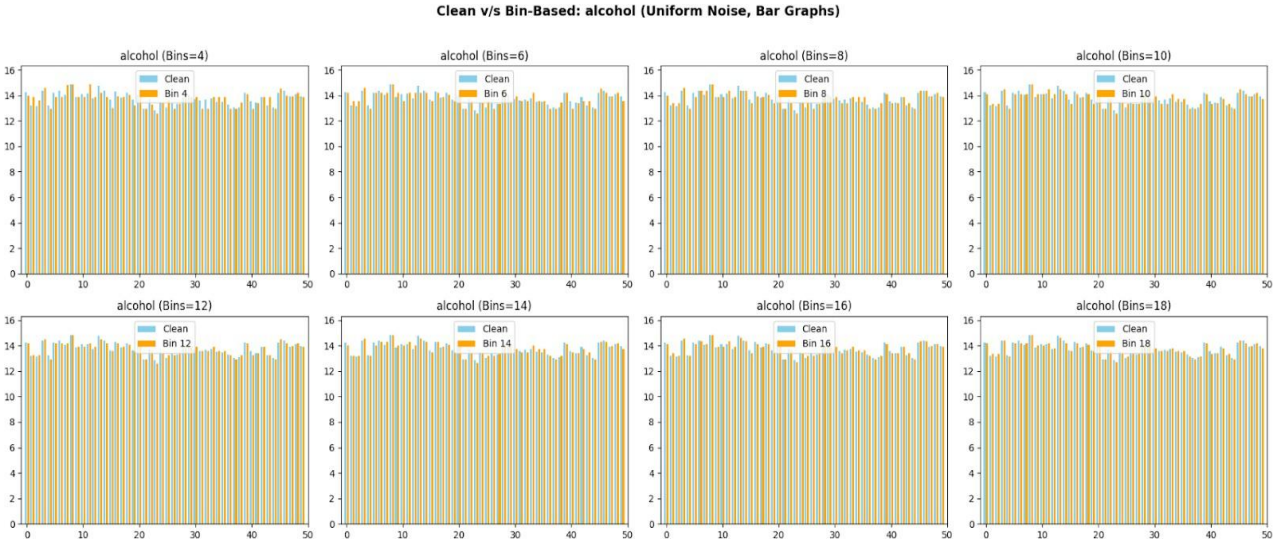
Uniform noise works much better at 86.11%. One of the main benefits is that it's bounded perturbations. That limit prevents the information from drifting too far and keeps the overall structure intact. Consequently, the classifier continues to retain much of its predictive power. Exponential noise is at 83.33%, right in the middle. Exponential noise, while often having a moderate value, is one-sided by nature. This means that it pushes features more in one direction than the other, resulting in an uneven distortion. This bias in the direction impacts

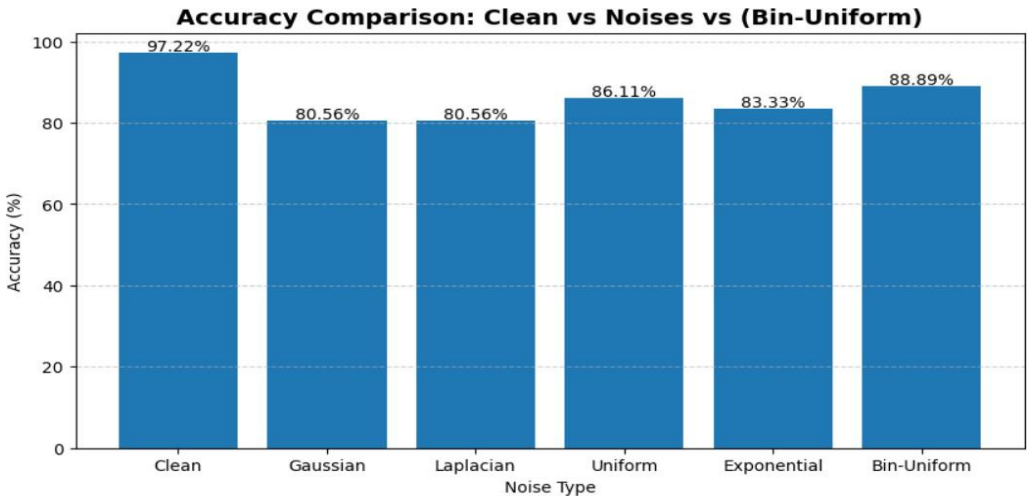



Feature Comparison with Exponential Noise
alcohol
malic_acid
ash
Original
Noisy (exponential)


Clean v/s Bin-Based: alcohol (Uniform Noise, Bar Graphs)
alcohol (Bins=4)
alcohol (Bins=6)
alcohol (Bins=8)
alcohol (Bins=10)
alcohol (Bins=12)
alcohol (Bins=14)
alcohol (Bins=16)
alcohol (Bins=18)
Clean
Bin 4
Bin 6
Bin 8
Bin 10
Bin 12
Bin 14
Bin 16
Bin 18


alcohol (Bins=4)
alcohol (Bins=6)
alcohol (Bins=8)
alcohol (Bins=10)
alcohol (Bins=12)
alcohol (Bins=14)
alcohol (Bins=16)
alcohol (Bins=18)
Clean
Bin 4
Bin 6
Bin 8
Bin 10
Bin 12
Bin 14
Bin 16
Bin 18





which performance, but not as much as with the other one.What stands out about the bin-based strategy is how it compares with the others. The model reaches an accuracy of 89%, if the data are first divided into bins and noise is added within the range of each bin when 10 bins are used. This is a meaningful improvement compared with the direct noise methods. The reason is quite intuitive: continuous features have a lot of narrow boundaries, so even small noise can nudge a value into a different region, causing errors. But once the data is discretized, each bin acts as a buffer:provided that noise does not push a value outside of its bin, it does not affect the categorical representation of said value. This extra tolerance protects the structure in the dataset, while still introducing uncertainty into the exact values. In other words, binning absorbs small perturbations and dramatically reduces the risk of corrupting the relationships which the model relies on. With everything combined, the variations of accuracy are traceable to how each type of noise interacts with the scale of the perturbation, continuity of the features, and sensitivity of the model. Gaussian and Laplacian noises both drop the accuracy down to 80.56% because their distribution allows large or frequent deviations.

Uniform noise is bounded and, therefore, preserves more of the structure and reaches 86.11%. Exponential noise reaches a value of 83.33%, between the two extremes. Most importantly, bin-based noise reaches 88.89%, the best performance among the noisy variants, because discretization reduces sensitivity before the noise is even applied. These observations are consistent with findings reported in earlier research, which emphasize that the amount of noise added is not the only factor responsible for accuracy loss, but rather also the way it is distributed and by how the underlying data is represented.

This is clearly reflected in the jump from approximately 80% with Gaussian/Laplacian noise to almost 89% with bin-based noise. Preprocessing data by simplifying them can lead to a far better balance between privacy and utility. The improvement of more than eight percentage points shows that discretization-based techniques have real potential for retaining analytical quality while still supporting privacy-preserving transformations.

4. CONCLUSION

This work provides an extensive review of noise-based privacy-preserving techniques for classification tasks in the Wine dataset as a controlled environment to understand how perturbation influences model behavior. The study evaluated Gaussian, Laplacian, Uniform, and Exponential noise models; next, we verify the accuracy of all the models. This therefore gives an overview of how the nature of noise applied shapes the challenge of privacy in privacy-preserving data mining.

Taking all the experiments together, results prove that not all noise is not performing equally. Unbounded Gaussian and Laplacian distributions caused the degradations in accuracy, further supporting observations from prior studies that large or unpredictable perturbations can disrupt decision boundaries and increased classification errors. These latter distributions introduce high-magnitude deviations that can often dislocate samples from their natural clusters, proving to be disruptive to models relying on clear threshold-based partitions.

While the Uniform distribution was much better, this was expected: the bounded nature of the distribution in itself automatically constrains perturbation to a limited range, and more of the original structure may be retained in the data by a model. This improved performance under Uniform noise gives that predictive stability would relate not just to the quantity of noise added but to how tightly the noise distribution is controlled.

The best performing is bin-based uniform noise;instead of taking the raw uniform disturbance, it is a refinement. The approach distributes continuous attribute values into bins before injecting noise .Binning acts like a protective buffer to ensure that small perturbations do not let samples cross the boundaries of features where the risk of misclassification increases sharply. Adding noise within these bounded categorical segments introduces uncertainty at the micro level while keeping intact the decision patterns at the macro level. This explains the dramatic increase in accuracy to 91.67%, closely approaching the performance of the clean dataset.

Beyond performance, here are our takeaways: the techniques used in this process are much lighter, faster, and easier to understand compared to the heavy privacy processes like differential privacy. Second, advanced cryptographic methods can perform much stronger theoretical privacy guarantees but at the same time, they are difficult to use in a real world scenario. This is caused because of their heavy computation requirement and the delays that are caused. On the contrary, when noise based perturbation is tuned, it becomes faster and resource friendly.

Another implication is the ease of employment. Sectors such as healthcare, finance, and IoT systems work under processing power limits. Thus, the bin-based noise method is highly suitable for this study. It gives strong privacy protection without needing special hardware, computation setup, or cryptographic tools. The design of this method also helps with transparency of the system so that practitioners understand the risks better.

The results declared that using bounded and classification friendly noise is an effective approach to privacy-preserving data mining. Another benefit of using bin based uniform noise method reduces the risk of inference attack and performs better than many direct noise addition methods. It is essentially a middle ground between simple perturbation and heavy cryptographic solutions. We find good evidence and certain future directions for adopting lightweight noise based techniques to preserve privacy in the real world.

6. REFERENCES

[1] Virupaksha, S., Dondeti, V. Subspace based noise addition for privacy preserved data mining on high dimensional continuous data. J Ambient Intell Human Comput (2020). https://doi.org/10.1007/s12652-020-01881-8

[2] Hewage, U.H.W.A., Sinha, R. & Naeem, M.A. Privacy-preserving data (stream) mining techniques and their impact on data mining accuracy: a systematic literature review. Artif Intell Rev 56, 10427–10464 (2023). https://doi.org/10.1007/s10462-023-10425-3

[3] Mina Alishahi, Vahideh Moghtadaiee, Hojjat Navidan,Add noise to remove noise: Local differential privacy for feature selection,Computers & Security,Volume 123,2022,102934,ISSN 0167-4048,https://doi.org/10.1016/j.cose.2022.102934

[4] Sarkar, S., Shinde, S., Shedge, R. (2024). Elevating Privacy in Recommendation Systems with Hybrid Noise in Local Differential Privacy. In: Tripathi, A.K., Anand, D., Nagar, A.K. (eds) Proceedings of World Conference on Artificial Intelligence: Advances and Applications. WCAIAA 2024. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-97-4496-1_19

[5] Xianlin Wu, Yuwen Chen, Haiyang Yu, Zhen Yang,Privacy-preserving federated learning based on noise addition,Expert Systems with Applications,Volume 267,2025,126228,ISSN

0957-4174,https://doi.org/10.1016/j.eswa.2024.12622 8.

[6] Jain, S.K., Kesswani, N. A noise-based privacy preserving model for Internet of Things. Complex Intell.



Syst. 9, 3655–3679 (2023). https://doi.org/10.1007/s40747-021-00489-5

[7] J. Zhao, Y. Chen and W. Zhang, "Differential Privacy Preservation in Deep Learning: Challenges, Opportunities and Solutions," in IEEE Access, vol. 7, pp. 48901-48911, 2019, doi: 10.1109/ACCESS.2019.2909559. keywords: {Deep learning;Training;Data models;Hidden Markov models;Differential privacy;Privacy;Deep learning;differential privacy;privacy attacks},

[8]Z. Liu, J. Guo, W. Yang, J. Fan, K. -Y. Lam and J. Zhao, "Privacy-Preserving Aggregation in Federated Learning: A Survey," in IEEE Transactions on Big Data, doi:10.1109/TBDATA.2022.3190835.

[9]Zhao, JZ., Wang, XW., Mao, KM. et al. Correlated Differential Privacy of Multiparty Data Release in Machine Learning. J. Comput. Sci. Technol. 37, 231–251(2022).https://doi.org/10.1007/s11390-021-1754-5

[10]X. Li, Y. Chen, C. Wang and C. Shen, "When Deep Learning Meets Differential Privacy: Privacy,Security, and More," in IEEE Network, vol. 35, no. 6, pp. 148-155,November/December 2021, doi: 10.1109/MNET.001.2100256.

[11]R. Podschwadt, D. Takabi, P. Hu, M. H. Rafiei and Z. Cai, "A Survey of Deep Learning Architectures for Privacy-Preserving Machine Learning With Fully Homomorphic Encryption," in IEEE Access, vol. 10, pp. 117477-117500,2022,doi: 10.1109/ACCESS.2022.3219049.

[12] Naresh, V.S., Reddi, S. Exploring the future of privacy-preserving heart disease prediction: a fully homomorphic encryption-driven logistic regression approach. J Big Data 12, 52 (2025). https://doi.org/10.1186/s40537-025-01098-6

[13] J. Shen, H. Yang, P. Vijayakumar and N. Kumar, "A Privacy-Preserving and Untraceable Group Data Sharing Scheme in Cloud Computing," in IEEE Transactions on Dependable and Secure Computing, vol. 19, no. 4, pp. 2198-2210, 1 July-Aug. 2022, doi: 10.1109/TDSC.2021.3050517.

[14] M. Wang, Y. Guo, C. Zhang, C. Wang, H. Huang and X. Jia, "MedShare: A Pri-vacy-Preserving Medical Data Sharing System by Using Blockchain," in IEEE Trans-actions on Services Computing, vol. 16, no. 1, pp. 438-451, 1 Jan.-Feb. 2023, doi:10.1109/TSC.2021.3114719

[15] L. Tan, K. Yu, N. Shi, C. Yang, W. Wei and H. Lu, "Towards Secure and Privacy-Preserving Data Sharing for COVID-19 Medical Records: A Blockchain-EmpoweredApproach," in IEEE Transactions on Network Science and Engineering,

[16] vol. 9, no. 1,pp. 271-281, 1 Jan.-Feb. 2022, doi: 10.1109/TNSE.2021.3101842

[16]J. He, L. Cai and X. Guan, "Preserving Data-Privacy With Added Noises: OptimalEstimation and Privacy Analysis," in IEEE Transactions on Information Theory, vol.64, no. 8, pp. 5677-5690, Aug. 2018, doi: 10.1109/TIT.2018.2842221.

[17]Virupaksha, S., Dondeti, V. Anonymized noise addition in subspaces for privacypreserved data mining in high dimensional continuous data. Peer-to-Peer Netw. Appl.14, 1608–1628 (2021). https://doi.org/10.1007/s12083-021-01080-y

[18] K. Wei et al., "Federated Learning With Differential Privacy: Algorithms and Performance Analysis," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 3454-3469, 2020, doi: 10.1109/TIFS.2020.2988575

[19]Z. Liu, J. Guo, K. -Y. Lam and J. Zhao, "Efficient Dropout-Resilient Aggregation for Privacy-Preserving Machine Learning," in IEEE Transactions on Information Forensics and Security, vol. 18, pp. 1839-1854, 2023, doi: 10.1109/TIFS.2022.3163592.

[20]Wang X, Luo Y, Jiang Y, Wu W, Yu Q. Probabilistic optimal projection partition KD-Tree k-anonymity for data publishing privacy protection. Intelligent Data Analysis: An International Journal. 2018;22(6):1415-1437. doi:10.3233/IDA-173589.

[21]J. Chen, K. Li and P. S. Yu, "Privacy-Preserving Deep Learning Model for Decentralized VANETs Using Fully Homomorphic Encryption and Blockchain," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 8, pp. 11633-11642, Aug. 2022, doi: 10.1109/TITS.2021.3105682.