Privacy-Preserving Machine Learning for Fraudulent Website Detection Using Differentially Private Decision Trees

Muhammad Ali Fauzi
Faculty of Computer Science
Universitas Brawijaya
Malang, Indonesia
moch.ali.fauzi@ub.ac.id

Yudhistira
Faculty of Computer Science
Universitas Brawijaya
Malang, Indonesia
yudhisthereal@student.ub.ac.id

Bian Yang
Department of Information Security and Communication Technology
Norwegian University of Science and Technology (NTNU)
Gjøvik, Norway
bian.yang@ntnu.no

***Abstract*—Detecting fraudulent websites is essential for cybersecurity, as these sites often serve as vehicles for phishing, identity theft, and malware distribution. Machine learning models, particularly Decision Trees, have shown strong effectiveness in identifying patterns indicative of fraudulent sites. However, the need to protect sensitive user data introduces challenges, necessitating privacy-preserving approaches like Differential Privacy (DP). This study investigates the application of DP in Decision Tree classifiers to assess the trade-offs between privacy and utility in detecting fraudulent websites. Using a publicly available dataset, we compare a traditional Decision Tree with a differentially private version across various values of the privacy budget, epsilon (ϵ). Our results show that while the original model achieves optimal performance with an accuracy of 95.4%, the DP model at an optimal $\epsilon = 4.09$ maintains high utility, achieving 83.7% accuracy and an F1-score of 82.7%, demonstrating its suitability for privacy-sensitive applications. This study highlights that with moderate ϵ values, DP Decision Trees can provide effective privacy protection with minimal performance loss, making them viable for real-world applications where both privacy and predictive accuracy are critical.**

***Index Terms*—Fraudulent website detection; machine learning; Decision Tree; differential privacy; privacy-preserving technology; epsilon; cybersecurity; privacy-utility trade-off.**

I. Introduction

Fraudulent websites are a significant threat in cyberspace, often used for phishing, identity theft, and malware distribution. Detecting these websites accurately and efficiently is essential for protecting users from online threats. Machine learning has become a popular approach for this task due to its capability to analyze patterns and detect anomalies in web data that may indicate fraud [1], [2].

While machine learning models are powerful, they often require access to sensitive user data, raising concerns about data privacy [3]. Privacy-preserving technologies, such as Differential Privacy (DP), address this issue by ensuring that individual user data cannot be inferred from the model's output, even if an attacker has access to the model [4], [5]. Differential Privacy introduces noise to the training process, balancing model accuracy with privacy. This paper investigates the application of DP in Decision Trees, comparing a traditional Decision Tree with a differentially private version and analyzing the impact of varying levels of the privacy budget, epsilon, on model performance.

II. Materials and Methods

A. Dataset

The dataset used in this study is sourced from a public dataset [6]. This dataset comprises several features that characterize websites, including elements such as URL length, domain age, and the presence of HTTPS. These features are essential in distinguishing legitimate websites from fraudulent ones. The dataset is labeled, with each instance marked as either fraudulent or non-fraudulent, which allows supervised learning models to be trained and evaluated on this binary classification task.

B. Methodology

The methodology involves two main steps: training the models and evaluating their performance. We use two types of Decision Tree classifiers: a standard Decision Tree and a Differential Privacy-based Decision Tree (DP Decision Tree). The DP Decision Tree incorporates privacy-preserving noise to ensure compliance with Differential Privacy standards, balancing privacy and performance.

1) Standard Decision Tree Classifier: The Decision Tree classifier is a widely used machine learning model for classification tasks, known for its interpretability and effectiveness in identifying decision rules based on input features. In this study, the standard Decision Tree is trained on the dataset without any privacy constraints. This model serves as a baseline for comparison, providing insights into the maximum achievable performance when no privacy-preserving mechanisms are applied.

2) Differentially Private Decision Tree Classifier: The Differential Privacy-based Decision Tree, provided by IBM's *diffprivlib* library, is a modified version of the Decision Tree classifier that integrates Differential Privacy. This model incorporates a privacy budget parameter, epsilon (ϵ), which controls the level of privacy. Lower values of ϵ introduce more noise to the model, ensuring stronger privacy but potentially impacting accuracy. In this study, we evaluate the performance of the DP Decision Tree with various values of ϵ (e.g., 0.01, 0.1, 1.0, 4.09, 10) to observe how privacy constraints affect model performance metrics such as accuracy, precision, recall, and F1-score.

3) Cross-Validation and Performance Metrics: To ensure robust evaluation, we employ 5-fold cross-validation for both models. Cross-validation divides the dataset into five subsets, or folds, training the model on four folds and testing on the fifth. This process is repeated five times, with each fold serving as the test set once. The average of the results from these folds provides an unbiased estimate of the model's performance [7].

The performance metrics used in this study are:

- **Accuracy**: Accuracy is the proportion of correctly classified instances out of the total instances and is calculated as:
$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$
where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively [8].
- **Precision**: Precision, or the positive predictive value, indicates the accuracy of positive predictions and is given by:
$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$
- **Recall**: Recall, also known as sensitivity or true positive rate, measures the model's ability to correctly identify positive instances. It is calculated as:
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$
- **F1-Score**: The F1-score is the harmonic mean of precision and recall, providing a balanced measure when both precision and recall are important. It is calculated as:
$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

4) Analysis of Epsilon Effect: The study evaluates the impact of different ϵ values on the performance of the DP Decision Tree. By observing how accuracy, precision, recall, and F1-score vary across a range of ϵ values, we gain insight into the trade-offs between privacy and utility. For each ϵ value, the average performance across the 5-fold cross-validation is recorded and compared with the results of the standard Decision Tree model.

The best-performing ϵ value, determined by the highest F1-score, is identified, and its results are presented alongside those of the standard Decision Tree to highlight the balance point between privacy and model performance. This comparison informs practitioners of the epsilon settings at which the model provides effective privacy protection with minimal loss in accuracy, precision, recall, and F1-score.

III. Results

A. Analysis of Results with Varying Epsilon Values

In differential privacy, the privacy budget, represented by epsilon (ϵ), determines the trade-off between privacy and utility. The performance of the differentially private Decision Tree model was evaluated across a range of epsilon values. The results across epsilon values can be seen in Figure 1 and 2. The following is the analysis of the results:

1) Accuracy: Accuracy generally increases as ϵ increases, peaking at higher epsilon values, such as 4.09, 10.48, and 100.0. At very low epsilon values (e.g., 0.01 to 0.03), accuracy is lower, indicating that a stricter privacy constraint (lower epsilon) may hinder model performance.

2) Precision, Recall, and F1-Score: **Precision** follows a similar trend to accuracy, increasing with epsilon. The highest precision values are observed at higher epsilon values (e.g., 4.09 and 10.48), indicating the model's improved ability to correctly predict positive samples as privacy constraints are relaxed.

Recall improves with increasing epsilon, though it fluctuates more than precision. This suggests that while the model's sensitivity improves with a higher epsilon, its effect on recall is less consistent.

F1-Score, balancing precision and recall, mirrors these trends, generally increasing with epsilon. Higher F1-scores are observed around epsilon values of 4.09, 10.48, and 100, confirming that a less strict privacy budget yields better model performance.

3) Optimal Epsilon Range: Epsilon values around 4 to 10 appear to provide a reasonable trade-off, offering high accuracy, precision, and F1-scores without compromising privacy significantly. For applications where balancing privacy and performance is crucial, this range may be most suitable.

4) Best Epsilon Value: The best epsilon value for the Differential Privacy-based Decision Tree was identified as $\epsilon = 4.09$, where the model achieved the highest F1-score. The performance metrics at this epsilon are as follows:

- **Accuracy**: 0.8369
- **Precision**: 0.8716
- **Recall**: 0.8369
- **F1-Score**: 0.8270

This epsilon value represents the optimal balance between privacy and performance for the DP model, as it provides a high F1-score, which is essential for applications where both precision and recall are important.

B. Performance Variability Across Epsilon Values

The standard deviation for each metric across all epsilon values indicates the variability of the model as epsilon changes:

- **Accuracy**: 0.0334

Fig. 1. Accuracy Across Epsilon Values

TABLE I
PERFORMANCE COMPARISON OF ORIGINAL AND DIFFERENTIALLY PRIVATE DECISION TREE (EPSILON = 4.09)

Metric	Original Decision Tree	DP Decision Tree
Accuracy	0.9537	0.8369
Precision	0.9537	0.8716
Recall	0.9537	0.8369
F1-Score	0.9537	0.8270

- **Precision**: 0.0257
- **Recall**: 0.0334
- **F1-Score**: 0.0547

These values reveal that the F1-score exhibits the highest variability, suggesting that it is most sensitive to changes in epsilon, likely due to the balance it maintains between precision and recall.

C. Comparison of Original and Differentially Private Decision Tree Models

To compare the performance of the original Decision Tree model with the differentially private version, we selected the best epsilon value (4.09) for the DP model. Table I presents the metrics for both models, while Figure 3 provides a visual comparison.

1) Analysis: **Accuracy**: The original Decision Tree achieves significantly higher accuracy (0.95) compared to the best-performing DP model (0.84). This suggests that the privacy-preserving noise introduced by DP constraints does impact overall accuracy, though the effect is limited in practical terms.

Precision: The original model demonstrates slightly better precision (0.95) than the differentially private version (0.87), indicating that the DP model may slightly underperform in correctly identifying positive instances. This precision drop could be due to privacy-induced noise, impacting the model's confidence in identifying positive cases accurately.

Recall: Both models perform comparably in recall, though the original model again shows a slight advantage. This suggests that differential privacy has a smaller effect on recall than other metrics, with the DP model still able to identify relevant positive instances consistently.

F1-Score: The original model's F1-score (0.95) is higher than that of the differentially private model (0.83). This difference implies that the original model achieves a more balanced trade-off between precision and recall.

IV. DISCUSSION

The results reveal that the original Decision Tree outperforms the differentially private Decision Tree across all

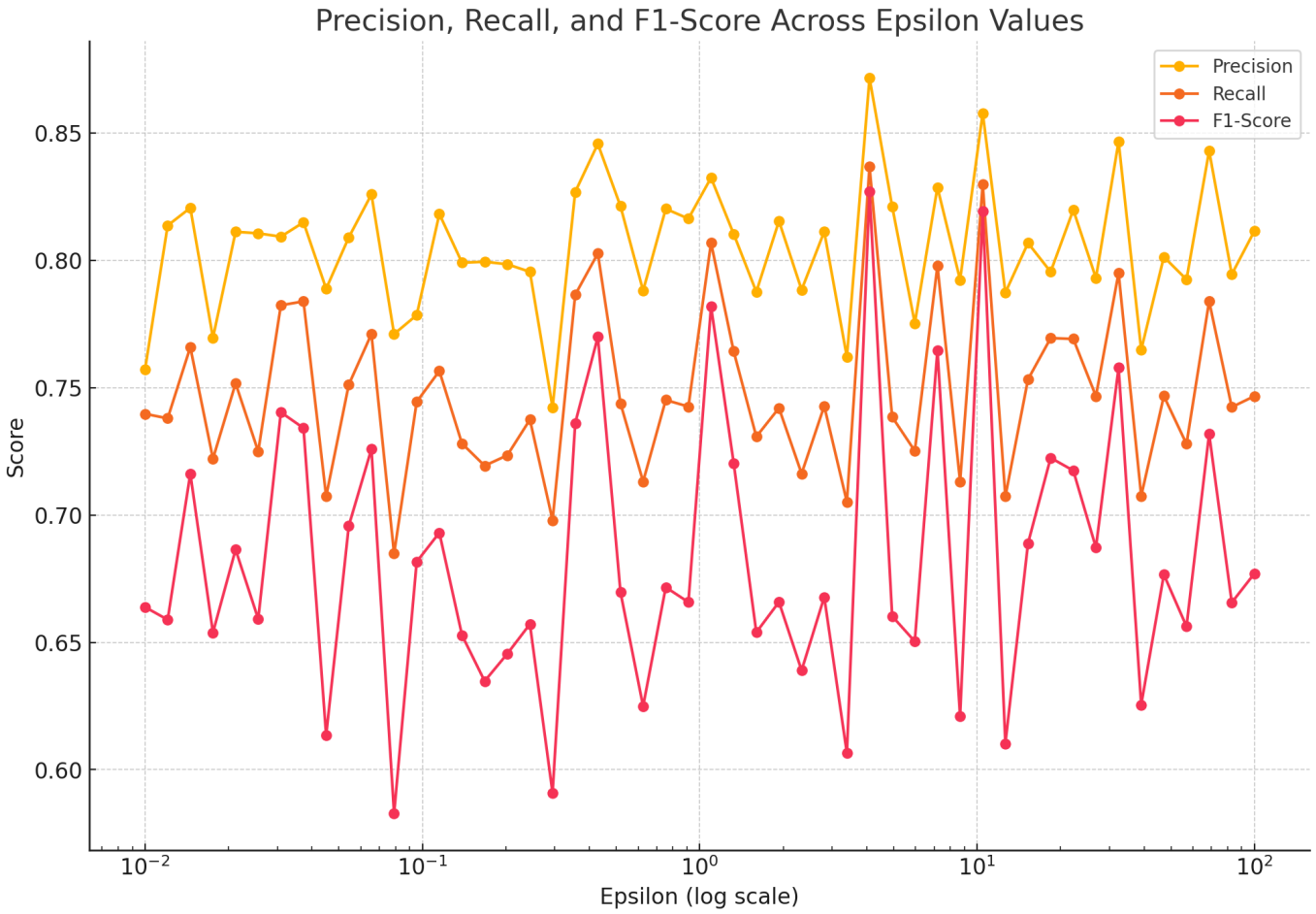

Fig. 2. Precision, Recall, And F1-Score Across Epsilon Values

metrics. However, the differentially private model with epsilon = 4.09 achieves reasonably high performance, particularly in precision and F1-score, which are critical for reducing false positives and false negatives.

The choice of epsilon significantly impacts model performance: low epsilon values introduce more noise, protecting privacy but reducing utility, while higher values reduce noise, enhancing utility but weakening privacy. For applications where both accuracy and privacy are essential, a balanced epsilon value (e.g., 4.09) appears to provide a good trade-off.

These findings highlight that differential privacy, when properly tuned, can enable effective fraudulent website detection while respecting user privacy. Practitioners should carefully consider epsilon settings based on the privacy requirements of their specific application.

V. Conclusion

In this study, we compared a traditional Decision Tree classifier and a differentially private version for detecting fraudulent websites. While the original model achieves optimal performance, the differentially private Decision Tree provides comparable results at higher epsilon values, such as 4.09, demonstrating that privacy-preserving machine learning can be both effective and secure. Future work may explore other privacy-preserving machine learning techniques and their impact on various models for fraud detection.

References

[1] M. Maktabar, A. Zainal, M. A. Maarof, and M. N. Kassim, "Content based fraudulent website detection using supervised machine learning techniques," in *Hybrid Intelligent Systems: 17th International Conference on Hybrid Intelligent Systems (HIS 2017) held in Delhi, India, December 14-16, 2017*. Springer, 2018, pp. 294–304.

[2] P. Saraswathi, J. Anchitaalagammai, and R. Kavitha, "A system review on fraudulent website detection using machine learning technique," *SN Computer Science*, vol. 4, no. 6, p. 702, 2023.

[3] M. A. P. Chamikara, P. Bertok, I. Khalil, D. Liu, and S. Camtepe, "Privacy preserving face recognition utilizing differential privacy," *Computers & Security*, vol. 97, p. 101951, 2020.

[4] F. K. Dankar and K. El Emam, "Practicing differential privacy in health care: A review." *Trans. Data Priv.*, vol. 6, no. 1, pp. 35–67, 2013.

[5] C. Dwork, "Differential privacy: A survey of results," in *International conference on theory and applications of models of computation*. Springer, 2008, pp. 1–19.

[6] G. Vrbančič, I. Fister Jr, and V. Podgorelec, "Datasets for phishing websites detection," *Data in Brief*, vol. 33, p. 106438, 2020.

[7] M. A. Fauzi, B. Yang, and P. Yeng, "Improving stress detection using weighted score-level fusion of multiple sensor," in *Proceedings of the 7th International Conference on Sustainable Information Engineering and Technology*, 2022, pp. 65–71.

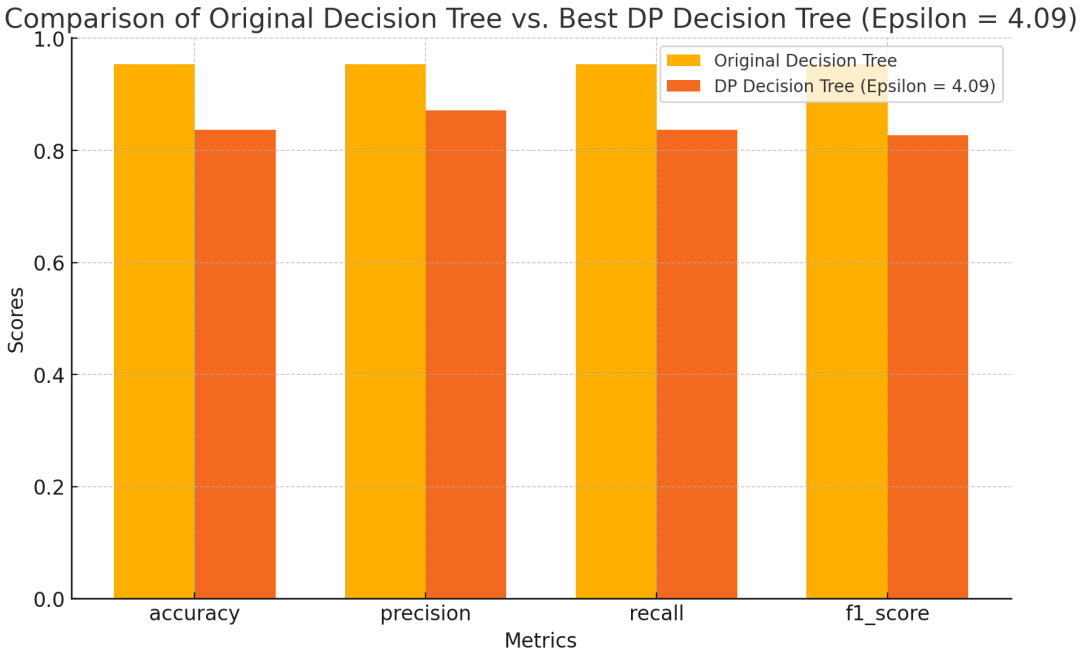

Fig. 3. Bar chart comparing Original Decision Tree and DP Decision Tree (Epsilon = 4.09)

[8] M. A. Fauzi and P. Bours, "Ensemble method for sexual predators identification in online chats," in *2020 8th international workshop on biometrics and forensics (IWBF)*. IEEE, 2020, pp. 1–6.